The genetics of selective breeding in Thoroughbred horses

Evelyn T. Todd

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Faculty of Science

The University of Sydney

Australia

2020

Table of Contents

Acknowledgements

I would like to extend a special thank you to my supervisors Peter Thomson, Natasha Hamilton and Brandon Velie for their support and patience throughout my PhD. I would also like to thank my honours supervisor Simon Ho for helping with my first publication. I am also grateful to my friends and family for their help and support throughout my candidature. Finally I would like to acknowledge Racing Australia, The Australian Studbook, private breeders and collaborators from the Swedish University of Agricultural Sciences and the Norwegian School of Veterinary Medicine who contributed data to my PhD research.

Declaration

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references given.

Evelyn T. Todd

20th of August 2020

A note on style

This thesis is presented as a series of three manuscripts, all of which have been published in peer reviewed scientific journals. Submitted manuscripts are formatted according to the guidelines for each journal. An introductory literature review and a general discussion are included as the first and final chapter of the thesis, respectively. An additional published paper is included in the supplementary information.

Manuscripts included in the main body of this thesis:

1. **E.T. Todd**, S.Y.W. Ho, P.C. Thomson, R.A. Ang, B.D. Velie & N.A. Hamilton. Founder-specific inbreeding depression affects racing performance in Thoroughbred horses. *Scientific Reports* **8**, 6167, doi:10.1038/s41598-018-24663-x (2018).
2 **E.T. Todd**, N.A. Hamilton, B.D. Velie & P.C. Thomson. The effects of inbreeding on covering success, gestation length and foal sex ratio in Australian Thoroughbred horses. *BMC Genetics* **21**, 41, https://doi.org/10.1186/s12863-020-00847-1 (2020).
3 **E.T. Todd**, P.C. Thomson, N.A. Hamilton, R.A. Ang, G. Lindgren, Å. Viklund, S. Eriksson, S. Mikko, E. Strand & B.D. Velie. A genome-wide scan for candidate lethal variants in Thoroughbred horses. *Scientific Reports* **10**, 13153, https://doi.org/10.1038/s41598-020-68946-8 (2020).

Manuscripts included only in the supplementary information:

4. **E.T. Todd**, K. Jäderkvist Fegraeus, P.C. Thomson, C.F. Ihler, E. Strand, G. Lindgren & B.D. Velie. Premie race participation is associated with increased career longevity and prize money earnings in Norwegian-Swedish Coldblooded Trotters. *Acta Agriculturae Scandinavica, Section A - Animal Science* **68**, 112-116, doi:10.1080/09064702.2018.1563211 (2018).

Author contributions

The ideas, development and writing up of all the papers in this thesis were conducted by the candidate working under the supervision of Dr Peter Thomson, Dr Natasha Hamilton and Dr Brandon Velie. The inclusion of co-authors in the published works reflects the active collaboration between researchers.

I contributed to the project design, primary analyses and writing up of the publication entitled "Founder-specific inbreeding depression affects racing performance in Thoroughbred horses". Initial data collection and cleaning was done by Dr Brandon Velie and Ms Rachel Ang. Drs Simon Ho, Peter Thomson and Natasha Hamilton assisted in the design of the project and provided assistance with finalizing the manuscript prior to publication.

I contributed to the project design, data collection, primary analyses and writing up of the publication entitled "The effects of inbreeding on covering success, gestation length and foal sex ratio in Australian Thoroughbred horses". Drs Peter Thomson and Natasha Hamilton assisted with the design of the project. Dr Peter Thomson, Natasha Hamilton and Brandon Velie provided assistance with finalizing the manuscript prior to publication.

I contributed to the project design, data collection, primary analyses and writing up of the manuscript entitled "A genome-wide scan for candidate lethal variants in Thoroughbred horses". Ms Rachel Ang and Drs Natasha Hamilton, Gabriella Lindgren, Åsa Viklund, Susanne Eriksson, Sofia Mikko and Eric Strand all contributed to some of the genotype data that was used in the data analysis of this chapter. Drs Peter Thomson, Brandon Velie and Natasha Hamilton contributed to study design and assisted in finalizing the manuscript prior to publication.

I contributed to the primary analyses and writing up of the publication entitled "Premie race participation is associated with increased career longevity and prize money earnings in Norwegian-Swedish Coldblooded Trotters". The data collection was conducted by Prof Gabriella Lindgren and Dr Brandon Velie. Prof Gabriella Lindgren, Drs Kim Jäderkvist Fegraeus, Peter Thomson, Carl Ihler, Eric Strand and Brandon Velie assisted in finalizing the manuscript prior to completion.

Confirmation of co-authorship of published work

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Assoc. Prof. Peter Thomson

20th of August, 2020

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Dr. Brandon Velie

20th of August, 2020

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Dr. Natasha Hamilton

20th of August, 2020

Conference presentations

1. **E.T. Todd**, N.A. Hamilton, B.D. Velie & P.C. Thomson. “Using genetics to improve athletic performance and fertility in Thoroughbred horses” School of Life and Environmental Sciences Seminar Series, Sydney, Australia (2019). Oral Presentation
2. **E.T. Todd**, N.A. Hamilton, B.D. Velie & P.C. Thomson. “Using genetics to improve athletic performance in Thoroughbred horses.” The Royal Society of New South Wales, Sydney, Australia (2019). Oral Presentation.
3. **E. T. Todd**, S.Y.W. Ho, P.C. Thomson & N.A. Hamilton. “Founder-specific inbreeding depression affects racing performance in Thoroughbred horses.” Society for Molecular Biology and Evolution Conference, Yokohama, Japan (2018). Poster Presentation.
4. **E. T. Todd**, S.Y.W. Ho, P.C. Thomson & N.A. Hamilton. “Does inbreeding affect racing performance in Thoroughbred horses?” School of Life and Environmental Sciences HDR Showcase, Sydney, Australia (2017). Oral Presentation.
5. **E. T. Todd**, S.Y.W. Ho & N.A. Hamilton. “Does inbreeding affect Thoroughbred racetrack performance?” The Genetics Society of Australasia, Adelaide, Australia (2015). Poster Presentation.

Awards and funding

Research Training Program (RTP) Stipend Scholarship | 2017-2020

The University of Sydney Postgraduate Research Prize for Outstanding Academic Achievement in the Faculty of Science | 2019

The Royal Society of NSW Scholarship | 2019

Faculty of Science Postgraduate Research Support Scheme Funding | 2017 ,2018, 2020

The Sydney Informatics Hub Publication Incentive Award | 2018

Society for Molecular Biology and Evolution (SMBE) Registration Award | 2018

Abstract

This thesis examines the effects of selective breeding practices on the Thoroughbred horse population. The Thoroughbred horse breed was founded in the 18th century and the studbook has been closed since 1791, resulting in contemporary Thoroughbred horses being the result of many generations of inbreeding for the intensive selection of athletic performance. Selective breeding can result in genetic improvement through increasing the frequency of variants pertaining to desirable traits and may also potentially remove harmful alleles from a population. However, an increasing body of evidence has shown that many domestic animal populations harbour high levels of deleterious variants as a result of population bottlenecks and low effective population sizes from selective breeding practices. Until now, no studies have reported the effects of these breeding practices on the Thoroughbred horse population.

The findings of this thesis provide important insights into the effects of selective breeding on genetic load and genetic gain in the Thoroughbred horse population. These insights can be used to assist breeding decisions in Thoroughbreds to increase genetic improvement in the population. The findings of this thesis not only have important economic implications for one of the largest domestic animal populations in the world, but also have the potential to improve welfare standards in an industry under increasing scrutiny. More broadly, this thesis provides insights into the effects of selection, inbreeding and population bottlenecks on the health and fitness of domestic breeds. The methods used in this thesis can be applied to other animal populations to improve breeding and population management decisions.

Chapter 1: Introduction

1.1 Inbreeding and genetic load

1.1.1 The principles of inbreeding

Inbreeding is the result of mating between related individuals and has well documented negative effects on the health and fitness of individuals and populations[1-3]. This phenomenon, known as inbreeding depression and the mechanisms underlying its negative phenotypic effects are among the most extensively researched topics in biology. The detrimental effects of inbreeding are commonly a major concern for endangered species with small population sizes[4,5]. However, inbreeding depression is also relevant for domestic animal populations, due to the economic repercussions of reduced health and fitness[2,6-8]. In commercial breeding programs, some inbreeding for the selection of desirable traits is necessary for genetic improvement[9]. However, high levels of inbreeding could also have unexpected unfavourable effects on the fitness of a population in future generations[10].

Inbreeding depression is caused by an increase in homozygosity due to the inheritance of identical sections of DNA from a common ancestor[11]. Characteristically, homozygosity that is a result of inbreeding presents as long tracts of a chromosome that have not been broken by recombination which are considered to be "identical by descent"[12]. It is also possible to inherit identical alleles by chance (known as "identical by state") particularly when they are at high frequencies in a population. However, alleles that are identical by state are usually not present as long homozygous regions.

There are two main mechanisms that have been proposed to explain how increased homozygosity caused by inbreeding can lower fitness. According to the "over-dominance" hypothesis, the decline in fitness is due to a reduction in favourable heterozygosity, which is normally associated with superior fitness[11,13,14]. In contrast, the "partial dominance" hypothesis predicts that the fitness decrease is due to progeny inheriting two copies of a recessive mutation from their common ancestor. Inheritance of two copies of a recessive deleterious mutation can lead to early embryonic death, reduced post birth survival, stunted growth or a decrease in the general fitness characteristics of an individual. Recessive deleterious alleles often persist in diploid populations at low frequencies, known as genetic load (see Section 1.1.2). Under the partial dominance hypothesis, a population with high levels of inbreeding will either become extinct, or recover from inbreeding depression because the deleterious alleles are removed by selection (purging, see Section 1.1.2)[11,15,16]. The partial dominance hypothesis is currently

favoured as the explanation of the majority of the fitness decline associated with inbreeding, although a combination of the two mechanisms is likely responsible[15,17-19].

There is also growing evidence that the genetic mechanisms of inbreeding may be more complex than initially believed. The importance of epistatic interacting gene networks in inbreeding depression has been recognised more recently[20,21]. The complex genetic basis of many traits (quantitative traits) means that the fitness of an inbred population could decline in a nonlinear fashion, in contrast to a steady decline expected under the partial or over-dominance theories[22]. Thus, an understanding of the molecular mechanisms involved in inbreeding depression is essential in predicting the impacts of inbreeding on the fitness of an individual or population.

1.1.2 Genetic load and the cost of domestication

Genetic load is the reduction in fitness compared to the optimal genotype, and is often discussed in the context of a whole population[23,24]. All animal populations carry some genetic load which is concealed as rare recessive mutations found at low frequencies (see section 1.1.3)[25]. Genetic load is only expressed in homozygous form, which can occur due to inbreeding increasing the frequency of these mutations in the population[24-26]. Individuals can carry different amounts of genetic load in their genome, so the effects of inbreeding on fitness can vary between individuals and populations[17,20,27,28]. For example, humans are estimated to carry between one and two recessive lethal alleles in their genomes[29,30].

Population bottlenecks from the domestication process often lead to increased genetic load in animal populations. This process is known as the "cost of domestication" and has been widely observed in many species of animals and plants[31-35]. In this process, deleterious variants hitchhike on selective sweep regions, thereby increasing their frequency in the population[34]. Many animal populations have undergone severe population bottlenecks during domestication or breed formation, resulting in high levels of genetic load being present[33,35]. Future inbreeding events, particularly due to selective breeding practices, can inadvertently further increase the frequency of these deleterious variants in the population, thereby also increasing genetic load[36].

Population bottlenecks and selective breeding practices decrease the number of breeding individuals in a population, which is measured as effective population size (N_e)[37,38]. N_e estimates the size of an idealised population that would show the same rate of loss of genetic diversity assuming random mating, equal genetic contribution and generational intervals between each individual[39]. In selective breeding programs, N_e can be small even with a large consensus population size due to extremely

skewed genetic contributions of some individuals with desirable traits[31,39,40]. Small N_e reduces the genetic diversity in a population, which can increase the levels of inbreeding in future generations[41]. The reduced genetic diversity associated with small N_e can also result in deleterious alleles drifting to high frequencies or fixation, resulting in increased genetic load in a population[4,26,38].

1.1.3 Lethal genetic variation

Genetic variants that cause embryonic lethality are expected to be under strong negative selection in populations due to their deleterious effects on phenotype[42]. Lethal recessive variants are often found at low frequencies in populations and commonly have no impact on health and viability due to individuals being extremely unlikely to inherit two copies of the same mutation by chance[43-45]. However, inbreeding increases the likelihood that two copies of the same lethal mutation will be inherited from a common ancestor[4,11,46]. Populations with high rates of inbreeding and small N_e are at particular risk of these recessive lethal mutations drifting to high frequencies (see section 1.1.2). High frequencies of recessive lethal variants in a population can have a negative impact on fertility rates. For example, in a 10% increase in inbreeding level is associated with a 1% decrease in pregnancy rates in cattle[47].

In recent times, a number of lethal variants have been identified at high frequencies in domestic animal populations, particularly cattle and pig breeds[48-56]. In some cases, these variants have reached high frequencies in the population as they provide a phenotypic advantage in heterozygous state. For example, a 600kb deletion in cattle that causes embryonic lethality in homozygous state may have reached high frequencies due heterozygotes producing an increased milk yield[57]. Similarly, a deletion in the Bardet-Biedl syndrome 9 (*BBS9*) gene in pigs increases growth rates in heterozygotes, but results in lethality when homozygous[58]. However, other variants may not provide a heterozygote advantage but have drifted to high frequencies in animal populations due to being carried by an individual (often a sire) that has made a large genetic contribution to the population[56].

The presence of recessive lethal variants in domestic animal populations are commonly identified by scanning large samples of genotypic data for haplotypes that deviate from the Hardy-Weinberg equilibrium[49]. Such studies are highly feasible for populations with large census population sizes and substantial amounts of genotypic data available. However, smaller populations are also likely to be susceptible to recessive lethal mutations drifting to high frequencies. Such analyses are often not possible for these populations due to the lack of genotypic data available or the associated costs.

1.1.4 Purging

As the frequencies of deleterious alleles increase with inbreeding or drift, selection against these alleles also becomes stronger[59,60]. This makes it possible for these mutations to be removed from the population over time, alleviating genetic load[24]. This rebound in fitness, known as purging, can result in future inbreeding events having a neutral or even positive effect on fitness[60]. Purging is likely to be more effective for alleles that have highly lethal effects as selection against them will be stronger[61,62]. Mildly deleterious alleles may be more difficult to purge[63], which could result in only a partial rebound in fitness relative to the original outbred population[17,24,60]. Genetic drift caused by high rates of inbreeding increases the probability of deleterious mutations becoming fixed in a population, so purging is more likely to be effective in populations with slow rates of inbreeding[24,64]. These reasons likely explain why many studies in animal populations have found that purging is not effective in measurably improving fitness[19,63,65].

Some animal populations have shown no evidence of a relationship between inbreeding levels and fitness. The most notable example is the Chillingham cattle population that show no indication of inbreeding depression despite being so highly inbred that they are almost clonal[66,67]. Additionally, a wild fox population and The Chatham Island finch population also show no evidence of inbreeding depression[68,69]. However, strong evidence of inbreeding depression has been found in many animal populations with population bottlenecks and many generations of inbreeding[19,65].

A further complication in predicting the likelihood of purging is that it may occur under certain conditions and only for certain traits. Selective breeding may increase the chance of purging in a population, for example, inbreeding of superior mates in the Cuvier's gazelle population has resulted in improved juvenile survival[70]. However, under these conditions purging may occur for some traits but not others. For example, selective breeding in dairy cattle has resulted in evidence of purging in milk production traits, but not fertility[71]. Deleterious variants that are linked to selective sweep regions are also less likely to be purged from a population[31,32,34].

Additionally, inbreeding depression may manifest only under certain conditions. For example, the highly inbred Chatham Island Black Robin still shows strong evidence of inbreeding depression, although inbred chicks from inbred dams show increased juvenile survival rates[19,72]. These findings show that the exact effects of purging are hard to predict. Evidence for purging may also depend on the method by which it

is measured[73]. Consequently, deliberate inbreeding to induce purging is not a strategy recommended or practiced in breeding programs.

1.1.5 Selective breeding and genetic gain

In selective breeding programs, some level of inbreeding between individuals with desirable traits is needed to increase the frequency of the associated alleles in the population. However, a small number of superior individuals making large genetic contributions will reduce diversity in the population[9]. This practice is likely to increase the rate of inbreeding in subsequent generations as their descendants will have limited outbreeding opportunities[9]. Therefore, selective breeding programs not only need to take into account the inbreeding level of the parents and progeny, but also consider the overall genetic contributions that each individual is making to the population[9]. Successful selective breeding programs aim to maximise genetic gain whilst minimising rates of inbreeding in a population[74].

For many centuries, selective breeding schemes were based on breeding animals with desirable traits and hoping that their offspring inherited the same phenotype[75]. However, an increased understanding of quantitative genetic concepts in the 20th century led to the development of more complex selection schemes, based on relationships between animals and statistical probabilities of inheritance[74,76]. This concept was first introduced by James and McBride in 1958 where they estimated the contributions from ancestors on the phenotypic changes in their descendants[77]. Best linear unbiased predictions (BLUPs) are a commonly used method that utilizes phenotypic data from relatives to give an accurate prediction of genetic merit known as estimated breeding values (EBVs) for tested and as yet untested individuals[76,78,79]. This method is particularly useful for animal populations with large and complex pedigrees as the relationship between individuals can be accurately determined[78,80,81]. Best linear unbiased predictions to maximise genetic gain are widely used in many livestock species, including cattle, pigs and chickens[6,82-86].

Traditionally, pedigree structures have been used to evaluate the genetic relationship between individuals. However, the increasing availability of genomic information in the 21st century has allowed for the incorporation of genotypic information into selective breeding schemes which is particularly useful for populations with unknown or unreliable pedigrees[9]. The use of genomic selection methods can assist in achieving homozygosity of favourable alleles in a population, which will result in greater levels of inbreeding at these loci compared to other parts of the genome[9].

1.2 The domestic horse

1.2.1 The foundation of the domestic horse population

Horses (*Equus caballus*) were first domesticated from their wild ancestors at around 3500B.C[87]. Many of the paternal lines that existed in early domestic horse populations are no longer found in modern horse breeds, probably due to a combination of selective breeding practices and genetic drift[88-90]. In contrast, the broad diversity of mitochondrial DNA in domestic horses indicates that many maternal lines have been maintained in the population[87,90,91]. These findings provide evidence that the early domestication process involved the backcrossing of local wild mares to already domesticated stallions[87,90,91]. Investigations of genetic load in domestic horses showed an accumulation of deleterious alleles compared to their wild ancestors, probably as a result of this population bottleneck at the foundation of the population[35].

Comparison of prehistoric horse DNA with modern domestic breeds shows evidence of positive selection in at least 125 candidate genes[35]. These loci provide an indication of the traits selected for during the process of domestication, in particular the increasing energetic demands, requirement for tameness and larger size[35]. Furthermore, there is evidence that selective breeding practices targeted coat colour in early domestic horses, as a high proportion of horse remains analysed from the Bronze Age had the allele for appaloosa spotted coat colour. There was also evidence in these horses of selection against the wild dun coat colour[92]. Consequently, it can be assumed that early domestic horses were selected for stamina, appearance and docility.

Approximately 500 different horse breeds are descended from the original domesticated horses, most of which have been developed within the last several hundred years[87]. Differing selective pressures have led to a wide range of phenotypic diversity between the different horse breeds[93]. However, selective breeding practices and severe population bottlenecks have also resulted in some horse populations having extremely low within-breed diversity, some equivalent to endangered species on the brink of extinction[36,89,93]. In particular there are a number of horse breeds with extremely low N_e such as the Sorraia and draft horse breeds[36,94]. Population bottlenecks due to breed formation or near extinction events have led to more recent increases in the mutational load of domestic horses[36]. Genetic variation has also decreased within the last 200 years by approximately 16%, which can be explained by dramatic reductions in effective population sizes of many breeds due to the replacement of horses by machinery for agriculture and transportation[89,95,96]. As such, increasing mutational loads and decreasing genetic

diversity has the potential to affect the health, fitness and viability of horse breeds in future generations. The effects of these past events and the impacts of current breeding practices on the genetic diversity of future generations is a topical concern for domestic horse breeders.

1.2.2 Reproduction and fertility in domestic horses

Horses are seasonal breeders, with the increased photoperiod at the start of spring inducing hormonal changes which are critical for oestrous cycling and parturition[97-102]. Reduced melatonin secretion in the longer daylight hours of spring increases the levels of gonadotrophins released, stimulating the production of sex hormones[98]. The oestrus cycles of some domestic horse breeds can start in late winter, in contrast to wild populations that may not breed until late spring[97,103]. Early oestrus cycling may be the result of selective breeding favouring horses born earlier in the season, which indicates that the trait may have a genetic basis.

Horses have an average gestation length of 342 days (approximately 11 months), also giving birth mostly during the spring months. The survival of twins is rare, so most mares only produce one foal in a year[104]. Gestation length differs between some horse breeds, indicating that it may be influenced by selective breeding strategies[105]. However, an understanding of the genetic factors that influence gestation length is complicated by the strong influence of a number of environmental factors. Increasing photoperiod plays an important role in inducing parturition in mares, resulting in foals that are born at the end of the spring/summer having shorter gestation lengths than those born in winter/early spring[99,100,105-108]. Male foals have a significantly longer gestation length than female foals by about 2-3 days, which is likely due to hormonal interactions between the mare and the foal[106-110].

A number of studies have also found that the age of a mare impacts her gestation length. However, trends differ between studies with some finding a linear increase with mare age[106,107,111], where others found that younger mares also have longer gestation lengths[107,108,112]. The changing physiology of older mares would explain why they have a longer gestation length[106]. The longer gestation length of younger mares may have been due to parity status as mares that had not produced a foal the year before have increased gestation lengths[108].

In natural horse populations, horses live in herds of several mares with their young offspring and a harem stallion[97]. Male horses in natural populations make uneven contributions to the next generation, with dominant stallions producing many offspring and bachelor males very few[97]. This is exaggerated in domestic breeds, due to a large number of males being castrated and small numbers of stallions with

desirable traits producing large numbers of offspring[113]. Consequently, high levels of stallion fertility are economically important for commercial breeding. A number of genes have been associated with stallion fertility and mating success in different domestic horse breeds[114-121]. However, very little is understood about the genetic variation underlying mare fertility rates and gestation length, probably because they provide a far smaller genetic contribution to the next generation. Mare fertility rates and gestation length both show evidence of significant heritable components in some breeds[100,107,108]. Mares from endangered breeds such as the Sorraia and the Przewalski's horse show abnormal oestrus cycles and low covering success, possibly as a result of severe population bottlenecks and high levels of inbreeding[122,123]. Good fertility output in mares is important for commercial breeding programs and may also be essential for the survival of smaller horse breeds[103,124].

1.2.3 Genomics in the domestic horse

A reference genome for the domestic horse (*Equus caballus*) was first published in 2009 and has led to an increased understanding of history, health and disease[125]. The reference genome was subsequently used to develop commercial equine genotyping arrays, which are composed of common single nucleotide polymorphisms (SNPs) evenly dispersed throughout the equine genome in order to capture genome-wide trends[95,126]. These arrays have been useful in understanding genetic diversity in different horse populations and genome-wide association studies (GWAS) to identify areas of the genome that may be targets of selection or harbour disease alleles[93,127-134].

Following the publication of the reference genome, equine genes were annotated and characterised based on motifs, start/stop codons and known sequences for human and mouse genes[135-137]. Annotation of genes in the domestic horse genome has been essential in the fine mapping of variants underlying many conditions in horses, and understanding the changes that mutations may have on protein function[35,138-144]. The annotation of the genome has made it possible to conduct transcriptomic analysis in equine tissues[136,137,145-147]. Such analyses have improved our understanding of differential expression patterns in response to environmental stimuli (such as exercise)[148,149] and physiological changes due to diseases[150,151].

1.3 The Thoroughbred horse breed

1.3.1 The foundation of the Thoroughbred horse breed

The Thoroughbred horse breed was developed for the purpose of producing horses with good racing ability. The foundation of the breed dates back to the early 18th century in England with the mating of stallions from the Middle East to local English mares[88]. Although horse racing had been practiced for many centuries prior to this, formal rules and regulations were implemented for the first time in 18th century England. The results of races were recorded and weight handicaps for horses of different abilities were implemented[152]. The breeding information for each horse was also recorded, resulting in the beginning of the first formal studbook[152]. Private breeding information was collated into the first official Thoroughbred stud book which was published in 1791[153]. From that point onwards, only Thoroughbred horses with parents registered in the studbook were permitted to race, resulting in a closed population. As a result, all modern Thoroughbred horses can trace their ancestry back to this time.

Many of the original horses used for breeding and racing in the 18th century quickly disappeared from the pedigree of subsequent generations, resulting in all modern Thoroughbred horses in the world being descended from a small number of foundation individuals[152,154]. A reflection of the small foundation population was highlighted in a 2001 study on Thoroughbreds in Britain showing that the top 20 founders have contributed to 65% to the genetic diversity in the modern population[155]. This finding is likely to be similar for Thoroughbred horses bred in other countries, due to continuous importation and admixture over the past 200 years.

As a result of being descended from a small foundation population, all Thoroughbred horses can trace their paternal ancestry directly back to three stallions imported from the Middle East: The Darley Arabian, The Godolphin Barb and the Byerley Turk[155]. Over 95% of horses sampled in the 1990s traced their paternal line back to the Darley Arabian, but this number could be higher 20 years later if selective breeding has continued to favour this line. This finding agrees with a more recent study, which reports that 96.5% of male Thoroughbreds share the same Y chromosome haplotype, which traces back to the Darley Arabian[88]. Recently, some inconsistencies have found between Y chromosome haplotypes and breeding records in Thoroughbreds, indicating that there may be some errors in early studbooks entries[156].

On the other hand, a larger pool of mitochondrial haplotypes in modern Thoroughbreds indicates greater female founder diversity[157,158]. These haplotypes are most closely related to the modern day Middle Eastern Barb, Irish Draft horse and Connemara, confirming pedigree records that a combination of native British and imported mares from the Middle East founded the Thoroughbred breed[152,158,159]. Similar to findings from male lineages, some mitochondrial haplotypes in modern Thoroughbred horses show discrepancies with pedigree records[157]. Many of these errors can be attributed to private records being collated into the first studbook. The lack of understanding of genetics and inheritance in the 18th century led to the identity of many mares being poorly recorded, such that some mares had no name or an unrecorded change of name with a change of ownership[159]. Additionally, mares could have been misidentified when they were named only after their sire (e.g. Pot8os Mare1 and Pot8os Mare2), which can lead to paternal half-sisters being recorded as the same horse[157,160].

1.3.2 Selective breeding in the Thoroughbred horse

Early Thoroughbred races were often held over 8 miles (13 km) with multiple heats on the same day[152,159]. At this time, most horses started racing at around 5 years of age, with some continuing to race up until 20 years old. In the 18th and 19th century, horses were the primary mode of transport, so Thoroughbreds were often expected to walk to the racetrack to participate, resulting in races mostly including local contestants[159]. In the 19th century, racing quickly evolved to horses participating over one- or two-mile races in a single race per day[152]. The 20th century saw the increasing popularity of races over shorter distances (less than one mile), and horses starting their racing career at younger ages[152,161].

The selection for racing performance over many generations has resulted in Thoroughbred horses becoming physically distinct from other breeds[152]. Selective sweep comparison of Thoroughbred horses compared to other breeds has identified a number of target regions that may contribute to their superior athletic performance[138]. Genes that play key roles in energy metabolism, insulin pathways, bone formation and muscle and cardiac growth show strong evidence of positive selection in the Thoroughbred breed[138,162].

A number of neurological genes have also been associated with trainability and racing performance of Thoroughbreds. Docility and tractability are important traits in Thoroughbred horses to allow them to be successfully trained and managed[163]. For example, a single SNP in the serotonin receptor HtrA serine peptidase 1 *(HTRA1*) was found to be strongly associated with anxiety and tractability in Thoroughbred

horses[163]. Increased anxiety was seen in heterozygotes, but few homozygous Thoroughbreds were found, possibly due to selection against these individuals. The association with this SNP and tractability was more pronounced in female horses potentially due to sex differences in the regulation of anxiety[163].

Variation in the functionally related neurotrimin (*NTM*) and opioid binding protein/cell adhesion molecule like (*OPCML*) gene also showed strong associations with racing performance[164]. These genes are thought to regulate neural growth, although their exact function is unknown. Variation in *NTM* has been associated with childhood aggressiveness in humans and was identified as a selective target during the domestication of the horse[35], so may be important for the behavioural adaptations required for racing and training[164]. Racing ability was also associated with variation in the prolylcarboxypeptidase (*PRCP*) gene, which has been associated with voluntary wheel exercise in lab mice[164]. The association of these behavioural genes with racing performance suggests that a horse's attitude towards training regimes may play an important role in contributing to their racing performance. These findings suggest that selective breeding over many generations in the Thoroughbred population may have increased the frequency of variants associated with behaviour in addition to physical characteristics.

1.3.3 Reproduction in the Thoroughbred horse

Reproductive technologies including artificial insemination, embryo transfer and cloning are prohibited in Thoroughbred horse breeding worldwide, making it essential to maintain high levels of natural fertility in the population. Breeders aim for foals to be born in late winter or the start of spring so that they have a maturing advantage over their peers. This has become particularly important in Thoroughbreds because of the increasing emphasis on horses racing at two years old. For foals to be consistently born at the same time each year, a mare needs to conceive within 25 days of foaling down[165]. In most cases, this does not happen, with mares foaling down at a later date each season until they reach the end of the season and have to wait until the next year[110]. Consequently, mares that have low covering success and increased gestation length will produce less offspring throughout their lifetime. Despite the importance of maintaining good fertility rates in the Thoroughbred population, currently there has been little investigation into the underlying genetic factors and genetic improvement of these traits. Fertility studies have mainly focussed on optimising veterinary treatments (e.g. hormonal supplements) and management techniques (e.g. light masks), rather than genetic selection[99,166].

Most male Thoroughbreds are gelded (about 90%), and only a small proportion of elite performing males provide large genetic contributions to the population. Consequently, good fertility rates are

required in these males as they are expected to cover large numbers of mares at stud. Stallions with low levels of covering success are often gelded and returned to racing, although the underlying basis of their subfertility is rarely investigated. Currently, a SNP in the FK506 binding protein 6 (*FKBP6*) has been associated with stallion subfertility in Thoroughbreds[119], although it is likely that other as yet unknown variants at high frequencies also influence male fertility rates in the breed. An improved understanding of the genetic basis of fertility traits could assist in mating and management decisions to improve breeding outcomes in the Thoroughbred population.

1.3.4 Common diseases in the Thoroughbred breed

There are a number of common conditions that can have negative and sometimes debilitating effects on a Thoroughbred horse's racing career. Recurrent laryngeal neuropathy is a condition in Thoroughbreds which results in the paralysis of one or both laryngeal nerves[167]. This disorder results in a horse having difficulty breathing under intensive exercise and a characteristic "roaring" sound can be heard under these conditions[168]. The nerve can be tied back in corrective surgery, although these horses often still show reduced racing performance[168]. Historically, it was noted that taller horses were more likely to be affected. Genome-wide association studies have since identified a link between the incidence of roaring and an allele in the ligand dependent nuclear receptor corepressor (*LCORL*) gene. This gene is associated with height, which is suggested to lead to longer laryngeal nerves[169]. However, the Belgian draft horse breed is fixed for this allele but still show a higher risk of roaring in taller horses, indicating that other unknown loci also contribute to the development of this disease[129,170]. It seems likely that selection for bigger, stronger Thoroughbreds has inadvertently selected for higher incidences of roaring in the population.

Osteochrondrosis (OC) is a disease characterised by abnormal cartilage growth or lesions in joints. Disturbance of growing cartilage leads to the development of a cartilage core, which can result in fractures, cyst-like lesions, cartilage flaps or cracks. These abnormalities are associated with lameness which reduces a horse's racing career[171,172]. There also may be an association of OC with fracture risk and suspensory ligament injury in Thoroughbreds[173]. In horses, OC is most commonly found in leg joints particularly the fetlock, stifle or hock[171]. The prevalence of OC varies between different horse breeds, potentially due to different physiology, management and selective breeding techniques[174]. Many studies have found a significant genetic component, although heritability differs between breeds and joints, with estimates between 0.05-0.46[132,134,171,175,176]. Heritability estimates in Thoroughbreds range between 0.10-0.22 depending on the joint[132,177].

A number of GWAS have attempted to find causal genetic variants underlying the incidence of OC, but associated regions differ between studies[134,171,178,179]. Currently, it appears that the underlying genetic basis of OC also differs between breeds and joints, and that larger sample sizes are needed for more conclusive results[171]. A GWAS study in Thoroughbreds identified UDP-glucose dehydrogenase (*UGDH*) on chromosome 3 as a potential candidate for harbouring causal variants[132]. This gene contributes to glycosaminoglycan synthesis in cells lining the cartilage surface[132]. Environmental factors such as the weight of the mare and insulin resistance status also appear to influence the prevalence of the disease[177].

Limb fracture in Thoroughbred horses is a condition that has attracted recent widespread media attention, as it often results in euthanasia. A study conducted on Thoroughbreds racing in Britain found that the heritability of distal limb fractures resulting in euthanasia was 0.48[180]. Three regions showed genome-wide significance levels with fracture risk, which encapsulated a number of genes associated with bone formation[180]. Further fine mapping of these regions has not yet been conducted, but could assist in selective breeding strategies to reduce the incidence of limb fractures in Thoroughbred horses.

Exercise induced pulmonary haemorrhage results in burst capillaries in the lungs as a result of strenuous exercise, resulting in epistaxis (bleeding from the nose)[181]. Under the rules of Australian racing, epistaxis from both nostrils results a three month ban for the first incident and then a lifetime ban of a horse from racing at the second incident[182]. However, the regulations surrounding epistaxis differ between racing jurisdictions. The occurrence of epistaxis in Thoroughbreds is relatively common (2.1%) and has been found to have a significantly heritable component[181,183]. So far, SNPs found in an ectonucleotidase family (*CD39*), which have been implicated to prevent bleeding and limit inflammation by binding to receptors on blood cells are potential causal variants for this condition[184].

Although many studies have found that these conditions are heritable, and have identified potential candidate genes, the exact causal variants associated with their increased prevalence in Thoroughbreds are currently unknown. Further fine mapping to identify such variants may assist in optimizing breeding strategies to reduce the incidence of these conditions in the population. However, many of these conditions appear to have a complex genetic basis with multiple contributing loci and only occur under certain environmental conditions, making the selection of appropriate cases and controls difficult. A better understanding of these conditions would greatly benefit the Thoroughbred population as breeding and management strategies to counteract the negative effects of these conditions could be improved.

1.3.5 Bottlenecks, inbreeding and genetic diversity in the Thoroughbred horse breed

From pedigree records and genetic findings, we can assume that there was a significant population bottleneck at the foundation of the Thoroughbred breed[131,153,155]. This bottleneck, combined with selective breeding practices, could have increased deleterious variation through hitchhiking on selective sweep regions[34]. The closed pedigree structure and selective breeding strategies used from the 18th century until the present day has resulted in contemporary Thoroughbred horses being the product of many generations of inbreeding. The Thoroughbred horse has been found to have high levels of genomic inbreeding when compared to other horse breeds[93,160]. As a result of the genetic bottleneck at the foundation of the breed and inbreeding for the selection of racing performance, Thoroughbred horses have higher mutational loads compared with other closely related breeds such as the Arabian and Akhal-Teke horses[36].

The global Thoroughbred population has an estimated effective population size of 330; however, individual country estimates are between 93-197 (although this could be influenced by sample sizes)[132]. There is strong evidence that N_e and genetic diversity has been decreasing in recent years as a result of selective breeding practices heavily favouring popular stallions[113,131]. Analysis of Thoroughbred genomes show evidence for purging, which suggests that selective breeding for racing performance over many generations may have removed some deleterious variation from the population[36]. As seen in domestic cattle populations, selective breeding practices could lead to purging for some traits and not others[7]. Due to the recent reduction in genetic diversity and the closed population structure, the effects of inbreeding on the health and fitness of the Thoroughbred population is a highly topical concern. However, to date no studies have attempted to quantify this. A previous study on cattle found evidence for purging only on traits that were targets of selection[7], so the effects of inbreeding may also vary between different traits in Thoroughbred horses. Understanding how inbreeding affects the Thoroughbred population would assist in creating breeding strategies to improve the performance of the population and assist in ensuring its continued viability in future generations.

1.3.6 The Thoroughbred breed in Australia

The first Thoroughbred horses were imported to Australia in the early 19th century. Thoroughbred racing quickly became a popular sport in Australia, with many racetracks opened throughout the major cities of Sydney and Melbourne throughout the 18th century[154]. The popularity of horse racing in Australia has led to the growth and development of an industry which makes large contributions to the economy each

year[113]. Australia currently has the second largest Thoroughbred population in the world (behind the United States of America)[113].

Thoroughbred breeding has become increasingly commercialised in the past 30 years, resulting in popular stallions covering large numbers of mares over a number of seasons at stud. Elite stallions are extremely valuable, often commanding large service fees, for example, the covering charge for the most expensive stallion in Australia in 2019, "I Am Invincible", was $AU247,500. Improved air transport has allowed popular stallions to shuttle between the Northern and Southern hemispheres to cover mares across both breeding seasons. This practice has the potential to provide new sources of genetic diversity and outbreeding opportunities to different Thoroughbred subpopulations; however it could also reduce the genetic diversity in the worldwide Thoroughbred breed. A reduction in genetic diversity is likely to result in an unavoidable increase in inbreeding of future generations and provide less outbreeding opportunities. Considering the closed nature of the Thoroughbred population, a loss of genetic diversity cannot be recuperated in future generations.

1.4 Aims of this thesis

Despite the extensive phenotypic and pedigree data available to Thoroughbred horses, the effects of selective breeding practices on the population are currently undetermined. Evidence of the negative effects of inbreeding and population bottlenecks in other animal populations makes the potential impact of such practices on the health and fitness of Thoroughbred horses a concern. The decreasing effective population size of the Thoroughbred breed in recent generations makes the effects of increased inbreeding particularly topical and relevant.

The over-arching aim of this thesis is to understand the effects of many generations of selective breeding on the Thoroughbred population. To address this, the three main aims of this thesis are:

1. To examine the effects of inbreeding for selection on racing performance traits in Australian Thoroughbred horses.
2. To examine the effects of inbreeding for selection on fertility traits in Australian Thoroughbred horses.
3. To identify common genetic variants at high frequencies in the Australian Thoroughbred horse population which are associated with embryonic lethality.

The broader aims of this research is to understand how the many generations of breeding for the selection of racing performance have impacted Thoroughbred horses in the 21st century. This research aims to assist in the understanding of the potential impacts of current breeding decisions on the Thoroughbred horse population in future generations. The findings of this thesis can provide insights for breeding decisions to increase the chance of a positive outcome (i.e. a live foal and elite racehorse) and improve the overall health and fitness of Thoroughbred horses. Ultimately, increasing the chance of a mare producing a healthy foal, the horse making it to the racetrack, increasing trainability and reducing injury risk is of benefit to all participants in the industry.

The Thoroughbred horse population has a large amount of documented phenotypic data available in the form of breeding and racing records. Additionally, the Thoroughbred breed has one of the largest and most detailed pedigrees of any animal population in the world. This allows for large scale analyses with high statistical confidence and the examination of whole population trends over many generations. Such studies are not possible for many animal populations, so insights from this research could be useful where such information is not available. More broadly, the research done in this thesis will also assist in understanding the overarching effects of animal domestication and selective breeding practices on genetic load and genetic improvement in populations. Such populations include (numerically) small horse and domestic animal breeds, and possibly endangered species where critical conservation management is required.

1.5 References

1 Kristensen, T. N. & Sørensen, A. C. Inbreeding - lessons from animal breeding, evolutionary biology and conservation genetics. *Anim. Sci.* **80**, 121-133, doi:10.1079/ASC41960121 (2005).

2 Leroy, G. Inbreeding depression in livestock species: review and meta-analysis. *Anim. Genet.* **45**, 618-628, doi:10.1111/age.12178 (2014).

3 Chapman, J. R., Nakagawa, S., Coltman, D. W., Slate, J. & Sheldon, B. C. A quantitative review of heterozygosity-fitness correlations in animal populations. *Mol. Ecol.* **18**, 2746-2765, doi:10.1111/j.1365-294X.2009.04247.x (2009).

4 Keller, L. F. & Waller, D. M. Inbreeding effects in wild populations. *Trends Ecol. Evol.* **17**, 230-241, doi:10.1016/S0169-5347(02)02489-8 (2002).

5 Frankham, R. Genetics and extinction. *Biol. Conserv.* **126**, 131-140, doi:10.1016/j.biocon.2005.05.002 (2005).

6 Gizaw, S. *et al.* Optimization of selection for growth in Menz Sheep while minimizing inbreeding depression in fitness traits. *Genet. Select. Evol.* **45**, 20, doi:10.1186/1297-9686-45-20 (2013).

7 Mc Parland, S., Kearney, F. & Berry, D. P. Purging of inbreeding depression within the Irish Holstein-Friesian population. *Genet. Select. Evol.* **41**, 16, doi:10.1186/1297-9686-41-16 (2009).

8 Croquet, C., Mayeres, P., Gillon, A., Vanderick, S. & Gengler, N. Inbreeding depression for global and partial economic indexes, production, type, and functional traits. *J. Dairy Sci.* **89**, 2257-2267, doi:10.3168/jds.S0022-0302(06)72297-4 (2006).

9 Woolliams, J. A., Berg, P., Dagnachew, B. S. & Meuwissen, T. H. E. Genetic contributions and their optimization. *J. Anim. Breed. Genet.* **132**, 89-99, doi:10.1111/jbg.12148 (2015).

10 Villanueva, B., Bijma, P., Daetwyler, H. D. & Woolliams, J. A. Inbreeding in genome-wide selection. *J. Anim. Breed. Genet.* **124**, 369-376, doi:10.1111/j.1439-0388.2007.00693.x (2007).

11 Charlesworth, D. & Willis, J. H. The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**, 783-796, doi:10.1038/nrg2664 (2009).

12 Browning, S. & Browning, B. Identity by descent between distant relatives: detection and applications. *Annu. Rev. Genet.* **46**, 617 (2012).

13 Charlesworth, D., Charlesworth, B. & Morgan, M. T. Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. *Evolution* **44**, 1469-1489 (1990).

14 Shull, G. H. The composition of a field of maize. *J. Hered.* **4**, 296-301, doi:10.1093/jhered/os-4.1.296 (1908).

15 Wright, L. I., Tregenza, T. & Hosken, D. J. Inbreeding, inbreeding depression and extinction. *Conserv. Genet.* **9**, 833-843, doi:10.1007/s10592-007-9405-0 (2008).

16 Davenport, C. B. Degeneration, albinism and inbreeding. *Science* **28**, 454-455, doi:10.1126/science.28.718.454-b (1908).

17 Ayroles, J. F. *et al.* A genomewide assessment of inbreeding depression: gene number, function, and mode of action. *Conserv. Biol.* **23**, 920-930, doi:10.1111/j.1523-1739.2009.01186.x (2009).

18 Charlesworth, B. & Charlesworth, D. The genetic basis of inbreeding depression. *Genet. Res.* **74**, 329-340, doi:10.1017/S0016672399004152 (1999).

19 Kennedy, E. S., Grueber, C. E., Duncan, R. P. & Jamieson, I. G. Severe inbreeding depression and no evidence of purging in an extremely inbred wild species- the Chatham Island black robin. *Evolution* **68**, 987-995, doi:10.1111/evo.12315 (2014).

20 Casellas, J. *et al.* Epistasis for founder-specific inbreeding depression in rabbits. *J. Hered.* **102**, 157-164, doi:10.1093/jhered/esq117 (2011).

21 Noguera, J. L. *et al.* A bi-dimensional genome scan for prolificacy traits in pigs shows the existence of multiple epistatic QTL. *BMC Genomics* **10**, 636-636, doi:10.1186/1471-2164-10-636 (2009).

22 Curik, I., Sölkner, J. & Stipic, N. The influence of selection and epistasis on inbreeding depression estimates. *J. Anim. Breed. Genet.* **118**, 247-262, doi:10.1046/j.1439-0388.2001.00284.x (2001).

23 Haldane, J. The effect of variation on fitness. *Am. Nat.* **71**, 337 (1937).

24 Hedrick, P. W. & Garcia-Dorado, A. Understanding inbreeding depression, purging, and genetic rescue. *Trends Ecol. Evol.* **31**, 940-952, doi:10.1016/j.tree.2016.09.005 (2016).

25 García-Dorado, A. Understanding and predicting the fitness decline of shrunk populations: Inbreeding, purging, mutation, and standard selection. *Genetics* **190**, 1461-1476, doi:10.1534/genetics.111.135541 (2012).

26 Henn, B. M., Botigue, L. R., Bustamante, C. D., Clark, A. G. & Gravel, S. Estimating the mutation load in human genomes. *Nat. Rev. Genet.* **16**, 333-343, doi:10.1038/nrg3931 (2015).

27 Casellas, J., Varona, L., Ibáñez-Escriche, N., Quintanilla, R. & Noguera, J. L. Skew distribution of founder-specific inbreeding depression effects on the longevity of Landrace sows. *Genet. Res. (Camb)* **90**, 499-508, doi:10.1017/S0016672308009907 (2008).

28 Casellas, J., Piedrafita, J., Caja, G. & Varona, L. Analysis of founder-specific inbreeding depression on birth weight in Ripollesa lambs. *J. Anim. Sci.* **87**, 72-79, doi:10.2527/jas.2008-0897 (2009).

29 Gao, Z., Waggoner, D., Stephens, M., Ober, C. & Przeworski, M. An estimate of the average number of recessive lethal mutations carried by humans. *Genetics* **199**, 1243-1254, doi:10.1534/genetics.114.173351 (2015).

30 Narasimhan, V. M. *et al.* Health and population effects of rare gene knockouts in adult humans with related parents. *Science* **352**, 474-477, doi:10.1126/science.aac8624 (2016).

31 Moyers, B. T., Morrell, P. L. & McKay, J. K. Genetic costs of domestication and improvement. *J. Hered.* **109**, 103-116, doi:10.1093/jhered/esx069 (2017).

32 Wang, G. D., Xie, H. B., Peng, M. S., Irwin, D. & Zhang, Y. P. Domestication genomics: evidence from animals. *Annu. Rev. Anim. BioSci.* **2**, 65-84, doi:10.1146/annurev-animal-022513-114129 (2014).

33 Marsden, C. D. *et al.* Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 152-157, doi:10.1073/pnas.1512501113 (2016).

34 Lu, J. *et al.* The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends Genet.* **22**, 126-131, doi:10.1016/j.tig.2006.01.004 (2006).

35 Schubert, M. *et al.* Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E5661-E5669, doi:10.1073/pnas.1416991111 (2014).

36 Orlando, L. & Librado, P. Origin and evolution of deleterious mutations in horses. *Genes* **10**, 649, doi:10.3390/genes10090649 (2019).

37 Wright, S. Coefficients of inbreeding and relationship. *Am. Nat.* **56**, 330-338, doi:10.1086/279872 (1922).

38 Kristensen, T. N., Hoffmann, A. A., Pertoldi, C. & Stronen, A. V. What can livestock breeders learn from conservation genetics and vice versa? *Front. Genet.* **6**, 38, doi:10.3389/fgene.2015.00038 (2015).

39 Husemann, M., Zachos, F. E., Paxton, R. J. & Habel, J. C. Effective population size in ecology and evolution. *Heredity* **117**, doi:10.1038/hdy.2016.75 (2016).

40 Leroy, G. *et al.* Methods to estimate effective population size using pedigree data: examples in dog, sheep, cattle and horse. *Genet. Select. Evol.* **45**, 1-1, doi:10.1186/1297-9686-45-1 (2013).

41 Bouzat, J. L. Conservation genetics of population bottlenecks: the role of chance, selection, and history. *Conserv. Genet.* **11**, 463-478, doi:10.1007/s10592-010-0049-0 (2010).

42 Amorim, C. E. G. *et al.* The population genetics of human disease: The case of recessive, lethal mutations. *PLoS Genet.* **13**, e1006915, doi:10.1371/journal.pgen.1006915 (2017).

43 Ballinger, M. A. & Noor, M. A. F. Are lethal alleles too abundant in humans? *Trends Genet.* **34**, 87-89, doi:10.1016/j.tig.2017.12.013 (2018).

44 Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. Mutation rate and dominance of genes affecting viability in Drosophila melanogaster. *Genetics* **72**, 335 (1972).

45 Sulem, P. *et al.* Identification of a large set of rare complete human knockouts. *Nat. Genet.* **47**, 448-452, doi:10.1038/ng.3243 (2015).

46 Curik, I., Ferencakovic, M. & Solkner, J. Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest. Sci.* **166**, 26-34, doi:10.1016/j.livsci.2014.05.034 (2014).

47 Vanraden, P. M. & Miller, R. H. Effects of nonadditive genetic interactions, inbreeding, and recessive defects on embryo and fetal loss by seventy days. *J. Dairy Sci.* **89**, 2716-2721, doi:10.3168/jds.S0022-0302(06)72347-5 (2006).

48 Hoff, J. L., Decker, J. E., Schnabel, R. D. & Taylor, J. F. Candidate lethal haplotypes and causal mutations in Angus cattle. *BMC Genomics* **18**, doi:10.1186/s12864-017-4196-2 (2017).

49 VanRaden, P. M., Olson, K. M., Null, D. J. & Hutchison, J. L. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *J. Dairy Sci.* **94**, 6153-6161, doi:10.3168/jds.2011-4624 (2011).

50 Fasquelle, C. *et al.* Balancing selection of a frame-shift mutation in the MRC2 gene accounts for the outbreak of the crooked tail syndrome in Belgian Blue cattle. *PLoS Genet.* **5**, e1000666, doi:10.1371/journal.pgen.1000666 (2009).

51 Zhang, C., MacNeil, M. D., Kemp, R. A., Dyck, M. K. & Plastow, G. S. Putative loci causing early embryonic mortality in Duroc swine. *Front. Genet.* **17**, 655, doi:10.3389/fgene.2018.00655 (2018).

52 Charlier, C. *et al.* NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. *Genome Res.* **26**, 1333-1341, doi:10.1101/gr.207076.116 (2016).

53 Sonstegard, T. S. *et al.* Identification of a nonsense mutation in CWC15 associated with decreased reproductive efficiency in Jersey cattle. *PLOS ONE* **8**, e54872, doi:10.1371/journal.pone.0054872 (2013).

54 Derks, M. F. L. *et al.* A systematic survey to identify lethal recessive variation in highly managed pig populations. *BMC Genomics* **18**, doi:10.1186/s12864-017-4278-1 (2017).

55 Derks, M. F. L. *et al.* Loss of function mutations in essential genes cause embryonic lethality in pigs. *PLoS Genet.* **15**, e1008055, doi:10.1371/journal.pgen.1008055 (2019).

56 Bourneuf, E. *et al.* Rapid discovery of de novo deleterious mutations in cattle enhances the value of livestock as model species. *Sci. Rep.* **7**, 11466, doi:10.1038/s41598-017-11523-3 (2017).

57 Kadri, N. K. *et al.* A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic Red cattle: additional evidence for the common cccurrence of balancing selection in livestock. *PLoS Genet.* **10**, e1004049, doi:10.1371/journal.pgen.1004049 (2014).

58 Derks, M. F. L. *et al.* Balancing selection on a recessive lethal deletion with pleiotropic effects on two neighboring genes in the porcine genome. *PLoS Genet.* **14**, e1007661, doi:10.1371/journal.pgen.1007661 (2018).

59 Crow, J. F. *An introduction to population genetics theory*. (Harper & Row, 1970).

60 Ballou, J. D. Ancestral inbreeding only minimally affects inbreeding depression in mammalian populations. *J. Hered.* **88**, 169-178, doi:10.1093/oxfordjournals.jhered.a023085 (1997).

61 Boakes, E. & Wang, J. A simulation study on detecting purging of inbreeding depression in captive populations. *Genet. Res.* **86**, 139-148, doi:10.1017/S001667230500772X (2005).

62 Benesh, D. P., Weinreich, F., Kalbe, M. & Milinski, M. Lifetime inbreeding depression, purging, and mating system evolution in a simultaneous hermaphrodite tapeworm. *Evolution* **68**, 1762-1774, doi:10.1111/evo.12388 (2014).

63 Boakes, E. H., Wang, J. & Amos, W. An investigation of inbreeding depression and purging in captive pedigreed populations. *Heredity* **98**, 172-182, doi:10.1038/sj.hdy.6800923 (2007).

64 Frankham, R., Gilligan, D. M., Morris, D. & Briscoe, D. A. Inbreeding and extinction: effects of purging. *Conserv. Genet.* **2**, 279-284, doi:10.1023/A:1012299230482 (2001).

65 Norén, K. *et al.* Inbreeding depression in a critically endangered carnivore. *Mol. Ecol.* **25**, 3309-3318, doi:10.1111/mec.13674 (2016).

66 Williams, J. L., Smith, D., Visscher, P. M. & Hall, S. J. G. A viable herd of genetically uniform cattle. *Nature* **409**, 303-303, doi:10.1038/35053160 (2001).

67 Williams, J. L. *et al.* Inbreeding and purging at the genomic level: the Chillingham cattle reveal extensive, non-random SNP heterozygosity. *Anim. Genet.* **47**, 19-27, doi:10.1111/age.12376 (2016).

68 Robinson, J. A., Brown, C., Kim, B. Y., Lohmueller, K. E. & Wayne, R. K. Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. *Curr. Biol.* **28**, 3487-3494.e3484, doi:10.1016/j.cub.2018.08.066 (2018).

69 Laws, R. J. & Jamieson, I. G. Is lack of evidence of inbreeding depression in a threatened New Zealand robin indicative of reduced genetic load? *Anim. Conserv.* **14**, 47-55, doi:10.1111/j.1469-1795.2010.00388.x (2011).

70 Moreno, E., Pérez-González, J., Carranza, J. & Moya-Laraño, J. Better fitness in the Capture Cuvier's gazelle despite inbreeding increase: evidence for purging? *PLOS ONE* **10**, doi:10.1371/journal.pone.0145111 (2015).

71 Mc Parland, S., Kearney, J. F., Rath, M. & Berry, D. P. Inbreeding effects on milk production, calving performance, fertility, and conformation in Irish Holstein-Friesians. *J. Dairy Sci.* **90**, 4411-4419, doi:10.3168/jds.2007-0227 (2007).

72 Weiser, E. L., Grueber, C. E., Kennedy, E. S. & Jamieson, I. G. Unexpected positive and negative effects of continuing inbreeding in one of the world's most inbred wild animals. *Evolution* **70**, 154-166, doi:10.1111/evo.12840 (2016).

73 Kalinowski, S. T., Hedrick, P. W. & Miller, P. S. Inbreeding depression in the Speke's gazelle captive breeding program. *Conserv. Biol.* **14**, 1375-1384, doi:10.1046/j.1523-1739.2000.98209.x (2000).

74 Sonesson, A. K. & Meuwissen, T. H. E. Mating schemes for optimum contribution selection with constrained rates of inbreeding. *Genet. Select. Evol.* **32**, 231-248, doi:10.1186/1297-9686-32-3-231 (2000).

75 Panigrahi, M. & Parida, S. Genomic selection - revolutionary breeding practice in domestic animals. *Vet. World* **5**, 433-436, doi:10.5455/vetworld.2012.433-436 (2012).

76 Henderson, C. R. Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423-447 (1975).

77 James, J. W. & McBride, G. The spread of genes by natural and artificial selection in closed poultry flock. *J. Genet.* **56**, 55-62, doi:10.1007/BF02984720 (1958).

78 Bijma, P. & Woolliams, J. A. Prediction of rates of inbreeding in populations selected on best linear unbiased prediction of breeding value. *Genetics* **156**, 361A-3373 (2000).

79 Boichard, D., Ducrocq, V., Croiseau, P. & Fritz, S. Genomic selection in domestic animals: principles, applications and perspectives. *C. R. Biol.* **339**, 274-277, doi:10.1016/j.crvi.2016.04.007 (2016).

80 Lourenco, D. A. L. *et al.* Are evaluations on young genotyped animals benefiting from the past generations? *J. Dairy Sci.* **97**, 3930-3942, doi:10.3168/jds.2013-7769 (2014).

81 Henderson, C. R. *Applications of linear models in animal breeding*. (University of Guelph, 1984).

82 Gulisija, D., Gianola, D., Weigel, K. A. & Toro, M. A. Between-founder heterogeneity in inbreeding depression for production in Jersey cows. *Livest. Sci.* **104**, 244-253, doi:10.1016/j.livsci.2006.04.007 (2006).

83 Cuyabano, B. C. D., Su, G. S. & Lund, M. S. Genomic prediction of genetic merit using LD-based haplotypes in the Nordic Holstein population. *BMC Genomics* **15**, 1171, doi:10.1186/1471-2164-15-1171 (2014).

84 Masuda, Y. *et al.* Implementation of genomic recursions in single-step genomic best linear unbiased predictor for US Holsteins with a large number of genotyped animals. *J. Dairy Sci.* **99**, 1968-1974, doi:10.3168/jds.2015-10540 (2016).

85 Meuwissen, T., Hayes, B. & Goddard, M. Accelerating improvement of livestock with genomic selection. *Annu. Rev. Anim. BioSci.* **1**, 221-237, doi:10.1146/annurev-animal-031412-103705 (2013).

86 Olsen, H. F., Meuwissen, T. & Klemetsdal, G. Optimal contribution selection applied to the Norwegian and the North-Swedish cold-blooded trotter - a feasibility study: optimal contribution selection in cold-blooded trotters. *J. Anim. Breed. Genet.* **130**, 170-177, doi:10.1111/j.1439-0388.2012.01005.x (2013).

87 Warmuth, V. *et al.* Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 8202-8206, doi:10.1073/pnas.1111122109 (2012).

88 Wallner, B. *et al.* Identification of genetic variation on the horse Y chromosome and the tracing of male founder lineages in modern breeds *PLOS ONE* **8**, e60015, doi:10.1371/journal.pone.0060015 (2013).

89 Fages, A. *et al.* Tracking five millennia of horse management with extensive ancient genome time series. *Cell* **177**, 1419-1435.e1431, doi:10.1016/j.cell.2019.03.049 (2019).

90 Lindgren, G. *et al.* Limited number of patrilines in horse domestication. *Nat. Genet.* **36**, 335-336, doi:10.1038/ng1326 (2004).

91 Achilli, A. *et al.* Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2449-2454, doi:10.1073/pnas.1111637109 (2012).

92 Ludwig, A. *et al.* Twenty-five thousand years of fluctuating selection on leopard complex spotting and congenital night blindness in horse. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20130386, doi:10.1098/rstb.2013.0386 (2015).

93 Petersen, J. L. *et al.* Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLOS ONE* **8**, e54997, doi:10.1371/journal.pone.0054997 (2013).

94 Luís, C., Cothran, E. G. & Oom, M. d. M. Inbreeding and genetic structure in the endangered Sorraia horse breed: implications for its conservation and management. *J. Hered.* **98**, 232-237, doi:10.1093/jhered/esm009 (2007).

95 McCue, M. E. *et al.* A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. *PLoS Genet.* **8**, e1002451, doi:10.1371/journal.pgen.1002451 (2012).

96 Librado, P. *et al.* The evolutionary origin and genetic makeup of domestic horses. *Genetics* **204**, 423-434, doi:10.1534/genetics.116.194860 (2016).

97 Aurich, C. Reproductive cycles of horses. *Anim. Reprod. Sci.* **124**, 220-228, doi:10.1016/j.anireprosci.2011.02.005 (2011).

98 Tortonese, D. J. Intrapituitary mechanisms underlying the control of fertility: key players in seasonal breeding. *Domest. Anim. Endocrinol.* **56**, S191-S203, doi:10.1016/j.domaniend.2016.01.002 (2016).

99 Nolan, M. B. *et al.* Artificially extended photoperiod administered to pre-partum mares via blue light to a single eye: observations on gestation length, foal birth weight and foal hair coat at birth. *Theriogenology* **100**, 126-133, doi:10.1016/j.theriogenology.2017.06.012 (2017).

100 Bene, S., Benedek, Z., Nagy, S., Szabo, F. & Polgar, P. J. Some effects on gestation length of traditional horse breeds in Hungary. *J. Cent. Eur. Agric.* **15**, 1, doi:10.5513/JCEA01/15.1.1402 (2014).

101 Turner, J. W. & Kirkpatrick, J. F. Hormones and reproduction in feral horses. *J. Equine Vet. Sci.* **6**, 250-258, doi:10.1016/S0737-0806(86)80051-X (1986).

102 Beythien, E., Aurich, C., Wulf, M. & Aurich, J. Effects of season on placental, foetal and neonatal development in horses. *Theriogenology* **97**, 98-103, doi:10.1016/j.theriogenology.2017.04.027 (2017).

103 Liu, G. *et al.* Evaluating the reintroduction project of Przewalski's horse in China using genetic and pedigree data. *Biol. Conserv.* **171**, 288-298, doi:10.1016/j.biocon.2013.11.022 (2014).

104 Wolc, A., Bresinska, A. & Szwaczkowski, T. Genetic and permanent environmental variability of twinning in Thoroughbred horses estimated via three threshold models. *J. Anim. Breed. Genet.* **123**, 186, doi:10.1111/j.1439-0388.2006.00575.x (2006).

105 Heck, L., Clauss, M. & Sánchez-Villagra, M. R. Gestation length variation in domesticated horses and its relation to breed and body size diversity. *Mamm. Biol.* **84**, 44-51, doi:10.1016/j.mambio.2017.01.002 (2017).

106 Ewert, M. *et al.* Determinants of gestation length in Thoroughbred mares on German stud farms. *Anim. Reprod. Sci.* **191**, 22-33, doi:10.1016/j.anireprosci.2018.02.001 (2018).

107 Valera, M., Blesa, F., Santos, R. d. & Molina, A. Genetic study of gestation length in Andalusian and Arabian mares. *Anim. Reprod. Sci.* **95**, 75-96, doi:10.1016/j.anireprosci.2005.09.008 (2006).

108 Christmann, A., Sieme, H., Martinsson, G. & Distl, O. Genetic and environmental factors influencing gestation length and parturition conception interval in Hanoverian Warmblood. *Livest. Sci.* **199**, 63-68, doi:10.1016/j.livsci.2017.03.011 (2017).

109 Wulf, M. *et al.* Effects of foal sex on some perinatal characteristics in the immediate neonatal period in the horse. *J. Vet. Behav.* **18**, 37-42, doi:10.1016/j.jveb.2016.12.010 (2017).

110 Van Rijssen, J. D., Gee, E. K., C.W, R. & O'Brien, J. R. Factors influencing gestation length and interval from foaling to conception in mares managed on a commercial Thoroughbred stud farm *Proc. N. Z. Soc. Anim. Prod.* **70**, 146-148 (2010).

111 Langlois, B. & Blouin, C. Genetic parameters for gestation length in French horse breeds. *Livest. Sci.* **146**, 133-139, doi:10.1016/j.livsci.2012.02.028 (2012).

112 Kuhl, J., Stock, K. F., Wulf, M. & Aurich, C. Maternal lineage of Warmblood mares contributes to variation of gestation length and bias of foal sex ratio. *PLOS ONE* **10**, e0139358, doi:10.1371/journal.pone.0139358 (2015).

113 Australian Racing Board. *2015/2016 Australian racing fact book*. (2016).

114 Schrimpf, R. *et al.* Genome-wide association study identifies phospholipase C zeta 1 (PLCZ1) as a stallion fertility locus in Hanoverian Warmblood horses. *PLOS ONE* **9**, doi:10.1371/journal.pone.0109675 (2014).

115 Giesecke, K. *et al.* Evaluation of ACE, SP17, and FSHB as candidates for stallion fertility in Hanoverian Warmblood horses. *Anim. Reprod. Sci.* **126**, 200-206, doi:10.1016/j.anireprosci.2011.05.007 (2011).

116 Giesecke, K. *et al.* Evaluation of SPATA1 -associated markers for stallion fertility. *Anim. Genet.* **40**, 359-365, doi:10.1111/j.1365-2052.2008.01844.x (2009).

117 Giesecke, K., Hamann, H., Sieme, H. & Distl, O. INHBA associated markers as candidates for stallion fertility. *Reprod. Domest. Anim.* **45**, 342-347, doi:10.1111/j.1439-0531.2008.01325.x (2010).

118 Hamann, H. *et al.* A polymorphism within the equine CRISP3 gene is associated with stallion fertility in Hanoverian warmblood horses. *Anim. Genet.* **38**, 259-264, doi:10.1111/j.1365-2052.2007.01594.x (2007).

119 Terje, R. *et al.* Genome-wide association study implicates testis-sperm specific FKBP6 as a susceptibility locus for impaired acrosome reaction in stallions. *PLoS Genet.* **8**, e1003139, doi:10.1371/journal.pgen.1003139 (2012).

120 Janečka, J. E. *et al.* Horse Y chromosome assembly displays unique evolutionary features and putative stallion fertility genes. *Nat. Commun.* **9**, 2945-2945, doi:10.1038/s41467-018-05290-6 (2018).

121 Schrimpf, R. *et al.* Screening of whole genome sequences identified high-impact variants for stallion fertility. *BMC Genomics* **17**, doi:10.1186/s12864-016-2608-3 (2016).

122 Collins, C. W. *et al.* Abnormal reproductive patterns in Przewalski's mares are associated with a loss in gene diversity. *Biol. Reprod.* **86**, doi:10.1095/biolreprod.111.092676 (2012).

123 Kjöllerström, H. J., Gama, L. T. & Oom, M. M. Impact of inbreeding on fitness-related traits in the highly threatened Sorraia horse breed. *Livest. Sci.* **180**, 84-89, doi:10.1016/j.livsci.2015.08.001 (2015).

124 Der Sarkissian, C. *et al.* Evolutionary genomics and conservation of the endangered Przewalski's horse. *Curr. Biol.* **25**, 2577-2583, doi:10.1016/j.cub.2015.08.032 (2015).

125 Wade, C. M. *et al.* Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* **326**, 865-867, doi:10.1126/science.1178158 (2009).

126 Schaefer, R. J. *et al.* Developing a 670k genotyping array to tag ~2M SNPs across 24 horse breeds. *BMC Genomics* **18**, doi:10.1186/s12864-017-3943-8 (2017).

127 Petersen, J. L., Valberg, S. J., Mickelson, J. R. & McCue, M. E. Haplotype diversity in the equine myostatin gene with focus on variants associated with race distance propensity and muscle fibre type proportions. *Anim. Genet.* **45**, 827-835, doi:10.1111/age.12205 (2014).

128 Petersen, J. L. *et al.* Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genet.* **9**, e1003211, doi:10.1371/journal.pgen.1003211 (2013).

129 Dupuis, M. C. *et al.* Results of a haplotype-based GWAS for recurrent laryngeal neuropathy in the horse. *Mamm. Genome* **22**, 613-620, doi:10.1007/s00335-011-9337-3 (2011).

130 Corbin, L. J. *et al.* The utility of low-density genotyping for imputation in the Thoroughbred horse. *Genet. Select. Evol.* **46**, 9-9, doi:10.1186/1297-9686-46-9 (2014).

131 Corbin, L. J. *et al.* Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Anim. Genet.* **41**, 8-15, doi:10.1111/j.1365-2052.2010.02092.x (2010).

132 Corbin, L. J. *et al.* A genome-wide association study of osteochondritis dissecans in the Thoroughbred. *Mamm. Genome* **23**, 294-303, doi:10.1007/s00335-011-9363-1 (2012).

133 Fritz, K. L., McCue, M. E., Valberg, S. J., Rendahl, A. K. & Mickelson, J. R. Genetic mapping of recurrent exertional rhabdomyolysis in a population of North American Thoroughbreds. *Anim. Genet.* **43**, 730-738, doi:10.1111/j.1365-2052.2012.02351.x (2012).

134 Govindarajan, P. *et al.* Genome-wide association study of osteochondrosis in the tarsocrural joint of Dutch Warmblood horses identifies susceptibility loci on chromosomes 3 and 10. *Anim. Genet.* **44**, 408-412, doi:10.1111/age.12016 (2013).

135 Hestand, M. S. *et al.* Annotation of the protein coding regions of the equine genome. *PLOS ONE* **10**, e0124375, doi:10.1371/journal.pone.0124375 (2015).

136 Mansour, T. A. *et al.* Tissue resolved, gene structure refined equine transcriptome. *BMC Genomics* **18**, 103, doi:10.1186/s12864-016-3451-2 (2017).

137 Coleman, S. J. *et al.* Structural annotation of equine protein-coding genes determined by mRNA sequencing. *Anim. Genet.* **41**, 121-130, doi:10.1111/j.1365-2052.2010.02118.x (2010).

138 Gu, J. *et al.* A genome scan for positive selection in Thoroughbred horses. *PLOS ONE* **4**, e5767, doi:10.1371/journal.pone.0005767 (2009).

139 Hill, E. W., Gu, J., McGivney, B. A. & MacHugh, D. E. Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. *Anim. Genet.* **41**, 56-63, doi:10.1111/j.1365-2052.2010.02104.x (2010).

140 Hauswirth, R. *et al.* Mutations in MITF and PAX3 cause "splashed white" and other white spotting phenotypes in horses. *PLoS Genet.* **8**, 404-412, doi:10.1371/journal.pgen.1002653 (2012).

141 McCoy, A. M. *et al.* Evidence of positive selection for a glycogen synthase (GYS1) mutation in domestic horse populations. *J. Hered.* **105**, 163-172, doi:10.1093/jhered/est075 (2014).

142 Metzger, J. *et al.* Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genomics* **16**, 764, doi:10.1186/s12864-015-1977-3 (2015).

143 Tryon, R. C. *et al.* Evaluation of allele frequencies of inherited disease genes in subgroups of American Quarter Horses. *J. Am. Vet. Med. Assoc.* **234**, 120-125, doi:10.2460/javma.234.1.120 (2009).

144 Musiał, A. D., Ropka-Molik, K., Piórkowska, K., Jaworska, J. & Stefaniuk-Szmukier, M. ACTN3 genotype distribution across horses representing different utility types and breeds. *Mol. Biol. Rep.*, doi:10.1007/s11033-019-05013-0 (2019).

145 Pacholewska, A. *et al.* The transcriptome of equine peripheral blood mononuclear cells. *PLOS ONE* **10**, e0122011, doi:10.1371/journal.pone.0122011 (2015).

146 Holl, H. M., Gao, S., Fei, Z., Andrews, C. & Brooks, S. A. Generation of a de novo transcriptome from equine lamellar tissue. *BMC Genomics* **16**, 739, doi:10.1186/s12864-015-1948-8 (2015).

147 Smits, K. *et al.* The equine embryo influences immune-related gene expression in the oviduct. *Biol. Reprod.* **94**, 36, doi:10.1095/biolreprod.115.136432 (2016).

148 Tozaki, T. *et al.* Profiling of exercise-induced transcripts in the peripheral blood cells of Thoroughbred horses. *J. Equine Sci.* **27**, 157-164, doi:10.1294/jes.27.157 (2016).

149 McGivney, B. A. *et al.* Characterization of the equine skeletal muscle transcriptome identifies novel functional responses to exercise training. *BMC Genomics* **11**, 398-398, doi:10.1186/1471-2164-11-398 (2010).

150 Barrey, E. *et al.* Transcriptome analysis of muscle in horses suffering from recurrent exertional rhabdomyolysis revealed energetic pathway alterations and disruption in the cytosolic calcium regulation. *Anim. Genet.* **43**, 271-281, doi:10.1111/j.1365-2052.2011.02246.x (2012).

151 Bellone, R. R. *et al.* Evidence for a retroviral insertion in TRPM1 as the cause of congenital stationary night blindness and leopard complex spotting in the horse. *PLOS ONE* **8**, e78280, doi:10.1371/journal.pone.0078280 (2013).

152 Tesio, F. *Breeding the racehorse*. Vol. 1 (J.A.Allen, 1958).

153 Weatherby and Sons. *An Introduction to the General Stud Book*. (Weatherby and Sons, 1791).

154 Ford, M. *The Australian Stud Book History*. Vol. 1 (The Australian Stud Book, 2006).

155 Cunningham, E. P., Dooley, J. J., Splan, R. K. & Bradley, D. G. Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to Throughbred horses. *Anim. Genet.* **32**, 360, doi:https://doi.org/10.1046/j.1365-2052.2001.00785.x (2001).

156 Sabine, F. *et al.* The horse Y chromosome as an informative marker for tracing sire lines. *Sci. Rep.* **9**, 1-12, doi:10.1038/s41598-019-42640-w (2019).

157 Bower, M. A. *et al.* Thoroughbred racehorse mitochondrial DNA demonstrates closer than expected links between maternal genetic history and pedigree records. *J. Anim. Breed. Genet.* **130**, 227-235, doi:10.1111/j.1439-0388.2012.01018.x (2013).

158 Bower, M. A. *et al.* The cosmopolitan maternal heritage of the Thoroughbred racehorse breed shows a significant contribution from British and Irish native mares. *Biol. Lett.* **7**, 316-320, doi:10.1098/rsbl.2010.0800 (2011).

159 Cassidy, R. in *The sport of kings: kinship, class and thoroughbred breeding in Newmarket* 140-157 (Cambridge University Press, 2002).

160 Bower, M. A. *et al.* Truth in the bones: resolving the identity of the founding elite Thoroughbred racehorses. *Archaeometry* **54**, 916-925, doi:10.1111/j.1475-4754.2012.00666.x (2012).

161 Bower, M. A. *et al.* The genetic origin and history of speed in the Thoroughbred racehorse. *Nat. Commun.* **3**, 643, doi:10.1038/ncomms1644 (2012).

162 Thomas, K. C., Hamilton, N. A., North, K. N. & Houweling, P. J. Sequence analysis of the equine ACTN3 gene in Australian horse breeds. *Gene* **538**, 88-93, doi:10.1016/j.gene.2014.01.014 (2014).

163 Hori, Y. *et al.* Evidence for the effect of serotonin receptor 1A gene (HTR1A) polymorphism on tractability in Thoroughbred horses. *Anim. Genet.* **47**, 62-67, doi:10.1111/age.12384 (2016).

164 McGivney, B. A. *et al.* A genomic prediction model for racecourse starts in the Thoroughbred horse. *Anim. Genet.* **50**, 347-357, doi:10.1111/age.12798 (2019).

165 Bosh, K. A., Powell, D., Shelton, B. & Zent, W. Reproductive performance measures among Thoroughbred mares in central Kentucky, during the 2004 mating season. *Equine Vet. J.* **41**, 883-888, doi:10.2746/042516409X456068 (2009).

166 Allen, W. R. & Wilsher, S. Half a century of equine reproduction research and application: A veterinary tour de force. *Equine Vet. J.* **50**, 10-21, doi:10.1111/evj.12762 (2018).

167 López-Plana, C., Sautet, J. Y., Pons, J. & Navarro, G. Morphometric study of the recurrent laryngeal nerve in young 'normal' horses. *Res. Vet. Sci.* **55**, 333-337, doi:10.1016/0034-5288(93)90103-M (1993).

168 Raffetto, J. A., Wearn, J. G. & Fischer, A. T. Racing performance following prosthetic laryngoplasty using a polyurethane prosthesis combined with a laser-assisted ventriculocordectomy for treatment of recurrent laryngeal neuropathy in 78 Thoroughbred racehorses. *Equine Vet. J.* **47**, 60-64, doi:10.1111/evj.12237 (2015).

169 Boyko, A. R. *et al.* Genomic analysis establishes correlation between growth and laryngeal neuropathy in Thoroughbreds. *BMC Genomics* **15**, 259-259, doi:10.1186/1471-2164-15-259 (2014).

170 Brooks, S. A. *et al.* Identification of loci affecting sexually dimorphic patterns for height and recurrent laryngeal neuropathy risk in American Belgian Draft horses. *Physiol. Genomics* **50**, 1051, doi:10.1152/physiolgenomics.00068.2018 (2018).

171 Distl, O. The genetics of equine osteochondrosis. *Vet. J.* **197**, 13-18, doi:10.1016/j.tvjl.2013.03.036 (2013).

172 Clarke, K. L., Reardon, R. & Russell, T. Treatment of osteochondrosis dissecans in the stifle and tarsus of juvenile Thoroughbred horses. *Vet. Surg.* **44**, 297-303, doi:10.1111/j.1532-950X.2014.12277.x (2015).

173 Welsh, C. E. *et al.* Preliminary genetic analyses of important musculoskeletal conditions of Thoroughbred racehorses in Hong Kong. *Vet. J.* **198**, 611-615, doi:10.1016/j.tvjl.2013.05.002 (2013).

174 Naccache, F., Metzger, J. & Distl, O. Genetic risk factors for osteochondrosis in various horse breeds. *Equine Vet. J.* **50**, 556-563, doi:10.1111/evj.12824 (2018).

175 Hilla, D. & Distl, O. Heritabilities and genetic correlations between fetlock, hock and stifle osteochondrosis and fetlock osteochondral fragments in Hanoverian Warmblood horses. *J. Anim. Breed. Genet.* **131**, 71-81, doi:10.1111/jbg.12062 (2014).

176 Jonsson, L. *et al.* Equine hospital data as a source for study of prevalence and heritability of osteochondrosis and palmar/plantar osseous fragments of Swedish Warmblood horses. *Equine Vet. J.* **43**, 695-700, doi:10.1111/j.2042-3306.2010.00354.x (2011).

177 Russell, J., Matika, O., Russell, T. & Reardon, R. J. M. Heritability and prevalence of selected osteochondrosis lesions in yearling Thoroughbred horses. *Equine Vet. J.* **49**, 282-287, doi:10.1111/evj.12613 (2017).

178 Lampe, V., Dierks, C. & Distl, O. Refinement of a quantitative gene locus on equine chromosome 16 responsible for osteochondrosis in Hanoverian Warmblood horses. *animal* **3**, 1224-1231, doi:10.1017/S1751731109004765 (2009).

179 Teyssedre, S. *et al.* Genome-wide association studies for osteochondrosis in French Trotter horses. *J. Anim. Sci.* **90**, 45-53, doi:10.2527/jas.2011-4031 (2012).

180 Blott, S. C. *et al.* A genome-wide association study demonstrates significant genetic variation for fracture risk in Thoroughbred racehorses. *BMC Genomics* **15**, 147-147, doi:10.1186/1471-2164-15-147 (2014).

181 Langford, J. L., Thomson, P. C. & Knight, P. K. Epistaxis in racehorses: risk factors and effects on career. *Aust. Vet. J.* **91**, 198-203, doi:10.1111/avj.12049 (2013).

182 Racing Australia. *The Australian rules of racing*. (Racing Australia, 2020).

183 Velie, B. D., Raadsma, H. W., Wade, C. M., Knight, P. K. & Hamilton, N. A. Heritability of epistaxis in the Australian Thoroughbred racehorse population. *Vet. J.* **202**, 274-278, doi:10.1016/j.tvjl.2014.06.010 (2014).

184 Boudreaux, M. K., Koehler, J., Habecker, P. L. & Piero, F. D. Evaluation of the genes encoding CD39/NTPDase-1 and CD39L1/NTPDase-2 in horses with and without abnormal hemorrhage and in horses with pathologic evidence of exercise-induced pulmonary hemorrhage. *Vet. Clin. Pathol.* **44**, 617-625, doi:10.1111/vcp.12303 (2015).

Chapter 2: The effects of inbreeding and selection on racing performance in the Thoroughbred horse population

2.1 Synopsis

This chapter consists of a published manuscript and associated supplementary material:

E.T. Todd, S.Y.W. Ho, P.C. Thomson, R.A. Ang, B.D. Velie & N.A. Hamilton. Founder-specific inbreeding depression affects racing performance in Thoroughbred horses. *Scientific Reports* **8**, 6167, doi:10.1038/s41598-018-24663-x (2018).

This publication has been formatted for consistency with this thesis. The original manuscript is available in Appendix 2.

This chapter investigates the effects of inbreeding for the selection of racing performance in Thoroughbred horses. This involved constructing linear mixed models to analyse racing performance data from 135,572 horses and their associated pedigree dating back to the founders of the population (*n* = 257,249). Analyses in this chapter show that higher levels of inbreeding were associated with decreased racing performance in Australian Thoroughbred horses. However, there is evidence of purging and selection for favourable alleles in the population over time which is associated with improved racing performance. It was also found that the closed population structure has resulted in some ancestors making large genetic contributions to contemporary Thoroughbred horses. Inbreeding attributed to these individuals has variable effects on the racing performance of their descendants, providing evidence that the effects of inbreeding can vary between individuals.

I designed this project and conducted the analyses using a number of population genetics software programs and custom scripts in R and Perl. Dr Brandon Velie assisted in cleaning and restructuring the raw data. Dr Natasha Hamilton and Ms Rachel Ang assisted in the collection and genotyping of the DNA samples. Professor Simon Ho, Dr Natasha Hamilton and Dr Peter Thomson provided guidance with the overall design of the project and in finalizing the manuscript.

2.2 Main article

Founder-specific inbreeding depression affects racing performance in Thoroughbred horses.

Evelyn T. Todd[1,*], Simon Y.W. Ho[1], Peter C. Thomson[1], Rachel A. Ang[1], Brandon D. Velie[2] and Natasha A. Hamilton[1].

[1]School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia

[2]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, 75007 Uppsala, Sweden.

Abstract

The Thoroughbred horse has played an important role in both sporting and economic aspects of society since the establishment of the breed in the 1700s. The extensive pedigree and phenotypic information available for the Thoroughbred horse population provides a unique opportunity to examine the effects of 300 years of selective breeding on genetic load. By analysing the relationship between inbreeding and racing performance of 135,572 individuals, we found that selective breeding has not efficiently alleviated the Australian Thoroughbred population of its genetic load. However, we found evidence for purging in the population that might have improved racing performance over time. Over 80% of inbreeding in the contemporary population is accounted for by a small number of ancestors from the foundation of the breed. Inbreeding to these ancestors has variable effects on fitness, demonstrating that an understanding of the distribution of genetic load is important in improving the phenotypic value of a population in the future. Our findings hold value not only for Thoroughbred and other domestic breeds, but also for small and endangered populations where such comprehensive information is not available.

Introduction

The Thoroughbred horse population is one of the largest closed populations of animals in the world. Thoroughbreds are extremely valuable because of the large amount of prizemoney on offer and the high residual value of superior athletes. All Thoroughbred horses trace their ancestry back to three paternal lines, due to the narrow bottleneck at the foundation of the population[1-3]. More than 300 years of breeding practices have produced signatures of selection in the 21st century Thoroughbred population, contributing to the superior athleticism of the breed[4,5]. At the same time, these practices have increased

levels of inbreeding and reduced the genetic diversity of Thoroughbreds compared with other domestic horse breeds[3,6,7].

To our knowledge, there has been no detailed examination of the effects of inbreeding on the racing performance of Thoroughbred horses and the genetic load of the population. Genetic load, the presence of unfavourable genetic material, is a reflection of a population's fitness because a higher genetic load leads to a lower mean fitness level[8]. A large proportion of genetic load consists of recessive deleterious mutations, known as mutational load. Inbreeding can expose mutational load because it increases an individual's chance of inheriting two copies of recessive deleterious alleles from a common ancestor[8,9]. The subsequent decrease in fitness caused by these expressed recessive deleterious mutations is thought to be a major cause of inbreeding depression[10]. Other mechanisms believed to contribute to inbreeding depression include epistatic interactions and reductions in favourable heterozygosity[10,11].

The inevitable effect of selection in a closed population is an increase in the level of inbreeding[12,13]. There is some evidence that continued inbreeding for selection can purge a population of some or all of its genetic load, such that new inbreeding events have negligible or even positive effects on phenotype[9]. Although some domestic and wild populations show signs of purging[14-16], others still show strong signs of inbreeding depression even after multiple population bottlenecks and inbreeding events[17-19]. Purging is most likely to occur in populations under strong selection and slow rates of inbreeding, allowing deleterious alleles to be effectively eliminated rather than fixed by genetic drift[11,20]. Additionally, inbreeding for favourable phenotypic characteristics can have unexpected negative implications through deleterious alleles hitchhiking on regions of the genome under positive selection, thereby increasing their frequency in the population[21-23].

Understanding the effects of selection is further complicated by the uneven distribution of genetic load in a population. Inbreeding to different ancestors can have varying effects on fitness, such that the total proportion of alleles identical by descent (IBD) might not be an accurate reflection of mutational load[24-26]. This raises the possibility that inbreeding in different pedigree lines has variable effects on genetic load in the Thoroughbred population.

The availability of extensive phenotypic and pedigree records, dating back to the late 18th century, makes the Thoroughbred population ideal for studying the long-term, population-wide effects of selection on performance and genetic load. Here, we examine the effects of inbreeding on racing performance and mutational load in the Australian Thoroughbred population. Australia has the second-

largest racing and breeding population in the world, containing approximately 15% of all Thoroughbreds[27].

We analyse a sample of 135,572 individuals, representing all Thoroughbred horses that had one or more race starts in Australia between 2000 and 2011. A genealogy of these individuals, dating back to the founders of the population (*n* = 257,249), is also included in our analyses. Although some lines of pedigree are incomplete, we have comprehensive pedigree information for all individuals in the racing performance data set, making our inbreeding estimates highly accurate. The availability of extensive pedigree records not only allows us to study broad population trends over time, but also to determine whether the selection for optimal racing performance has alleviated mutational load. We use these data to measure inbreeding and ancestral coefficients for all individuals. We also identify the ancestors that have made the greatest genetic contributions, in order to understand better the distribution of mutational load in the population. For a representative subset of individuals, we perform high-density genotyping to determine whether inbreeding load is reflected at the genomic level.

Results and Discussion

The effects of inbreeding and purging on racing performance

Our analysis of data from 135,572 Thoroughbred horses revealed a strong negative relationship (all $P<0.001$, Figure 1) between Wright's inbreeding coefficient, *F*, and five measures of racing performance that encompass a range of factors that contribute to exercise performance[28,29]. These included two measures that are based on the assumption that more successful individuals earn more prizemoney: cumulative prizemoney earnings and prizemoney earnings per start. We also included two measures of constitutional soundness: total number of race starts and career length. Finally, we accounted for consistency of performance with the measure winning strike rate.

The negative relationship between *F* and performance can be explained by a genetic load of partially deleterious alleles still being carried by the population. We expect that the alleles causing the observed inbreeding depression are more difficult to select out of the population than those with lethal or debilitating effects on juvenile or embryonic survival[10,21,30-32]. Population bottlenecks that occurred during the ancestry of the Thoroughbred, including the domestication of the horse[33], and the foundation of the breed[2,3], might have increased the frequency of deleterious alleles through genetic drift. It is also possible that continued inbreeding of the Thoroughbred population over the past 300 years has

inadvertently increased the frequency of deleterious variants in the population, potentially through hitchhiking on selective sweep regions[13,21,23]. As a result of many generations of inbreeding, the average *F* of the 21st century Thoroughbred population is 0.139 (*s* = 0.011).

In contrast with the results from Wright's inbreeding coefficient, the ancestral history coefficient, A_{HC}, showed a strong positive association with racing performance (all *P*<0.001, Figure1). This statistic, described by Baumung, et al.[34], counts the number of times that an allele has been IBD in an individual's pedigree, thus providing a comprehensive reflection of selection for favourable traits over time. The A_{HC} statistic is based on the assumption that an allele that has been IBD multiple times in an individual's pedigree is likely to have a neutral or positive effect on fitness. In contrast, an allele that is IBD for the first time is more likely to have a negative effect on fitness. Therefore, individuals with higher A_{HC} are more likely to contain larger proportions of alleles in their genomes that have been positively selected over many generations. It is possible for an individual with a comprehensive pedigree to have an A_{HC} greater than 1. As a consequence of the comprehensive and inbred pedigree, the reference population had average A_{HC} of 1.973 (*s* = 0.089).

The positive relationship between A_{HC} and all measures of racing performance is possibly due to the many generations of selective breeding that have increased the frequency of alleles associated with positive improvements with exercise physiology. These alleles will appear IBD more times in the pedigrees of each subsequent generation, thus driving up A_{HC} (Appendix S2). Our results indicate that inbreeding for selection has effectively increased the frequencies of favourable alleles, but has not completely eliminated genetic load from the population. Considering this finding, it is unsurprising that parts of the Thoroughbred genome show signatures of selective sweeps linked to genes related to athletic performance, including formation of muscular fibres, upregulation of mitochondrial activity, angiogenesis, brown adipose tissue formation, and lipid metabolism[5,35]. In agreement with our results, there is some evidence for selection improving racing performance in another horse breed, the Norwegian cold-blooded trotter[32].

Both *F* and A_{HC} showed the strongest associations with cumulative earnings and earnings per start (Figure 1). We expect that this is because these measures reflect not only talent, but also good constitution because horses that race more are more likely to win more prizemoney. The smallest regression coefficient was for winning strike rate, probably because this measure is a crude estimate of

consistency and does not reflect the race class, or the finishing order of a horse on non-winning occasions.

The estimated breeding values of the population over time

We found that selective breeding practices have not increased the overall performance levels of the population over time. We implemented a numerator relationship matrix in conjunction with a linear mixed model to account for additive genetic relationships between animals in the pedigree (Materials and Methods). Based on the racing performance of contemporary individuals (*n*=135,572), we used this relationship matrix to calculate the estimated breeding values (EBVs) of all individuals in their pedigree (*n*=257,249)[36,37]. The large increase in EBVs at the foundation of the population indicates that early selection events resulted in an initial jump in the frequency of favourable alleles (Figure 2). After this initial increase, the distribution of EBVs remains constant; demonstrating that selective breeding from the early 19th century was not effective in improving the racing performance of the population. The level of *F* has increased constantly during this time (Figure S8), so we conclude that inbreeding has not effectively removed mutational load from the population. This explains why we observe strong inbreeding depression persisting in the contemporary population. We expect that this is due in part to a change in racing and training regimes over time that, in turn, has changed selection pressures on the population[38]. In the 18th and early 19th century, Thoroughbred races were held over a distance of several miles, with each horse participating in multiple heats on the same day. In the 20th century, focus shifted to breeding sprinters and early developers for two-year-old racing[39]. Similarly, there was very little increase over time in the EBVs of Polish Warmblood horses despite selection for performance, indicating that intensive selection might be necessary to improve the mean value of complex quantitative traits in a population[40].

The dip in EBVs between 1930 and 1980 can also be partly attributed to an increased number of individuals with unknown pedigree information, as shown in red on Figure 2. This, together with the increased variability of EBVs during this period, could also be due to the presence of less successful pedigree lines that have not been purged from the modern population. We expect that the increase in the average EBV from 1980 onwards is partly due to the introduction of parental testing in the 1980s, leading to complete pedigrees for all registered individuals. The increasing trend in EBVs over recent generations indicates a possibility for future improvement in the population's overall phenotypic quality.

The uneven ancestral genetic contribution in the contemporary Thoroughbred population

Selective breeding practices are likely to result in uneven ancestral genetic contributions, favouring ancestors carrying beneficial alleles and leading to the extinction of less successful ancestral lines[25,41,42]. We found that a small number of ancestors in the early years of the breed formation accounted for much of the inbreeding coefficient in the modern Australian Thoroughbred population.

We found that 10 ancestors accounted for, on average, over 80% of the IBD alleles in the modern Australian Thoroughbreds (Table 1). Almost 20% of the IBD alleles in the contemporary population were attributed to a single individual, Herod. We selected these 10 ancestors because they provided the greatest marginal contributions to the individuals in our racing performance data (Appendix S4). The greatest marginal contributors are selected by first identifying the single ancestor with the greatest contribution to the population, and then subsequently finding the other ancestors that provide the greatest genetic contributions not accounted for by previously selected ancestors[43] (Appendix S4). We then estimated the proportion of F (pF_i) and A_{HC} (pA_{HCi}) for each individual in our data set that is attributed to each of these ancestors[34,44].

We identified these individuals as superior athletes that were also highly successful at stud. Historical records show that most of these individuals are closely related to each other (Figure S7). One of them, Godolphin Barb, was one of the three foundation stallions of the breed in the early 18th century[1]. He has been reported to contribute to 13.8% of the genetic makeup of British Thoroughbred horses[3]. Another of the foundation stallions, Eclipse, was identified as the source of a Y chromosome mutation that is near fixation in the modern Thoroughbred population[45].

The 10 notable ancestors accounted for over 82% of the A_{HC} coefficient in their modern descendants (Table 1). We expected this relationship because these individuals appear many generations back in the pedigree of modern horses. For alleles inherited from them to have such a large contribution to F, they must appear IBD many times in the pedigrees of their descendants. In concordance with the principle of the A_{HC} coefficient, alleles that are found IBD multiple times in the pedigree are likely to have neutral or beneficial effects on fitness. These findings are reflected in the positive trends in F and A_{HC} over time in the population (Figure S8).

Uneven distribution of genetic load between different ancestors

We found evidence that founder-specific inbreeding depression differentially affects racing performance in the Australian Thoroughbred population (Figure 3). We determined the distribution of genetic load between the 10 dominant ancestors by using linear mixed models to examine the relationship between partial inbreeding coefficients and racing performance. Genetic load may be unevenly distributed between different ancestors, such that inbreeding to different individuals can have a variable effect on fitness[24-26]. If inbreeding to a particular ancestor results in a reduction in the racing performance of their descendants, a higher proportion of the genetic load in the population can be attributed to them. The variation in genetic load between different ancestors indicates that inbreeding depression in Thoroughbreds is due to a small number of loci that have large effects on performance[24,25,46].

We found that inbreeding resulting from four ancestors had significant effects on racing performance. Individuals with more IBD alleles attributed to Herod had greater cumulative earnings, earnings per start, and career length. This does not mean that increased inbreeding to Herod has had no negative effects on the phenotypic value of his descendants, but that overall they exhibit less inbreeding depression than other, equally inbred individuals[25]. Conversely, inbreeding to Eclipse, Stockwell, and Touchstone had negative effects on the racing performance of their descendants. We propose that these negative effects are partly due to the "cost of domestication"[23], whereby inbreeding these individuals has inadvertently selected for deleterious alleles linked to sites that have undergone selective sweeps[21,22].

Additionally, historical reports describe these stallions to be potential carriers of disease alleles, which may have predisposed their descendants to common conditions that reduce racing performance. Touchstone was reported by his contemporaries to have a number of conformational and behavioural issues[47], which might also have contributed to the reduced level of performance in his descendants. Although Eclipse was a superior racehorse, his grandsire suffered from exercise-induced pulmonary haemorrhage (bleeding from the lungs)[48]. This hereditary condition reduces racing success[49,50], and recurrent episodes result in a horse's permanent ban from racing in Australia. Inheritance of this condition might be a contributing factor to the reduced career lengths of Eclipse's descendants. Individuals with higher levels of inbreeding to Stockwell show reduced winning strike rates, although this might be a statistical abnormality because P=0.04. However, Stockwell's mother suffered from the

congenital condition of laryngeal neuropathy (paralysis of the larynx)[51-53], which may partly explain the observed reduction in performance.

We expect that most of these ancestors have passed on a mix of alleles with both positive and negative effects, such that inbreeding to them has the same effect as inbreeding to other individuals in the pedigree. However, it is also possible that two individuals inbred to the same ancestor could have inherited different sets of loci from different ancestral paths, making this ancestor's effect on fitness variable between different descendants[46]. An instance of this has been found in cattle, where the occurrence of ectodermal dysplasia in a number of calves from unaffected parents was traced back to a *de novo* mutation in one bull[54]. The condition was only revealed through inbreeding of his descendants, when some of their progeny inherited two copies of the disease allele. This example demonstrates that inbreeding to a particular individual can have highly variable effects on fitness levels between their different descendants.

Considering the strong evidence for an uneven distribution in genetic load, we conclude that the majority of inbreeding depression is only due to small proportion of IBD alleles[25,42]. Consequently, we suggest that simply measuring the proportion of IBD alleles in the genome does not provide a comprehensive reflection of a population's genetic load. Understanding the heterogeneous distribution of genetic load is important in assisting breeding decisions to minimize inbreeding to ancestors that negatively affect fitness[42].

Relationship between genome-based inbreeding coefficients and racing performance

In contrast with the pedigree-based estimates of inbreeding, we found that genomic measures of inbreeding showed no overall relationship with any measure of racing performance (Table 2). For a representative subset of the population (n=122), we estimated genomic inbreeding levels as the proportion of the genome consisting of runs of homozygosity (F_{ROH}). This method reflects inbreeding levels by capturing long, homologous tracts of DNA inherited from a common ancestor that have not been broken by recombination[55-59]. For our analyses, we selected minimal length thresholds of 5 Mb (F_{ROH_5}) and 12 Mb (F_{ROH_12}) to correspond to old and new inbreeding, respectively (Appendix S2).

For this smaller data set, however, we also found that the F and A_{HC} coefficients for these individuals also showed no relationship with performance (Table 2). Considering that this relationship was significant for a larger sample size, we conclude that a sample size of 122 was not sufficient to capture

the relationship between inbreeding and performance. Our models were unable to account for a number of confounding environmental factors that could affect racing performance (such as training regime, jockey success, and foal-rearing process), so a large sample size is needed to tease out the underlying relationship between inbreeding and performance. There is also a large continuum between the best- and worst-performing individuals in such a large population that might not be captured by a small subset of individuals. Our findings indicate that caution should be exercised in studies of smaller populations.

Molecular estimates of inbreeding are often considered to be superior to genealogical measures because they account for the unpredictable nature of recombination and inaccurate pedigree-recording information. However, the parameters of F_{ROH} measurements should also be chosen carefully, so that they accurately reflect inbreeding levels. The accuracy of these estimations might be affected by inadequate SNP density[56,60] and long tracts of ROH persisting in areas of low recombination[61,62] (Appendix S2). Many studies use different parameters for genotyping densities, data trimming, and ROH, making comparisons between them difficult to draw.

We found that the correlation between F_{ROH} and F in our data set (Figure S3) was lower than that reported in other domestic species[63,64], which may partly explain the contrasting results. We found that a large proportion of the inbreeding coefficient in the Australian Thoroughbred population was accounted for by ancestors many generations back in the pedigree. Inbreeding to distant ancestors results in shorter ROH regions that might not be captured by the SNP density used in our analysis (Appendix S2).

For these reasons, we believe that for large populations with comprehensive pedigrees, genealogical measures of inbreeding can provide important inferences if the size of the pedigree is much larger than the number of individuals genotyped. The use of pedigree data allows inferences to be made for deceased individuals, for which genotyping might not be possible. Additionally, using a pedigree to analyse trends over time can be advantageous because it might not be possible to obtain molecular data for deceased individuals (such as the founders of the population). Pedigrees also provide the opportunity to estimate the effects of specific individuals over time on the fitness of their descendants.

Conclusions

In this study, we have presented the effects of inbreeding and selection in a very large population with extensive phenotypic and pedigree records. Our analyses have shown that genetic load can still persist in a population even after many generations of inbreeding. However, we have also found evidence that multiple generations of inbreeding for selection can have positive effects on the overall genetic value of a population. We suggest that using EBVs whilst managing inbreeding levels will increase the efficiency of selection to reduce inbreeding depression in subsequent generations. Further, our findings highlight the need for caution in studies with small sample sizes because they can lead to inaccurate inferences about the effects of inbreeding.

We have also found evidence that the genetic load is unevenly distributed in the Thoroughbred population. This indicates that studies of inbreeding need to account for heterogeneity between different ancestors, because the total proportion of IBD alleles might not accurately reflect genetic load. Understanding the distribution of genetic load in the population will assist in breeding decisions to reduce disease alleles and improve the overall fitness of the population in future generations. Our findings open the possibility of evaluating the effects of particular individuals on the fitness of the population in order to improve phenotypic quality and reduce genetic load in the future.

Materials and Methods

Calculating pedigree-based inbreeding coefficients

Racing Australia provided race records for all individuals that had participated in a race start in Australia between 2000 and 2010 (*n*=135,572). A genealogy of all horses born after 1970, dating back to the founders of the population, was provided by the Australian Stud Book (*n*=500,477) (Appendix S1). We trimmed the pedigree file so that it only included the ancestors of the individuals in our data set, leaving a pedigree size of 257,249. We found that all individuals included in our analysis had a comprehensively recorded pedigree (an average of 24.60 discrete generational equivalents of known pedigree[65,66]). Before 1980, however, a small number of individuals appear in the stud book with no recorded pedigree[67,68] (Appendix S6). These individuals accounted for 1.4% of the total ancestors included in our genealogy file, and mostly appear more than 6 generations back in the pedigree.

We estimated inbreeding levels for all individuals in the data set using Wright's inbreeding coefficient (F)[69]. We used this traditional measure of quantifying inbreeding to allow our results to be compared with those from previous studies. We also used the pedigree data to estimate several ancestral inbreeding coefficients that account for genetic load (SI Materials and Methods, Appendix S1, S2)[18,34,70]. We selected the ancestral history coefficient (A_{HC}) for further analysis because this measure counts the number of times that an allele has been IBD in an individual's pedigree, thus providing a comprehensive reflection of selection for favourable traits over time[34].

We calculated F and A_{HC} for all individuals in the pedigree using 10^6 replications of simulated gene drops in GRain 1.0[34] (Appendix S1). This method uses Mendelian segregation rules to simulate gene flow through a population by flagging each allele as it runs through the pedigree. These data are then used to estimate the probability-based inbreeding coefficients[71].The accuracy of the results depends on the number of replications performed, which is proportional to the number of unlinked loci calculated in the analysis[34]. We checked the accuracy of our output by comparing F estimations using GRain with a deterministic approach as implemented by PEDIG[66]. Estimates from the two methods had a correlation coefficient of 0.99, indicating high accuracy of the inbreeding estimations by GRain.

We identified the 20 ancestors that provided the greatest marginal contributions to the population of 135,572 individuals by using iterations in the *prog_orig.f* program in PEDIG[43,66]. We then used GRain to calculate pF and pA_{HC} of each ancestor for each individual in our data set. Ten ancestors were chosen for further analysis, and their identities were determined using the Australian Stud Book and the online pedigree database (pedigreequery.com).

Estimating inbreeding from genomic data

We selected a representative subset of individuals for high-density genotyping (n=128). These individuals were selected to provide a reflection of different bloodlines in the population and a continuum of racing successes. We used these data to estimate the proportion of the genome consisting of runs of homozygosity (F_{ROH}).

To estimate genome-based levels of inbreeding, we first extracted DNA from hair samples (collected under approval from University of Sydney Ethics Committee N00-2009-3-5109) using the Qiagen Gentra® Puregene® Tissue Kit (Qiagen, Redwood City, CA, USA). We genotyped 105 individuals using the Equine SNP70 BeadChip (Illumina, San Diego, CA, USA), which consists of 65,102 SNPs evenly distributed

throughout the equine genome. Additionally, we typed 23 individuals on the Axiom Affymetrix SNP Chip (670,671 SNPs), when this higher density array became available at a later date. We used custom Perl scripts to extract only the SNPs that were common to these two panels, which we then used in further analyses.

SNP data were edited and analysed using PLINK 1.07[72]. The data were trimmed to be in concordance using the following parameters: minor allele frequency >0.01; individual call rate >0.9; and SNP call rate >0.9[6,73]. This process yielded a final data set comprising 45,451 SNPs for each of 122 individuals. Additionally, we only analysed autosomal SNPs in order to exclude any bias between male and female.

We used these data to estimate the proportion of the genome consisting of runs of homozygosity (F_{ROH}). To define the parameters of an ROH, we set the minimum density to 0.05 Mb/SNP and the largest gap to 1 Mb, in accordance with the settings used by Goddard, et al.[74] and Silió, et al.[63]. We set the minimal number of SNPs in each ROH to 20, because our SNP coverage was approximately 1 SNP every 50 Mb, making this sufficient to distinguish an ROH of 1 Mb. ROH lengths were calculated as a proportion of total ROH length in relation to the total equine autosome size of 2,242,879,462 bp.

Measuring racing performance

We selected five different measures of racing performance that account for talent, consistency, and constitutional soundness[28,29,75]. These measures were: cumulative earnings ($AU), earnings per start ($AU), career length (months), total number of race starts, and winning strike rate. Cumulative earnings and earnings per start favour talented individuals, because more prestigious races carry larger prizemoney purses. Career length and total starts favour individuals with good constitutions; individuals with health and conformational defects are unable to race for extended periods. Winning strike rate accounts for consistency in horses, because more talented horses are expected to win a higher proportion of their race starts (Appendix S3).

Statistical analysis

The relationship between each measure of inbreeding and racing performance was analysed using (generalized) linear mixed models in ASReml-R 3.0[76]. We used the five measures of racing performance as outcome variables. Cumulative earnings, earnings per start and career length were analysed with linear mixed models. These variables were log-transformed; to accommodate zero-value in these

measure, $100 was added to all career earnings and earnings per starts, and 1 month to all career length values. Total starts was analysed using a Poisson generalized linear mixed model and winning strike rate using a binomial generalized linear mixed model.

Each model included a predictor variable of either *F*, A_{HC}, or *F* partitioned into partial coefficients for each of the 10 important ancestors, making a total of 15 models. Sex and year of birth were also included as predictor variables in each model. We also included a random animal effect that was associated with the numerator relationship matrix derived from the pedigree (n=257, 249).

The significance of fixed effects was assessed using Wald tests. To allow comparisons of regression coefficients across different traits, the regression coefficients were divided by the standard deviation of their respective traits. EBVs were obtained from the fitted models. We summarised the EBV distributions over time in 10-year bins, which approximately represents one generation interval. We calculated the average generational interval to be 10.5 years using the *intgen.f* program from PEDIG[66].

References

1 Weatherby and Sons. *An Introduction to the General Stud Book*. (Weatherby and Sons, 1791).

2 Corbin, L. J. *et al.* Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Anim. Genet.* **41**, 8-15 (2010).

3 Cunningham, E. P., Dooley, J. J., Splan, R. K. & Bradley, D. G. Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to Throughbred horses. *Anim. Genet.* **32**, 360 (2001).

4 Petersen, J. L. *et al.* Genome-wide analysis reveals selection for important traits in domestic Horse Breeds. *PLoS Genet.* **9**, e1003211 (2013).

5 Gu, J. *et al.* A genome scan for positive selection in Thoroughbred horses. *PLOS ONE* **4**, e5767 (2009).

6 Petersen, J. L. *et al.* Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLOS ONE* **8**, e54997 (2013).

7 Metzger, J. *et al.* Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genomics* **16**, 764 (2015).

8 Hedrick, P. W. & Garcia-Dorado, A. Understanding inbreeding depression, purging, and genetic rescue. *Trends in Ecol. Evol.* **31**, 940-952 (2016).

9 Crnokrak, P. & Spencer, C. H. B. Perspective: purging the genetic load: a review of the experimental evidence. *Evolution* **56**, 2347-2358 (2002).

10 Charlesworth, D. & Willis, J. H. The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**, 783-796 (2009).

11 Leroy, G. Inbreeding depression in livestock species: review and meta-analysis. *Anim. Genet.* **45**, 618-628 (2014).

12 Woolliams, J. A., Berg, P., Dagnachew, B. S. & Meuwissen, T. H. E. Genetic contributions and their optimization. *J. Anim. Breed. Genet.* **132**, 89-99 (2015).

13 Zhang, Q., Guldbrandtsen, B., Bosse, M., Lund, M. S. & Sahana, G. Runs of homozygosity and distribution of functional variants in the cattle genome. *BMC Genomics* **16**, 542 (2015).

14 Mc Parland, S., Kearney, F. & Berry, D. P. Purging of inbreeding depression within the Irish Holstein-Friesian population. *Genet. Select. Evol.* **41**, 16 (2009).

15 Moreno, E., Pérez-González, J., Carranza, J. & Moya-Laraño, J. Better fitness in the Capture Cuvier's Gazelle despite inbreeding increase: evidence for purging? *PLOS ONE* **10** (2015).

16 Laws, R. J. & Jamieson, I. G. Is lack of evidence of inbreeding depression in a threatened New Zealand robin indicative of reduced genetic load? *Anim. Conserv.* **14**, 47-55 (2011).

17 Kennedy, E. S., Grueber, C. E., Duncan, R. P. & Jamieson, I. G. Severe inbreeding depression and no evidence of purging in an extremely inbred wild species- the Chatham Island black robin. *Evolution* **68**, 987-995 (2014).

18 Kalinowski, S. T., Hedrick, P. W. & Miller, P. S. Inbreeding depression in the Speke's gazelle captive breeding program. *Conserv. Biol.* **14**, 1375-1384 (2000).

19 Boakes, E. H., Wang, J. & Amos, W. An investigation of inbreeding depression and purging in captive pedigreed populations. *Heredity* **98**, 172-182 (2007).

20 Kristensen, T. N. & Sørensen, A. C. Inbreeding - lessons from animal breeding, evolutionary biology and conservation genetics. *Anim. Sci.* **80**, 121-133 (2005).

21 Marsden, C. D. *et al.* Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 152-157 (2016).

22 Cruz, F. *et al.* The Legacy of domestication: accumulation of deleterious mutations in the dog genome. *Mol. Biol. Evol.* **25**, 2331-2336 (2008).

23 Lu, J. *et al.* The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends Genet.* **22**, 126-131 (2006).

24 Lacy, R. C., Alaks, G. & Walsh, A. Hierarchical analysis of inbreeding depression in Peromyscus polionotus. *Evolution* **50**, 2187-2200 (1996).

25 Gulisija, D., Gianola, D., Weigel, K. A. & Toro, M. A. Between-founder heterogeneity in inbreeding depression for production in Jersey cows. *Livest. Sci.* **104**, 244-253 (2006).

26 Casellas, J., Varona, L., Ibáñez-Escriche, N., Quintanilla, R. & Noguera, J. L. Skew distribution of founder-specific inbreeding depression effects on the longevity of Landrace sows. *Genet. Res. (Camb)* **90**, 499-508 (2008).

27 Australian Racing Board. *2015/2016 Australian racing fact book*. (Australian Racing Board, 2016).

28 Velie, B. D., Hamilton, N. A. & Wade, C. M. Heritability of racing performance in the Australian Thoroughbred racing population. *Anim. Genet.* **46**, 23-29 (2015).

29 Velie, B. D., Hamilton, N. A. & Wade, C. M. Performance selection for Thoroughbreds racing in Hong Kong. *Equine Vet. J.* **47**, 43-47 (2015).

30 Charlesworth, B. & Charlesworth, D. The genetic basis of inbreeding depression. *Genet. Res.* **74**, 329-340 (1999).

31 Hinrichs, D. *et al.* Analysis of inbreeding depression in the first litter size of mice in a long-term selection experiment with respect to the age of the inbreeding. *Heredity* **99**, 81-88 (2007).

32 Klemetsdal, G. The effect of inbreeding on racing performance in Norwegian Cold-Blooded Trotters. *Genet. Select. Evol.* **30**, 351-366 (1998).

33 Schubert, M. *et al.* Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E5661-E5669 (2014).

34 Baumung, R. *et al.* GRain: a computer program to calculate ancestral and partial inbreeding coefficients using a gene dropping approach. *J. Anim. Breed. Genet.* **132**, 100-108 (2015).

35 Hill, E. W., Gu, J., McGivney, B. A. & MacHugh, D. E. Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. *Anim. Genet.* **41**, 56-63 (2010).

36 Henderson, C. R. Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423-447 (1975).

37 Bijma, P. & Woolliams, J. A. Prediction of rates of inbreeding in populations selected on best linear unbiased prediction of breeding value. *Genetics* **156**, 361A-3373 (2000).

38 Cassidy, R. Blood will tell in *The sport of kings: kinship, class and thoroughbred breeding in Newmarket* 140- 157 (Cambridge University Press, 2002).

39 Vamplew, W. Development of racing in *The turf : a social and economic history of horse racing* Ch. 1, 1-28 (Allen Lane, 1976).

40 Borowska, A., Wolc, A. & Szwaczkowski, T. Genetic variability of traits recorded during 100-day stationary performance test and inbreeding level in Polish Warmblood stallions. *Archiv fur Tierzucht* **54**, 327-337 (2011).

41 Dadar, M., Mahyari, S. A., Rokouei, M. & Edriss, M. A. Rates of inbreeding and genetic diversity in Iranian Holstein Cattle: Inbreeding, Genetic Diversity. *Anim. Sci. J.* **85**, 888-894 (2014).

42 Rodrigáñez, J., Toro, M. A., Rodriguez, M. C. & Silió, L. Effect of founder allele survival and inbreeding depression on litter size in a closed line of Large White pigs. *Anim. Sci.* **67**, 573-582 (1998).

43 Boichard, D., Maignel, L. & Verrier, É. The value of using probabilities of gene origin to measure genetic variability in a population. *Genet. Select. Evol.* **29**, 5-23 (1997).

44 Nagy, I. *et al.* Genetic diversity and population structure of the synthetic Pannon White rabbit revealed by pedigree analyses. *J. Anim. Sci.* **88**, 1267-1275 (2010).

45 Wallner, B. *et al.* Identification of genetic variation on the horse Y chromosome and the tracing of male founder lineages in modern breeds *PLOS ONE* **8**, e60015 (2013).

46 Casellas, J., Piedrafita, J., Caja, G. & Varona, L. Analysis of founder-specific inbreeding depression on birth weight in Ripollesa lambs. *J. Anim. Sci.* **87**, 72-79 (2009).

47 Fairfax-Blakeborough, J. *Northern turf history, vol. III* (J.A. Allen & Co, 1970).

48 Erickson, H. H., Kindig, C. A. & Poole, D. C. Exercise-induced pulmonary hemorrhage: a new concept for prevention. *J Equine Vet Sci.* **20**, 164-167 (2000).

49 Morley, P. S., Bromberek, J. L., Saulez, M. N., Hinchcliff, K. W. & Guthrie, A. J. Exercise-induced pulmonary haemorrhage impairs racing performance in Thoroughbred racehorses. *Equine Vet. J.* **47**, 358-365 (2015).

50 Velie, B. D., Raadsma, H. W., Wade, C. M., Knight, P. K. & Hamilton, N. A. Heritability of epistaxis in the Australian Thoroughbred racehorse population. *Vet. J.* **202**, 274-278 (2014).

51 Ewart, J. C. Horse-breeding and horse-racing. *Nature* **99**, 346-347 (1917).

52 Boyko, A. R. *et al.* Genomic analysis establishes correlation between growth and laryngeal neuropathy in Thoroughbreds. *BMC Genomics* **15**, 259-259 (2014).

53 Dupuis, M. C. *et al.* Detection of copy number variants in the horse genome and examination of their association with recurrent laryngeal neuropathy. *Anim. Genet.* **44**, 206-208 (2013).

54 Bourneuf, E. *et al.* Rapid discovery of de novo deleterious mutations in cattle enhances the value of livestock as model species. *Sci. Rep.* **7**, 11466 (2017).

55 Keller, M. C., Visscher, P. M. & Goddard, M. E. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* **189**, 237-249 (2011).

56 Purfield, D. C., Berry, D. P., McParland, S. & Bradley, D. G. Runs of homozygosity and population history in cattle. *BMC Genet.* **13**, 70 (2012).

57 Leutenegger, A.-L. *et al.* Estimation of the inbreeding coefficient through use of genomic data. *Am. J. Hum. Genet.* **73**, 516-523 (2003).

58 McQuillan, R. *et al.* Runs of homozygosity in European populations. *Am. J. Hum. Genet.* **83**, 359-372 (2008).

59 Kristensen, T. N., Pedersen, K. S., Vermeulen, C. J. & Loeschcke, V. Research on inbreeding in the 'omic' era. *Trends Ecol. Evol.* **25**, 44-52 (2010).

60 Marras, G. *et al.* Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Anim. Genet.* **46**, 110-121 (2015).

61 Gibson, J., Morton, N. E. & Collins, A. Extended tracts of homozygosity in outbred human populations. *Hum. Mol. Genet.* **15**, 789-795 (2006).

62 Pemberton, T. J. *et al.* Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.* **91**, 275-292 (2012).

63 Silió, L. *et al.* Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. *J. Anim. Breed. Genet.* **130**, 349-360 (2013).

64 Kim, E.-S. *et al.* Effect of artificial selection on runs of homozygosity in U.S Holstein cattle. *PLOS ONE* **8**, e80813 (2013).

65 Woolliams, J. A. & Mäntysaari, E. A. Genetic contributions of Finnish Ayrshire bulls over four generations. *Anim. Sci.* **61**, 177-187 (1995).

66 Boichard, D. in *7th World Congress on Genetics Applied to Livestock Production* Vol. paper 28-13 (Montpellier, France, 2002).

67 O'Loghlen, F. *Champions of the Turf: By 'Eurythmic' (Frank O'Loghlen)*. (F. H. Johnston Publishing Company, 1945).

68 Ford, M. *The Australian stud book history*. Vol. 1 (The Australian Stud Book, 2006).

69 Wright, S. Coefficients of inbreeding and relationship. *Am. Nat.* **56**, 330-338 (1922).

70 Ballou, J. D. Ancestral inbreeding only minimally affects inbreeding depression in mammalian populations. *J. Hered.* **88**, 169-178 (1997).

71 Suwanlee, S., Baumung, R., Sölkner, J. & Curik, I. Evaluation of ancestral inbreeding coefficients: Ballou's formula versus gene dropping. *Conserv. Genet.* **8**, 489-495 (2007).

72 Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575 (2007).

73 Ferencakovic, M., Solkner, J. & Curik, I. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet. Select. Evol.* **45**, 42-42 (2013).

74 Goddard, M. E., Powell, J. E. & Visscher, P. M. Reconciling the analysis of IBD and IBS in complex trait studies. *Nat. Rev. Genet.* **11**, 800-805 (2010).

75 Velie, B. D., Wade, C. M. & Hamilton, N. A. Profiling the careers of Thoroughbred horses racing in Australia between 2000 and 2010. *Equine Vet. J.* **45**, 182-186 (2013).

76 Gilmour, A. R., Gogel, B. J., Cullis, B. R. & Thompson, R. *ASReml User Guide Release 3.0*. (VSN International Ltd, 2009).

2.3 Figures and tables

Table 1: The average partial *F* (pF_i) and A_{HC} (pA_{HCi}) coefficients of the contemporary population for the 10 ancestors with the greatest marginal contributions to the modern Australian Thoroughbred population (*n*=135,572).

Ancestor name	**Year of birth**	**Percentage contribution by each ancestor**	
		***p*F**	***p*A_{HC}**
Herod	1758	19.87	25.13
Eclipse	1764	11.5	12.97
St Simon	1881	8.74	4.58
Godolphin Arabian	1724	8.34	10.34
Touchstone	1831	7.73	5.79
Stockwell	1849	7.15	4.76
Rachel	1763	5.75	6.32
Snap	1750	5.41	5.77
Partner	1718	3.62	12.97
Roxana	1718	2.28	2.49
Total contribution		80.40	82.18

The final pair of columns shows the total average contribution of all 10 ancestors to the *F* and A_{HC} coefficients. All values are expressed as a percentage of the total *F* or A_{HC} value.

Table 2: Regression coefficients of linear mixed estimating the association between five measures of racing performance and pedigree-based and genomic coefficients (*n*=122).

	F_{ROH_5}	F_{ROH_12}	F	A_{HC}
Cumulative earnings	1.95 (11.62)	5.05 (14.35)	-10.56 (19.74)	3.04 (3.51)
Earnings per start	0.63 (8.53)	2.30 (10.54)	-10.10 (14.55)	1.61 (2.58)
Career length	-0.69 (2.32)	-0.84 (2.86)	-3.26 (3.64)	0.54 (0.69)
Total starts	-10.61 (87.49)	-24.01 (107.94)	16.32 (142.13)	38.81 (25.81)
Winning strike rate	-2.17 (4.16)	-5.08 (5.14)	-17.37 (6.61)*	-0.43 (1.15)

Sex and year of birth were added as fixed effects and a numerator relationship matrix as a random effect in each model. Cumulative earning, earnings per start, and career length were log transformed for a normal distribution and analysed with a linear mixed model. Total starts was analysed using a Poisson generalized linear mixed model and winning strike rate using a binomial generalized linear mixed model. Inbreeding was measured using the pedigree measures of: Wright's inbreeding coefficient (F) and the ancestral history coefficient (A_{HC}). Genealogical inbreeding was measured as the proportion of runs of homozygosity (ROH) in the genome with the minimal lengths of 5MB (F_{ROH_5}) and 12MB (F_{ROH_12}). Standard errors are shown in parentheses. *P<0.05; **P<0.001.

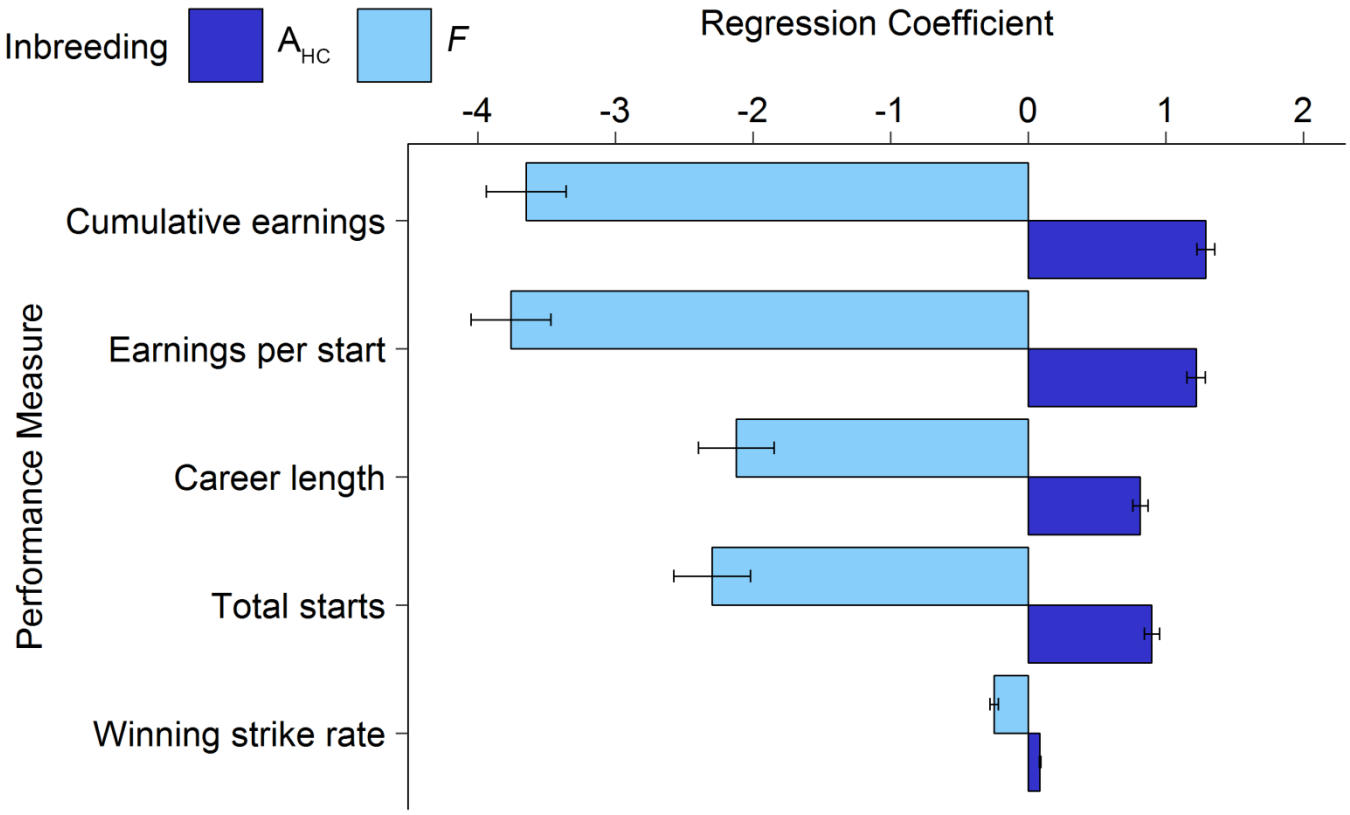

Figure 1: Regression coefficients showing the relationship between measures of racing performance and inbreeding in Thoroughbred horses (*n*=135,572). All measures of racing performance have a negative relationship with *F* but a positive association with A_{HC}. Error bars represent 1 standard error around the mean. Regression coefficients and standard errors were divided by the standard error of their respective traits. The relationship between each measure of inbreeding and racing performance was highly significant (P<0.001).

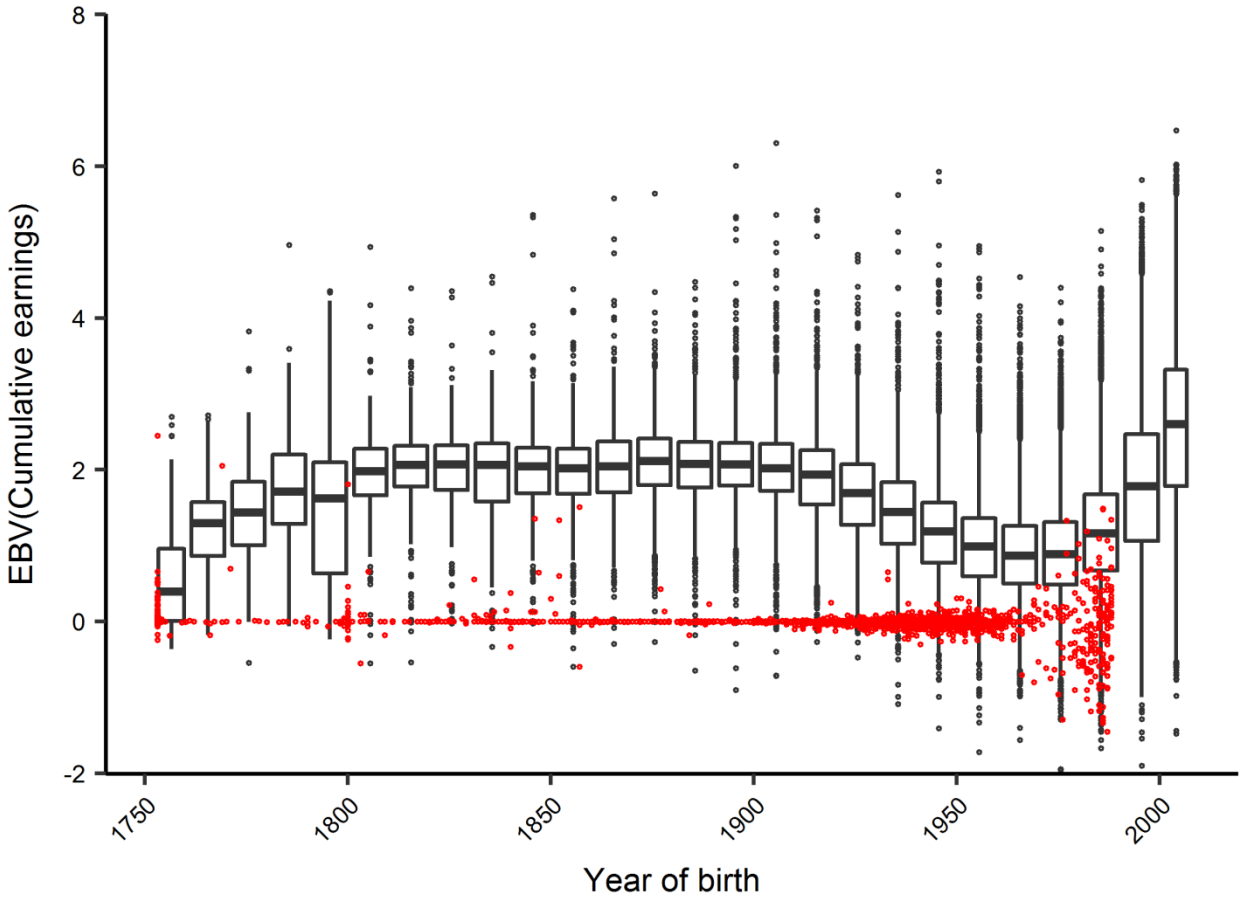

Figure 2: The distribution of estimated breeding values (EBVs) over time for Australian Thoroughbred horses (*n*=257, 249), based on the cumulative earnings of 135,572 individuals that raced between 2000 and 2010. Bins were calculated over intervals of 0.2, with each bin representing a 10-year period. Individuals with unknown parents are shown in red. The EBV results for the other measures of racing performance follow the same trends and are included in the Appendix.

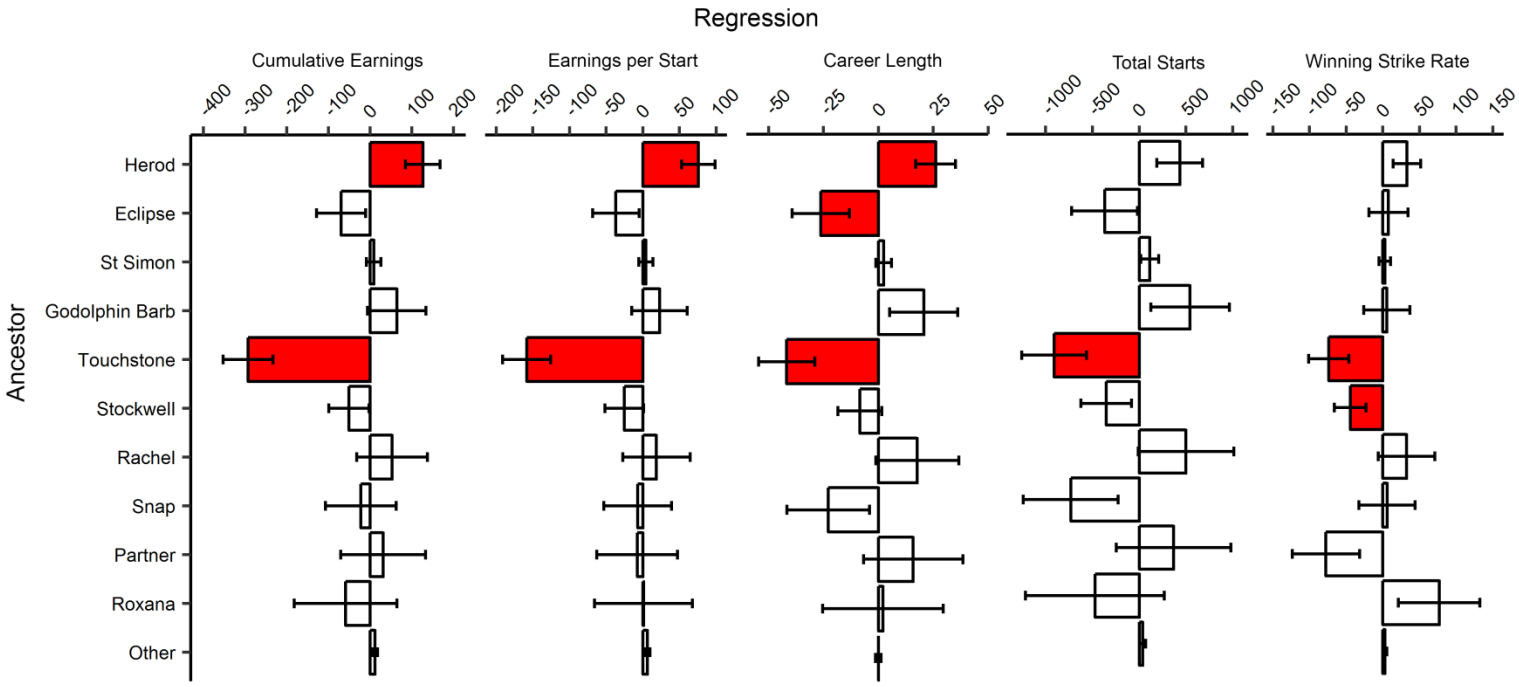

Figure 3: Inbreeding to different ancestors has variable effects on five measures of racing performance in modern Australian Thoroughbred horses. Partial inbreeding coefficients were calculated for the 10 ancestors with the greatest marginal contributions to the contemporary Australian Thoroughbred population. The relationship between each partial coefficient and inbreeding was analysed using regression coefficients from restricted maximum likelihood models. Error bars represent 1 standard error from the mean. This plot uses the same data set as in Figure 1, but with each inbreeding coefficient split into partials. Red bars denote significant relationships.

2.4 Supplementary information

S1: Materials and Methods

Calculating inbreeding coefficients

Racing data for all horses who had participated in a race start between 2000 and 2010 was provided by Racing Australia (*n*=138,996). We pruned this data to exclude individuals who did not have their complete race records outlined in the data (e.g. individuals that might have just started their racing career in 2010), leaving a sample size of 135,572.

The genealogy of each individual born after 1970, dating back to the founders of the population was also provided. The pedigree file consisted of 500,477 individuals and listed the sex, year of birth, and sire and dam identification for each individual. Pedigree loops were identified using *verif_ped* from PEDIG (1), and fixed by assigning the individual as a founder. F, F_{a_BAL}, F_{a_KAL} and A_{HC} coefficients were calculated by utilizing Grain 1.0 (2) to run 10^6 stochastic gene drops for each individual.

Calculating inbreeding from genomic data

We first estimated the proportion of homozygous SNP's in each individual's genome (F_H). We then calculated the proportion of each individual's genome made up of runs of homozygosity ($F_{\_ROH}$). We set the minimum number of SNP's in each ROH to 20, because our SNP coverage was approximately 1 SNP every 50Mb, making this sufficient to distinguish an ROH of 1Mb. We used the minimal length parameters of 1, 2, 5, 8, 12 and 16 Mb, which correspond to approximately 50, 25, 10, 8, 5 and 3 generations respectively (3). Correlation statistics were generated between all measures of inbreeding. Correlations were measured using Pearson's correlation coefficient in R.

Results and Discussion

S2: The comparison between different methods of measuring inbreeding and purging

Comparison between genealogical measures of inbreeding

There are a number of ways to account for inbreeding and genetic load in a population. We compared a number of genealogical and genomic measures to determine the optimal coefficients to analyse inbreeding trends in individuals and populations. We used the stochastic gene dropping program GRain

to calculate genealogical inbreeding values (2) (See SI Materials and Methods). This simulation method allows for accurate calculation of a number of ancestral inbreeding coefficients by accounting for alleles which are IBD multiple times in the pedigree.

First, we calculated Wright's classical inbreeding coefficient (*F*) (4). Although *F* is the traditional method for calculating inbreeding in a population, it does not always reflect the genetic load of the population (5). There are a number of coefficients that can be calculated from pedigree data to account for purging of load and favourable selection. These measures rely on the principle that selection will remove deleterious alleles from the population, so alleles which are IBD multiple times in a pedigree are more likely to have neutral or positive effects on fitness.

Theoretically, a population can be purged of some or all of its genetic load, so that individuals with a high inbreeding coefficient may show little or no evidence of inbreeding depression (6-8). For this reason, accounting for potential purging provides a better reflection of genetic load than simply measuring *F* (9). A number of measures have been proposed, which all operate under the assumption that an allele which has been IBD more than once in an individual's pedigree is less likely to have a deleterious effect on phenotype than one which is IBD for the first time.

The first of these coefficients was proposed by Ballou (5) that measures the proportion of the genome which has been IBD one or more times in previous generations: $F_{\text{a\_BAL}} = \frac{[F_{a(s)}+(1-F_{a(s)})F_{(s)}+F_{a(d)}+(1-F_{a(d)})F_{(d)}]}{2}$

where $F_{a(s)}$ is the ancestral inbreeding coefficient of the sire, $F_{(s)}$ is the inbreeding coefficient of the sire, $F_{a(d)}$ is the ancestral inbreeding coefficient of the dam, and $F_{(d)}$ is the inbreeding coefficient of the dam. Importantly, this measure accounts fr alleles that are IBD multiple times in each parents pedigree, so $F_{\text{a\_BAL}}$ can be greater than 0 for individuals with $F = 0$ (8).

In contrast, Kalinowoski's coefficient ($F_{\text{a\_KAL}}$) only accounts for alleles that are currently IBD, and have also been IBD in the past at least once (10). Hence, when F=0, $F_{\text{a\_KAL}}$ is also 0, resulting in a strong correlation between the two measures (Figure S1). A major shortcoming of $F_{\text{a\_BAL}}$ and $F_{\text{a\_KAL}}$ is that they only measure for alleles being IBD one or more times, so only account high lethal, recessive alleles (11).

The ancestral history coefficient (A_{HC}) differs from $F_{\text{a\_BAL}}$ and $F_{\text{a\_KAL}}$ in that it accounts for the number of times an allele has been IBD in an individual's pedigree (2). This calculation is based on the assumption

that the more times an allele has been IBD in an individual's pedigree, the more likely it is to have a neutral or beneficial effect on phenotype. It is therefore possible for an individual with a comprehensive and inbred pedigree to have an A_{HC}>1. A_{HC} is closely correlated with F_{a_BAL} because both measures account for all inbreeding events throughout the pedigree, although the former always holds a higher value because it quantifies the number of times an allele has been IBD (Figure S1). This measure provides the most comprehensive and accurate reflection of purging in inbred individuals.

Comparison between genomic measures of inbreeding

The accuracy of genealogical based inbreeding measures are highly reliant on the base population used. Each pedigree estimate assumes that the founders are unrelated, making them highly inaccurate for populations without reliable and comprehensive records. Our population has a complex pedigree with an average of 24.70 generations, allowing us to assume these estimates are fairly accurate. However, the stochastic nature of recombination and the increase of allele frequencies through selection can lead to variability between probability based and actual levels of autozygosity (approximately 2.43%) (12, 13). For this reason there are increasing numbers of studies using high density genomic information for inbreeding estimates (14). Genomic measures of inbreeding assume that increased levels of alleles IBD from common ancestors will result in higher levels of increased homozygosity (F_H) (14, 15).

To more accurately distinguish between alleles that are IBS and IBD, inbreeding levels are now often measured using runs of homozygosity (ROH). These long homozygous segments show evidence of a common ancestor, as they have not been broken down by recombination in meiosis (12, 16). The stochastic process of recombination means that shorter ROH segments correspond to ancient inbreeding, whereas larger ones correspond to recent consanguineous events (12, 17, 18). To differentiate between new and old inbreeding, we measured ROH with the minimal thresholds 1Mb, 2Mb, 5Mb, 8Mb, 12Mb and 16Mb, which correspond to 50, 25, 10, 6, 5 and 3 generations, respectively (19). We also measured each individual's genomic level of homozygosity (F_H), because it does not distinguish between alleles that are IBS and IBD.

Pairwise relationships showed close correlations between each ROH threshold. There was very little difference between $F_{\_ROH1}$ and $F_{\_ROH2}$ (Figure S2), possibly because the 45,451 SNP panel did not provide adequate density to distinguish between these thresholds (3, 20, 21). There is increasing evidence that high density panels are needed for accurate estimations of smaller ROH lengths (3, 20, 21). Some

individuals in our dataset showed no evidence of recent inbreeding: six individuals did not have an ROH over 12Mb (five generations), and twenty three with none over 16Mb (three generations).

Comparison between genealogical and genomic based inbreeding estimates

As we expected, there are some correlations between pedigree and genomic based measures. *F* showed the closest correlation with $F_{\_ROH16}$ (0.35, Figure S3), probably reflecting recent inbreeding events (12, 22). Shorter ROH regions may not always correspond to autozygous segments, or may not be detected due to insufficient SNP coverage, explaining their lower correlations with *F*. (21). Interestingly, our studies show a lower correlation between F and $F_{\_ROH}$ than many other studies, with reported correlations of ~0.7 (22, 23). This is probably because much of *F* in the Thoroughbred population is attributed to individuals many generations back. The SNP density used in our analysis was probably not comprehensive enough to capture these distant inbreeding events.

The close relationship between F_{a_Kal} and *F* measures has resulted in similar correlations with ROH coefficients (Figure S3). Since A_{HC} and F_{a_Bal} reflect levels of purging, rather than the proportion of the genome which is IBD, it is unsurprising they show no significant relationship with any genomic inbreeding measures. F_H had no relationship with any pedigree based-inbreeding, indicating that its failure to distinguish between IBD and IBS alleles makes it a crude measure of inbreeding (Figure S3).

Which measure of inbreeding is best?

One major shortcoming that we have identified in using ROH estimations to quantify inbreeding is the lack of concordance in the parameters used to define an ROH (24), making it difficult to compare results between studies. For example, some studies allow for one heterozygous SNP in an ROH because of genotyping errors, whereas others believe that it makes estimations inaccurate, particularly if the heterozygous SNP is at the end of the ROH (24). Additionally, a 50k SNP panel may not be sufficiently dense to accurately detect ROH, and increasing panel density can identify different ROH regions (3, 20, 21). Additionally, the discovery of long ROH segments in outbred human populations (25) brings into question their accuracy for capturing levels of inbreeding. Recombination 'hotspots' and high LD are proposed to account for these long IBD segments persisting for many generations (21, 25, 26). Considering the Thoroughbred population has one of the highest LD rates of any domestic animal population (27), this makes the accuracy of using ROH measurements to reflect inbreeding levels in the population questionable.

In contrast, *F* is a well-known and widely used method, so making comparisons between different studies are much easier to make (although the accuracy and number of generations of the pedigree can confound these results). For populations with deep and complex pedigrees, such as the Thoroughbred population, *F* allows us to estimate whole population inbreeding levels, as well as those for deceased individuals. In this respect, *F* is highly useful to study inbreeding trends over time and predict future implications for the population. However *F* does not account for factors such as genetic drift and selection in the population, making ROH advantageous in this respect (17). For this reason, using $F_{\_ROH}$ measures may be the most accurate way of measuring individual inbreeding levels in the Thoroughbred population.

Additionally, pedigree information can be used to estimate levels of purging and account for the selection of favourable alleles in a population. However, F_{a_Bal} and F_{a_Kal} do not effectively detect the purging of mildly deleterious alleles (11, 28). Therefore we propose that the best coefficient for quantifying purging is A_{HC}. In a deep and complex pedigree (such as the Thoroughbred pedigree), this coefficient best captures the effectiveness of selective breeding practices in increasing the frequency of favourable alleles, and the purging of highly and mildly deleterious alleles.

Consequently, we have chosen *F* and A_{HC} measures for the further analysis of the effects of inbreeding on fitness. We have also chosen ROH thresholds of 5 and 12Mb (corresponding to 10 and 5 generations, respectively). We chose the 5 Mb threshold to account for old inbreeding- with the SNP panel used in our study, any ROH estimate below 5Mb may not truly reflect ROH coverage. We have also chosen the 12Mb threshold to reflect new inbreeding, as 10% of the individuals in our dataset did not have any SNP's in the 16Mb category.

S3: Further output from linear mixed models.

In our analyses we implemented multiple measures of racing performance to account for talent, consistency, and constitutional soundness. The first measure, cumulative earnings, accounts for the amount of prizemoney a Thoroughbred earns throughout their racing career, and is based on the assumption that an individual's ability will be reflected by the amount of prizemoney that they earn. However, this measure can favour individuals that perform inconsistently, but win one big prizemoney race in their career. For this reason, we included earnings per start, which favours individuals that perform consistently well in high class races. Individuals that only contest in one or two race starts in their lifetime could still have high cumulative earnings and earnings per start, so we also included career

length as a performance measure. Career length does not account for long breaks in between race starts on account of injuries or poor recovery, so total number of starts was also included in the analysis. Lastly a measure of a horse's consistency was also included as winning strike race, because top class individuals should win most, if not all of their races. Consequently, these measures are all correlated (Figure S4).

We found that sex and year of birth significantly affected all measures of racing performance. In all models, female horses had lower levels of performance. For prizemoney measures, this discrepancy is probably due to bigger, stronger males being able to win more races with higher prizemoney. This probably also accounts for males having a higher winning strike rate, as females racing against males will have less chance of winning. The lower total starts and career length observed for female horses can be accounted for female horses being retired earlier for breeding purposes. Although some stallion prospects may also be retired early to stud, castrated males will continue racing for longer due to having no residual breeding value.

We used the linear mixed models to estimate the predicted values of each racing performance measure over a range of *F* and A_{HC} coefficients (Figure S5). These predictions follow the trends seen in the regression coefficients, with increasing *F* decreasing performance, and increasing A_{HC} enhancing it.

Using the numerator relationship matrix incorporated into the linear mixed models, we calculated the estimated breeding values (EBVs) for all individuals in the pedigree. The extensive pedigree information available for each individual in our sample increases the accuracy of our EBV information. However, it has been reported that phenotypic information over multiple generations will also increase the accuracy of EBV estimates (29).

We found that the distribution of EBVs in the population has changed over time (Figure S6). The EBVs for each phenotypic measure showed similar trends, probably because of the correlations between measures. For this reason, we only present EBVs for cumulative earnings in the main body of the paper.

S4: The greatest ancestral contributions to the Thoroughbred population.

We calculated the 20 ancestors with the greatest marginal contributions to the current population using iterations as implemented by PEDIG 1.0 (30). Calculating ancestral contributions rather than only contributions from founders accounts for bottlenecks in the pedigree, and is particularly advantageous

in our population because it allows us to estimate and understand the contributions of particularly successful and popular breeders to the current Thoroughbred population.

Marginal contributions account for relationships between ancestors by finding the greatest contributor to the population, then finding the contributions of other individuals not accounted for by the already selected individuals (1). Founders and individuals near the top of the pedigree will be favoured in marginal contribution estimations, as individuals further down in the pedigree will have their contributions diluted by redundancy if their ancestors have also made large genetic contributions to the population, leading to a large difference between their raw and marginal contributions.

In this respect, marginal contributions are advantageous when modelling the heterogeneity of inbreeding depression. However, it is important to note that an ancestor will pass on different sets of genes to each of its descendent (31), so marginal contributions could overestimate redundancies in the pedigree, as close relations may have inherited completely different genetic information from the same common ancestor. We found large differences in the raw and marginal contributions of some ancestors in our dataset (Figure S7, Table S1) such that selecting individuals based on raw contributions would result in a largely different sample space.

All of the greatest marginal contributors in our dataset are closely related, some of them sharing common ancestors, and others that were mated to produce a number of successful and influential progeny (Figure S8). The large contributions made by these close relatives demonstrate the narrow population bottleneck from which the breed has originated, mirrored in the initial large increase in the F coefficient at the foundation of the breed (Figure S9).

We found that many of these individuals were reported to be highly influential sires and dams in the early days of the Thoroughbred breed formation. Historically, there are considered to be four great sire lines responsible for the early formation of the Thoroughbred breed. Two of these, Herod and Eclipse, both feature as the first and third greatest marginal contributors. The third, Highflyer is the son of Herod and Rachel, and the fourth, Matchem, is the grandson of Godolphin Arabian, Roxana and Brown Farewell. Highflyer's remaining grandparent Croft's Partner was not featured in the top ten contributors because his contribution has been diluted by the inclusion of his son, Partner.

Of the 20 ancestors analysed for their partial contributions, five contributed to over 5% of the genomes of the current population and ten contributed to over 2% (Figure S1). We selected the top ten individuals for further analysis because they all had a p*F*i of over 0.005.

S5: Whole-population inbreeding trends over time.

As expected in a closed population with a small number of dominant founders (32), *F* has increased consistently since the foundation of the Thoroughbred population (Figure S9a). A_{HC} values have also increased in the population over time (Figure S9b). The exponential increase of A_{HC} over time agrees with our previous findings that selection for alleles contributing favourably to performance has increased their frequency over time. This result is in concordance with the positive relationship found between AHC and performance, indicating that an individual with a higher AHC has a greater accumulation of these favourable alleles in their genome.

Of the 135,572 individuals included in our racing performance analysis, the average *F* was 0.139 and the average A_{HC} was 1.973. The large difference in these values further demonstrates the selection for favourable alleles derived from the individuals in the early breed formation has increased their frequency over time.

We found that *F* levels rapidly increased after the bottleneck at the foundation of the breed. *F* then increased slowly in relation to A_{HC} over later generations (Figure S9c). There is a collection of individuals from 1930 onwards that have a lower *F* level than the majority of the population. These are the result of a parent that has an unknown pedigree or from an outbreeding event, often with one parent originating from a different continent. Although the *F* of these individuals is low, or 0, most of them have an A_{HC} value above 0. This indicates that both parents have inbreeding events in their own pedigree, which is not captured in *F* (see S2). We suggest that analysing both *F* and A_{HC} is needed to thoroughly examine inbreeding trends and effects in a population over time.

S6: Pedigree structure and missing ancestors

Although the individuals in our analysis trace back most of their pedigree lines to the founders of the population, a small number of ancestors have incomplete pedigree information. There are a number of reasons why their pedigrees are not completely recorded [33]. Firstly, an ancestor may not be registered in the stud book because its owner could not pay the stud book fees. This was particularly relevant during the Great Depression in the 1930s. Before 1980, it was acceptable in Australia for horses to be registered for racing without having a complete pedigree in the stud book based on the assumption that to be competitive in Thoroughbred races, these horses would have to be of Thoroughbred origin. Pedigree records of some horses were lost when they were shipped from England to Australia in the early 19th century. For horses with American bloodlines, many pedigree records were lost during the civil

war. Additionally, when DNA testing was introduced in the late 20th century, an individual was found to have false parentage. The proportion of ancestors with missing pedigree information by year is modelled in Figure S10. These individuals accounted for 1.4% of the total ancestors in our pedigree file.

Additionally, we estimated the proportion of missing ancestors by generation for all individuals used in our racing performance analysis (Figure S11). No individuals have missing parents, and most individuals (80%) have a complete ancestry up to 6 generations (Figure S11). Considering that the majority of *F* is captured in the first 6 generations of a pedigree[22], we consider that this pedigree structure makes the inbreeding estimates used in our analyses are highly accurate.

Additional References

1. Boichard, D., Maignel, L., & Verrier, É. (1997) The value of using probabilities of gene origin to measure genetic variability in a population. *Genet. Select. Evol.* 29(1):5-23.
2. Baumung, R. *et al.* (2015) GRain: a computer program to calculate ancestral and partial inbreeding coefficients using a gene dropping approach. *J. Anim. Breed. Genet.* 132(2):100-108.
3. Ferencakovic, M., Solkner, J., & Curik, I. (2013) Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet. Select. Evol.* 45(1):42-42.
4. Wright, S. (1922) Coefficients of inbreeding and relationship. *Am. Nat.* 56(645):330-338.
5. Ballou, J. D. (1997) Ancestral inbreeding only minimally affects inbreeding depression in mammalian populations. *J. Hered* 88(3):169-178.
6. Laws, R. J. & Jamieson, I. G. (2011) Is lack of evidence of inbreeding depression in a threatened New Zealand robin indicative of reduced genetic load? *Anim.* 14(1):47-55.
7. Moreno, E., Pérez-González, J., Carranza, J., & Moya-Laraño, J. (2015) Better fitness in the Capture Cuvier's Gazelle despite inbreeding increase: evidence for purging? *PLOS ONE* 10(12).
8. Mc Parland, S., Kearney, F., & Berry, D. P. (2009) Purging of inbreeding depression within the Irish Holstein-Friesian population. *Genet. Select. Evol.* 41(1):16.
9. Woolliams, J. A., Berg, P., Dagnachew, B. S., & Meuwissen, T.H.E. (2015) Genetic contributions and their optimization. *J. Anim. Breed. Genet.* 132(2):89-99.
10. Kalinowski, S. T., Hedrick, P. W., & Miller, P. S. (2000) Inbreeding depression in the Speke's gazelle captive breeding program. *Conserv. Biol.* 14(5):1375-1384.
11. Boakes, E. & Wang, J. (2005) A simulation study on detecting purging of inbreeding depression in captive populations. *Genet. Res. (Camb.)* 86(2):139-148.

12. Keller, M. C., Visscher, P. M. & Goddard, M. E. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* **189**, 237-249 (2011).
13. Carothers A. D. *et al.* (2006) Estimating human inbreeding coefficients: comparison of genealogical and marker heterozygosity approaches. *Ann. Hum. Genet.* 70(5):666-676.
14. Kristensen, T. N., Pedersen, K. S., Vermeulen, C. J. & Loeschcke, V. Research on inbreeding in the 'omic' era. *Trends Ecol. Evol.* **25**, 44-52 (2010).
15. Leutenegger, A.-L. *et al.* Estimation of the inbreeding coefficient through use of genomic data. *Am. J. Hum. Genet.* **73**, 516-523 (2003).
16. McQuillan, R. *et al.* Runs of homozygosity in European populations. *Am. J. Hum. Genet.* **83**, 359-372 (2008).
17. Curik, I., Ferencakovic, M., & Solkner, J. (2014) Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest. Sci.* 166(1):26-34.
18. Lencz, T. *et al.* (2007) Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci U S A* 104(50):19942-19947.
19. Ferenčaković, M. *et al.* (2013) Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *J Anim Breed Genet* 130(4):286-293.
20. Purfield, D. C., Berry, D. P., McParland, S., & Bradley, D. G. (2012) Runs of homozygosity and population history in cattle. *BMC Genet* 13:70.
21. Marras, G. *et al.* (2015) Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Anim Genet* 46(2):110-121.
22. Silió, L. *et al.* (2013) Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. *J Anim Breed Genet* 130(5):349-360.
23. Kim, E .S. *et al.* (2013) Effect of artificial selection on runs of homozygosity in U.S Holstein cattle. *PLOS ONE* 8(11):e80813.
24. Gurgul, A. *et al.* (2014) The application of genome-wide SNP genotyping methods in studies on livestock genomes. *J. Appl. Genet.* 55(2):197-208.
25. Gibson, J., Morton, N. E., & Collins, A. (2006) Extended tracts of homozygosity in outbred human populations. *Hum. Mol. Genet* .15(5):789-795.
26. Pemberton, T. J. *et al.* (2012) Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.* 91(2):275-292.

27. Corbin, L. J. *et al.* (2010) Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Anim. Genet.* 41(2):8-15.
28. Boakes, E. H., Wang, J., & Amos, W. (2007) An investigation of inbreeding depression and purging in captive pedigreed populations. *Heredity* 98(3):172-182.
29. Yang, H. & Su, G. (2016) Impact of phenotypic information of previous generations and depth of pedigree on estimates of genetic parameters and breeding values. *Livest. Sci.* 187:61-67.
30. Boichard, D. (2002) Pedig: a fortran package for pedigree analysis suited to large populations. in *7th World Congress on Genetics Applied to Livestock Production* (Montpellier, France).
31. Casellas, J., Piedrafita, J., Caja, G., & Varona, L. (2009) Analysis of founder-specific inbreeding depression on birth weight in Ripollesa lambs. *J. Anim. Sci.* 87(1):72-79.
32. Hedrick, P. W., Hoeck, P. E. A., Fleischer, R. C., Farabaugh, S., & Masuda, B. M. (2016) The influence of captive breeding management on founder representation and inbreeding in the ‘Alalā, the Hawaiian crow. *Conserv. Genet.* 17(2):369-378.
33. Ford, M. *The Australian Stud Book History*. Vol. 1 (The Australian Stud Book, 2006).

Table S1: The marginal and raw contributions for the top 20 ancestors for the current population (*n*=135, 572), selected based on their marginal contributions.

Name	Year of Birth	Raw	Marginal
Herod	1758	0.1811	0.1811
Godolphin Barb	1724	0.1383	0.1383
Eclipse	1764	0.133	0.1164
Snap	1718	0.0718	0.0718
Partner	1718	0.0966	0.0514
St Simon	1881	0.0996	0.0423
Rachel	1763	0.065	0.0364
Touchstone	1831	0.0813	0.0313
Stockwell	1849	0.0822	0.0307
Roxana	1718	0.0305	0.0224
Crab	1722	0.0268	0.0202
Trumpator	1782	0.0511	0.0196
Bartlet's Childers	1716	0.0447	0.0147
Grey Robinson	1723	0.0442	0.0145
Miss Slamerkin	1729	0.0182	0.0133
Bay Bolton	1705	0.036	0.0132
Brown Farewell	1753	0.0212	0.0116
Termagant	1772	0.0249	0.0112
Grecian Princess	1770	0.0256	0.0105
Flying Whigg	1715	0.0204	0.009

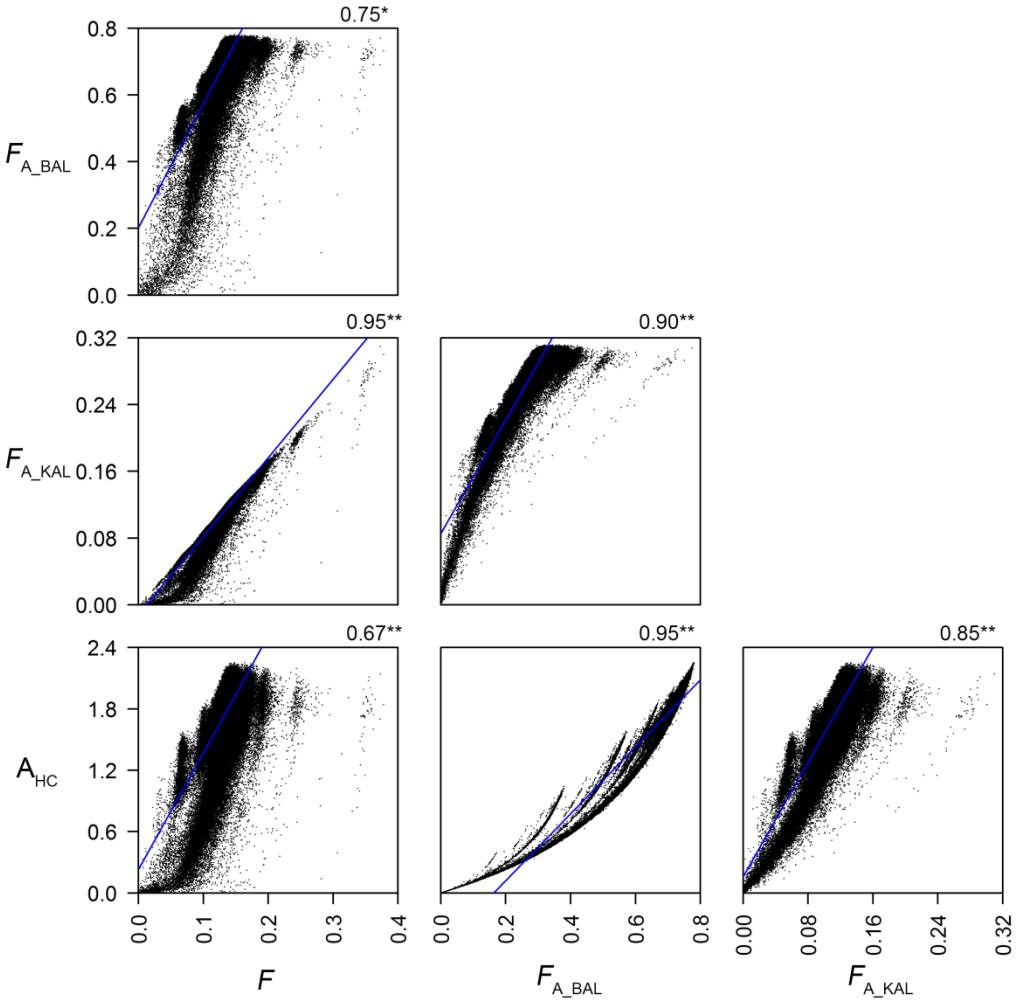

Figure S1: Correlations between pedigree based inbreeding measures estimated for Australian Thoroughbred racehorses (*n*=500,477). The inbreeding measures compared were: F, F_{A_BAL}, F_{A_KAL} and A_{HC}. Linear regressions are represented by blue lines. Pearson correlation coefficients are presented at the top right hand corner of each graph. P values <0.05 are marked with a *, and P values <0.001 with **.

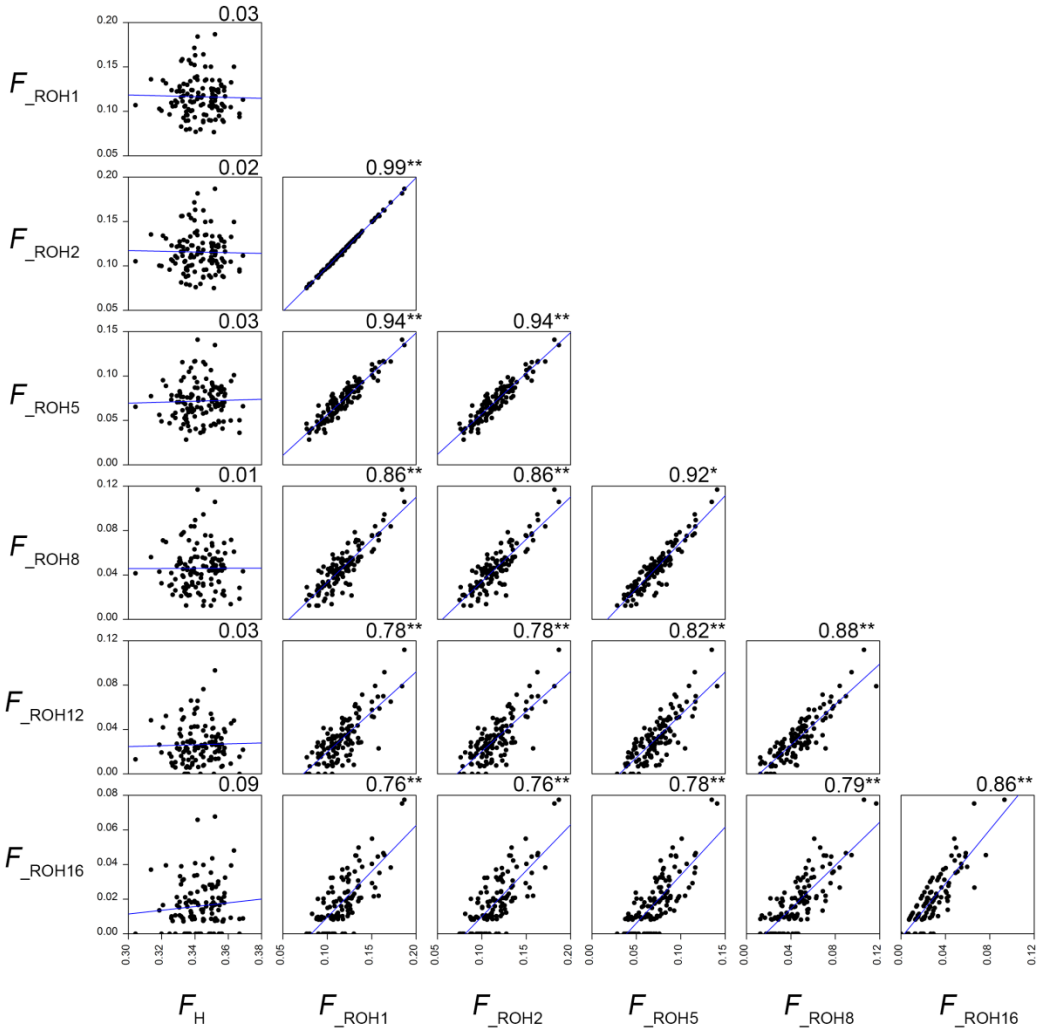

Figure S2: Correlations between SNP based inbreeding measures estimated for a representative subset of the modern Australian Thoroughbred (*n*=122). The genomic based inbreeding levels compared were: F_{H}, $F_{\_ROH1}$, $F_{\_ROH2}$, $F_{\_ROH5}$, $F_{\_ROH8}$, $F_{\_ROH12}$, and $F_{\_ROH16}$. Linear regressions are represented by blue lines. Pearson correlation coefficients are presented at the top right hand corner of each graph. P values <0.05 are marked with a *.

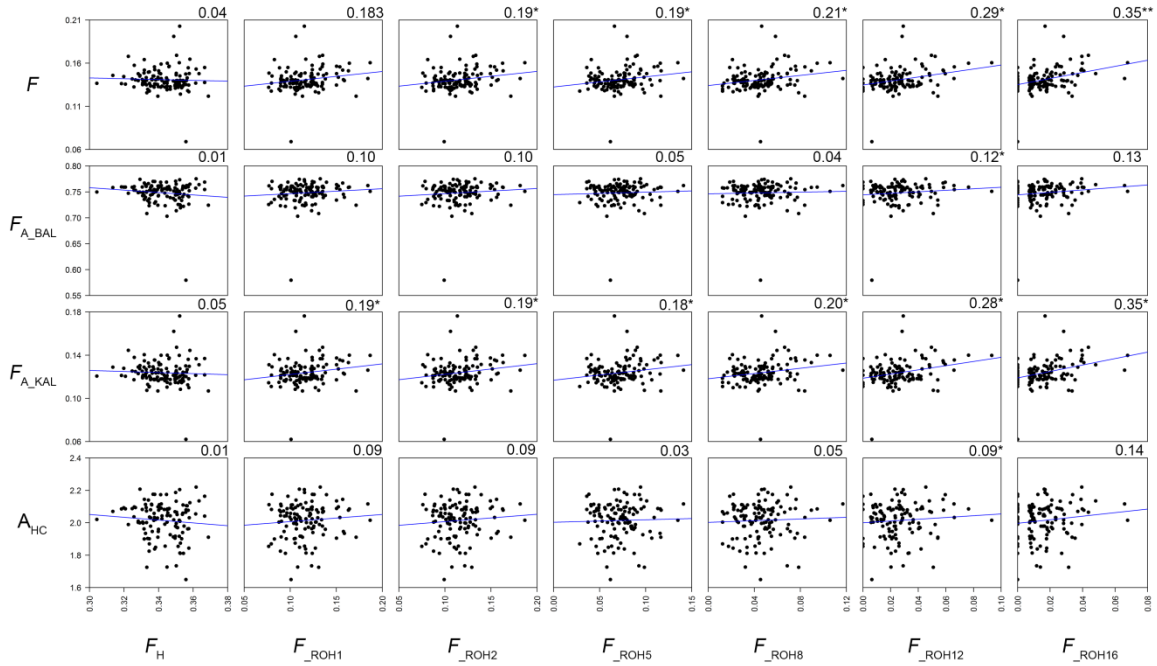

Figure S3: Correlations between SNP and pedigree based inbreeding measures for a representative subset of the modern Australian Thoroughbred population (*n*=122). The pedigree based inbreeding measures compared were: *F*, F_{A_BAL}, F_{A_KAL} and A_{HC}. The genomic based inbreeding levels compared were: F_H, $F_{\_ROH1}$, $F_{\_ROH2}$, $F_{\_ROH5}$, $F_{\_ROH8}$, $F_{\_ROH12}$, and $F_{\_ROH16}$. Linear regressions are represented by blue lines. Pearson correlation coefficients are presented at the top right hand corner of each graph. P values <0.05 are marked with a *, and P values <0.001 with **.

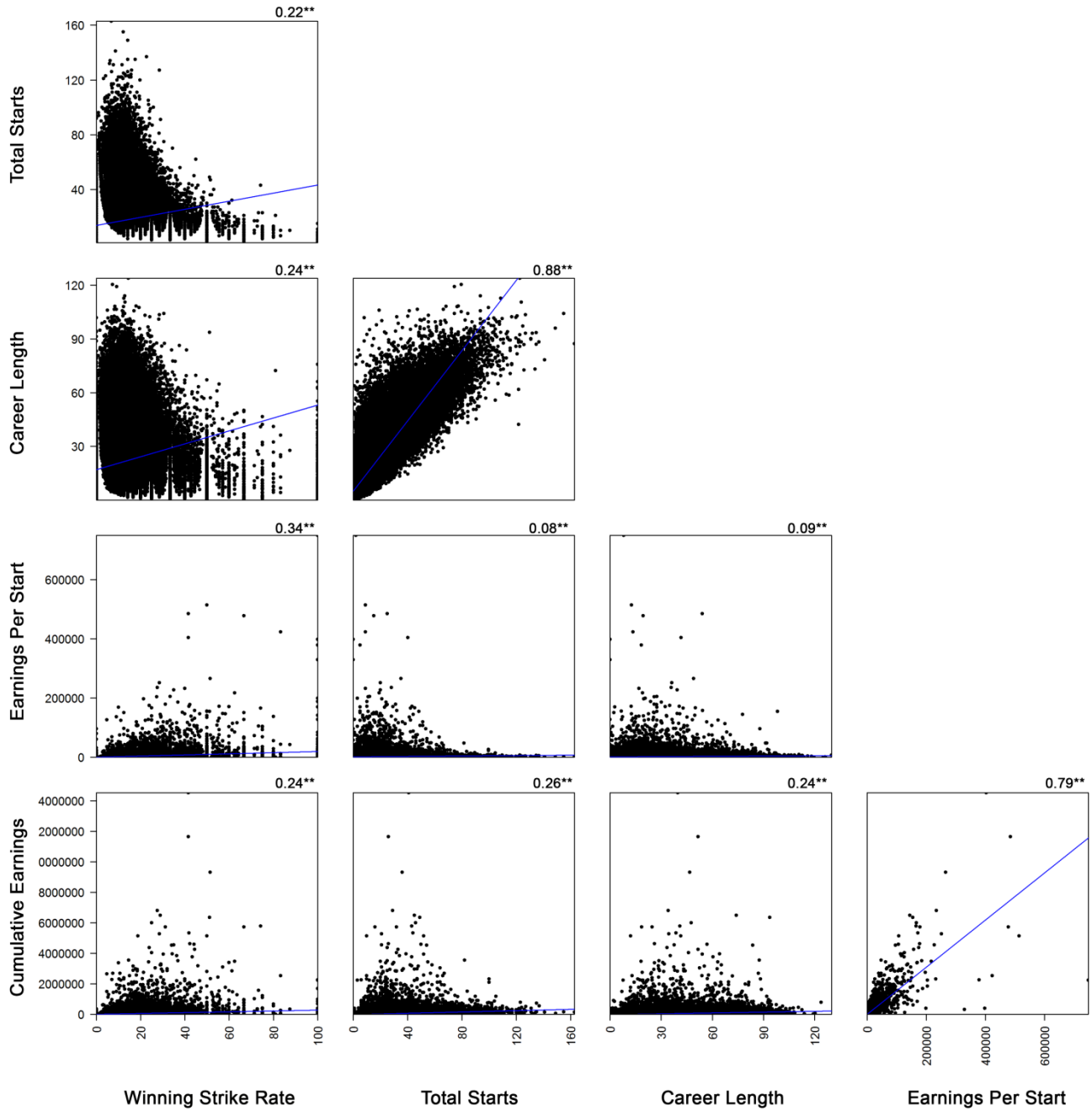

Figure S4: Correlations between the five measures of racing performance used in the mixed linear model analysis (*n*= 135,572). The measured analysed are: cumulative earnings ($AU), earnings per start ($AU), career length (in months), total number of starts and winning strike rate. Linear regressions are represented by the blue lines. Pearson correlation coefficients are present at the top right hand corner of the graph. P values<0.001 are marked with **.

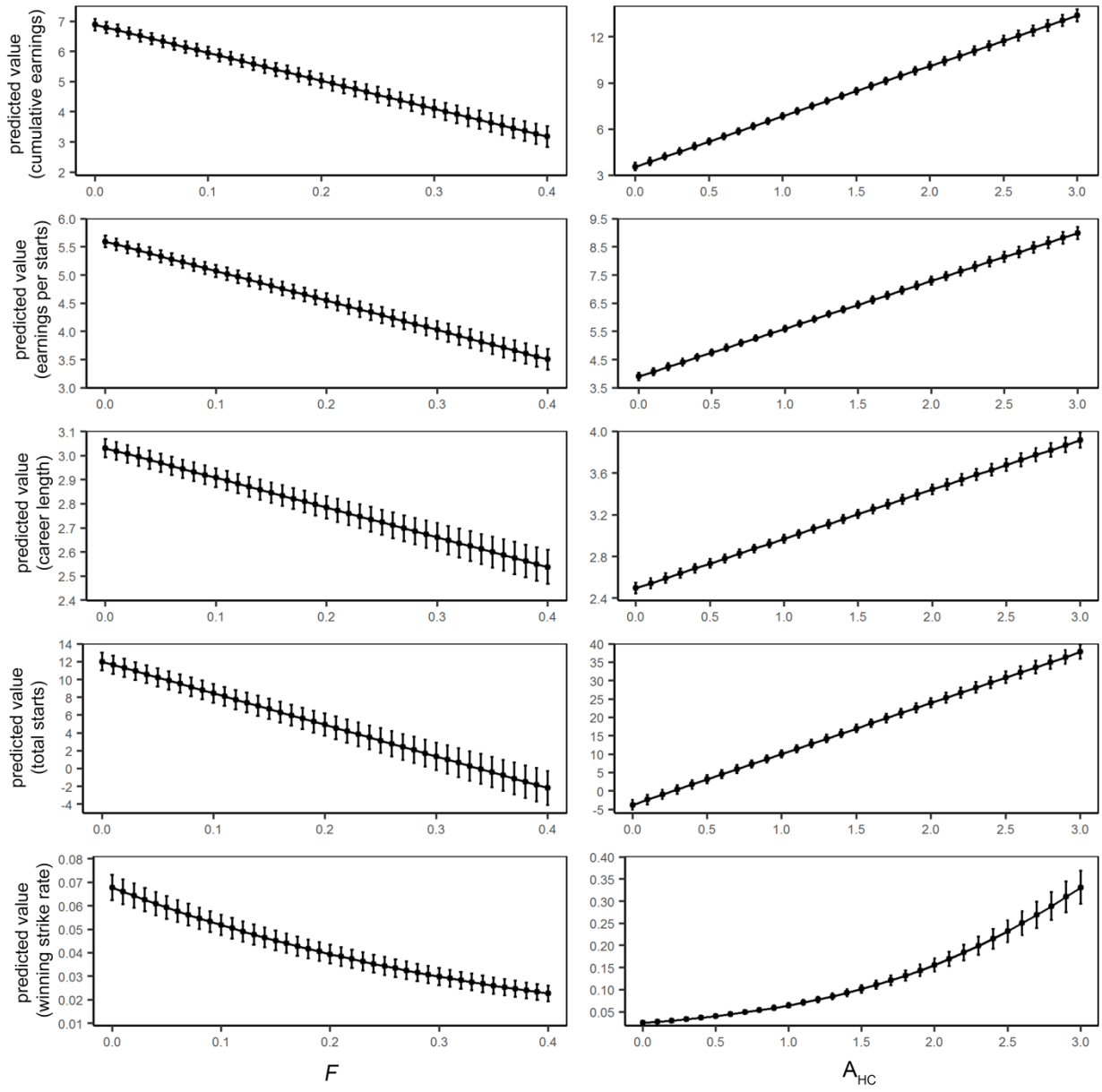

Figure S5: Predicted values of the effects of *F* and A_{HC} on performance (*n*=135, 572). Estimates derived from REML models. Error bars represent 1 standard error deviation from the mean.

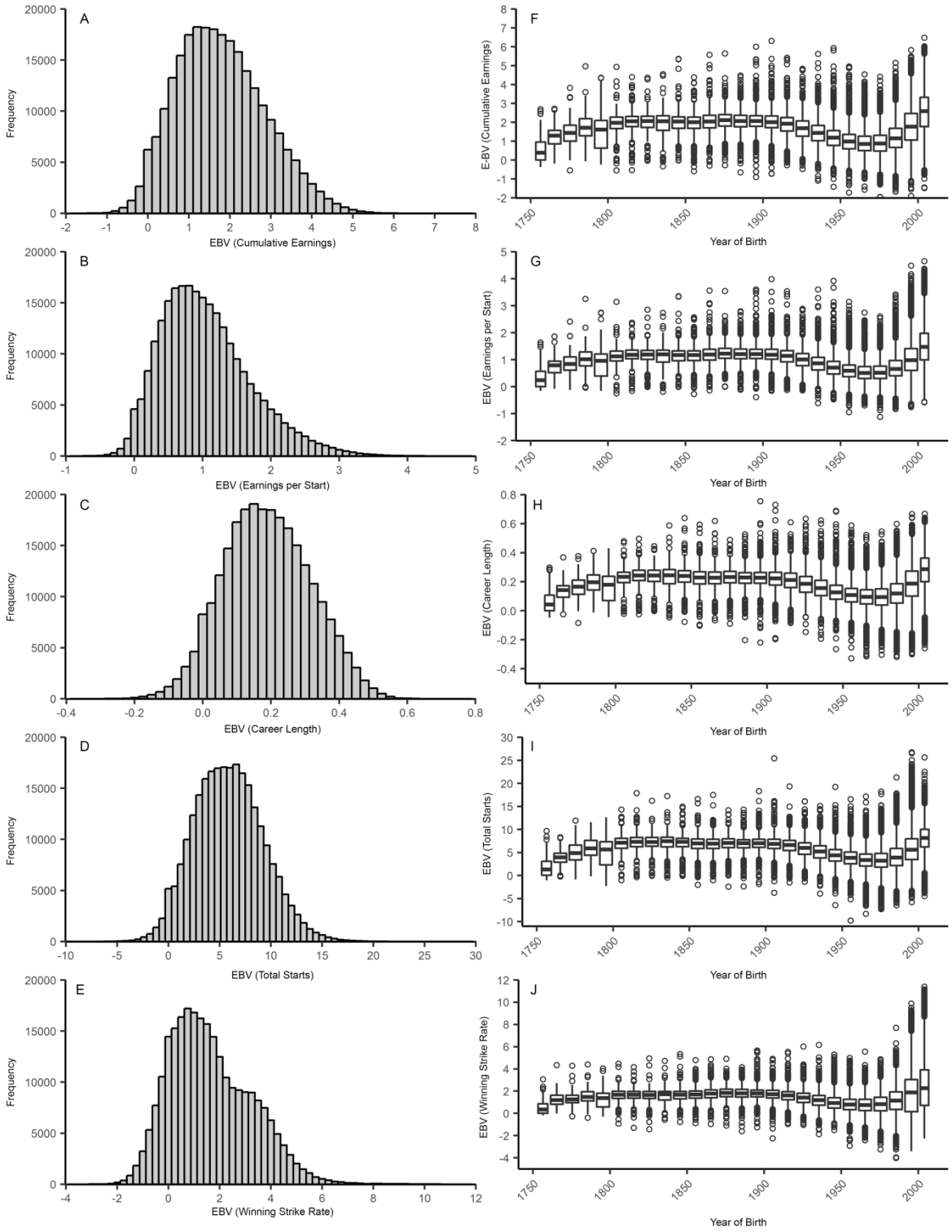

Figure S6: Selective breeding has increased EBVs in the Australian Thoroughbred population (*n*=257, 249). Frequencies distributions depict the distribution of EBV's for cumulative earnings (A), earnings per start (B), career length (C), total starts (D) and winning strike rate (E) measures. Box and whisker plot model the changing distributions of EBV's over time are also modelled for cumulative earnings (F), earnings per start (G), career length (H), total starts (I) and winning strike rate (J).

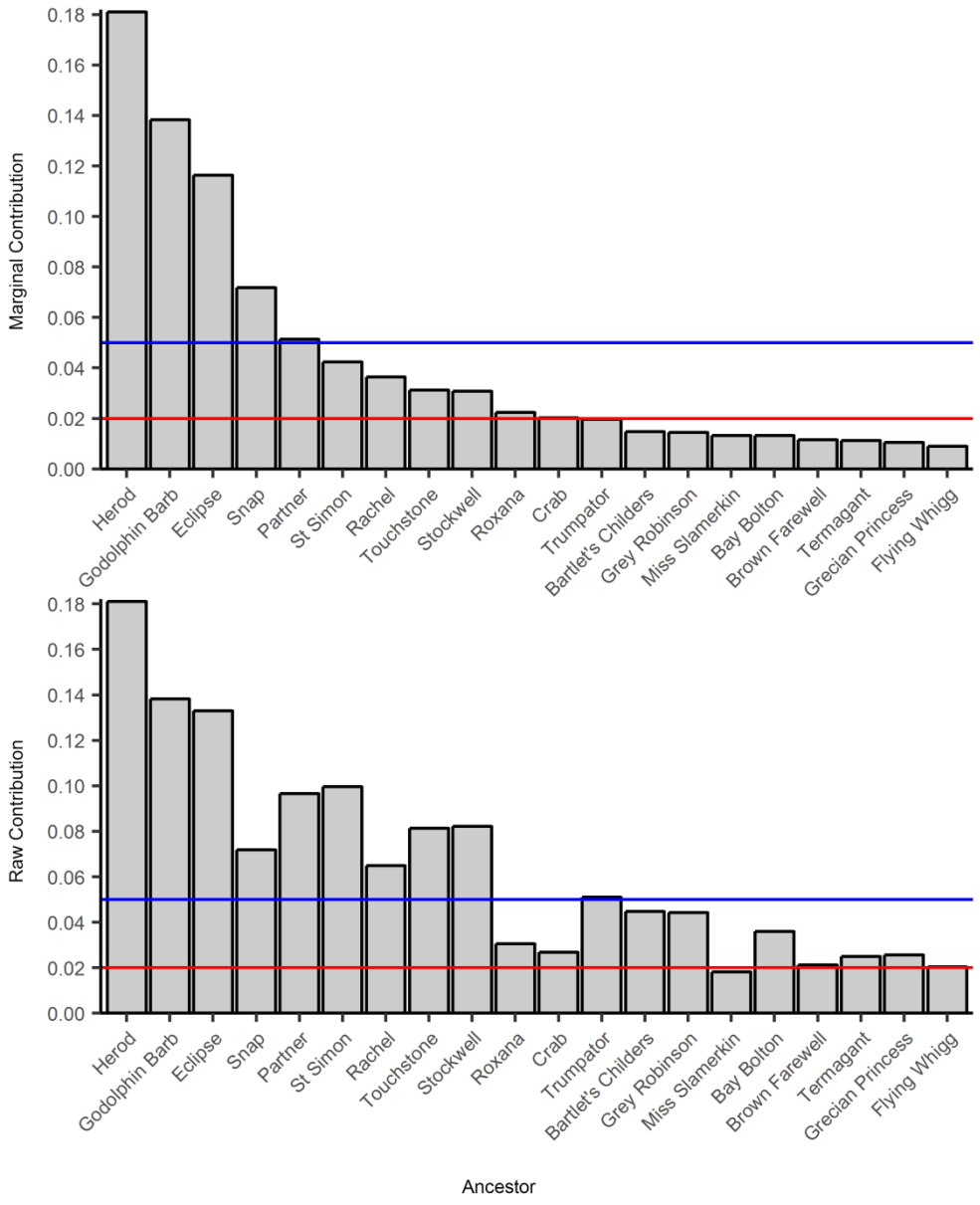

Figure S7: The top 20 marginal contributors to the contemporary Thoroughbred population (*n*=135,572). These ancestors were selected using their marginal contributions (A), which can deviate from their raw contribution (B). The blue line indicates the 5% marginal contribution level and the red line represents the 2% level.

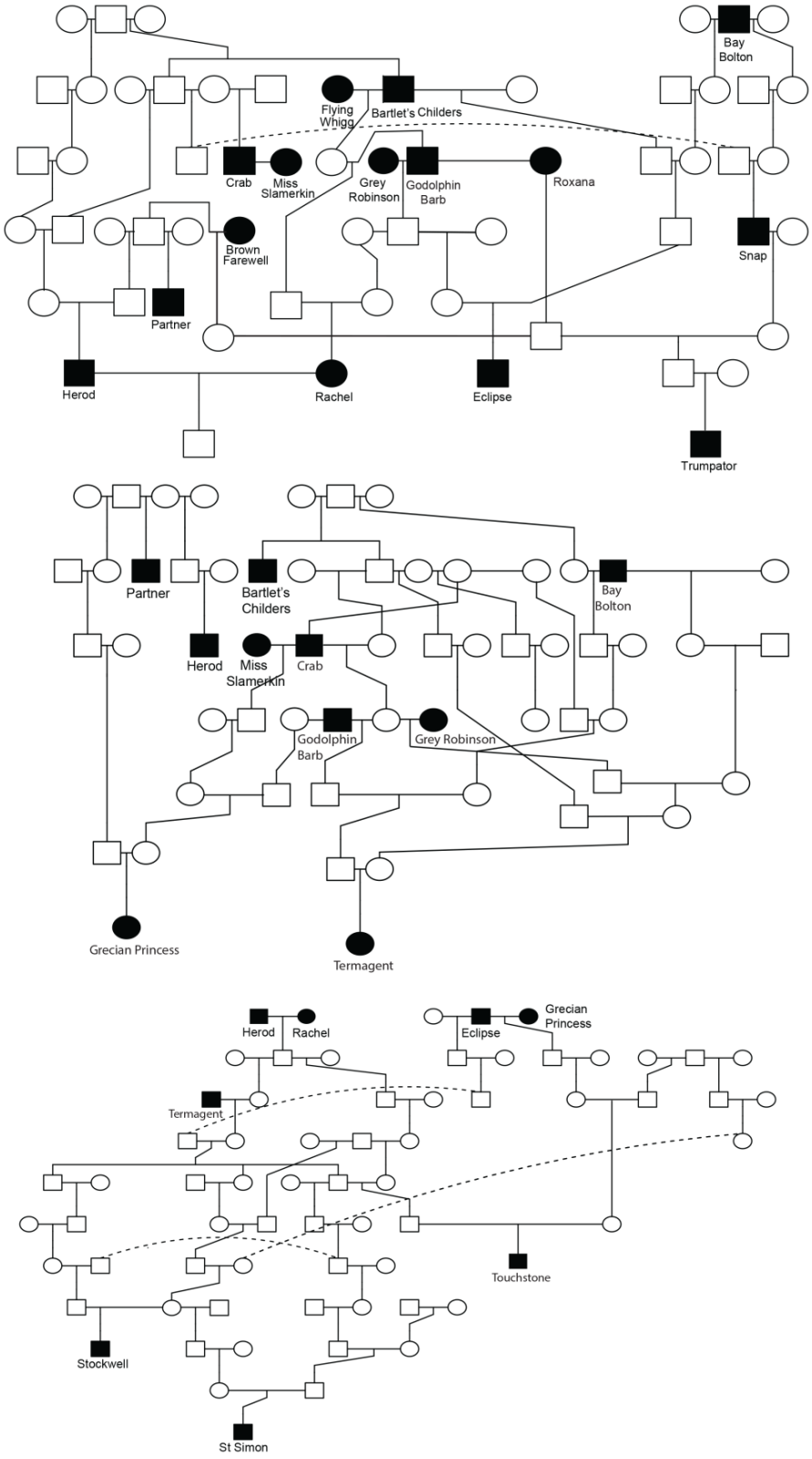

Figure S8: Pedigree illustrations visualising the relationships between the 20 greatest marginal contributors to the Australian Thoroughbred population

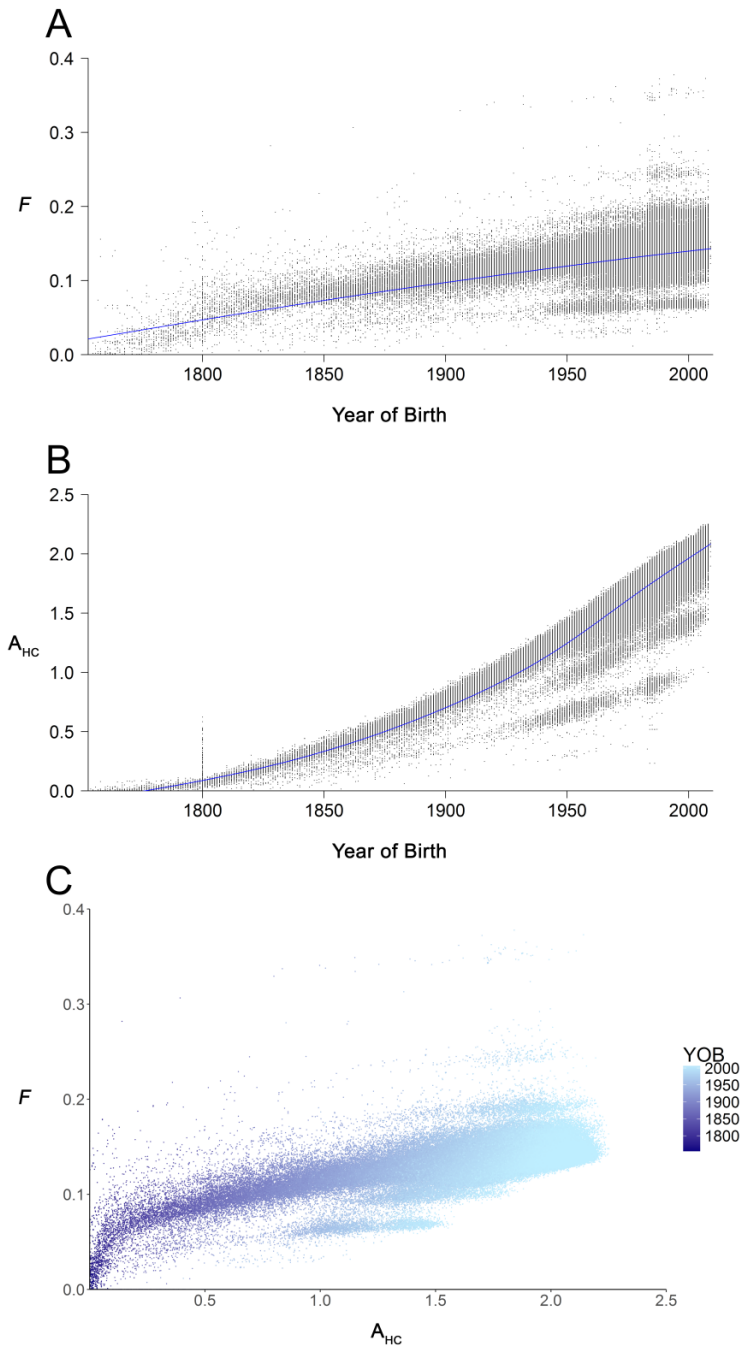

Figure S9: Inbreeding and purging trends in the Australian Thoroughbred population dating from the foundation of the breed in 1753 to the modern 21st century population (*n*=500,477). (A) Changes in the levels of *F* over time in the population. The regression line shows the increasing trend in inbreeding over time. (B) The distribution of A_{HC} levels over time in the population. The regression line shows the accumulation of alleles identical by descent multiple times in the pedigree. (C) F verses AHC with year of birth included as a sliding colour scale.

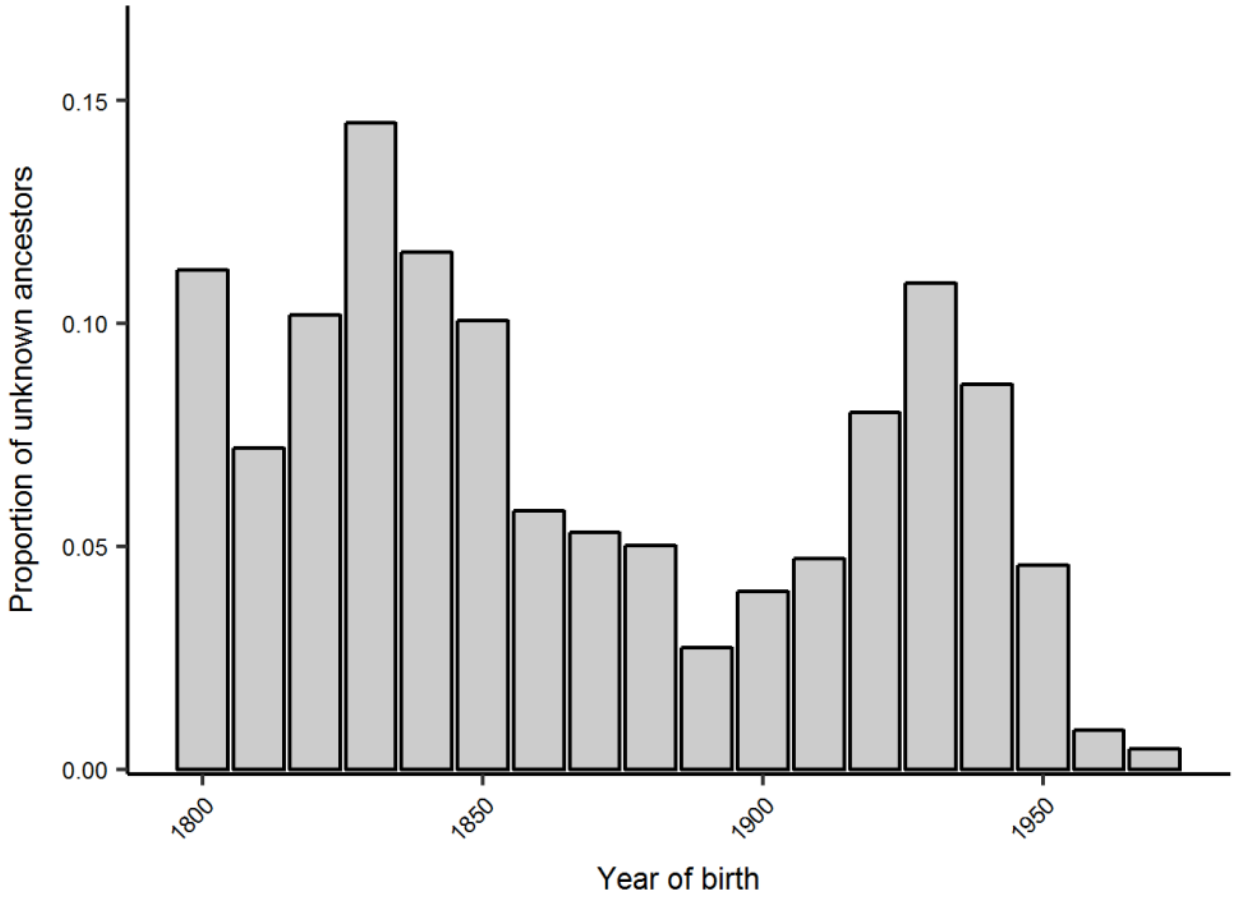

Figure S10: Proportion of ancestors with incomplete pedigree information from the genealogy of the Australian Thoroughbred horses that were used in the racing performance analysis (*n*= 119, 637). The date of birth for these individuals was listed as after the studbook was closed (1792), so are not considered to be founders of the Thoroughbred population. Data is expressed as the proportion of individuals with missing pedigree information from all individuals born over a 10 year period.

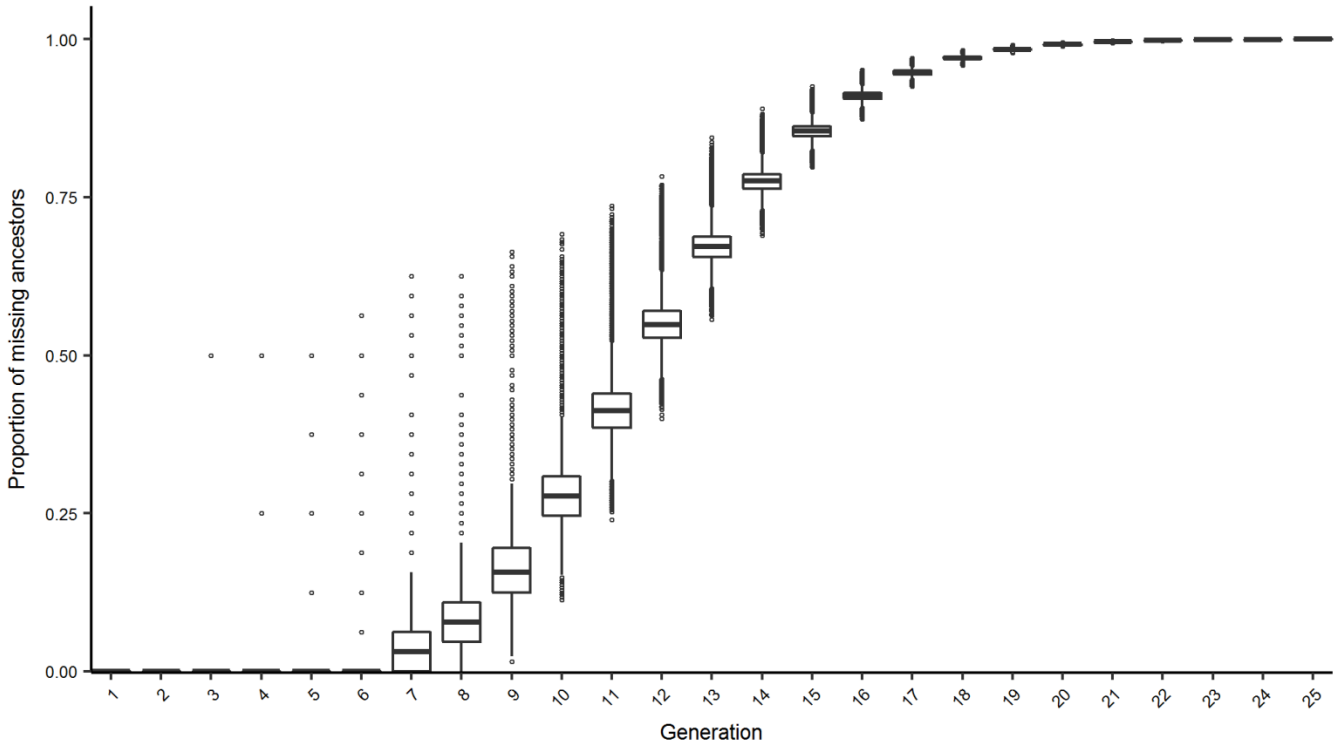

Figure S11: A box and whisker plot modelling the proportion of unknown ancestors over 25 generations for all Thoroughbred horses that were used in the analysis of racing performance (*n* = 135, 572). Proportions were estimated as the number of individuals with pedigree information listed as 0, divided by the total number of individuals found in each generation.

Chapter 3: The effects of inbreeding on fertility traits in the Thoroughbred horse population

3.1 Synopsis

This chapter consists of a manuscript:

E.T. Todd, N.A. Hamilton, B.D. Velie & P.C. Thomson. The effects of inbreeding on covering success, gestation length and foal sex ratio in Australian Thoroughbred horses. *BMC Genetics* **21**, 41, https://doi.org/10.1186/s12863-020-00847-1 (2020).

This publication has been formatted for consistency with this thesis. The original manuscript is available in Appendix 3.

This chapter examines the relationship between inbreeding levels and fertility traits in the Thoroughbred population. Fertility data from 27,262 breeding records from Australian Thoroughbred horses with an associated pedigree of 92,852 horses and these data were analysed by fitting of linear and generalised linear mixed models. The fertility traits analysed in this chapter included covering success, gestation length and sex ratio. These analyses showed that there was no association between inbreeding levels and fertility traits in Thoroughbred horses. This is in contrast to the strong relationship found between racing performance and inbreeding found in Chapter 2. However, all fertility traits analysed in this study had a measurable heritable component from either the sire or the dam. Additionally, this study examines the effects of environmental factors on fertility traits in Thoroughbred horses and provides insights into optimizing breeding decisions to improve fertility levels.

I designed this project and carried out the analyses using custom scripts in R. Drs Peter Thomson, Natasha Hamilton and Brandon Velie assisted in the overall design for the project and finalizing the manuscript.

3.2 Main article

The effects of inbreeding on covering success, gestation length and foal sex ratio in Australian Thoroughbred horses.

Evelyn T. Todd[1*], Natasha A. Hamilton[2], Brandon D. Velie[1] and Peter C. Thomson[1].

[1] School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia.
[2]Racing Australia Equine Genetics Research Centre, Racing Australia, Sydney, NSW 2000, Australia.

Abstract

Background

Horses produce only one foal from an eleven-month gestation period, making the maintenance of high reproductive rates essential. Genetic bottlenecks and inbreeding can increase the frequency of deleterious variants, resulting in reduced reproductive levels in a population. In this study we examined the influence of inbreeding levels on foaling rate, gestation length and secondary sex ratio in Australian Thoroughbred mares. We also investigated the genetic change in these traits throughout the history of the breed. Phenotypic data were obtained from 27,262 breeding records of Thoroughbred mares provided by three Australian stud farms. Inbreeding was estimated using the pedigree of each individual dating back to the foundation of the breed in the 18th century.

Results

While both gestation length and foaling rate were heritable, no measurable effect of inbreeding on either trait was found. However, we did find that the genetic value for both traits had decreased within recent generations. A number of environmental factors also had significant effects on foaling rate and gestation length. Secondary sex ratio had only an extremely small paternal heritable effect and was not susceptible to environmental influences.

Conclusions

In contrast to racing performance, inbreeding had no measurable effect on foaling rate or gestation length in Australian Thoroughbred horses. This could be because the level of inbreeding in the population examined is not high enough to show a discernible effect on reproductive traits. Populations that experience higher levels of inbreeding due to use of artificial reproductive technologies or extremely small population sizes may show a more pronounced reduction in natural foaling rate or

gestation length. It is also possible that the intensive management techniques used in the Thoroughbred population masks any negative effects of inbreeding. The decrease in the genetic value of foaling rate is likely to be because horses with unfavourable genetic potential have not yet been selected out of the population. The change in genetic value of gestation length may be due to selective breeding favouring horses with shorter pregnancies. We also found that prioritising the mating of older mares, and avoiding out of season mating could lead to an increased breeding success.

Background

Horses have an eleven month gestation period, both conceiving and giving birth in the spring and summer[1]. Like most spring seasonal-breeding animals, increased photoperiod induces ovulation cycles in mares. Survival of twins is rare, so mares can only produce one foal a year. The long gestation period and short breeding season make the maintenance of good fertility rates in horse populations imperative to provide commercial returns for domestic breeds, and to increase the size of endangered populations[2,3].

Deleterious genetic variants have accumulated in the genomes of modern horses as a result of population bottlenecks during domestication and breed foundation events[4,5]. In this process, also known as the "cost of domestication", deleterious mutations increase in frequency by "hitchhiking" on selective sweep regions [6]. These mutations can also increase in frequency through inbreeding, selective breeding and genetic drift in a population[7]. The presence of these variants in a population can have negative consequences for overall fitness, including a decrease in fertility rates. However, it is also possible that selective breeding over a number of generations may have removed some or all of these deleterious variants from contemporary horse populations. Evidence of positive selection in regions harbouring genes related to conceptus development have been found in domestic horse breeds[8,9], indicating that fertility rates may have been targeted and improved by breeding practices.

The effects of inbreeding on reproductive traits vary between studies. Increased inbreeding levels were associated with reduced fertility in some domestic and wild horse populations[2,3,10]. Impaired ovarian function resulting from high levels of inbreeding was reported in the Przewalski's horse, the most closely related species to the domestic horse[3]. Conversely, a number of studies in other horse breeds have shown no relationship between inbreeding levels and reproductive traits[11-13]. Varying relationships between inbreeding and reproductive performance also exist for a number of other domestic animal

populations[14-18]. It is possible that the effects of inbreeding on fertility may vary between different populations depending on the rate of increase in inbreeding, selective pressures and genetic diversity.

As well as affecting fertility, there is some evidence that increased inbreeding can skew secondary sex ratios (the sex ratio at birth) in animal populations[19]. Variations in sex ratio exist due to an increased chance of early conceptus loss of one sex under different conditions[20]. As maternal condition declines due to environmental stresses or inbreeding, the chance of producing a viable male conceptus may also decrease[19,21]. Early female horse conceptuses produce more insulin like growth factor-1 than males, which may promote their survival in adverse conditions[20]. It is possible that the environment at the time of conception or levels of inbreeding in horses may favour the survival of one sex.

Maintaining high reproductive rates is particularly important for the Thoroughbred horse breed. The Thoroughbred population has been closed since the 18th century, resulting in reduced levels of genetic diversity in current individuals[22-24]. Although selective breeding in Thoroughbred horses focusses on improving racetrack performance, the prohibition on reproductive technologies also makes it essential to maintain good fertility rates in the population. Thoroughbred foals that are born early in the spring season are assumed to have a size and maturing advantage over their peers. Mares that can conceive within 30 days after parturition will give birth at the same time next year. However, mares often take more than one covering each season, leading to their parturition date being delayed until later in the spring every year[25]. When a mare's parturition date reaches the end of spring, she cannot be covered until the next year. This will result in a two-year period in which she does not produce a foal.

Despite many generations of selective breeding for athletic ability, increased inbreeding is associated with reduced racing performance in Thoroughbred horses[24]. In this study we examine the effects of inbreeding levels on foaling rate, gestation length and secondary sex ratio in Thoroughbred mares. We use the pedigree data of 21st century Thoroughbred horses to estimate the heritability and the effects of inbreeding on these three reproductive traits. We also evaluate the environmental effects on these reproductive traits in Australian Thoroughbred mares. Additionally, we used estimated breeding values to measure the genetic change in foaling rate and gestation length since the foundation of the Thoroughbred population in the 18th century. Estimated breeding values (or genetic values) can measure the genetic potential of an animal for a trait based on the phenotypic information of themselves and their relatives. Genetic values are not commonly used for Thoroughbred horses but are utilized to assist in breeding management for other horse and livestock breeds. Although phenotypic data are not available for previous generations, we utilize the comprehensive pedigree information available for

Thoroughbred horses to calculate the genetic values for the ancestors in the pedigree, based on the reproductive trait data available for their modern day descendants.

Results

Data were available for 27,296 coverings of 12,922 mares bred to 131 stallions between 2000 and 2017. The pedigree of these individuals dating back to the founders of the population consisted of 92,852 individuals. The conceptuses, mares and stallions in the dataset had an average pedigree depth of 29 generations and a mean level of inbreeding of 0.156 ± 0.012 (mean ± SD).

The overall proportion of mares with a positive 15 day scan each season was 81.65% (22,287 out of 27,296). Based on the estimated variances from the mixed models, the maternal heritability of foaling rate was 0.058 (± 0.015), and 0.00 (± 0.00) for paternal heritability. Estimated breeding values for all individuals in the pedigree based on the foaling rate of their modern descendants showed a decrease in recent generations from a mean of -0.002 (± 0.092) in 1990 to -0.133 (± 0.138) by 2017 (Figure 1). Variation in the estimated breeding values of foaling rate was also estimated to increase dramatically from 1990 onwards (between -1.05 and 0.502) from previous years (between -0.468 to 0.264) (Figure 1). Mares that were covered later in the season showed a significant reduction in foaling success ($P < 0.001$) (Figure 2). There was an overall decrease in foaling rate with increasing mare age ($P < 0.001$) (Figure 3). Foaling rate had no relationship with the sire, dam or conceptus inbreeding level ($P = 0.142$, 0.788 and 0.701 respectively).

The sex of 7,578 live foal births were recorded in the dataset, 3,785 of which were colts (49.95%) and 3,793 were fillies (50.05%). Secondary sex ratio did not have an estimated heritable component for maternal genetic estimates (0.005 (± 0.026)), but had a small paternal heritability estimate of 0.011 (± 0.005). The sex ratio was not influenced by the sire, dam or foal inbreeding level ($P = 0.637$, 0.746 and 0.899, respectively). Environmental variables of mare age and month of birth also had no significant relationship with sex ratio ($P = 0.495$ and 0.337, respectively).

Gestation length data were available for 764 foals from 152 mares covered by 89 stallions. Gestation length was normally distributed, with a mean of 341 days (± 8.633), a minimum of 311 and a maximum of 376. Based on the estimated variances from the mixed model, the maternal heritability of gestation length was found to be 0.562 (± 0.042) and the paternal heritability was 0.004 (± 0.001). The average estimated breeding value in the pedigree remained at a mean of 0.03 (± 0.241) and showed little variation between -4.695 and 3.959 from the foundation of the population until 1990 (Figure 4). Since

2000, the average estimated breeding value for gestation length has decreased to -0.7 (± 1.91) and variation has increased, with a minimum of -11.82 and a maximum of 15.41. There was a significant decrease in gestation length in the later months of the season ($P < 0.001$) (Figure 5). Gestation length increased linearly with mare age ($P < 0.001$), going from a mean of 342 days at 2 years old to a mean of over 354 days by 24 years old (Figure 6). Male foals had a significantly longer predicted gestation length (349 days) than female foals (346 days) ($P < 0.001$). Gestation length had no significant association with the sire ($P = 0.087$), dam ($P = 0.419$) or foal ($P = 0.062$) inbreeding level.

Discussion

We found that inbreeding had no measurable effect on foaling rate, sex ratio or gestation length in Australian Thoroughbred horses. This may be because intensive management techniques used on commercial Thoroughbred stud farms have masked any negative effects of inbreeding on these traits. Some studies have similarly shown that inbreeding has no effect on gestation length[12,13], or foaling rate[11] in domestic horses. However, high levels of inbreeding in the endangered Przewalski and Sorraia horse breeds are associated with decreased birthing rates[2,3]. It is possible that inbreeding has no measurable effects on reproductive traits until it reaches a very high level. Both the Sorraia and the Przewalski's horse population have extremely high inbreeding levels (0.21 and 0.38, respectively) due to small effective population sizes and recent severe population bottlenecks[2,3,26,27]. In contrast, the relatively high average inbreeding coefficient (0.156) found for Thoroughbreds in this study is due to many generations of slow inbreeding. A large increase in the rate of inbreeding may lead to more noticeable effects on the reproductive traits of Thoroughbred horses in the future. The prohibition of artificial reproductive technologies (e.g. artificial insemination, cloning and embryo transfer) likely limits the rate of increase in inbreeding in the Thoroughbred population. The use of these technologies in other horse populations has been associated with increases in the rate of inbreeding[28,29], which could show more measurable effects on reproductive traits. It is also important for Thoroughbred breeders to be vigilant in their selection of sires and dams to avoid overbreeding to successful families as it may increase inbreeding rates in future generations and have unexpected negative effects on the population.

The maternal heritability estimate for gestation length based on our models was 0.562, higher than previous estimates in horses of 0.18-0.39[12,30,31]. This increased heritability estimate may be because all mares in this study have been intensively managed in the same way, minimising the amount of environmental variation that can reduce such estimates. Sires were also estimated to have a smaller, but still significant genetic effect on the gestation length of the foal. Stallions have similarly been found

to influence gestation length in other horse breeds[30,31], and such knowledge could assist in breeding management decisions. For example, avoiding the mating of mares and stallions genetically predisposed to longer gestation lengths, particularly towards the end of the season, could aid in avoiding delayed parturition dates.

Foaling rate had a lower but still significant maternal heritability estimate of 0.058, which similarly to gestation length was slightly higher than previous estimates (0.013-0.024)[32]. Unexpectedly, paternal heritability estimates from our models found no genetic influence of the sire on foaling rate. Only 131 sires were included in this study, so it is possible that a larger sample size would reveal significant heritable effects. Additionally, Thoroughbred stallions that show suboptimal fertility may be gelded and returned to racing. Analyses that include these individuals may reveal more measurable paternal genetic effects on covering success.

In contrast to gestation length and covering success, we found that secondary sex ratio had a negligible maternal heritable component (0.005) with a high standard error, indicating that genetic variation in the mare has no influence on the sex of the foal. However, paternal heritability estimates revealed an extremely small but nonetheless significant effect of sire genetic variation on foal sex ratios (0.011). Other studies in mammals have also found evidence of male-driven sex ratio bias[33,34]. However, in contrast to these studies, we found no association between the inbreeding level of the sire and offspring sex ratio[33,34]. It is postulated that higher quality, less inbred fathers produce more male offspring[35]. The artificial breeding practices used in the Thoroughbred population that only allow a small proportion of high-quality males to breed may mitigate any effects of inbreeding on offspring sex ratios, and result only a small measurable heritable effect.

In this study there was found to be little improvement in the estimated breeding value (genetic value) for reproductive traits in Thoroughbred horses from the 18th century until present day. The horses for which these values are obtained only include the direct ancestors to the horses with reproductive trait records in our study. Some individuals in previous generations may have had greater variation in genetic value but do not appear in the pedigree of modern Thoroughbreds. We found no change in the mean genetic value for foaling rate until the most recent three generations (Figure 1). Selection may favour mares that produce many offspring, making them more likely to be present in the pedigree of future generations. On the other hand, female families with low fertility have less chance of appearing in the pedigree of Thoroughbreds in future generations. It is possible that if we had been able to evaluate the genetic value for all horses in the studbook from previous generations, we would have found a greater

spread of values. Another reason for the minimal variation in the genetic value of earlier generations may be due to the lack of information conveyed in the binary trait of foaling rate. The near zero values may represent regression to the mean when there is limited information from previous generations.

In the past 30 years (approximately three generations), the average genetic value for foaling rate has decreased, and variation has increased. The decreasing mean may be because horses with a lower genetic value have not yet been selected out of the current Thoroughbred population. These horses may not appear in the pedigree of future Thoroughbreds, such that the values from their generation will show minimal variation. The decreasing average genetic value conflicts with an increased foaling rate reported in recent years[36]. It is likely that the improving fertility rates in the population are due to better management techniques rather than genetic gain. Recent advances in veterinary treatments have led to widespread use of hormonal therapies to increase foaling rate[36]. Increasing commercial demand for mares with good fertility may explain the outlying individuals with high genetic potential in recent years. However, widespread use of intensive veterinary treatments could make selection against individuals with lower genetic value less efficient, resulting in long term reductions in natural reproductive levels of the Thoroughbred population.

Similarly to foaling rate, the genetic value for gestation length has shown little variation since the foundation of the breed (Figure 4). However, unlike foaling rate, gestation length is not a binary trait. Selection against horses with longer gestation lengths (corresponding to higher genetic values) is likely to occur because they can produce fewer foals throughout their lifetimes. Extremely short gestation lengths may also be selected against because foals born prematurely will have a lower survival rate. These factors will remove individuals with high and low values, resulting in minimal variation and a mean of zero in more distant ancestral generations. It is also possible that the low genetic diversity in previous generations of the pedigree has resulted in reduced variation of the genetic value of these individuals based on breeding records from the current population. In the past twenty years, the average genetic value has reduced, showing that gestation length in Thoroughbred mares has, on average, become shorter. The increase in commercial Thoroughbred breeding during this time may have favoured mares with slightly shorter gestation lengths, as they are more likely to be successfully covered again in the same season. Foals that are born earlier in the season can have a competitive advantage over their peers when racing at a young age, which also favours mares with a decreased gestation length. Variation in genetic values of gestation length has also increased in the recent generations of the

pedigree, indicating that there are increasing numbers of mares with genetic potential for very long and short gestation lengths. These individuals may be selected out of the population in future generations.

In contrast to foaling rate and gestation length, the genetic values for racing performance in the Thoroughbred population have increased in recent generations[24]. This may be because Thoroughbred horses are directly selected for good racing success rather than fertility. If racing performance traits are driven primarily by positive selection (selection for advantageous alleles), this would explain the increase in genetic values for these traits over the past few generations. On the other hand, reproductive traits may be driven by mostly negative selection against disadvantageous alleles, explaining the decrease in genetic values of foaling rate in recent generations because there has not been an opportunity for the population to be purged of these genes. Thoroughbred horses have been selectively bred for racing performance since the start of the 18th century. However, it is likely that domestic horses have been selected for reproductive traits for many generations prior to the foundation of the Thoroughbred breed. Common signatures of selection for fertility have been found in domestic horse breeds including the Thoroughbred [8,9]. This could explain why an increase in genetic values for racing performance traits, but not fertility measures, is seen at the foundation of the Thoroughbred breed.

We also found that a number of environmental effects have significant influences on both gestation length and foaling rate. Our results showed that mares foaling down later in the season had significantly shorter gestation lengths (Figure 5). Our findings agree with previous studies[12,30,32] and highlight the importance of photoperiod length for inducing parturition in horses[37]. This pattern may strongly depend on the mare's location, as photoperiod variation is highly dependent on the proximity to the equator.

We also found that mares who produced male foals had significantly longer gestation lengths than those who produced female foals, possibly because of differences in maternal-foetal hormonal interactions[12,13,32,38]. However, an average difference of 3 days between the gestation lengths of colts and fillies is unlikely to impact breeding management decisions. Gestation length also increased with mare age, which could be explained by changes in hormonal, nutritional and uterine changes as a female ages[13] (Figure 6). Some studies have found a similar linear increase with age[32], whereas others have found that gestation length is longer in both younger and older mares[12]. This pattern may be dependent on the veterinary management provided to maiden mares. Foaling rate declined with increasing mare age, with mares over the age of 20 having less than 70% success (Figure 3). Foaling rate reduced dramatically in November and December, most likely due to the accumulation of less fertile

mares at the end of the season (Figure 2). To optimise foaling rate, the breeding of older mares would need to be prioritized because they tend to have longer gestation lengths and lower foaling rates.

In contrast to foaling rate and gestation length, secondary sex ratio was not significantly influenced by any environmental effects included in these models. Mares in a poor nutritional condition at conception have been reported to have an increased chance of successfully carrying a female foetus, with reports of female foal ratios up to 80%[39,40]. In the Mangalarga Marchador horse breed, higher ratios of female foals were found in older mares[21]. Additionally, increased dam inbreeding in cattle is associated with higher female birth rate[19]. However, we postulate that the high level of veterinary management and care provided to Thoroughbred horses in commercial stud farms examined in this study has resulted in no environmental or inbreeding factors having a measurable influence on foal sex ratios. Differing management of other horse populations (e.g. feed and hormonal supplements) may show more measurable effects on sex ratio. Wild animal populations with more variable environmental conditions and rates of inbreeding may also show different trends.

Conclusions

In this study we found that inbreeding had no measureable effect on reproductive traits in Australian Thoroughbred horses. Although this contrasts with previous findings in racing performance traits, we postulate the inbreeding levels are not yet high enough to have a measurable effect on the Thoroughbred reproduction traits examined in the current study. The effects of artificial reproductive technologies on natural fertility rates in other horse populations should be examined to ensure that such practices do not result in long-term reductions in natural fertility levels. Genomic scans for reproduction traits in Thoroughbred and other breeds may assist in understanding genetic variation that influences fertility. We also found that unlike racing performance, there has been little increase in the breeding value of reproductive traits in Thoroughbred horses. Breeding values of foaling rates have decreased in recent generations, possibly because these traits are primarily governed by negative rather than positive selection. Further monitoring of these traits in future generations would assist in understanding the selective forces influencing these traits.

Methods

Reproductive trait data were provided by three large Australian Thoroughbred stud farms that provide a representative sample of the population as a whole. These data included the mating records of 12,922 mares bred to 131 stallions between 2000 and 2017. The scan status of each mare covering (negative or positive at the first scan 15 days after covering) was transformed into a binary trait. The sex of each live foal recorded in the dataset (*n* = 7,578) was also transformed into a binary trait (female=1, male=0). Additionally, more detailed reproductive trait data were available for 152 mares mated to 89 stallions over multiple seasons (*n* = 764 foals), including the date of the foal birth. The gestation length of each live foal birth was calculated from these data.

The pedigree for all the mares, stallions and conceptuses included in the study dating back to the founders of the population consisted of 92,852 records. We used the CFC program (version 1) to reorder the pedigree to ensure that each individual was listed after their parents and to estimate Wright's inbreeding coefficient for all individuals [41]. We also used CFC to estimate the overall average of the average number of generations in the pedigree for each mare and stallion[41].

The genetic and environmental influences on the foaling rate, gestation length and sex ratio were estimated in ASReml-R[42], using a linear mixed model for gestation length, and a generalised liner mixed model for the binary traits of foaling rate and foal sex. Details of the fixed and random effects in each model are included in Table 1. The outcome variables were foaling rate, the sex of the foal and the gestation length. The fixed effects included in each model were: inbreeding coefficient (of the mare, the stallion and the foal), the month of covering, the year of covering, the age of the mare and the stud farm. An animal model was implemented through an inverse relationship matrix using the pedigree in ASReml-R[42]. Multiple coverings of the same mare in the same season were accounted for in the models by inclusion of a permanent environment effect of the mare, as a random effect.

The heritability of each outcome variable was estimated using the variance component estimates of the fitted models. Significance of model terms was evaluated through Wald statistics and the estimated value of each fixed effect tabulated. Estimated breeding values (i.e. genetic values) for foaling rate and gestation length were calculated as best linear unbiased predictions for each individual in the pedigree using the models fitted in ASReml-R. This method uses the available phenotypic data and the associated pedigree structure in the models to provide genetic value estimates for all animals in the pedigree.

References

1 Aurich, C. Reproductive cycles of horses. *Anim. Reprod. Sci.* **124**, 220-228, doi:10.1016/j.anireprosci.2011.02.005 (2011).

2 Kjöllerström, H. J., Gama, L. T. & Oom, M. M. Impact of inbreeding on fitness-related traits in the highly threatened Sorraia horse breed. *Livest. Sci.* **180**, 84-89, doi:10.1016/j.livsci.2015.08.001 (2015).

3 Collins, C. W. *et al.* Abnormal reproductive patterns in Przewalski's mares are associated with a loss in gene diversity. *Biol. Reprod.* **86**, doi:10.1095/biolreprod.111.092676 (2012).

4 Schubert, M. *et al.* Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E5661-E5669, doi:10.1073/pnas.1416991111 (2014).

5 Librado, P. *et al.* The evolutionary origin and genetic makeup of domestic horses. *Genetics* **204**, 423-434, doi:10.1534/genetics.116.194860 (2016).

6 Marsden, C. D. *et al.* Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 152-157, doi:10.1073/pnas.1512501113 (2016).

7 Moyers, B. T., Morrell, P. L. & McKay, J. K. Genetic costs of domestication and improvement. *J. Hered.* **109**, 103-116, doi:10.1093/jhered/esx069 (2017).

8 Metzger, J. *et al.* Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genomics* **16**, 764, doi:10.1186/s12864-015-1977-3 (2015).

9 Gu, J. *et al.* A genome scan for positive selection in Thoroughbred horses. *PLOS ONE* **4**, e5767, doi:10.1371/journal.pone.0005767 (2009).

10 Klemetsdal, G. & Johnson, M. Effect of inbreeding on fertility in Norwegian trotter. *Livest. Prod. Sci.* **21**, 263-272, doi:10.1016/0301-6226(89)90055-9 (1989).

11 Muller-Unterberg, M., Wallmann, S. & Distl, O. Effects of inbreeding and other systematic effects on fertility of Black Forest Draught horses in Germany. *Acta Vet. Scand.* **59**, doi:10.1186/s13028-017-0338-4 (2017).

12 Valera, M., Blesa, F., Santos, R. d. & Molina, A. Genetic study of gestation length in Andalusian and Arabian mares. *Anim. Reprod. Sci.* **95**, 75-96, doi:10.1016/j.anireprosci.2005.09.008 (2006).

13 Ewert, M. *et al.* Determinants of gestation length in Thoroughbred mares on German stud farms. *Anim. Reprod. Sci.* **191**, 22-33, doi:10.1016/j.anireprosci.2018.02.001 (2018).

14 Lazzari, G. *et al.* Embryonic genotype and inbreeding affect preimplantation development in cattle. *Reproduction* **141**, 625-632, doi:10.1530/REP-10-0282 (2011).

15 Leroy, G., Phocas, F., Hedan, B., Verrier, E. & Rognon, X. Inbreeding impact on litter size and survival in selected canine breeds. *Vet. J.* **203**, 74-78, doi:10.1016/j.tvjl.2014.11.008 (2015).

16 Martikainen, K., Tyrisevä, A. M., Matilainen, K., Pösö, J. & Uimari, P. Estimation of inbreeding depression on female fertility in the Finnish Ayrshire population. *J. Anim. Breed. Genet.* **134**, 383-392, doi:10.1111/jbg.12285 (2017).

17 Saura, M. *et al.* Detecting inbreeding depression for reproductive traits in Iberian pigs using genome-wide data. *Genet. Select. Evol.* **47**, 1, doi:10.1186/s12711-014-0081-5 (2015).

18 Zhang, C., MacNeil, M. D., Kemp, R. A., Dyck, M. K. & Plastow, G. S. Putative loci causing early embryonic mortality in Duroc Swine. *Front. Genet.* **17**, 655, doi:10.3389/fgene.2018.00655 (2018).

19 Mc Parland, S., Kearney, J. F., Rath, M. & Berry, D. P. Inbreeding effects on milk production, calving performance, fertility, and conformation in Irish Holstein-Friesians. *J. Dairy Sci.* **90**, 4411-4419, doi:10.3168/jds.2007-0227 (2007).

20 Beckelmann, J. *et al.* Sex-dependent insulin like growth factor-1 expression in preattachment equine embryos. *Theriogenology* **79**, 193-199, doi:https://doi.org/10.1016/j.theriogenology.2012.10.004 (2013).

21 Santos, M. M. *et al.* Sex ratio of equine offspring is affected by the ages of the mare and stallion. *Theriogenology* **84**, 1238-1245, doi:10.1016/j.theriogenology.2015.07.001 (2015).

22 Cunningham, E. P., Dooley, J. J., Splan, R. K. & Bradley, D. G. Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to Throughbred horses. *Anim. Genet.* **32**, 360 (2001).

23 Corbin, L. J. *et al.* Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Anim. Genet.* **41**, 8-15, doi:10.1111/j.1365-2052.2010.02092.x (2010).

24 Todd, E. T. *et al.* Founder-specific inbreeding depression affects racing performance in Thoroughbred horses. *Sci. Rep.* **8**, 6167, doi:10.1038/s41598-018-24663-x (2018).

25 Kuhl, J., Stock, K. F., Wulf, M. & Aurich, C. Maternal lineage of Warmblood mares contributes to variation of gestation length and bias of foal sex ratio. *PLOS ONE* **10**, e0139358, doi:10.1371/journal.pone.0139358 (2015).

26 Der Sarkissian, C. *et al.* Evolutionary genomics and conservation of the endangered Przewalski's horse. *Curr. Biol.* **25**, 2577-2583, doi:10.1016/j.cub.2015.08.032 (2015).

27 Liu, G. *et al.* Evaluating the reintroduction project of Przewalski's horse in China using genetic and pedigree data. *Biol. Conserv.* **171**, 288-298, doi:10.1016/j.biocon.2013.11.022 (2014).

28 Teegen, R., Edel, C. & Thaller, G. Population structure of the Trakehner hrse breed. *Animal* **3**, 6-15, doi:10.1017/S1751731108003273 (2009).

29 Roos, L., Hinrichs, D., Nissen, T. & Krieter, J. Investigations into genetic variability in Holstein horse breed using pedigree data. *Livest. Sci.* **177**, 25 (2015).

30 Bene, S., Benedek, Z., Nagy, S., Szabo, F. & Polgar, P. J. Some effects on gestation length of traditional horse breeds in Hungary. *J. Cent. Eur. Agric.* **15**, 1, doi:10.5513/JCEA01/15.1.1402 (2014).

31 Rodrigues, J. A., Gonçalves, A. R., Antunes, L., Bettencourt, E. V. & Gama, L. T. Genetic and environmental factors influencing gestation length in Lusitano horses. *J Equine Vet Sci.* **84**, 102850, doi: 10.1016/j.jevs.2019.102850 (2020).

32 Christmann, A., Sieme, H., Martinsson, G. & Distl, O. Genetic and environmental factors influencing gestation length and parturition conception interval in Hanoverian warmblood. *Livest. Sci.* **199**, 63-68, doi:10.1016/j.livsci.2017.03.011 (2017).

33 Malo, A. F., Gilbert, T. C. & Riordan, P. Drivers of sex ratio bias in the eastern bongo: lower inbreeding increases the probability of being born male. *Proc. Biol. Sci.* **286**, 20190345-20190345, doi:10.1098/rspb.2019.0345 (2019).

34 Malo, A. F. *et al.* A father effect explains sex-ratio bias. *Proc. Biol. Sci.* **284**, doi:10.1098/rspb.2017.1159 (2017).

35 Douhard, M., Festa-Bianchet, M., Coltman, D. & Pelletier, F. Paternal reproductive success drives sex allocation in a wild mammal. *Evolution* **70**, 358-368, doi:10.1111/evo.12860 (2016).

36 Allen, W. R. & Wilsher, S. Half a century of equine reproduction research and application: A veterinary tour de force. *Equine Vet. J.* **50**, 10-21, doi:10.1111/evj.12762 (2018).

37 Nolan, M. B. *et al.* Artificially extended photoperiod administered to pre-partum mares via blue light to a single eye: observations on gestation length, foal birth weight and foal hair coat at birth. *Theriogenology* **100**, 126-133, doi:10.1016/j.theriogenology.2017.06.012 (2017).

38 Wulf, M. *et al.* Effects of foal sex on some perinatal characteristics in the immediate neonatal period in the horse. *J. Vet. Behav.* **18**, 37-42, doi:10.1016/j.jveb.2016.12.010 (2017).

39 Aurich, C. & Schneider, J. Sex determination in horses-current status and future perspectives. *Anim. Reprod. Sci.* **146**, 34-41, doi:10.1016/j.anireprosci.2014.01.014 (2014).

40 Cameron, E. Z., Linklater, W. L., Stafford, K. J. & Veltman, C. J. Birth sex ratios relate to mare condition at conception in Kaimanawa horses. *Behav. Ecol.* **10**, 472-475, doi:10.1093/beheco/10.5.472 (1999).

41 Sargolzaei, M., Iwaisaki, H. & Colleau, J. J. CFC: a tool for monitoring genetic diversity. *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production* (2006).

42 Gilmour, A. R., Gogel, B. J., Cullis, B. R. & Thompson, R. *ASReml User Guide Release 3.0*. (VSN International Ltd, 2009).

3.3 Figures and tables

Table 1: Mixed models of each fertility trait for Australian Thoroughbred horses. Foal, sire and dam inbreeding were estimated in separate models. Models were also fitted separately for the full and five-generation pedigree. Model description are in R syntax, not full mathematical notation.

Outcome variable	**Model type**	**Model**
Gestation length	LMM	GL ~ mare.age + month + foal.sex + season + *F* + ped(mare) + ide(mare)
Covering success	Binomial GLMM	CS ~ mare.age + month + season + *F* + ped(mare) + ide(mare)
Foal sex ratio	Binomial GLMM	SR ~ mare.age + month + season + *F* + ped(mare) + ide(mare)

LMM = linear mixed model, GLMM = generalised linear mixed model, GL = gestation length, CS = covering success, SR = foal sex ratio, mare.age = age of the mare at the time of covering, month = month of covering, *F* = inbreeding coefficient of the sire, dam or foal (models were run separately for each), ped = pedigree of the mare, ide = permanent environmental effect of the mare.

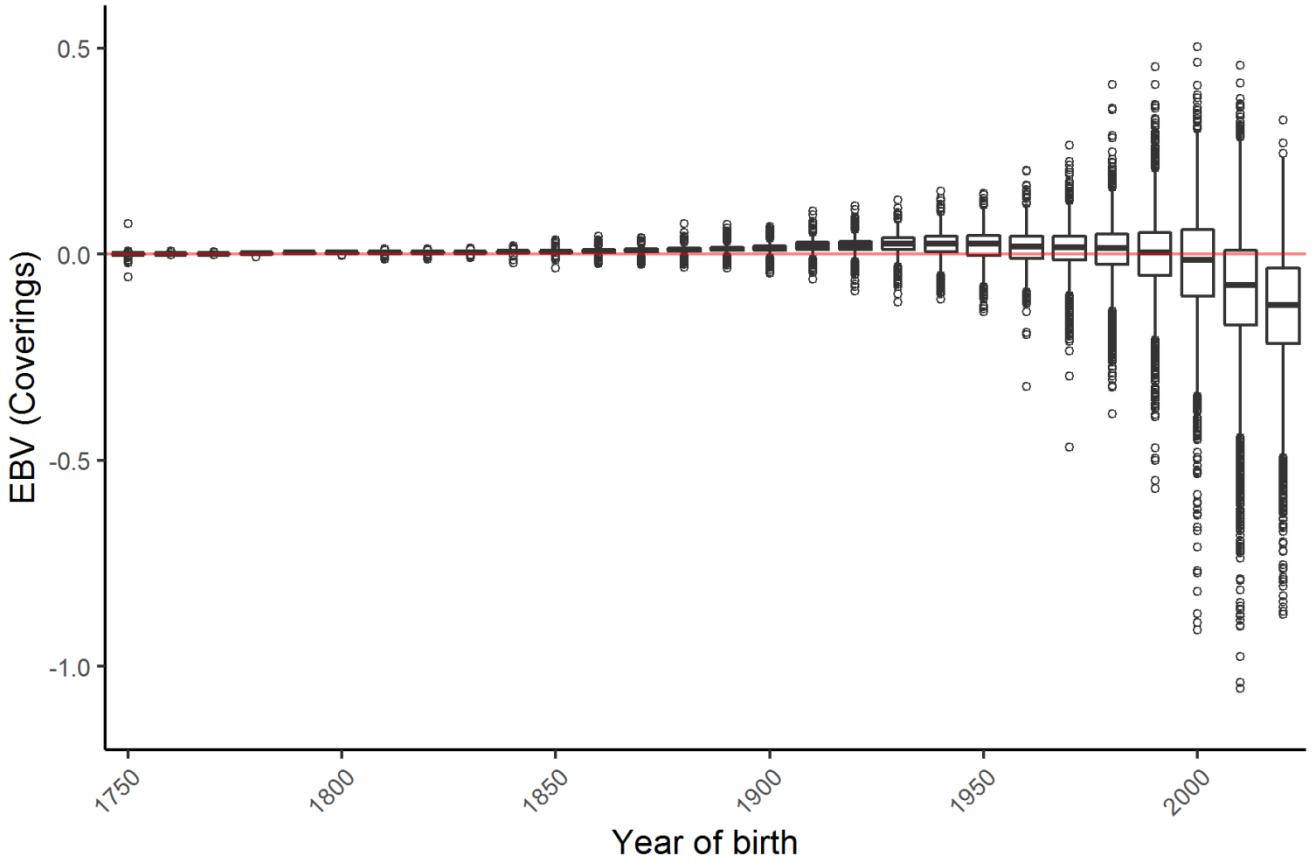

Figure 1: Boxplot of the distribution of estimated breeding values (EBVs) over time for Thoroughbred horses (*n* = 95,663), based on the covering success of 27,962 individuals bred between 2000 and 2017. Each bin represents a 10-year period.

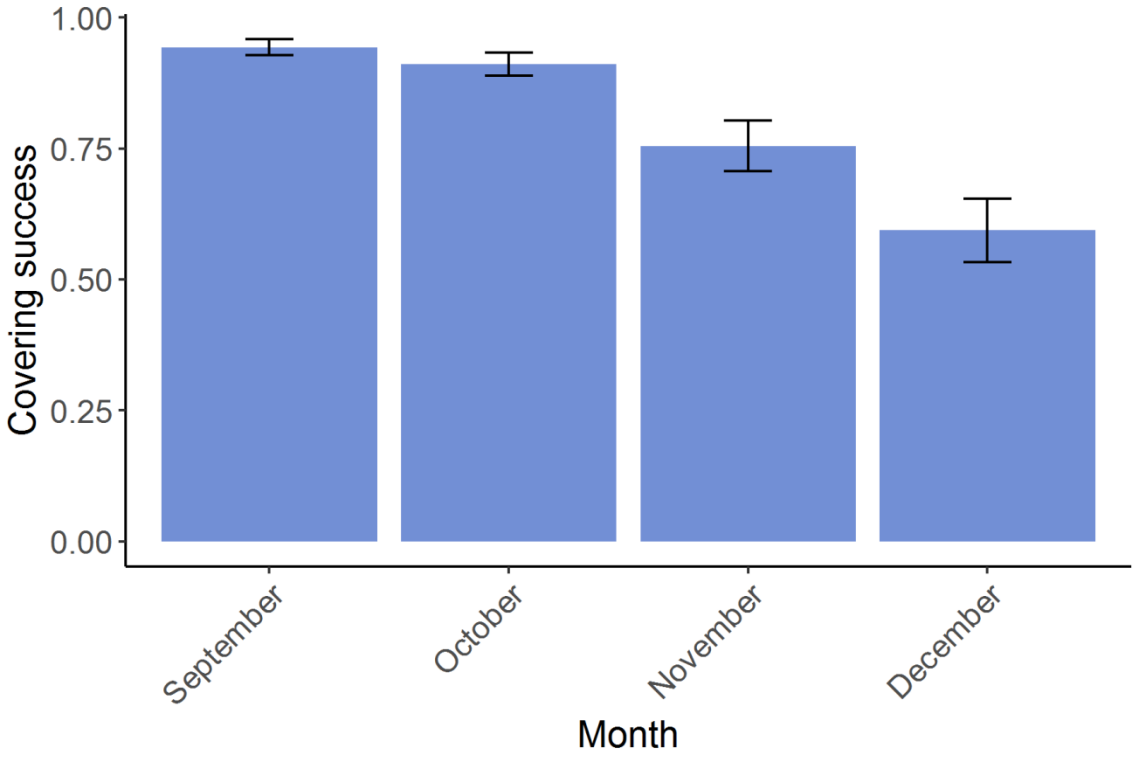

Figure 2: The relationship between the predicted values of covering success by month of covering for Australian Thoroughbred horses between 2000 and 2017 (*n* = 27,962). The error bars represents ± 1 standard error of the predicted value.

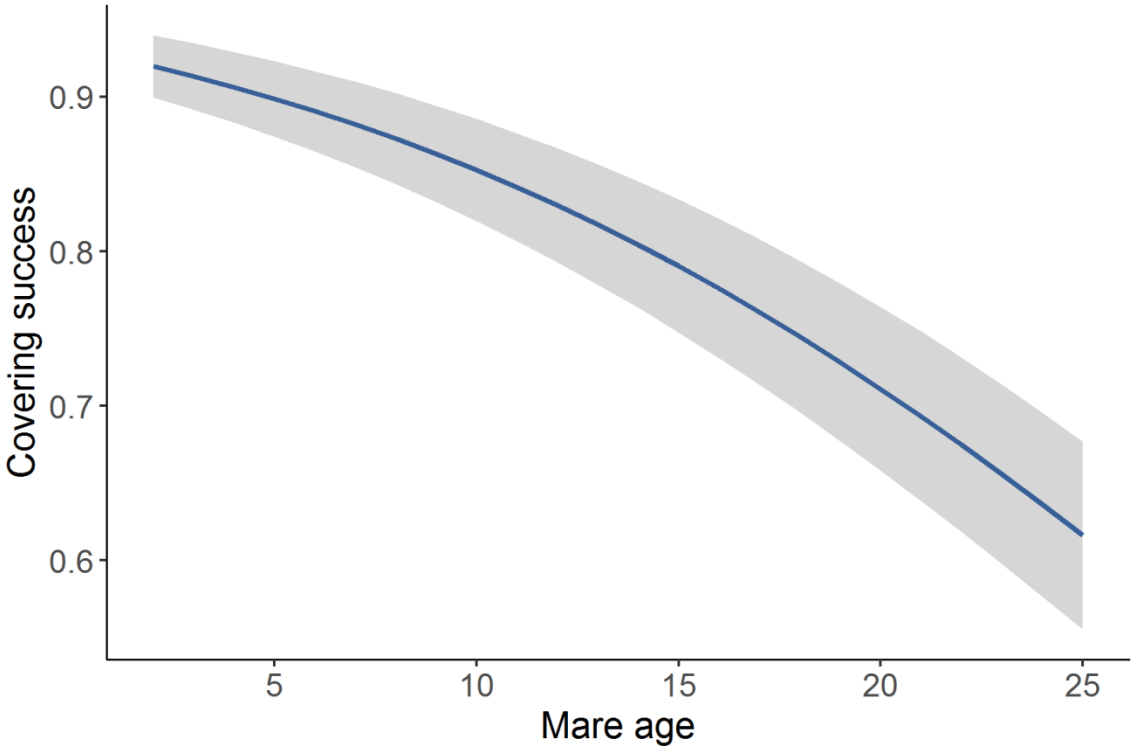

Figure 3: The relationship between the predicted values of covering success and mare age for Australian Thoroughbred horses between 2000 and 2017 (*n* = 27,962). The grey band represents ± standard error of the predicted value.

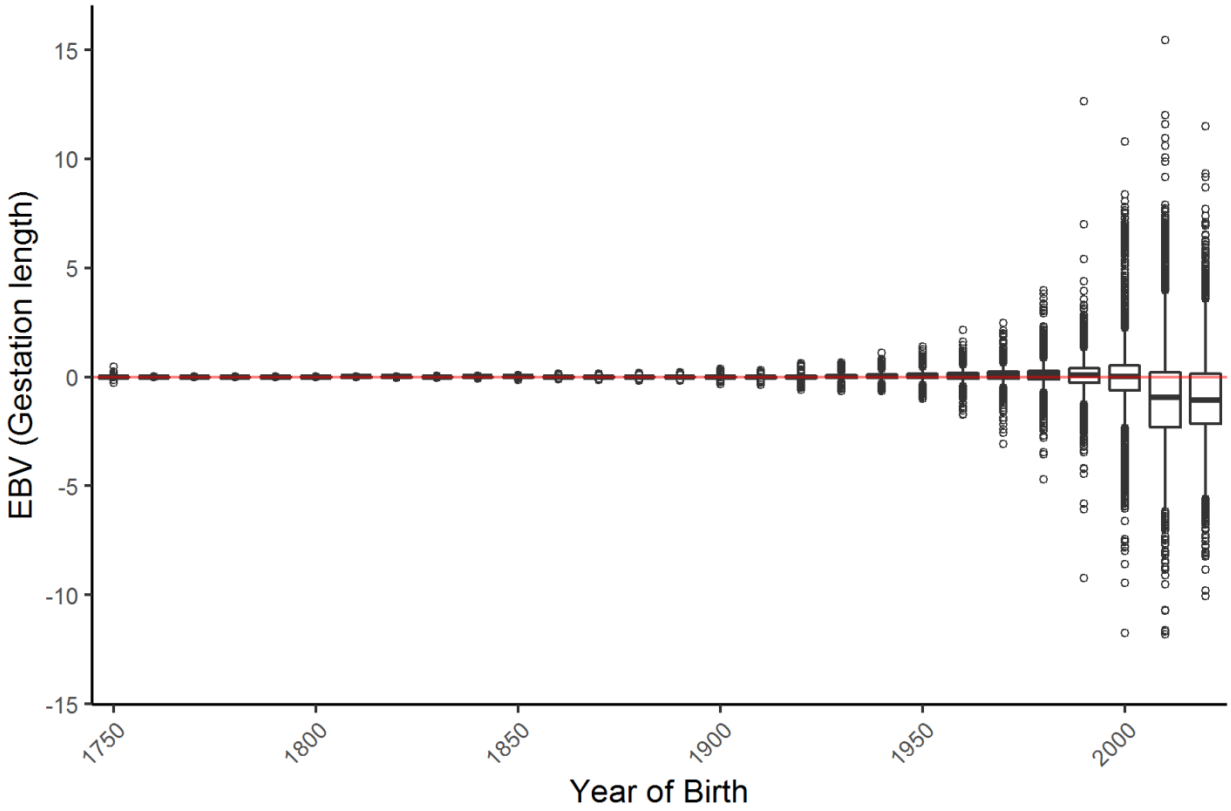

Figure 4: Boxplot of the distribution of estimated breeding values (EBVs) over time for Thoroughbred horses (*n* = 95,663), based on the gestation length of 764 individuals bred between 2000 and 2017. Each bin represents a 10-year period.

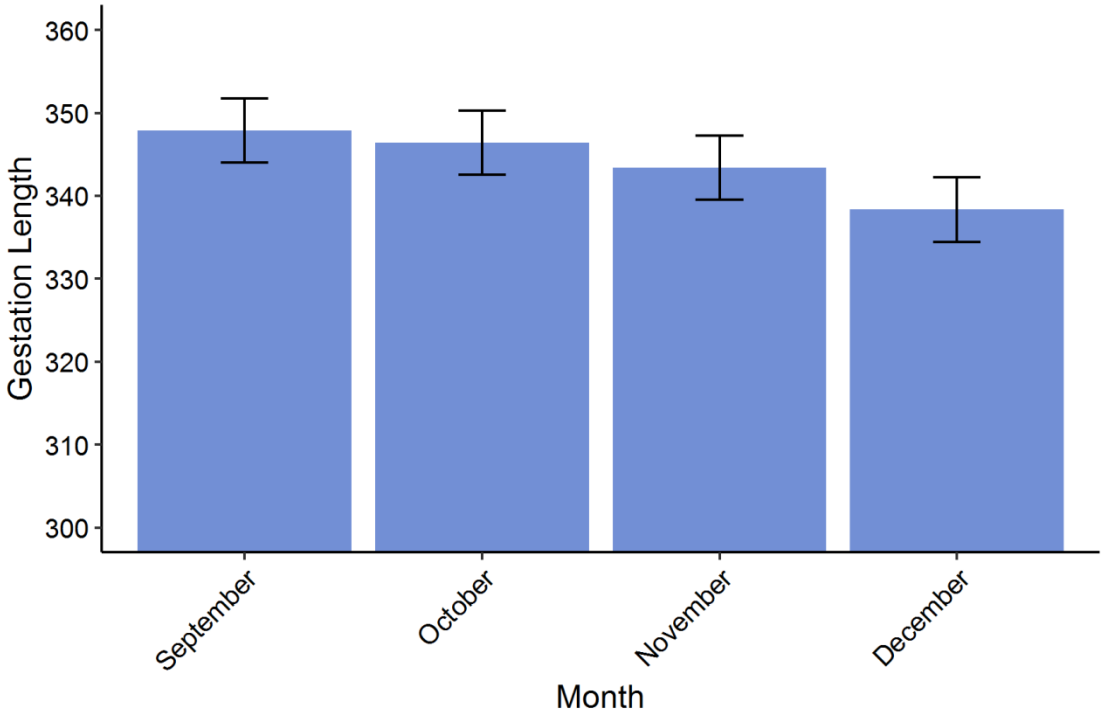

Figure 5: The relationship between the predicted values of gestation length by month of covering for Australian Thoroughbred horses between 2000 and 2017 (*n* = 764). The error bars represents ± 1 standard error of the predicted value.

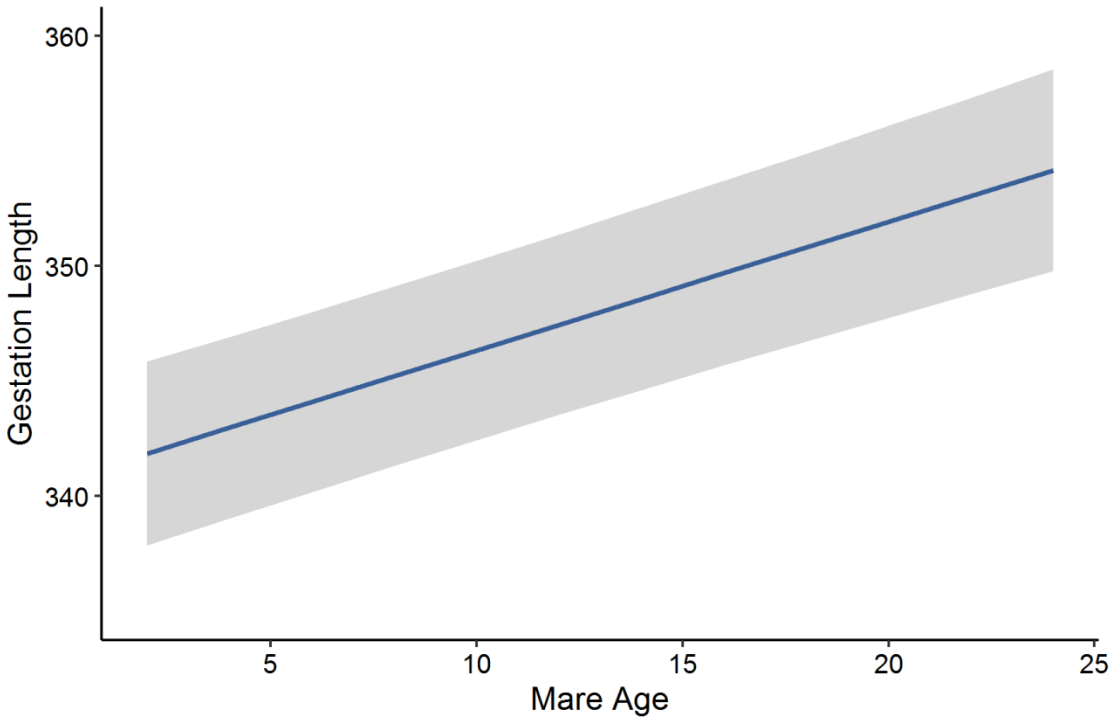

Figure 6: The relationship between the predicted values of gestation length by mare age for Australian Thoroughbred horses between 2000 and 2017 (*n* = 764). The grey band represents ± standard error of the predicted value.

Chapter 4: A genome-wide scan for candidate lethal variants in Thoroughbred horses

4.1 Synopsis

This chapter consists of a manuscript and associated supplementary material prepared for submission to *BMC Genomics*:

E.T. Todd, B.D. Velie, N.A. Hamilton, R.A. Ang, G. Lindgren, Å. Viklund, S. Eriksson, S. Mikko, E. Strand & P.C. Thomson. A genome-wide scan reveals a region in the LY49B gene that may harbour a recessive lethal variant in horses. *(BMC Genomics, under review).*

This manuscript has been formatted for consistency with this thesis.

This chapter aimed to identify haplotypes that may harbour recessive lethal variants and are found at high frequencies in the Thoroughbred horse population. SNP data from commercial genotyping arrays were used to identify variants deviating from the Hardy-Weinberg Equilibrium with a high frequency of heterozygotes and an absence of homozygotes for one allele. Two adjacent SNPs that fit such criteria were identified and mapped to an intronic region in the *LY49B* gene. A similar absence of homozygotes was found for these SNPs in other domestic horse breeds. A variant-calling pipeline was constructed for analysis of 90 publicly-available whole-genome sequences from domestic horses to further investigate this region. These data showed a similar absence of minor homozygotes for the candidate SNPs and identified thee linked variants that may cause loss of function in the *LY49B* gene. Additionally, analysis of transcriptomic data from early embryonic tissues indicated that the *LY49B* gene is expressed in trophectoderm tissue during placentation and may be crucial for implantation success.

I designed this project, carried out the analyses using custom scripts generated in R, PLINK and Linux and wrote the manuscript. Ms Rachel Ang and Drs Natasha Hamilton, Brandon Velie, Gabriella Lindgren, Åsa Viklund, Susanne Eriksson, Sofia Mikko and Eric Strand assisted in the collection and genotyping of the DNA samples. Drs Peter Thomson, Natasha Hamilton and Brandon Velie assisted in the project design and with finalizing the manuscript.

2.2 Main article

A genome-wide scan for candidate lethal variants in Thoroughbred horses.

Evelyn T. Todd[1*], Natasha A. Hamilton[2], Rachel A. Ang[1], Gabriella Lindgren[3], Åsa Viklund[3], Susanne Eriksson[3], Sofia Mikko[3], Eric Strand[4], Peter C. Thomson[1] and Brandon D. Velie[1].

[1]School of Life and Environmental Sciences, The University of Sydney, NSW 2006, Australia.

[2]Racing Australia Equine Genetics Research Centre, Racing Australia, NSW 2000, Australia.

[3]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden.

[4]Livestock Genetics, Department of Biosystems, KU Leuven, Leuven, Belgium.

[5]Department of Companion Animal Clinical Sciences, Faculty of Veterinary Medicine, Norwegian University of Life Sciences, Oslo, Norway.

Abstract

Background

Recessive lethal variants often segregate at low frequencies in animal populations, such that two randomly selected individuals are unlikely to carry the same mutation. However, the likelihood of an individual inheriting two copies of a recessive lethal mutation is dramatically increased by inbreeding events. Such occurrences are particularly common in domestic animal populations, which are often characterised by high rates of inbreeding and low effective population sizes. To date there have been no published investigations into the presence of specific variants at high frequencies in domestic horse populations. This study aimed to identify potential recessive lethal haplotypes in the Thoroughbred horse breed, a closed population that has been selectively bred for racing performance.

Results

In this study, we scanned genotype data from Thoroughbred horses (*n* = 526) for adjacent single nucleotide polymorphisms (SNPs) at high heterozygote frequencies, but with a complete absence of homozygotes. Two SNPs that matched these criteria were mapped to an intronic region in the *LY49B* gene, indicating that a closely linked mutation may cause lethality in homozygous state. Despite a complete absence of homozygotes, almost 35% of Thoroughbreds included in these analyses were heterozygous for both SNPs. A similar loss or absence of homozygotes was observed in genotype data from other domestic horse breeds (*n* = 2025). Variant analysis of whole-genome sequence data (*n* = 90)

identified two SNPs in the 3'UTR region of the *LY49B* gene that may result in loss of function. Analysis of transcriptomic data from equine embryonic tissue revealed that *LY49B* is expressed in the trophoblast during placentation stage of development.

Conclusions

In this study, a region in the *LY49B* gene was identified as a strong candidate for harbouring a variant causing lethality in homozygous state. These findings suggest that *LY49B* may have an essential, but as yet unknown function in the implantation stage of equine development. Further investigation of this region may allow for the development of a genetic test to improve fertility rates in horse populations. Identification of other lethal variants could assist in improving natural levels of fertility in horse populations.

Introduction

There is estimated to be a high rate of natural embryonic mortality in mammals. A large proportion of these embryonic losses occur soon after fertilisation, such that pregnancies often go undetected, with the only sign being reduced fertility[1]. Mutation screens in mice reveal that many genes are essential for development, with knockout of 29% of genes tested resulting in embryonic death by day 14[2,3]. Although mutations in these genes are expected to be under strong negative selection due to being completely deleterious, many species are estimated to carry between one and two recessive lethal mutations per genome[4]. However, single mutations are often uncommon in a population, such that unrelated individuals are unlikely to carry the same recessive lethal mutations[5-7].

The likelihood of an individual inheriting two copies of the same lethal mutation is dramatically increased by inbreeding events, whereby alleles that are identical by descent are inherited from a common ancestor[8-10]. Populations with high levels of inbreeding and low effective population size (N_e) are at risk of these otherwise rare lethal mutations drifting to high frequencies. Endangered and fragmented species are particularly susceptible due to their small population size and low genetic diversity[11]. Recessive lethal mutations can also been found at high frequencies in domestic animal populations despite their large census size due to inbreeding for the selection of desirable traits[12].

In recent years, a number of studies in livestock have identified embryonic lethal mutations at high frequencies due to intensive selective breeding practices[13-20]. This is often due to a limited number of sires with desirable characteristics making large genetic contributions to the population[13,21]. Moreover,

population bottlenecks due to domestication and breed formation have also resulted in increased deleterious mutation loads and diminished gene pools in many domestic breeds[12,22-24]. These processes lead to genetic variation and N_e being reduced in future generations, which increases the chance of drift and inbreeding events. Lethal mutations that have reached high frequencies are often detected by deviations from the Hardy-Weinberg equilibrium with a lack of homozygotes for one allele[14]. Characterisation of such mutations can assist in improving breeding decisions to increase fertility rates in these populations and prevent these mutations from drifting to higher frequencies[25,26].

The identification of high frequency lethal variants is of particular interest in domestic horse populations. Although a recent study has identified some candidate mutations[27], to date there has been no published comprehensive characterisation of common embryonic lethal alleles in horse populations. Despite the large variety of domestic horse breeds found throughout the world, many breeds suffer from low within-breed diversity and small N_e[28-30]. Genetic variation in most horse breeds has decreased markedly within the last 200 years, largely due to their replacement by machinery in agriculture and transport[31,32]. Some horse breeds with large consensus population sizes also experience low N_e and genetic diversity due to intense artificial selective breeding practices and closed population structures[29,30]. Maintaining good fertility rates is particularly important for horse populations due to the seasonal nature of breeding and the low individual fertility output, as mares produce only one foal from an eleven month gestation period[33]. Despite the extensive use of hormonal therapies to increase covering success in many domestic horse populations, per cycle pregnancy rates in some breeds only average around 65%, suggesting the presence of unknown variables that may reduce fertility[34].

In this study, we aimed to characterise variants at high frequencies that may cause lethality in the Thoroughbred horse population. The Thoroughbred breed is of particular interest due to the closed population structure since the foundation of the studbook in the 18th century[35]. The population has since been intensely selected for the improvement of athletic abilities[36,37], resulting in contemporary Thoroughbred horses being characterised by high levels of inbreeding and a small N_e[30,38-40]. Due to selective breeding practices, all Thoroughbred horses can trace their ancestry back to a small number of individuals from the foundation of the breed[38,39]. Genetic diversity in the Thoroughbred breed has been reduced in recent decades due to the increased commercialisation of popular stallions providing large genetic contributions to the population[41]. Although such practices are in line with selective breeding principles[42], they could also inadvertently increase the frequency of embryonic lethal variants in the population. Reproductive technologies such as artificial insemination are banned in the Thoroughbred

population, making the maintenance of high levels of natural fertility imperative. Additionally, Thoroughbred horses have been used as foundation stock for other popular horse breeds including The Quarter Horse, Standardbred, and Warmblood[29], thus identification of embryonic lethal variants in Thoroughbreds is also likely to assist in the breeding management of these populations. Therefore, we also aimed to determine the frequency of any potentially lethal variants identified in the Thoroughbred population in other horse breeds and examine their transcriptomic profile in embryonic tissue.

Results

Identifying candidate lethal SNPs at high frequencies in Thoroughbred horses

Analysis of genotype data from Thoroughbred horses (*n* = 156) identified only two adjacent, linked SNPs that significantly deviated from the Hardy-Weinberg equilibrium with an absence of homozygotes (Table 1). Under Hardy-Weinberg equilibrium, seven minor allele homozygotes were expected for both of these SNPs in the dataset. Genotype data from Japanese Thoroughbred horses (*n* = 370) also showed an absence of homozygotes for these SNPs (Table 1). In this dataset, the expected number of minor homozygotes for these SNPs under Hardy-Weinberg equilibrium was nine. Despite a complete absence of homozygotes, almost 35% of Thoroughbreds across both of these datasets were heterozygous for this two-SNP haplotype. These SNPs also showed an absence or reduction of minor homozygotes in genotype data for other domestic horse breeds (*n* = 2025, Table 2).

These two candidate SNPs mapped to the coordinates of 6:38278097 and 6:38278874, which are found in an intronic region of the *LY49B* gene on ECA6. This gene is part of the *LY49* gene family, which plays an important role in innate immunity. There are five functional members of the *LY49* gene family in *Equus caballus*, all of which closely grouped together on chromosome 6. Since both of the SNPs mapped to a non-coding region of the *LY49B* gene, the likelihood of either being a causal variant for lethality is low.

Phylogenetic origin of the candidate SNPs

According to the phylogenetic tree generated by Petersen et al[28,29], and their associated SNP data, the SNPs of interest were present in heterozygous state across most phylogenetic branches of domestic horse breeds. Of the 32 breeds in this dataset, 23 had at least one heterozygote for both SNPs of interest. Notably, this two-SNP haplotype was not found in genotype data from one branch of the tree which contains the North Swedish Horse (*n* = 19), Norwegian Fjord Horse (*n* = 21) and Exmoor Pony (*n* =

24) (Table S1). A larger sample of Exmoor Pony data (*n* = 274, Table 2) found only one heterozygote for this haplotype.

Frequency of the candidate SNPs in other breeds

Analysis of SNP data from other domestic breeds showed that heterozygotes for the SNPs of interest were at a particularly high frequency in the Quarter Horse population (71%) (Table 2, Table S2). The proportion of heterozygotes was also high in Swedish Warmbloods (*n* = 370) and Coldblooded Trotters (*n* = 641), being 20% and 40% respectively (Table 2). Smaller datasets also revealed that Belgian Draft (*n* = 19), French Trotter (*n* = 17), Paint (*n* = 15), Morgan (*n* = 19), Mongolian Paulista (*n* = 19) and Tuva (*n* = 15) breeds may also have a high proportion of heterozygotes for this haplotype in their populations (Table S1).

Identifying candidate causal variants using whole-genome sequence data

To further investigate SNP frequencies in this region, variants were called from whole-genome sequence data of 90 domestic horses. The two SNPs identified in the preliminary analysis showed a complete absence of homozygotes for their minor alleles in these individuals (Table S3). Additionally, a number of variants closely linked to these SNPs were identified (Table 3, Figure S1). Annotation of these loci using SIFT[43] identified three variants that may result in changes to protein structure or expression, so these represent the most likely candidates to cause lethality in homozygous state (Figure 1).

The first of these variants, 6:38282610G>A, was located in an intronic region of the *LY48B* gene and resulted in an amino acid change from a phenylalanine to a serine residue. This SNP is located next a tryptophan residue that appears to be highly conserved across members of the *LY49* family and across species. However, there is little conservation of the phenylalanine residue across taxa; some species have a phenylalanine and others a serine at this position. This SNP is annotated as being “tolerated” in SIFT.

Two other variants that were closely linked to the candidate SNPs (6:38276742A>T and 6:38276955G>A) were found within the 3’UTR region of the *LY49B* gene. Alignment of the 3’UTR region of the five functional *LY49* genes in *Equus caballus* revealed that the region containing the SNP 6:38276955G>A is highly conserved in all members of the *LY49* gene family (Table 4). This region may be important for mRNA stability and translation into a functional protein. The other variant was found in an AU-rich

region at the end of the *LY49B* mRNA transcript, which is often associated with polyadenylation and post translation stability.

Transcriptomic analysis of RNA sequence data

Measurable levels of *LY49B* mRNA were not detected in equine trophectoderm tissue collected on day 16 of development. However, *LY49B* mRNA was observed in trophectoderm tissue collected on days 23 and 24 of development (Table 5). Additionally, *LY49B* mRNA transcripts were detected in microarray data from equine chorion and chorionic girdle tissue between days 27 and 34 of development (Table S4). Inner cell mass tissue collected on days 15, 22 and 25 of development did not show any measurable transcription of *LY49B* (Table 5). The genotypes of the SNPs of interest in the samples included in the mRNA analysis are unknown.

Discussion

In this study, we aimed to identify variants at high frequencies in the Thoroughbred horse population that may be lethal in homozygous state. Analyses of genotype data from Thoroughbred horses identified only two adjacent SNPs that fit the strict filtering criteria (Table 1). Genotype data from other horse breeds showed a similar reduction or absence of homozygotes for these SNPs (Table 2). Therefore, this two-SNP haplotype is a strong candidate for harbouring a variant that causes lethality in homozygous state.

The SNPs identified in the preliminary analysis mapped to an intronic region in the *LY49B* gene on ECA6. The *LY49B* gene belongs to the *LY49* (Killer cell lectin-like receptor subfamily A) family of receptors, which consists of five functional members in *Equus caballus*[44]. Other species (including humans) have a functionally similar, but structurally different gene family called *KIR* (Killer cell immunoglobin line receptors)[45]. The *LY49/KIR* gene family are expressed across various types of immune cells, and mediate their function through bindings to MHC-1[46]. The *LY49B* gene is expressed in myeloid cells where it regulates their activity through an inhibitory effect, possibly to prevent their spontaneous activation[47]. Despite the important role that they play in immunity, the function of *LY49* genes in development is currently unknown. In humans, incompatibilities between foetal *KIR* and maternal *MHC* (*HLA*) genotypes are associated with an increased risk of miscarriage and preeclampsia[48-50]. Additionally, knockdown of *LY49* in mice showed a high rate of implantation failure[51,52]. These findings indicate that *LY49B* may play an important role in maternal/foetal compatibility and implantation success in horses.

Analysis of transcriptomic data found that *LY49B* was first expressed in equine trophoblast tissue during the placental development stage. The first evidence of *LY49B* expression was found on day 23-24 of development (Table 5), during which the glycoprotein capsule surrounding the embryo is broken down and placental tissue starts to develop[53]. Measurable expression of *LY49B* was also found in chorion and chorionic girdle tissues between days 27 and 34 of development (Table S4). During this time, trophoblast cells rapidly proliferate to form the chorionic girdle, which then invades the endometrium to form epithelial cups[54]. It is possible that *LY49B* is important for successful implantation of the embryo by mediating the action of MHC-1 which is expressed during this time[55,56]. However, it is also possible that loss of function in *LY49B* may result in post-natal juvenile death, which would explain the lack of homozygotes seen in our data. High rates of juvenile death would be more discernible in the population, so loss of function leading to embryonic death seems more likely. Further investigations into the role of *LY49B* in equine development would confirm whether impaired function causes lethality and the stage of development at which this occurs.

Variant calling in whole-genome sequence data from 90 domestic horses further confirmed an absence of minor homozygotes for the two SNPs of interest. Three variants closely linked to these SNPs were also identified in these data as the most likely candidates to cause loss of function in the *LY49B* gene and result in lethality in homozygous state (Table 3). One SNP was a missense mutation in the coding region of the *LY49B* gene that results in the substitution of a negatively charged serine for an aromatic phenylalanine residue. However, lack of conservation of this SNP in *LY49* genes across taxa makes it seem unlikely to be a causative mutation for embryonic lethality. Two other variants found in the 3'UTR region of the *LY49B* gene were also closely linked to the SNPs identified in the preliminary analysis, and seemed more likely candidates to cause embryonic lethality in homozygous state.

The 3'UTR region of a gene is responsible for transcriptional stability through the binding of miRNAs and RNA binding proteins[57]. The addition of the polyadenylation tail to the 3'UTR is also essential to ensure proper processing and translation of the mRNA strand[58]. Mutations in the 3'UTR region can lead to degradation of the mRNA, resulting in reduced or inhibited translation even when the gene is transcribed[59]. Variation in the 3'UTR region of genes are associated with a number of diseases including Huntington's and breast cancer in humans[60,61]. Additionally, SNPs in the 3'UTR region are associated with production traits in livestock including milk production in cows, muscularity in sheep and obesity in horses[59,62,63].

Despite the importance of the 3'UTR region for the mRNA stability and normal expression of a gene, little is known about how specific polymorphisms can affect post-transcriptional processing. This makes it difficult to identify how the 3'UTR variants identified in this study could affect the translation of *LY49B* mRNA into a functional protein. The 3'UTR variant 6:38276955G>A was identified as a possible candidate for embryonic lethality because it is highly conserved between all members of the *LY49B* family (Table 4) in horses, so may play an important role in mRNA stability. The other 3'UTR variant (6:38276742A>T) is found in an AU-rich region at the end of the transcript, so may be important in the addition of the polyadenylation tail. Further examination of the effects that these variants have on post-transcriptional processing would determine if they impact the normal expression of *LY49B* in horses.

Despite an absence of homozygotes, the two SNPs identified in this study were found at high heterozygote frequencies in the Thoroughbred population. Currently, there is no evidence that variation in the *LY49B* gene is associated with phenotypic advantages in horses. However, it is possible that one of the variants linked to these SNPs confers a heterozygote advantage, which could explain why they have reached such high frequencies in the breed. It is also possible that selective breeding practices favouring a limited number of stallion bloodlines are responsible for this potentially lethal haplotype drifting to high frequencies in the Thoroughbred population. This would be most likely to occur if a stallion that made a large genetic contribution to the population was a carrier. A similar instance has recently been documented in cattle, where a lethal mutation at high frequencies was traced back to a sire with an extensive genetic influence on a population[21].

The presence of this potentially lethal haplotype across many diverse breeds of domestic horses indicates that it may not be the result of a recent mutation present only in the Thoroughbred population. Rather, this haplotype may have been present in pre-domesticated horses as a rare variant, and has become more frequent in some domestic breeds as the result of population bottlenecks due to breed formations, selective breeding practices and potentially a heterozygote advantage. Domestication and breed formation events have been well documented to result in increased deleterious mutation loads in horses and other domestic species[24,30,31,64]. A high proportion of heterozygotes for this haplotype were found in some breeds closely related to the Thoroughbred including the Paint, French Trotter, Morgan and Quarter Horse. Notably, over 70% of Quarter Horse samples included in this study were heterozygous for these SNPs. The Quarter Horse has an open stud book, and higher genetic diversity than the Thoroughbred population[65], making the high frequency of a potentially lethal haplotype at first surprising. The Quarter Horse dataset reportedly did not contain full or half siblings[65],

but the collection of samples from one geographical area may not fully reflect the diversity of the worldwide population. An average relatedness analysis of these samples noted the large genetic influence of one particular Thoroughbred stallion[65], which may explain the high frequency of heterozygotes observed in this population. However, the extremely high frequency of heterozygotes in this breed is more likely to be due to a heterozygote advantage.

The Belgian Draft, Mangalarga Paulista and Tuva breeds also show a high proportion of heterozygotes, but are more distantly related to the Thoroughbred and to each other. Therefore, the high frequency of heterozygotes in these breeds may be due to independent genetic drift events. Heterozygotes for the SNPs of interest were notably absent from one branch of the tree containing small heavy horses from Northern Europe, which are more distantly related to the Thoroughbred. A larger dataset of Exmoor Pony samples from this phylogenetic branch revealed one heterozygote for this haplotype (Table 2). This could be due to a calling error, but it is also possible that these SNPs exist at very low frequencies in these breeds. The small sample size of the genotype data for many individual breeds in this study means that heterozygote frequencies across all subpopulations found throughout the world may deviate from that reported here. However, these data provide an indication of breeds with high proportions of heterozygotes for this region. Overall, our findings suggest that this region shows evidence harbouring a homozygous lethal variant, yet a high proportion of heterozygotes are found across many domestic horse breeds.

Conclusions

In this study, we identified a haplotype at high frequencies in the Thoroughbred horse population that is a strong candidate for harbouring a variant causing lethality in homozygous state. Similar analyses on larger datasets in other livestock populations have identified multiple lethal haplotypes, so it is likely that other such variants are present at high frequencies in the Thoroughbred population but were not captured in this study. Additionally, the use of commercial SNP arrays only allows for the identification of variants with high minor allele frequencies in populations. Analysis of larger sample sizes, and using higher density genotype data could allow for identification of other variants associated with lethality in domestic horses. The identification of this potentially lethal haplotype demonstrates the potential implications of heavily favouring a limited number of bloodlines in selective breeding practices. The association of this haplotype with lethality can be used to assist in breeding decisions to improve mating outcomes in the Thoroughbred population. Although there are a high proportion of heterozygotes in some domestic breeds, limiting the use of stallions that are carriers could reduce the frequency of this

haplotype in future generations. Additionally, further investigations into a possible heterozygote advantage could assist in understanding the high frequency of this haplotype across many domestic breeds. Further characterisation of lethal haplotypes in other breeds would also assist in breeding management to increase fertility rates in domestic horse populations.

Methods

Initial genotyping

Genotype data from a representative sample of Thoroughbreds were used to identify SNPs with a high proportion of heterozygotes, but an absence of homozygotes for one allele. Genome-wide SNP data were generated for 156 Australian Thoroughbred horses by genotyping them on either the Illumina 70K Chip (65,102 SNPs) (*n* =102) or the Affymetrix 670K Chip (670,796 SNPs) (*n* = 54). Common genotyped SNPs between the two arrays were scanned for deviations from the Hardy-Weinberg equilibrium with an absence of homozygotes for one allele using PLINK (version 1.9)[66]. The *P*-values were adjusted using a false discovery rate correction with the R package "qvalue"[67]. Since SNPs with an absence of homozygotes could indicate a calling error, the search was narrowed to only include adjacent SNPs that fit such criteria.

The frequencies of the candidate SNPs were then examined in publicly available genotype data from Japanese Thoroughbreds (*n* = 370) typed on the Affymetrix 670K Chip[68] and these were added the Thoroughbred sample. The SNP frequencies were then characterised from genotype data from Swedish Warmbloods (*n* = 380)[69] and Coldblooded Trotters (*n* = 670)[70] typed on the Affymetrix 670K Chip. Publicly available data from Exmoor Ponies (*n* = 285, typed on the Affymetrix 670K Chip)[70], Quarter Horses (*n* = 137, typed on the Illumina 70K Chip)[71] and horses of 32 different domestic breeds (*n* = 582, typed on the Illumina 50K Chip)[28,29] were also included in this preliminary scan for SNP frequencies. In these data, raw intensities were plotted to check for calling errors. If potential calling errors were detected, SNPs were recalled using a mixture model fitted with an expectation-maximization algorithm in R.

Variant discovery and mapping

Publicly available whole-genome sequence data were used to further examine the frequencies of the candidate SNPs identified in the initial genotype analysis, and to identify linked variants. Paired end whole-genome sequence data from 90 horses of different domestic breeds were used in this analysis

(Table S5). The whole-genome datasets were downloaded from the European Nucleotide Archives (ENA, https://www.ebi.ac.uk/ena) which included horses of different domestic breeds (PRJEB14779, *n* = 70) and additional Thoroughbred samples (PRJNA168142, *n* = 16 and PRJNA184688, *n* = 4) (Table S5).

The SNP array used in the initial genotyping analysis was developed based on coordinates of the EquCab2.0 reference genome. For consistency, we used the EquCab2.0 assembly as a reference for the whole-genome sequence analysis. The raw reads were mapped to the EquCab2.0 reference genome using BWA-MEM algorithm from Burrows-Wheeler Alignment Tool (version 0.7.17)[72]. Duplicate reads were flagged using Samblaster (version 0.1.22)[73], and base recalibration was performed using Genome Analysis Toolkit (GATK) (version 4.0.8.1)[74]. Variants (SNPs and INDELs (insertions and deletions)) were called using Haplotype Caller and then filtered using the standard hard filtering recommendations in GATK[75]. The individual SNPs were then filtered to only include high quality allele calls with an average filtered depth over 10 and a Phred score over 20.

Variants that were linked to the SNPs identified from the genotype data were produced using the LD function in PLINK (version 1.9) with a window size of 5 Mb[66]. Only SNPs with an r^2 value of over 0.8 and a D' value > 0.9 were shortlisted. The effects of each SNP on gene structure and function was characterised using SIFT (version 4G)[43]. The conservation of variants across taxa was analysed using the NCBI Conserved Domain Database Search[76].

Transcriptomic analysis

Publicly available RNA sequence data were used to examine expression levels of the genes of interest in embryonic tissue. The data included equine inner cell mass tissue (collected at day 15, 22 and 25, *n* = 3) and trophectoderm tissue (collected at day 16, 23 and 24, *n* = 3) from the Functional Annotation of ANimal Genomes (FAANG) equine biobank (available from ENA under the project name PRJNA223157)[77]. Adaptors were trimmed using bbduk from BBtools (version 37.98)[78]. Reads were aligned to the EquCab 2.0 genome using STAR (version 2.7.2b)[79]. Counts were generated using featurecounts from subread (version 1.5.1)[80], then quantified in fragments per kilobase/million (FPKM) using the R package "edgeR"[81] with the Equus_caballus_Ensembl_94 file used for annotation. Microarray data for chorion (*n* = 19) and chorionic girdle (*n* = 19) tissue collected from horse embryos between days 27-34 of development[82] were also examined for gene expression levels.

References

1 Spencer, T. E. Early pregnancy: concepts, challenges, and potential solutions. *Anim. Front.* **3**, 48-55, doi:10.2527/af.2013-0033 (2013).

2 White, J. K. *et al.* Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell* **154**, 452-464, doi:10.1016/j.cell.2013.06.022 (2013).

3 Mary, E. D. *et al.* High-throughput discovery of novel developmental phenotypes. *Nature* **537**, doi:10.1038/nature19356 (2016).

4 Gao, Z., Waggoner, D., Stephens, M., Ober, C. & Przeworski, M. An estimate of the average number of recessive lethal mutations carried by humans. *Genetics* **199**, 1243-1254, doi:10.1534/genetics.114.173351 (2015).

5 Ballinger, M. A. & Noor, M. A. F. Are lethal alleles too abundant in humans? *Trends Genet.* **34**, 87-89, doi:10.1016/j.tig.2017.12.013 (2018).

6 Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. Mutation rate and dominance of genes affecting viability in Drosophila melanogaster. *Genetics* **72**, 335 (1972).

7 Sulem, P. *et al.* Identification of a large set of rare complete human knockouts. *Nat. Genet.* **47**, 448-452, doi:10.1038/ng.3243 (2015).

8 Curik, I., Ferencakovic, M. & Solkner, J. Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest. Sci.* **166**, 26-34, doi:10.1016/j.livsci.2014.05.034 (2014).

9 Keller, L. F. & Waller, D. M. Inbreeding effects in wild populations. *Trends Ecol. Evol.* **17**, 230-241, doi:10.1016/S0169-5347(02)02489-8 (2002).

10 Charlesworth, D. & Willis, J. H. The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**, 783-796, doi:10.1038/nrg2664 (2009).

11 Trask, A. E. *et al.* Evidence of the phenotypic expression of a lethal recessive allele under inbreeding in a wild population of conservation concern. *J. Anim. Ecol.* **85**, 879-891, doi:10.1111/1365-2656.12503 (2016).

12 Bosse, M., Derks, M. & Groenen, M. Deleterious alleles in the context of domestication, inbreeding, and selection. *Evol. Appl.* **12**, 6-17, doi:10.1111/eva.12691 (2019).

13 Hoff, J. L., Decker, J. E., Schnabel, R. D. & Taylor, J. F. Candidate lethal haplotypes and causal mutations in Angus cattle. *BMC Genomics* **18**, doi:10.1186/s12864-017-4196-2 (2017).

14 VanRaden, P. M., Olson, K. M., Null, D. J. & Hutchison, J. L. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *J. Dairy Sci.* **94**, 6153-6161, doi:10.3168/jds.2011-4624 (2011).

15 Fasquelle, C. *et al.* Balancing selection of a frame-shift mutation in the MRC2 gene accounts for the outbreak of the crooked tail syndrome in Belgian blue cattle. *PLoS Genet.* **5**, e1000666, doi:10.1371/journal.pgen.1000666 (2009).

16 Zhang, C., MacNeil, M. D., Kemp, R. A., Dyck, M. K. & Plastow, G. S. Putative loci causing early embryonic mortality in Duroc swine. *Front. Genet.* **17**, 655, doi:10.3389/fgene.2018.00655 (2018).

17 Charlier, C. *et al.* NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. *Genome Res.* **26**, 1333-1341, doi:10.1101/gr.207076.116 (2016).

18 Sonstegard, T. S. *et al.* Identification of a nonsense mutation in CWC15 associated with decreased reproductive efficiency in Jersey cattle. *PLOS ONE* **8**, e54872, doi:10.1371/journal.pone.0054872 (2013).

19 Derks, M. F. L. *et al.* A systematic survey to identify lethal recessive variation in highly managed pig populations. *BMC Genomics* **18**, doi:10.1186/s12864-017-4278-1 (2017).

20 Derks, M. F. L. *et al.* Loss of function mutations in essential genes cause embryonic lethality in pigs. *PLoS Genet.* **15**, e1008055, doi:10.1371/journal.pgen.1008055 (2019).

21 Bourneuf, E. *et al.* Rapid discovery of de novo deleterious mutations in cattle enhances the value of livestock as model species. *Sci. Rep.* **7**, 11466, doi:10.1038/s41598-017-11523-3 (2017).

22 Schubert, M. *et al.* Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E5661-E5669, doi:10.1073/pnas.1416991111 (2014).

23 Cruz, F. *et al.* The legacy of domestication: accumulation of deleterious mutations in the dog genome. *Mol. Biol. Evol.* **25**, 2331-2336, doi:10.1093/molbev/msn177 (2008).

24 Marsden, C. D. *et al.* Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 152-157, doi:10.1073/pnas.1512501113 (2016).

25 García-Dorado, A. Understanding and predicting the fitness decline of shrunk populations: Inbreeding, purging, mutation, and standard selection. *Genetics* **190**, 1461-1476, doi:10.1534/genetics.111.135541 (2012).

26 Casas, E. & Kehrli, M. E. A review of selected genes with known effects on performance and health of cattle. *Front Vet Sci* **3**, doi:10.3389/fvets.2016.00113 (2016).

27 Jagannathan, V. *et al.* Comprehensive characterization of horse genome variation by whole-genome sequencing of 88 horses. *Anim. Genet.* **50**, 74-77, doi:10.1111/age.12753 (2019).

28 Petersen, J. L. *et al.* Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genet.* **9**, e1003211, doi:10.1371/journal.pgen.1003211 (2013).

29 Petersen, J. L. *et al.* Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLOS ONE* **8**, e54997, doi:10.1371/journal.pone.0054997 (2013).

30 Orlando, L. & Librado, P. Origin and evolution of deleterious mutations in horses. *Genes* **10**, 649 , doi: 10.3390/genes10090649 (2019).

31 Fages, A. *et al.* Tracking five millennia of horse management with extensive ancient genome time series. *Cell* **177**, 1419-1435.e1431, doi:10.1016/j.cell.2019.03.049 (2019).

32 Librado, P. *et al.* The evolutionary origin and genetic makeup of domestic horses. *Genetics* **204**, 423-434, doi:10.1534/genetics.116.194860 (2016).

33 Aurich, C. Reproductive cycles of horses. *Anim. Reprod. Sci.* **124**, 220-228, doi:10.1016/j.anireprosci.2011.02.005 (2011).

34 Allen, W. R. & Wilsher, S. Half a century of equine reproduction research and application: A veterinary tour de force. *Equine Vet. J.* **50**, 10-21, doi:10.1111/evj.12762 (2018).

35 Weatherby and Sons. *An Introduction to the General Stud Book*. (Weatherby and Sons, 1791).

36 Hill, E. W., Gu, J., McGivney, B. A. & MacHugh, D. E. Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. *Anim. Genet.* **41**, 56-63, doi:10.1111/j.1365-2052.2010.02104.x (2010).

37 Gu, J. *et al.* A genome scan for positive selection in Thoroughbred horses. *PLOS ONE* **4**, e5767, doi:10.1371/journal.pone.0005767 (2009).

38 Cunningham, E. P., Dooley, J. J., Splan, R. K. & Bradley, D. G. Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to Throughbred horses. *Anim. Genet.* **32**, 360, doi:https://doi.org/10.1046/j.1365-2052.2001.00785.x (2001).

39 Todd, E. T. *et al.* Founder-specific inbreeding depression affects racing performance in Thoroughbred horses. *Sci. Rep.* **8**, 6167, doi:10.1038/s41598-018-24663-x (2018).

40 Corbin, L. J. *et al.* Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Anim. Genet.* **41**, 8-15, doi:10.1111/j.1365-2052.2010.02092.x (2010).

41 McGivney, B. A. *et al.* Genomic inbreeding trends, influential sire lines and selection in the global Thoroughbred horse population. *Sci. Rep.* **10**, 466, doi:10.1038/s41598-019-57389-5 (2020).

42 Woolliams, J. A., Berg, P., Dagnachew, B. S. & Meuwissen, T. H. E. Genetic contributions and their optimization. *J. Anim. Breed. Genet.* **132**, 89-99, doi:10.1111/jbg.12148 (2015).

43 Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812-3814, doi:10.1093/nar/gkg509 (2003).

44 Futas, J. & Horin, P. Natural killer cell receptor genes in the family Equidae: not only Ly49. *PLOS ONE* **8**, e64736, doi:10.1371/journal.pone.0064736 (2013).

45 Kelley, J., Walter, L. & Trowsdale, J. Comparative genomics of natural killer cell receptor gene clusters. *PLoS Genet.* **1**, 129-139, doi:10.1371/journal.pgen.0010027 (2005).

46 Rahim, M. M. A. *et al.* Ly49 receptors: Innate and adaptive immune paradigms. *Front. Immunol.* **5**, 145, doi:10.3389/fimmu.2014.00145 (2014).

47 Gays, F. *et al.* Ly49B is expressed on multiple subpopulations of myeloid cells. *J Immunol Res* **177**, 5840-5851, doi:10.4049/jimmunol.177.9.5840 (2006).

48 Hiby, S. *et al.* Association of maternal killer-cell immunoglobulin-like receptors and parental HLA-C genotypes with recurrent miscarriage. *Hum Reprod* **23**, 972-976, doi:10.1093/humrep/den011 (2008).

49 Hiby, S. E. *et al.* Materal activating KIRs protect against human reproductive failure mediated by fetal HLA-C2. *J. Clin. Investig.* **120**, 4102, doi:10.1172/JCI43998 (2010).

50 Long, W. *et al.* Association of maternal KIR and fetal HLA-C genes with the risk of preeclampsia in the Chinese Han population. *Placenta* **36**, 433-437, doi:10.1016/j.placenta.2014.05.008 (2015).

51 Leon, L. *et al.* Ly49 knockdown in mice results in aberrant uterine crypt formation and impaired blastocyst implantation. *Placenta* **39**, 147-150, doi:10.1016/j.placenta.2016.01.013 (2016).

52 Lima, P. D. *et al.* Ly49 receptors activate angiogenic mouse DBA+ uterine natural killer cells. *Cellular and Molecular Immunology* **11**, 467-476, doi:10.1038/cmi.2014.44 (2014).

53 Allen, W. R. & Wilsher, S. A review of implantation and early placentation in the mare. *Placenta* **30**, 1005-1015, doi:10.1016/j.placenta.2009.09.007 (2009).

54 Noronha, L. E. & Antczak, D. F. Maternal Immune Responses to Trophoblast: The Contribution of the Horse to Pregnancy Immunology. *Am. J. Reprod. Immunol.* **64**, 231-244, doi:10.1111/j.1600-0897.2010.00895.x (2010).

55 Donaldson, W. L., Oriol, J. G., Pelkaus, C. L. & Antczak, D. F. Paternal and maternal major histocompatibility complex class I antigens are expressed co-dominantly by equine trophoblast. *Placenta* **15**, 123-135, doi:10.1016/S0143-4004(05)80449-7 (1994).

56 Bacon, S. J., Ellis, S. A. & Antczak, D. F. Control of expression of major histocompatibility complex genes in horse trophoblast. *Biol. Reprod.* **66**, 1612-1620, doi:10.1095/biolreprod66.6.1612 (2002).

57 Steri, M., Idda, M. L., Whalen, M. B. & Orrù, V. Genetic variants in mRNA untranslated regions. *Wiley Interdiscip. Rev. RNA* **9**, e1474, doi:10.1002/wrna.1474 (2018).

58 Di Giammartino, D. C., Nishida, K. & Manley, James l. Mechanisms and consequences of alternative polyadenylation. *Mol. Cell* **43**, 853-866, doi:10.1016/j.molcel.2011.08.017 (2011).

59 Lewis, S. L. *et al.* Genomewide association study reveals a risk locus for equine metabolic syndrome in the Arabian horse 1. *J. Anim. Sci.* **95**, 1071-1079, doi:10.2527/jas.2016.1221 (2017).

60 Dorairaj, J. *et al.* A germline mutation in the BRCA1 3'UTR predicts Stage IV breast cancer. *BMC Cancer* **14**, 421, doi:10.1186/1471-2407-14-421 (2014).

61 Kim, K. H. *et al.* Full sequence of mutant huntingtin 3’-untranslated region and modulation of its gene regulatory activity by endogenous microRNA. *J. Hum. Genet.* **64**, 995-1004, doi:10.1038/s10038-019-0639-8 (2019).

62 Hou, J. *et al.* Two mutations in the caprine MTHFR 3'UTR Regulated by microRNAs are associated with milk production traits. *PLOS ONE* **10**, e0133015-e0133015, doi:10.1371/journal.pone.0133015 (2015).

63 Alex, C. *et al.* A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat. Genet.* **38**, 813, doi:10.1038/ng1810 (2006).

64 Librado, P. *et al.* Ancient genomic changes associated with domestication of the horse. *Science* **356**, 442-445, doi:10.1126/science.aam5298 (2017).

65 Petersen, J. L., Mickelson, J. R., Cleary, K. D. & McCue, M. E. The American Quarter Horse: population structure and relationship to the Thoroughbred. *J. Hered.* **105**, 148-162, doi:10.1093/jhered/est079 (2014).

66 Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575, doi:10.1086/519795 (2007).

67 Storey, J., Bass, A., Dabney, A. & Robinson, D. qvalue: Q-value estimation for false discovery rate control v. R package version 2.18.0 (http://github.com/jdstorey/qvalue, 2019).

68 Fawcett, J. A. *et al.* Genome-wide SNP analysis of Japanese Thoroughbred racehorses. *PLOS ONE* **14**, doi:10.1371/journal.pone.0218407 (2019).

69 Ablondi, M., Viklund, A., Lindgren, G., Eriksson, S. & Mikko, S. Signatures of selection in the genome of Swedish Warmblood horses selected for sport performance. *BMC Genomics* **20**, doi:10.1186/s12864-019-6079-1 (2019).

70 Velie, B. D. *et al.* Using an inbred horse breed in a high density genome-wide scan for genetic risk factors of insect bite hypersensitivity (IBH). *PLOS ONE* **11**, doi:10.1371/journal.pone.0152966 (2016).

71 Petersen, J. L., Valberg, S. J., Mickelson, J. R. & McCue, M. E. Haplotype diversity in the equine myostatin gene with focus on variants associated with race distance propensity and muscle fibre type proportions. *Anim. Genet.* **45**, 827-835, doi:10.1111/age.12205 (2014).

72 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

73 Faust, G. & Hall, I. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics.* **30**, 2503-2505, doi:10.1093/bioinformatics/btu314 (2014).

74 McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data.(Report). *Genome Res.* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).

75 Depristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491, doi:10.1038/ng.806 (2011).

76 Lu, S. *et al.* CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* **48**, doi:10.1093/nar/gkz991 (2019).

77 Iqbal, K., Chitwood, J. L., Meyers-Brown, G. A., Roser, J. F. & Ross, P. J. RNA-Seq transcriptome profiling of equine inner cell mass and trophectoderm. *Biol. Reprod.* **90**, doi:10.1095/biolreprod.113.113928 (2014).

78 Bushnell, B. *Bbmap: a fast, accurate, splice-aware aligner* (Ernest Orlando Lawrence Berkeley National Laboratory, 2014)

79 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

80 Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).

81 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).

82 Read, J. E. *et al.* Dynamic changes in gene expression and signalling during trophoblast development in the horse. *Reproduction* **156**, doi:10.1530/REP-18-0270 (2018).

83 Velie, B. D. *et al.* A genome-wide association study for harness racing success in the Norwegian-Swedish coldblooded trotter reveals genes for learning and energy metabolism. *BMC Genet.* **19**, 1-13, doi:10.1186/s12863-018-0670-3 (2018).

4.3: Figures and tables

Table 1: The allele frequencies of two adjacent SNPs with an absence of minor homozygotes in genotype data from two Thoroughbred horse datasets. The expected number of minor homozygotes in each population was calculated under Hardy-Weinberg equilibrium.

Population	Sample size	Reference	6:38278097				6:38278874			
			Expected GG	GG	AG	AA	Expected CC	CC	AC	AA
Australian Thoroughbreds	156	Own data	7	0	66	90	7	0	66	90
Japanese Thoroughbreds	370	Fawcett et al., 2019[68]	9	0	117	253	9	0	117	253

Table 2: Allele frequencies of SNPs in domestic horses at positions 6:38278097 and 6:38278874 on the EquCab2.0 assembly. The total number of horses included in this analysis was 2025.

Breed	Sample size	Reference	6:38278097				6:38278874			
			Expected GG	GG	AG	AA	Expected CC	CC	AC	AA
Swedish Warmblood	380	Privately provided, Ablondi et al., 2019[69]	4	0	117	253	4	0	117	253
Coldblooded Trotter	641	Privately provided, Velie et al., 2018[83]	26	0	258	388	28	22	226	393
Quarter Horse	137	Petersen et al., 2014[65]	17	0	97	40	17	0	97	40
Exmoor Pony	285	Velie et al., 2016[70]	0	0	1	279	0	0	1	282
Mixed breeds	582	Petersen et al., 2013[29]	15	0	85	497	15	0	85	497

Table 3: Variants linked to the SNPs at 6:38278097 and 6:38278874 on the EquCab2.0 assembly. Variants were identified from whole-genome sequence data for 90 domestic horses of mixed breeds. Only variants within 5 Mb of the SNPs of interest, with a $r^2 < 0.8$ and a $D' < 0.9$ were included. Variants most likely to cause a loss of function in homozygous state are indicated with a star.

Variant	r^2	D'	**Distance from 6:38278097-6:38278874**	**Annotation**
6:38223012	0.83	1	55085	Intronic Variant LY49F
6:38264682	0.81	1	13415	Intronic Variant LY49F
6:38265144	0.85	0.92	12953	Intronic Variant LY49F
6:38273491	0.87	1	4606	Intronic Variant LY49F
6:38273498	0.87	1	4599	Intronic Variant LY49F
6:38273500	0.87	1	4597	Intronic Variant LY49F
6:38273759	0.93	1	4338	Intronic Variant LY49F
6:38274367	1	1	3730	Intronic Variant LY49F
6:38274942	1	1	3155	Intronic Variant LY49F
6:38276456	1	1	1641	Intronic Variant LY49F
6:38276742*	0.83	1	1355	3'UTR variant LY49B
6:38276955*	1	1	1142	3'UTR variant LY49B
6:38278097	1	1	0	Intronic Variant LY49B
6:38278874	0.93	1	0	Intronic Variant LY49B
6:38281733	1	1	2859	Intronic Variant LY49B
6:38282610*	1	1	3736	Missense variant LY49B
6:38284541	0.93	1	5667	Synonymous variant LY49B
6:38285848	0.85	0.92	6974	Intronic Variant LY49B
6:38290614	0.85	0.92	11740	Intergenic variant
6:38292923	0.86	0.93	14049	Intergenic variant
6:38294589	0.85	0.92	15715	Intergenic variant
6:38296564	0.85	0.92	17690	Intergenic variant
6:38297140	0.86	0.93	18266	Intronic Variant MAGOHB

6:38297149	0.86	0.93	18275	Intronic Variant MAGOHB
6:38304295	0.83	0.91	25421	Intronic Variant MAGOHB
6:38319283	0.81	0.9	40409	Intergenic variant
6:38334772	0.82	0.91	55898	Intergenic variant
6:38341477	0.86	0.93	62603	Intergenic variant
6:38341607	0.86	0.92	62733	Intergenic variant
6:38345686	0.86	0.93	66812	Intergenic variant
6:38349116	0.86	1	70242	Intronic Variant LY49C
6:38349743	0.85	1	70869	Intronic Variant LY49C
6:38351619	0.8	0.93	72745	Intronic Variant LY49C
6:38352677	0.89	1	73803	Intronic Variant LY49C
6:38353265	0.85	1	74391	Intronic Variant LY49C
6:38354110	0.86	0.93	75236	Intronic Variant LY49C
6:38358475	0.83	1	79601	Intergenic variant
6:38360920	0.87	1	82046	Intergenic variant
6:38399815	0.86	0.93	120941	Intergenic variant
6:38441535	0.8	0.93	162661	Intronic Variant LY49E

Table 4: Amino acid residue sequence in a conserved area of the 3′UTR found in all *Equus caballus LY49* genes as mapped in the EquCab2.0 assembly. The SNP position of 6:38276955G>A is highlighted in yellow.

Gene	Sequence
LY49B	AAAGACTTTCTCAGGGCCATTAAAGAGATGGGAAACTGCTTTCCAAAGAC
LY49C	AGAGAATTTCCCAGGGCCATTAAAGAGAAGAGCAACTGATTTCCAAAGAC
LY49D	AGAGAATTTCTCAGGGCCATTAAAGAGAAGGGCAACTGATTTCCAAAGAC
LY49E	AGAGAATTTCTCAGGGCCATTAAAGAGAAGGGCAACTGATTTCCAAAGAC
LY49F	AGAGAATTTTGCAGGGTCATTAAAGAGAGGGGTAACTGCTTTCCAAAGAC

Table 5: Gene counts from RNA sequence data of three trophectoderm and three inner cell mass tissue samples from equine embryos. Transcript counts are in fragments per kilobase/million (FPKM).

Tissue	Gene count (FPKM)		
Trophectoderm	Day 15	Day 22	Day 25
	0.000	0.031	0.024
Inner cell mass	Day 16	Day 23	Day 24
	0.00	0.00	0.00

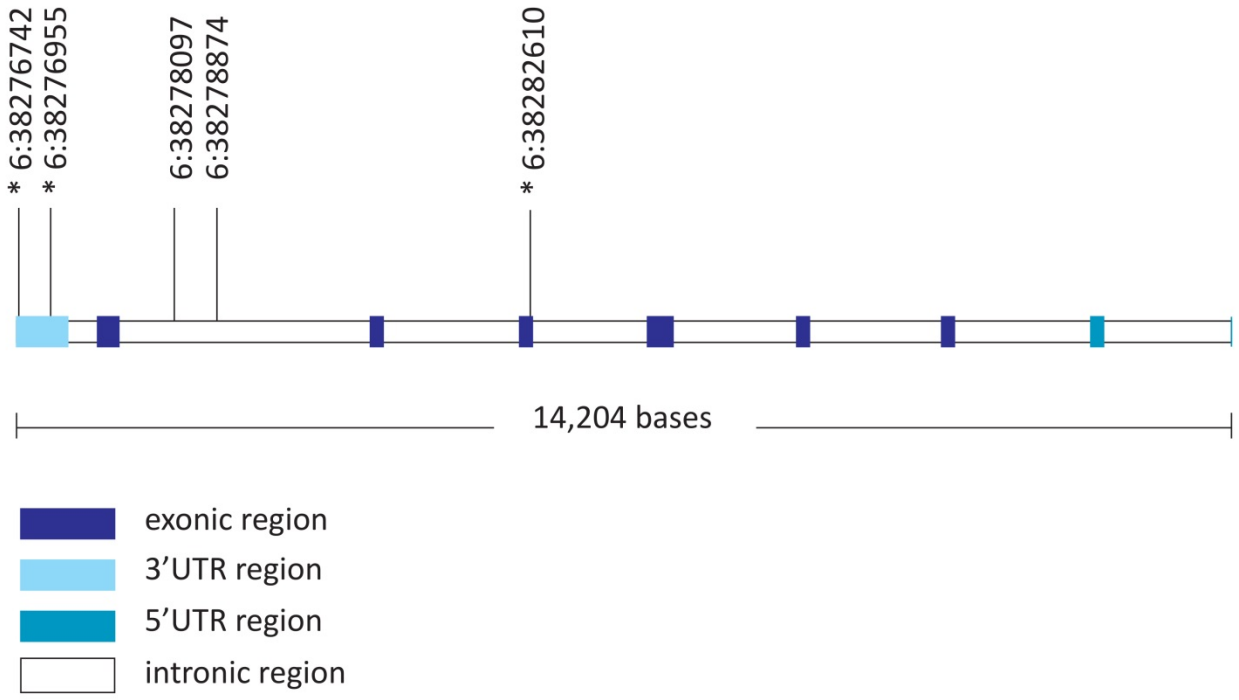

Figure 1: The equine *LY49B* gene structure and SNP positions. The two variants in the intronic region (6:38278097 and 6:38278874) were identified in preliminary analysis as showing a significant absence of homozygotes for one allele. The three variants marked with a * are in linkage disequilibrium to these SNPs and may cause a loss of function in homozygous state. The structure of the gene is based on the EquCab 2.0 reference genome where the *LY49B* gene is on the reverse strand.

4.4 Supplementary information

Table S1: Variant frequencies of SNPs 6:38278097 and 6:38278874 for mixed domestic horse from Petersen et al. (2013)[1]. The total number of individuals in this dataset is 582.

Breed	Sample Size	6.38278097			6.38278874		
		GG	AG	AA	CC	AC	AA
Akhal Teke	19	0	0	19	0	0	19
Andalusian	18	0	2	16	0	2	16
Arabian	19	0	1	18	0	1	18
Belgian	19	0	4	15	0	4	15
Caspian	16	0	0	16	0	0	16
Clyde	24	0	0	24	0	0	24
Exmoor	19	0	0	19	0	0	19
Fell	19	0	0	19	0	0	19
Finn	18	0	0	18	0	0	18
Fjord	19	0	0	19	0	0	19
Frenches-Montagnes	19	0	2	17	0	2	17
French Trotter	17	0	8	9	0	8	9
Hanovarian	15	0	1	14	0	1	14
Icelandic	19	0	3	16	0	3	16
Mini	18	0	2	16	0	2	16
Mongolian	19	0	3	16	0	3	16
Morgan	19	0	5	14	0	5	14
New Forest Pony	14	0	7	7	0	7	7
Paint	15	0	1	14	0	1	14
Puerto Rican Paso Fino	19	0	8	11	0	8	11
Mangalarga Paulista	19	0	5	14	0	5	14
Percheron	19	0	0	19	0	0	19
Peruvian Paso	19	0	2	17	0	2	17
Quarter Horse	19	0	12	7	0	12	7
Saddlebred	19	0	2	17	0	2	17
Shetland	19	0	2	17	0	2	17
Shire	19	0	2	17	0	2	17
Standardbred	19	0	3	16	0	3	16
North Swedish Horse	18	0	0	18	0	0	18
Swiss Warmblood	14	0	1	13	0	1	13
Thoroughbred	19	0	5	14	0	5	14
Tuva	15	0	4	11	0	4	11

Table S2: Variant frequencies of SNPs 6:38278097 and 6:38278874 in different populations of Quarter Horses (n=137). The SNP data was made publicly available from Petersen et al. (2014)[2].

Breed	6:38278097				6:38278874			
	SNP frequencies			Proportion of heterozygotes	SNP frequencies			Proportion of heterozygotes
	GG	AG	AA		CC	AC	AA	
Reining (*n*=23)	0	19	4	0.826087	0	19	4	0.826087
Pleasure (*n*=23)	0	16	7	0.695652	0	16	7	0.695652
Halter (*n*=23)	0	11	12	0.478261	0	11	12	0.478261
Working Cow (*n*=24)	0	21	3	0.875	0	21	3	0.875
Racing (*n*=23)	0	13	10	0.565217	0	13	10	0.565217
Cutting (*n*=21)	0	17	4	0.809524	0	17	4	0.809524
Total (*n*=137)	0	97	40	0.708029	0	97	40	0.708029

Table S3: Allele frequencies of shortlisted variants in high linkage equilibrium with the two SNPs of interest (of 6:38278097 and 6:38278874). Variants were identified from whole-genome sequence data for 90 domestic horses of mixed breeds.

Variant	Minor Allele (B)	Major Allele (A)	B/B	A/B	A/A	r^2	D'	Distance from 6:38278097-6:38278874	Annotation
6:38223012	A	AT	2	11	23	0.83	1	55085	Intronic Variant LY49F
6:38264682	T	C	1	16	51	0.81	1	13415	Intronic Variant LY49F
6:38265144	C	T	1	13	46	0.85	0.92	12953	Intronic Variant LY49F
6:38273491	G	A	0	18	50	0.87	1	4606	Intronic Variant LY49F
6:38273498	T	C	0	18	49	0.87	1	4599	Intronic Variant LY49F
6:38273500	T	C	0	18	50	0.87	1	4597	Intronic Variant LY49F
6:38273759	T	C	0	15	52	0.93	1	4338	Intronic Variant LY49F
6:38274367	A	C	0	16	56	1	1	3730	Intronic Variant LY49F
6:38274942	C	T	0	14	44	1	1	3155	Intronic Variant LY49F
6:38276456	C	T	0	15	46	1	1	1641	Intronic Variant LY49F
6:38276742	T	A	0	11	53	0.83	1	1355	3'UTR variant LY49B
6:38276955	A	G	0	15	54	1	1	1142	3'UTR variant LY49B
6:38278097	G	A	0	16	56	1	1	0	Intronic Variant LY49B
6:38278874	C	A	0	15	54	0.93	1	0	Intronic Variant LY49B
6:38281733	A	AG	0	16	51	1	1	2859	Intronic Variant LY49B
6:38282610	A	G	0	16	47	1	1	3736	Missense variant LY49B
6:38284541	C	T	0	16	51	0.93	1	5667	Synonymous variant LY49B
6:38285848	A	G	0	14	53	0.85	0.92	6974	Intronic Variant LY49B
6:38290614	T	A	1	13	49	0.85	0.92	11740	Intergenic variant

6:38292923	T	A	1	14	55	0.86	0.93	14049	Intergenic variant
6:38294589	T	C	1	13	51	0.85	0.92	15715	Intergenic variant
6:38296564	T	C	1	13	53	0.85	0.92	17690	Intergenic variant
6:38297140	A	G	1	14	42	0.86	0.93	18266	Intronic Variant MAGOHB
6:38297149	A	C	1	14	42	0.86	0.93	18275	Intronic Variant MAGOHB
6:38304295	T	A	1	11	43	0.83	0.91	25421	Intronic Variant MAGOHB
6:38319283	*	A	0	12	27	0.81	0.90	40409	Intergenic variant
6:38334772	T	A	0	12	47	0.82	0.91	55898	Intergenic variant
6:38341477	A	G	0	16	48	0.86	0.93	62603	Intergenic variant
6:38341607	A	G	0	15	58	0.86	0.92	62733	Intergenic variant
6:38345686	A	G	0	16	58	0.86	0.93	66812	Intergenic variant
6:38349116	C	G	0	9	12	0.86	1	70242	Intronic Variant LY49C
6:38349743	A	T	1	13	48	0.85	1	70869	Intronic Variant LY49C
6:38351619	C	G	1	15	57	0.80	0.93	72745	Intronic Variant LY49C
6:38352677	G	A	1	9	20	0.89	1	73803	Intronic Variant LY49C
6:38353265	A	G	1	13	51	0.85	1	74391	Intronic Variant LY49C
6:38354110	A	AGTTTAGTTTC	0	16	59	0.86	0.93	75236	Intronic Variant LY49C
6:38358475	AATAT	AAT	1	11	38	0.83	1	79601	Intergenic variant
6:38360920	A	T	1	15	47	0.87	1	82046	Intergenic variant
6:38399815	A	C	0	16	40	0.86	0.93	120941	Intergenic variant
6:38441535	G	C	1	15	52	0.80	0.93	162661	Intronic Variant LY49E

Table S4: Intensity values from microarray data of chorion and chorionic girdle equine embyronic tissue samples. Data from Read et al. (2018)[3]

Tissue	Day of development	Intensity value
Chorionic Girdle	27	6.602977
Chorionic Girdle	27	6.589985
Chorionic Girdle	27	6.62416
Chorionic Girdle	27	6.546489
Chorionic Girdle	30	6.611029
Chorionic Girdle	30	6.639321
Chorionic Girdle	30	6.656763
Chorionic Girdle	30	6.633977
Chorionic Girdle	30	6.561605
Chorionic Girdle	31	6.619057
Chorionic Girdle	31	6.637935
Chorionic Girdle	31	6.651646
Chorionic Girdle	31	6.607523
Chorionic Girdle	31	6.585455
Chorionic Girdle	34	6.568278
Chorionic Girdle	34	6.657548
Chorionic Girdle	34	6.624379
Chorionic Girdle	34	6.656019
Chorionic Girdle	34	6.659633
Chorion	27	6.620722
Chorion	27	6.558723
Chorion	27	6.672482
Chorion	27	6.570658
Chorion	30	6.57291
Chorion	30	6.602505
Chorion	30	6.55263
Chorion	30	6.58852

Chorion	30	6.638503
Chorion	31	6.582505
Chorion	31	6.587206
Chorion	31	6.639011
Chorion	31	6.616271
Chorion	31	6.539944
Chorion	34	6.630667
Chorion	34	6.642043
Chorion	34	6.569519
Chorion	34	6.624507
Chorion	34	6.588266

Table S5: Publicly available whole-genome sequence data for domestic horses used in variant calling analysis. The study accession and sample accession relate to the fields under which the sample is listed on The European Nucleotide Archive.

Sample	Breed	Study accession
ERR1527947	Holsteiner	PRJEB14779
ERR1527948	Akhal-Teke	PRJEB14779
ERR1527949	Akhal-Teke	PRJEB14779
ERR1527950	Akhal-Teke	PRJEB14779
ERR1527951	Arabian	PRJEB14779
ERR1527952	Franches-Montagnes	PRJEB14779
ERR1527953	Franches-Montagnes	PRJEB14779
ERR1527954	Franches-Montagnes	PRJEB14779
ERR1527955	Franches-Montagnes	PRJEB14779
ERR1527956	Franches-Montagnes	PRJEB14779
ERR1527957	Franches-Montagnes	PRJEB14779
ERR1527958	Franches-Montagnes	PRJEB14779
ERR1527959	Franches-Montagnes	PRJEB14779
ERR1527960	Franches-Montagnes	PRJEB14779
ERR1527961	Franches-Montagnes	PRJEB14779
ERR1527962	Franches-Montagnes	PRJEB14779
ERR1527963	Franches-Montagnes	PRJEB14779
ERR1527964	Franches-Montagnes	PRJEB14779
ERR1527965	Franches-Montagnes	PRJEB14779
ERR1527966	Haflinger	PRJEB14779
ERR1527967	Koninklijk Warmbloed Paard Nederland	PRJEB14779
ERR1527968	Quarter Horse	PRJEB14779
ERR1527969	Quarter Horse	PRJEB14779
ERR1527970	Quarter Horse	PRJEB14779
ERR1527971	Franches-Montagnes	PRJEB14779
ERR1527972	Swiss Warmblood	PRJEB14779

ERR1545178	Badenwürttembergisches Warmblut	PRJEB14779
ERR1545179	Bayrisches Warmblut	PRJEB14779
ERR1545180	Hannoveraner	PRJEB14779
ERR1545181	Holsteiner	PRJEB14779
ERR1545182	Holsteiner	PRJEB14779
ERR1545183	Oldenburger	PRJEB14779
ERR1545184	Oldenburger	PRJEB14779
ERR1545185	Trakehner	PRJEB14779
ERR1545186	Westfale	PRJEB14779
ERR1545187	Westfale	PRJEB14779
ERR1545188	Swiss Warmblood	PRJEB14779
ERR1545189	Swiss Warmblood	PRJEB14779
ERR1545190	Holsteiner	PRJEB14779
ERR1735862	Thoroughbred	PRJEB14779
ERR2179540	Franches-Montagnes	PRJEB14779
ERR2179541	Franches-Montagnes	PRJEB14779
ERR2179542	German Riding Pony	PRJEB14779
ERR2179543	Welsh Pony	PRJEB14779
ERR2179544	Polish Warmblood	PRJEB14779
ERR2179545	German Riding Pony	PRJEB14779
ERR2179546	Morgan Horse	PRJEB14779
ERR2179547	Holsteiner	PRJEB14779
ERR2179548	Oldenburger	PRJEB14779
ERR2179549	Hannoveraner	PRJEB14779
ERR2179550	Welsh Pony	PRJEB14779
ERR2179551	Arabian	PRJEB14779
ERR2179552	Noriker	PRJEB14779
ERR2179553	Haflinger	PRJEB14779
ERR2179554	Haflinger	PRJEB14779
ERR2179555	Haflinger	PRJEB14779
ERR2179556	Swiss Warmblood	PRJEB14779

ERR2203766	Trakehner	PRJEB14779
ERR2731055	American Paint Horse	PRJEB14779
ERR2731056	Shetland Pony	PRJEB14779
ERR2731057	Icelandic	PRJEB14779
ERR2731058	Holsteiner	PRJEB14779
ERR2731059	Holsteiner	PRJEB14779
ERR2731060	Holsteiner	PRJEB14779
ERR2731061	Akhal-Teke	PRJEB14779
ERR3317814	Deutsches Reitpony	PRJEB14779
ERR3628171	Traber	PRJEB14779
ERR3628172	Traber	PRJEB14779
ERR3628173	Traber	PRJEB14779
ERR3628174	Franches-Montagnes	PRJEB14779
SRR505867	Thoroughbred	PRJNA168142
SRR515202	Thoroughbred	PRJNA168142
SRR515203	Thoroughbred	PRJNA168142
SRR515204	Thoroughbred	PRJNA168142
SRR515205	Thoroughbred	PRJNA168142
SRR515206	Thoroughbred	PRJNA168142
SRR515208	Thoroughbred	PRJNA168142
SRR515209	Thoroughbred	PRJNA168142
SRR515211	Thoroughbred	PRJNA168142
SRR515212	Thoroughbred	PRJNA168142
SRR515213	Thoroughbred	PRJNA168142
SRR515214	Thoroughbred	PRJNA168142
SRR515215	Thoroughbred	PRJNA168142
SRR515216	Thoroughbred	PRJNA168142
SRR515217	Thoroughbred	PRJNA168142
SRR515218	Thoroughbred	PRJNA168142
SRR641364	Thoroughbred	PRJNA184688
SRR641365	Thoroughbred	PRJNA184688

SRR641366	Thoroughbred	PRJNA184688
SRR641367	Thoroughbred	PRJNA184688

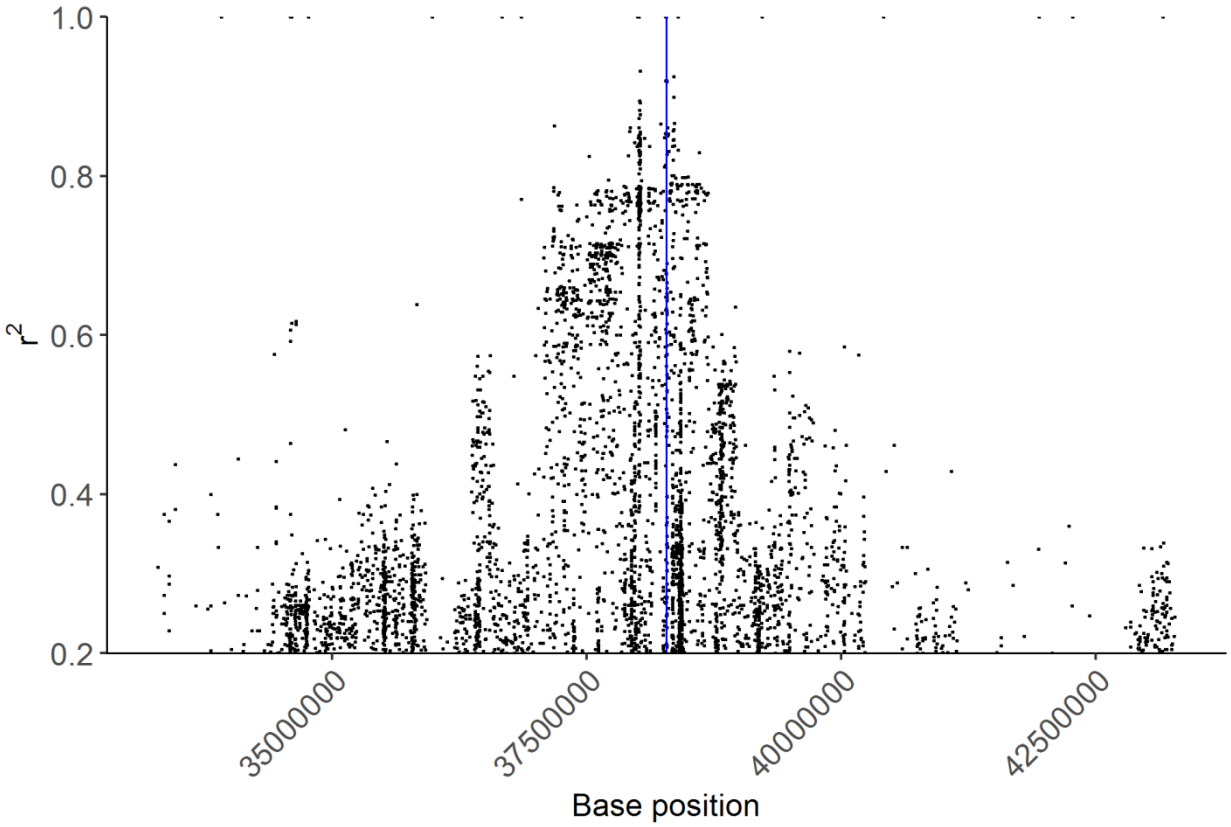

Figure S1: Linkage disequilibrium patterns relative to the markers 6.38278097 and 6.38278874. The positions of 6.38278097 and 6.38278874 are indicated with the blue line. The region spanning 5Mb on either side of the 6.38278097 and 6.38278874 was included in this plot. Variants were called from whole-genome sequence data of 90 horses.

References

1 Petersen, J. L. *et al.* Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLOS ONE* **8**, e54997, doi:10.1371/journal.pone.0054997 (2013).

2 Petersen, J. L., Mickelson, J. R., Cleary, K. D. & McCue, M. E. The American Quarter Horse: population structure and relationship to the Thoroughbred. *J. Hered.* **105**, 148-162, doi:10.1093/jhered/est079 (2014).

3 Read, J. E. *et al.* Dynamic changes in gene expression and signalling during trophoblast development in the horse. *Reproduction* **156**, doi:10.1530/REP-18-0270 (2018).

Chapter 5: General discussion

5.1 Main findings

In this thesis, the effects of selective breeding practices on racing performance, fertility and lethal haplotype frequencies in the Australian Thoroughbred horse population were examined. The key discoveries were:

1. Recent inbreeding events are negatively associated with racing performance in Australian Thoroughbred horses.
2. Genetic improvement for racing performance in contemporary Australian Thoroughbreds mainly occurred during the foundation of the population and in the most recent three generations.
3. Inbreeding to different ancestors has had variable effects on the racing performance of their modern day descendants.
4. Inbreeding has no measurable effect on foaling rate, gestation length or sex ratios in Australian Thoroughbred horses.
5. All fertility traits were heritable, but covering success was only influenced by maternal genetic variation and foal sex ratio by paternal genetic variation.
6. A potentially homozygous lethal haplotype was identified in the LY49B gene and may play an important role in maternal/foetal compatibility or implantation success.
7. This haplotype segregates at high frequencies in the Thoroughbred population, and other domestic horse breeds.

5.2 Summary of results

The first main aim of this thesis was to understand the effects of inbreeding on the racing performance traits in Thoroughbred horses. In Chapter 2, a strong relationship was found between high levels of inbreeding and reduced racing performance outcomes in Australian Thoroughbred horses. Although there was also strong evidence for genetic improvement through selection, estimated breeding values from racing performance traits showed that this mainly occurred at the foundation of the population and in more recent generations. Additionally, inbreeding to different individuals with large genetic influences on the breed had variable effects on the racing performance of modern day Australian Thoroughbreds, indicating that genetic load is not evenly distributed across the population.

The second main aim of this thesis was to understand the impacts of inbreeding on reproductive traits in Thoroughbred horses. In Chapter 3, no detectable relationship was found between inbreeding levels and covering success, gestation length or foal sex ratio in Australian Thoroughbred horses. This contrasts to the strong relationship found between inbreeding and racing performance traits in Chapter 2, which may possibly be due to individuals with low fertility output being selectively removed from the population. Individuals with low levels of fertility are more likely to quickly disappear from the pedigrees of future generations due to having smaller numbers of descendants. Thoroughbred breeders often use individuals with poor or no racetrack performance for breeding if their relatives show superior racing ability, which is likely to reduce the efficiency of selection against low athletic abilities. However, this practice is also likely to increase selection for female families with good fertility because a larger number of relatives may increase the chance of an individual being related to an elite performer. The lack of a relationship between inbreeding levels and fertility traits may also be due to the intensive veterinary management practices used in the Thoroughbred population, including hormonal therapies to induce ovulation and blue light masks to increase photoperiod signals[1,2]. However, an increase in the rate of inbreeding in future generations of the Thoroughbred breed could lead to a more discernible relationship with fertility traits.

The third and final main aim to this thesis was to identify recessive lethal variants at high frequencies in the Thoroughbred horse population. In Chapter 4, SNP data from Thoroughbreds and other domestic horse breeds were used to identify a haplotype in the *LY49B* gene on chromosome 6 with a high proportion of heterozygotes, but a significant absence of homozygotes for one allele. Analysis of whole genome sequencing data revealed three variants linked to this haplotype that were predicted to impact the structure and expression of the *LY49B* gene. The two most likely variants were found in the 3'UTR region of the *LY49B* gene, highlighting its potential importance in post-transcriptional processing and expression. Studies of RNA transcriptomic data showed that the *LY49B* gene was expressed in trophoblast tissue during the placental stage of embryogenesis, suggesting that it may be important in implantation success. This haplotype presents a strong candidate for an association with lethality in the homozygous state.

More broadly, this thesis aimed to understand the effects of selective breeding on the Thoroughbred horse population. Although there was evidence that selection has resulted in genetic improvement of racing performance over the 300 year history of the breed, there was also strong evidence that genetic load still persists in the Thoroughbred population. Additionally, selective breeding practices favouring a

limited number of popular stallions may have resulted in a recessive lethal haplotype in the *LY49B* gene drifting to high frequencies in the Thoroughbred horse population. The effects of inbreeding varied between different traits, highlighting the importance of evaluating multiple traits to understand the full effects of breeding practices in a population.

The findings of this thesis can be used to assist in making breeding decisions both on an individual and a population level scale in Thoroughbred horses. Aiming to promote genetic diversity in the population will avoid increases in the rate of inbreeding and loss of genetic diversity in future generations, which may impact the health and the viability of the Thoroughbred breed. Additionally, the identification of a potential recessive lethal mutation may assist in improving mating decisions to increase foaling rates in the Thoroughbred breed. More broadly, this research is also relevant to other animal populations. The techniques used in this thesis to evaluate the effects of selective breeding on genetic load and lethal mutation frequencies are also likely to hold value for other domestic and wildlife populations.

5.3 Future directions

In this thesis, a relationship between high levels of inbreeding and reduced racing performance was found in Chapter 2. This finding is of particular concern for the Thoroughbred industry considering the trend of reducing genetic diversity and effective population size in the breed[3]. The trend is likely to be the result of selective breeding practices in recent generations favouring a limited number of stallion bloodlines. Together, these findings indicate that current breeding practices could result in a population-wide reduction of racing performance in future generations. More recently, some Thoroughbred racing associations have considered limiting the number of mares covered by each stallion within a season to promote genetic diversity. Additionally, a limit on the number of progeny that a stallion can produce across their lifetime should also be considered to reduce extremely large genetic contributions being made by a single individual. Such measures could be effective in maintaining genetic diversity in future generations if they are implemented across all Thoroughbred racing jurisdictions. This is particularly important considering that some popular stallions serve mares in both hemispheres each year. Such measures may assist in maintaining levels of genetic diversity in the population in order to prevent reductions in the phenotypic value of future generations.

More comprehensive genotyping of Thoroughbred horses on a population-level scale could allow for advancements to improve breeding decisions and understand the underlying genetic basis of disease. Recently, a study of privately-available large-scale genomic data from the Thoroughbred population

worldwide revealed a trend of reduced genetic diversity and increased genomic inbreeding within the past five generations[3]. In this thesis, pedigree records were utilized because they allow for the analysis of larger sample sizes and deceased individuals that would not be possible with genomic data. However, errors in the Thoroughbred pedigree have been identified which may confound pedigree-based analyses[4,5]. Further population-scale genotyping of the Thoroughbred breed would assist in continued monitoring of genetic trends in the population and improving our understanding of the genetic basis of disease.

Evaluating the proportion of the genome comprised of runs of homozygosity is a common technique used to estimate genomic inbreeding levels[6,7]. Analysis of the association between different runs of homozygosity and performance traits may provide novel insights into differentiation between favourable and unfavourable inbreeding in Thoroughbred horses. In Chapter 2, high levels of inbreeding to Herod were found to have a positive relationship with racing performance traits. In contrast, inbreeding to Eclipse, Touchstone and Stockwell had a negative relationship with the racing performance of their modern day descendants. These findings show that although an individual and their immediate progeny may have a superior phenotype, providing large genetic contributions to a population may have an unexpected effect in future generations. Often these effects would not be seen until at least five generations into the future, when there is inbreeding between the descendants of an individual. An example of this is a recessive lethal mutation found at high frequencies in cattle populations due to being carried by a male used extensively in artificial insemination practices[8].

In addition, the availability of population-level genomic data could also lead to an increased understanding of the underlying genetic basis of many common diseases in the Thoroughbred horse population. As outlined in Chapter 1 (Section 1.3.3), although many of these conditions have a measurable heritable component, the underlying genetic basis of disease remains unknown. In Chapter 2, we suggested that the relationship between increased inbreeding to Eclipse and Stockwell with reduced racing performance may be due to a greater likelihood of inheriting congenital conditions from these individuals. It is likely that these conditions have a complex quantitative genetic basis and are also influenced by environmental stimuli. Consequently, large sample sizes are needed to identify the genetic variants associated with increased or decreased disease occurrence. Such findings would assist in improving breeding strategies to decrease the likelihood of an individual developing these conditions. An understanding of the underlying genetic basis of disease may also assist in improving techniques to manage these conditions (e.g. veterinary treatment, supplements or gene therapies).

In Chapter 3 of this thesis, covering success was found to be measurably influenced by genetic variation of the dam but not the sire. This finding is possibly because sub-fertile Thoroughbred stallions are often gelded and returned to racing. However, the underlying genetic causes of reduced fertility in these stallions are rarely investigated. A number of loci have been associated with stallion fertility in domestic horses, most notably FK506 binding protein 6 (*FKBP6*) in Thoroughbreds [9-14]. An improved understanding of the genetic basis of sub-fertility in individual stallions may assist in the management of male horses (e.g. it may be preferable to geld a male horse genetically pre-disposed to sub-fertility issues early in their life). Alternatively, such knowledge could also assist in managing stallions with reduced fertility rates (e.g. by using genetic or hormonal therapies to counteract the effects of these mutations).

Chapter 3 of this thesis also found that foal sex ratio had an extremely small, but still measurable heritable component from the sire. This finding is in agreement with previous studies, which hypothesise that superior males should produce more sons, whereas males with lower fitness are advantaged by produced a greater proportion of daughters[15,16]. In populations such as horses, the genetic contributions of males are highly skewed towards superior individuals. Under these circumstances, sons from an inferior male would be unlikely to produce a large offspring output, thus advantaging this male to produce a higher proportion of female offspring[16]. In the Thoroughbred population, artificial selection means that only superior males are allowed to breed, which may partly explain why there was only a small heritable component, with no significant effect of inbreeding or environmental influences on foal sex ratio. However, in this study, only data from large stud farms that mainly breed from top-quality stallions were analysed. Analysis of data from other horse breeds, where a larger proportion of males are allowed to breed may reveal more detectable inbreeding and environmental influences on foal sex ratios.

The endangered Suffolk Punch horse breed has experienced highly skewed sex ratios in recent generations. The breed is characterised by low population numbers and genetic diversity due to the replacement of horsepower by machinery since the industrial revolution. Currently, there are no published studies that have investigated the effects of genetic and environmental factors on the skewed sex ratios observed in this breed. Skewed sex ratios could have severe effects on an already small population by limiting breeding pairs and further reducing genetic diversity. Further investigation into the genetic basis of foal sex ratios could assist in the management of many domestic horse breeds, and other animal populations that may suffer from small N_e and low genetic diversity.

In Chapter 4 of this thesis, a haplotype in the *LY49B* gene segregating at high frequencies in the Thoroughbred population that may cause embryonic lethality was identified. Based on the results of similar studies in cattle and pig populations, it is highly likely that selective breeding practices have resulted in multiple recessive lethal variants reaching high frequencies in the Thoroughbred breed. Large population-wide genotype data over a number of generations were available for cattle and pig breeds used in these studies[17-24]. These data also allowed for the analysis of parent/offspring genotype ratios to identify deviations from the expected allele frequencies[20,23]. Availability of such comprehensive data for the Thoroughbred and other horse populations would dramatically increase the statistical power to detect lethal variants. Identification and characterisation of such variants would assist in making informed breeding decisions to improve foaling rates in the Thoroughbred population. This could assist in reducing the number of coverings that a mare requires to conceive each season, and may increase the foal output of a mare over her lifetime. Such research may also benefit stallions, as they may require fewer coverings per mare each season.

Chapter 4 of this thesis focussed on identifying lethal haplotypes at high frequencies in the Thoroughbred horse breed. However, other domestic horse breeds are also likely to have lethal haplotypes segregating at high frequencies due to population decline and selective breeding practices. Such studies are particularly important for populations with small census sizes, including Caspian, Florida Cracker and Sorraia breeds[25-27]. Additionally, some domestic horse populations such as the Clydesdale, Shire and Exmoor breeds have undergone severe population bottlenecks since World War II[25-27]. These breeds may be particularly susceptible to recessive lethal mutations drifting to high frequencies due to the limited number of mating opportunities in the population leading to some individuals making large genetic contributions. Identification of recessive lethal mutations may also be beneficial to other animal populations, including cat and dog breeds where few such studies are currently published. Additionally, genetic rescue programs for wildlife species may also benefit from similar studies to assist in managing breeding decisions.

The haplotype identified in Chapter 4 of this thesis in the immune gene *LY49B* has not previously been directly associated with embryonic lethality in any other species. However, the findings of this thesis suggest that *LY49B* may play a role in maternal/foetal compatibility and the successful implantation of the embryo during development. It is also possible that knockout of the *LY49B* gene results in post-natal death due to the individual being immunocompromised. More detailed transcriptomic studies of equine embryonic tissues would provide a more comprehensive expression profile for the *LY49B* gene and

other lethal haplotypes identified in future studies. Such investigations may improve our understanding of how genes function in development, which is often poorly understood. Additionally, knockout studies of potentially lethal variants in equine embryos could confirm whether genes are essential for embryogenesis and at which life stage loss of function results in death.

The Thoroughbred horse has been used in the foundation of many other popular breeds including the Quarter Horse, Standardbred and Warmblood. These breeds are also popular throughout the world for use in recreational riding as well as other horse sports. Examination of the origins of selective sweep regions and genetic load from founders of different breed origins may assist in guiding future breeding decisions. In particular, some of these breeds (e.g. The Quarter Horse) still maintain an open stud book. Improved understanding of population genomics in the breed may provide valuable insights for the inclusion of new individuals into the population. This technique would allow for the inclusion of new sources of genetic diversity into the population, whilst also facilitating genetic improvement.

Overall, this thesis has provided novel insights into the effects of selective breeding practices in the Thoroughbred horse population. These insights have to potential to provide valuable guidance for future population genetic studies and breeding management decisions to maximise genetic gain across a range of phenotypes. The findings of this thesis can be used to promote genetic improvement in the Thoroughbred population and reduce the production of individuals unsuitable for racing and breeding. This not only has productivity impacts for the breed, but also has the potential to reduce wastage which is currently a topical concern for the industry. The methods used in this thesis can also be applied to assist in optimising breeding management in other domestic and wild animal populations.

5.4 References

1 Allen, W. R. & Wilsher, S. Half a century of equine reproduction research and application: A veterinary tour de force. *Equine Vet. J.* **50**, 10-21, doi:10.1111/evj.12762 (2018).

2 Nolan, M. B. *et al.* Artificially extended photoperiod administered to pre-partum mares via blue light to a single eye: observations on gestation length, foal birth weight and foal hair coat at birth. *Theriogenology* **100**, 126-133, doi:10.1016/j.theriogenology.2017.06.012 (2017).

3 McGivney, B. A. *et al.* Genomic inbreeding trends, influential sire lines and selection in the global Thoroughbred horse population. *Sci. Rep.* **10**, 466, doi:10.1038/s41598-019-57389-5 (2020).

4 Felkel, S. *et al.* The horse Y chromosome as an informative marker for tracing sire lines. *Sci. Rep.* **9**, 6095-6095, doi:10.1038/s41598-019-42640-w (2019).

5 Bower, M. A. *et al.* Thoroughbred racehorse mitochondrial DNA demonstrates closer than expected links between maternal genetic history and pedigree records. *J. Anim. Breed. Genet.* **130**, 227-235, doi:10.1111/j.1439-0388.2012.01018.x (2013).

6 Curik, I., Ferencakovic, M. & Solkner, J. Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest. Sci.* **166**, 26-34, doi:10.1016/j.livsci.2014.05.034 (2014).

7 Gurgul, A. *et al.* The application of genome-wide SNP genotyping methods in studies on livestock genomes. *J. Appl. Genet.* **55**, 197-208, doi:10.1007/s13353-014-0202-4 (2014).

8 Bourneuf, E. *et al.* Rapid discovery of de novo deleterious mutations in cattle enhances the value of livestock as model species. *Sci. Rep.* **7**, 11466, doi:10.1038/s41598-017-11523-3 (2017).

9 Schrimpf, R. *et al.* Genome-wide association study identifies phospholipase C zeta 1 (PLCZ1) as a stallion fertility locus in Hanoverian Warmblood horses. *PLOS ONE* **9**, doi:10.1371/journal.pone.0109675 (2014).

10 Giesecke, K. *et al.* Evaluation of ACE, SP17, and FSHB as candidates for stallion fertility in Hanoverian Warmblood horses. *Anim. Reprod. Sci.* **126**, 200-206, doi:10.1016/j.anireprosci.2011.05.007 (2011).

11 Giesecke, K. *et al.* Evaluation of SPATA1 -associated markers for stallion fertility. *Anim. Genet.* **40**, 359-365, doi:10.1111/j.1365-2052.2008.01844.x (2009).

12 Giesecke, K., Hamann, H., Sieme, H. & Distl, O. INHBA associated markers as candidates for stallion fertility. *Reprod. Domest. Anim.* **45**, 342-347, doi:10.1111/j.1439-0531.2008.01325.x (2010).

13 Hamann, H. *et al.* A polymorphism within the equine CRISP3 gene is associated with stallion fertility in Hanoverian warmblood horses. *Anim. Genet.* **38**, 259-264, doi:10.1111/j.1365-2052.2007.01594.x (2007).

14 Terje, R. *et al.* Genome-wide association study implicates testis-sperm specific FKBP6 as a susceptibility locus for impaired acrosome reaction in stallions. *PLoS Genet.* **8**, e1003139, doi:10.1371/journal.pgen.1003139 (2012).

15 Malo, A. F., Gilbert, T. C. & Riordan, P. Drivers of sex ratio bias in the eastern bongo: lower inbreeding increases the probability of being born male. *Proc. Biol. Sci.* **286**, 20190345-20190345, doi:10.1098/rspb.2019.0345 (2019).

16 Malo, A. F. *et al.* A father effect explains sex-ratio bias. *Proc. Biol. Sci.* **284**, doi:10.1098/rspb.2017.1159 (2017).

17 Hoff, J. L., Decker, J. E., Schnabel, R. D. & Taylor, J. F. Candidate lethal haplotypes and causal mutations in Angus cattle. *BMC Genomics* **18**, doi:10.1186/s12864-017-4196-2 (2017).

18 VanRaden, P. M., Olson, K. M., Null, D. J. & Hutchison, J. L. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *J. Dairy Sci.* **94**, 6153-6161, doi:10.3168/jds.2011-4624 (2011).

19 Fasquelle, C. *et al.* Balancing selection of a frame-shift mutation in the MRC2 gene accounts for the outbreak of the crooked tail syndrome in Belgian Blue cattle. *PLoS Genet.* **5**, e1000666, doi:10.1371/journal.pgen.1000666 (2009).

20 Zhang, C., MacNeil, M. D., Kemp, R. A., Dyck, M. K. & Plastow, G. S. Putative loci causing early embryonic mortality in Duroc swine. *Front. Genet.* **17**, 655, doi:10.3389/fgene.2018.00655 (2018).

21 Charlier, C. *et al.* NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. *Genome Res.* **26**, 1333-1341, doi:10.1101/gr.207076.116 (2016).

22 Sonstegard, T. S. *et al.* Identification of a nonsense mutation in CWC15 associated with decreased reproductive efficiency in Jersey cattle. *PLOS ONE* **8**, e54872, doi:10.1371/journal.pone.0054872 (2013).

23 Derks, M. F. L. *et al.* A systematic survey to identify lethal recessive variation in highly managed pig populations. *BMC Genomics* **18**, doi:10.1186/s12864-017-4278-1 (2017).

24 Derks, M. F. L. *et al.* Loss of function mutations in essential genes cause embryonic lethality in pigs. *PLoS Genet.* **15**, e1008055, doi:10.1371/journal.pgen.1008055 (2019).

25 Kjöllerström, H. J., Gama, L. T. & Oom, M. M. Impact of inbreeding on fitness-related traits in the highly threatened Sorraia horse breed. *Livest. Sci.* **180**, 84-89, doi:10.1016/j.livsci.2015.08.001 (2015).

26 Petersen, J. L. *et al.* Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLOS ONE* **8**, e54997, doi:10.1371/journal.pone.0054997 (2013).

27 Orlando, L. & Librado, P. Origin and evolution of deleterious mutations in horses. *Genes* **10**, 649, doi:10.3390/genes10090649 (2019).

Appendix 1

ACTA AGRICULTURAE SCANDINAVICA, SECTION A — ANIMAL SCIENCE
2018, VOL. 68, NO. 2, 112–116
https://doi.org/10.1080/09064702.2018.1563211

Premie race participation is associated with increased career longevity and prize money earnings in Norwegian-Swedish Coldblooded Trotters

E. T. Todd[a], K. Jäderkvist Fegraeus[b], P. C. Thomson[a], C. F. Ihler[c], E. Strand[c], G. Lindgren[b,d] and B. D. Velie[a,b]

[a]School of Life and Environmental Sciences, The University of Sydney, Sydney, Australia; [b]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden; [c]Department of Companion Animal Clinical Sciences, Faculty of Veterinary Medicine, Norwegian University of Life Sciences, Oslo, Norway; [d]Department of Biosystems, KU Leuven, Leuven, Belgium

ABSTRACT

Norwegian-Swedish Coldblooded Trotters can participate in premie races as two year olds, where prize money is awarded for finishing the race within a specific time interval rather than winning. In this study, the association with premie race participation and future competitive racing success was evaluated. Analyses including all raced horses born between 2000 and 2009 ($n = 9350$) showed a relationship between premie race participation with reduced career longevity and prize money earnings. However, when analyses only included horses that raced competitively ($n = 7497$; i.e. horses with participation in only premie or qualification races were excluded), premie race participation was associated with increased racing success. These findings indicate that premie race fields consist of both horses with good racing ability as well as horses with limited talent. Overall this study shows that premie racing is likely to be beneficial for both the horse and the trainer.

ARTICLE HISTORY
Received 2 October 2018
Accepted 19 December 2018

KEYWORDS
Horse; harness racing; performance; career length; wellbeing

Introduction

The Norwegian- Swedish Coldblooded trotter (NSCT) is a racehorse breed originating from draft horses (Jäderkvist Fegraeus et al., 2018). Since the 1950's, the population has been selectively bred for performance in harness racing, a popular sport in Norway and Sweden (Jäderkvist Fegraeus et al., 2018). In Coldblooded trotters there is a strong relationship between an early age at first start and increased career length as well as prize money earned (Saastamoinen & Ojala, 1991a; Saastamoinen & Nylander, 1996; Velie et al., 2018a). This trend is also reflected in larger populations of Thoroughbred and Standardbred racehorses (Cheetham et al., 2010; Knight & Thomson, 2011; Velie et al., 2013). An early age at racing is often considered a desirable trait as any prize money earned provides a quicker return on investment to the owners and trainer of a horse (Saastamoinen & Nylander, 1996).

Although NSCTs may start training before 18 months of age (Revold et al., 2010), very few competitive races are available for two year olds. However, two year olds may participate in both premie and qualification races. Premie races, which are only available to two year olds, differ from competitive racing in that they award prize money to all horses that finish the race within a certain time interval (2:10–1:55 per kilometre; minutes:seconds). Thus, finishing first in a premie race provides no immediate financial benefits, compared to finishing in last place. Premie races mainly serve as an early stepping-stone for trainers and owners to assess the present physical and mental capabilities of their 2-year-old horses for future horse racing. Qualification races, on the other hand, similarly require horses to finish within a certain time, but are necessary in order for a horse to become eligible for entry in competitive racing (Velie et al., 2018a). Consequently, premie races can potentially provide an opportunity for young horses to prepare for qualification races, better equipping them to enter competitive racing at an earlier age. Premie races can also provide an opportunity for horses with limited racing ability to earn some prizemoney without having to qualify or race competitively. However, it should be noted that premie race prize money is not included in a horse's official career earnings. Previous studies that have evaluated the relationship between age at first start and performance in Coldblooded trotters have excluded premie races in their analyses (Saastamoinen, 1991a; Saastamoinen & Nylander, 1996; Velie et al., 2018a). As a result, little information is available as to the impact that premie race participation has on the future career of NSCTs. In this study, we examined the relationship between participation in premie races and future competitive racing success in NSCTs.

CONTACT B. D. Velie brandon.velie@sydney.edu.au School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia

Materials and methods

Description of data

Racing performance data for all NSCTs born between 2000 and 2009 that participated in at least one race start as of the 8th of February 2017 ($n = 9350$) was provided by the Norwegian and Swedish trotting associations (Det Norske Travselskap and Svensk Travsport). The NSCT's in this dataset were born between January and October. The majority of horses (8192) were born between April and July. Only two horses were born in September, and one in October. The number of horses born each year varied between 808 and 1067. Of the horses in the dataset, 7634 participated in premie races, and 5781 of these premie participants continued on to race competitively. The total number of individuals in the sample with competitive racing data was 7497, with 1716 of these horses not having participated in a premie race. The data does not specify whether the horse competed in more than one premie race, or completed the premie race within the required time interval.

A number of measures were used to evaluate each individual's competitive racing success. More talented horses tend to earn a greater amount of prizemoney, thus cumulative prizemoney earnings (CE) and earnings per start (EPS) were calculated for each individual. Prizemoney from races in Sweden were converted to Norwegian Krone (NOK) based on the average exchange rate in the year that the race occurred. Only prizemoney won in competitive races was included in these calculations. Career longevity is also an important trait for Coldblooded trotters (Saastamoinen & Nylander, 1996). Therefore, career starts (NCS) and career length in months (CLM) were also calculated for each individual. Horses can be disqualified in a race for using a gait that is not allowed (i.e. gallop or pace), so the number of career starts where the horse was not disqualified was also calculated

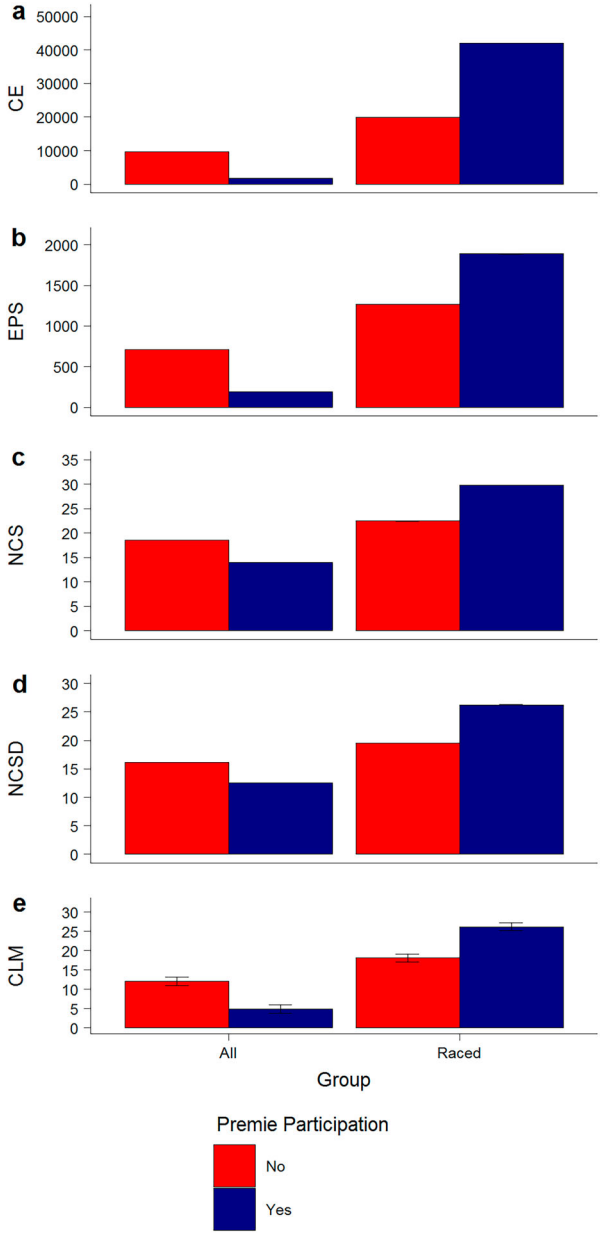

Figure 1. Predicted values from the fitted linear mixed models of the relationships between premie race participation and future competitive racing performance of Norwegian- Swedish Coldblooded trotters. Separate analyses were conducted including all horses ($n = 9350$) and only horses that raced competitively ($n = 7497$). Bar charts represent the relationship of premie race participation with: a) cumulative earnings in Norwegian Krone (CE), b) earnings per start in Norwegian Krone (EPS), c) number of competitive starts (NCS), d) number of competitive starts with no disqualifications (NSCD) and e) career length in months (CLM). Error bars represent one standard error of the model based from the mean.

Table 1. Linear mixed models for each outcome variable.

Outcome variable	Transformation	Model
CE	log	$\log(CE) = \mu + PP + YOB + sex + COB + sire + dam + \varepsilon$
EPS	log	$\log(EPS) = \mu + PP + YOB + sex + COB + sire + dam + \varepsilon$
BT	Power of 0.25	$(BT)^{0.25} = \mu + PP + YOB + sex + COB + sire + dam + \varepsilon$
CLM	log	$\log(CLM) = \mu + PP + YOB + sex + COB + sire + dam + \varepsilon$
NCS	Power of 0.25	$(NCS)^{0.25} = \mu + PP + YOB + sex + COB + sire + dam + \varepsilon$
NCSD	Power of 0.25	$(NCSD)^{0.25} = \mu + PP + YOB + sex + COB + sire + dam + \varepsilon$

Notes: Models were fitted using ASReml-R. PP, YOB, sex and COB were included as fixed effects in each model. Sire and dam were included as random effects. Models were log or power-transformed to achieve a normal distribution and constant residual variance.

CE = Cumulative earnings, EPS = Earnings per start, BT = Best time, CLM = career length (months), NCS = Number of competitive starts, NCSD = Number of competitive starts with no disqualifications, PP = premie participation (yes/no), YOB = year of birth, COB = country of birth.

(NCSD) to account for correct racing technique. Only competitive race starts were included in these measures. The career length of horses that only competed in one competitive race was set at 1 day (0.03 months). More details on the dataset can be found in Velie et al. (2018a).

Data analysis

The relationship between each racing performance measure and premie race participation was analyzed using linear mixed models in ASReml-R (Butler et al., 2009) (Table 1). Sex, year of birth and country of origin (either Sweden or Norway), were included as fixed effects in each model. These traits were significantly associated with each racing performance measure (all $P < 0.001$) and have previously been shown to influence the racing performance of NSCTs (Velie et al., 2018a). Additionally the sire and dam of each individual were included as random effects to account for the non-independence of data. Two analyses were performed using the same model structure for each analysis: the first including all horses ($n = 9350$), and the second including only horses that raced competitively ($n = 7497$). The significance of each fixed effect was assessed through Wald statistics. These resulting fitted models were used to estimate the predicted values of each racing performance measure based on premie race participation (Figure 1). Additionally, the relationship between premie race participation with month and year of birth were assessed using generalized linear mixed models.

Results and discussion

In analyses that included all horses, premie race participation in NSCTs showed a strong association with a reduction in all measures of racing performance (all $P < 0.001$) (Figure 1). Generally speaking, the main reasons that horses do not race competitively are lack of talent, bad character, and injuries (Saastamoinen 1991a, 1991b; Velie et al., 2018b). It is highly plausible that a portion of premie race participants are entered simply because their owners and trainers are taking the opportunity to earn prize money – all the while knowing that due to their horse's lack of talent or propensity for injury, competitive racing participation and success in the future is unlikely.

Conversely, the analyses that only included horses with a competitive racing record showed that premie race participation was associated with higher levels of competitive racing success in NSCTs (all $P < 0.001$) (Figure 1). These findings support previous studies that have shown a link between early racing age and increased career success (Saastamoinen & Ojala, 1991a; Saastamoinen & Nylander, 1996; Velie et al., 2018a). Over 75% of the premie participants in this dataset continued on to race competitively, suggesting that the vast majority of horses that participate in premie races do so because they have both a good constitution and have

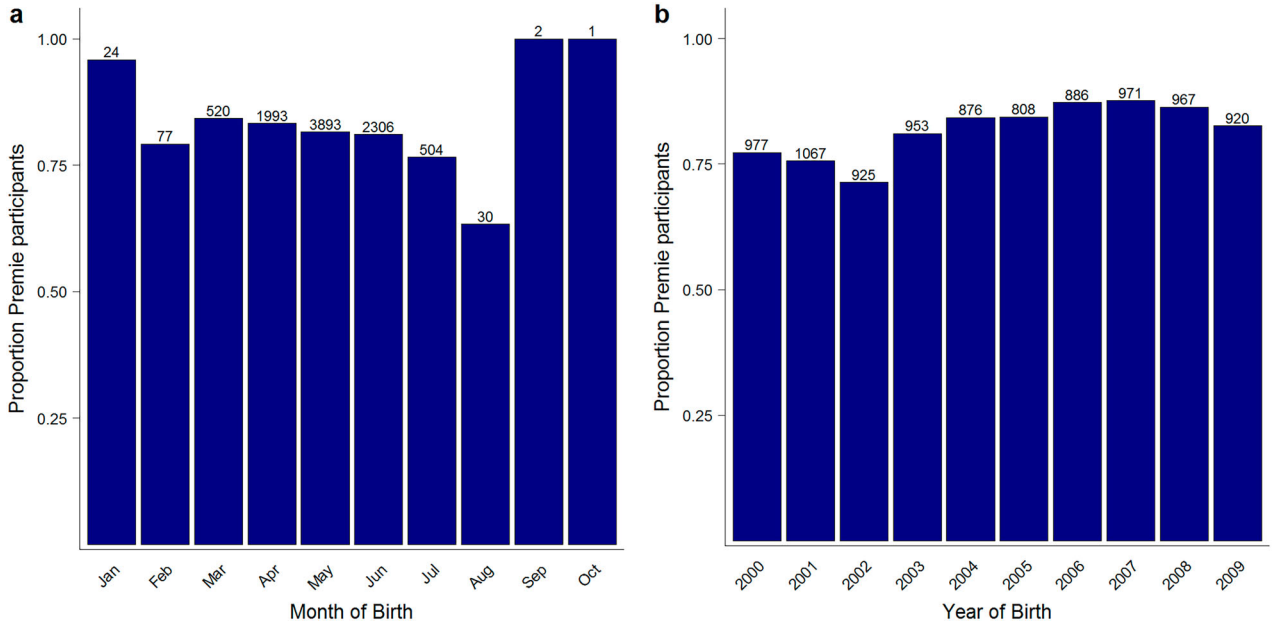

Figure 2. The proportion of premie race participants for each a) month and b) year of birth in Norwegian- Swedish Coldblooded trotters ($n = 9350$) born between 2000 and 2009. The total number of individuals for each month and year of birth is shown above the respective column.

shown their trainers that they have enough ability to start their career at an early age. Premie race participation also allows these horses to improve their fitness, experience race day conditions, and practice their trotting technique in order to qualify for competitive racing more quickly. As such, premie race participation and potential earlier qualification allows horses greater opportunities to earn prizemoney and increases their chance of a longer racing career.

Furthermore, NSCT's born earlier in the breeding season were more likely to participate in premie races ($P < 0.01$, Figure 2). These findings agree with a previous study on other trotter breeds (Saastamoinen & Ojala, 1991b). However, the month of birth showed no relationship with future competitive racing performance (all $P > 0.05$), a trend that was also found in Standardbred trotters (Physick-Sheard, 1986). Although there are some outliers, these findings suggest that more mature (early season) foals may be more likely to start training or show propensity for racing before 2 years of age. An increasing trend shows that a greater proportion of NSCTs born in the later years of the study participated in premie races ($P < 0.001$, Figure 2). The increasing amount of horses participating in premie races each year may be due to trainers observing the positive effects of race practice on future performance, or the benefits of early commercial returns to connections. It is also plausible that this may be a side effect of changing racing structures (e.g. more premie races available, increases/decreases in money awarded for premie races).

These findings provide evidence that many in the industry already suspect: premie races are made up of both horses with limited talent as well as horses with elite potential. It is highly likely that trainers, particularly professional trainers, understand a horses racing potential at an early age. For an NSCT to be physically fit and educated enough to successfully complete a premie race, the horse would have to start training before 18 months of age. While low intensity training regimes can be sufficient for trotters to participate in racing at a young age (Ringmark et al., 2015), some early training and practicing under race conditions can also assist in optimising the racing success of NSCTs, with premie races providing an ideal opportunity for trainers to assess a horses potential for future competitive racing. Additionally, these findings demonstrate that premie races provide an important function in filtering out many individuals that lack the physical and mental talent for racing.

Further research into the benefits of premie race participation would confirm whether horses that participate in premie races take fewer attempts to qualify for competitive racing. Horses may participate in more than one premie race, which could provide further advantage in early qualification. However, once a horse has finished within the specified time interval, they are not eligible to participate in future premie races. Additionally, comparing the career profiles of horses that finish premie races within the required time interval versus those that don't may provide additional insight in early determinants of career success in NSCTs. Future studies could also benefit from a more comprehensive breakdown of prizemoney over different age brackets. Overall, this study demonstrates that NSCTs that show propensity for racing at a young age have significantly longer and more successful careers.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Swedish-Norwegian Foundation for Equine Research, H-15-47-075, http://hastforskning.se (GL, ES). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Butler, D., Cullis, B.R., Gilmour, A.R. & Gogel, B.J. (2009). Asreml: asreml() fits the linear mixed model. R package version 3.0.

Cheetham, J., Riordan, A.S., Mohammed, H.O., McIlwraith, C.W. & Fortier, L.A. (2010). Relationships between race earnings and horse age, sex, gait, track surface and number of race starts for Thoroughbred and Standardbred racehorses in North America. *Equine Veterinary Journal* 42(4), 346–350. doi:10.1111/j.2042-3306.2010.00032.x

Jäderkvist Fegraeus, K., Velie, B.D., Axelsson, J., Ang, R., Hamilton, N.A., Andersson, L., Meadows, J.R.S. & Lindgren, G. (2018). A potential regulatory region near the EDN3 gene may control both harness racing performance and coat color variation in horses. *Physiological Reports* 6(10), e13700. doi:10.14814/phy2.13700

Knight, P.K. & Thomson, P.C. (2011). Age at first start and racing career of a cohort of Australian Standardbred horses. *Australian Veterinary Journal* 89(9), 325–330. doi:10.1111/j.1751-0813.2011.00816.x

Physick-Sheard, P.W. (1986). Career profile of the Canadian Standardbred II. Influence of age, gait and sex upon number of races, money won and race times. *Canadian Journal of Veterinary Research* 50(1), 457–470.

Revold, T., Larsen, S. & Ihler, C.F. (2010). Prediction of early race starts in Norwegian-Swedish Coldblooded Trotters. *Acta Veterinaria Scandinavica* 52(1), 52–53. doi:10.1186/1751-0147-52-53

Ringmark, S., Lindholm, A., Hedenström, U., Lindinger, M., Dahlborn, K., Kvart, C. & Jansson, A. (2015). Reduced high intensity training distance had no effect on VLa4 but attenuated heart rate response in 2–3-year-old Standardbred

horses. *Acta Veterinaria Scandinavica* 57(1), 17. doi:10.1186/s13028-015-0107-1

Saastamoinen, M.T. (1991a). Factors affecting age at onset of breaking, training, qualifying and first start in Finnish trotters. *Acta Agriculturae Scandinavica* 41(2), 137–145. doi:10.1080/00015129109438595

Saastamoinen, M.T. (1991b). Some factors of the time of breaking and training affecting racing performance in young trotters. *Journal of Agricultural Science in Finland* 63(5), 483–492. doi:10.1080/00015129109438595

Saastamoinen, M.T. & Nylander, A. (1996). Genetic and phenotypic parameters for age at starting to race and racing performance during early career in trotters. *Livestock Production Science* 45(1), 63–68. doi:10.1016/0301-6226(95)00080-1

Saastamoinen, M.T. & Ojala, M.J. (1991a). Estimates of genetic and phenotypic parameters for racing performance in young trotters. *Acta Agriculturae Scandinavica* 41(4), 427–436. doi:10.1080/00015129109439925

Saastamoinen, M.T. & Ojala, M.J. (1991b). Influence of birthmonth on age at first start and racing performance in young trotters. *Acta Agriculturae Scandinavica* 41(4), 437–445. doi:10.1080/00015129109439926

Velie, B.D., Knight, P.K., Thomson, P.C., Wade, C.M. & Hamilton, N.A. (2013). The association of age at first start with career length in the Australian Thoroughbred racehorse population. *Equine Veterinary Journal* 45(4), 410–413. doi:10.1111/j.2042-3306.2012.00651.x

Velie, B.D., Jäderkvist Fegraeus, K., Ihler, C.F., Lindgren, G. & Strand, E. (2018a). Competition lifespan survival analysis in the Norwegian-Swedish Coldblooded Trotter racehorse. *Equine Veterinary Journal*. doi:10.1111/evj.12989

Velie, B.D., Jäderkvist Fegraeus, K., Solé, M., Rosengren, M.K., Røed, K.H., Ihler, C.-F., Strand, E. & Lindgren, G. (2018b). A genome-wide association study for harness racing success in the Norwegian-Swedish coldblooded trotter reveals genes for learning and energy metabolism. *BMC Genetics* 19(1), 1–13. doi:10.1186/s12863-018-0670-3

Appendix 2

SCIENTIFIC REPORTS

OPEN

Founder-specific inbreeding depression affects racing performance in Thoroughbred horses

Received: 7 December 2017
Accepted: 5 April 2018
Published online: 18 April 2018

Evelyn T. Todd[1], Simon Y. W. Ho[1], Peter C. Thomson[1], Rachel A. Ang[1], Brandon D. Velie[2] & Natasha A. Hamilton[1]

The Thoroughbred horse has played an important role in both sporting and economic aspects of society since the establishment of the breed in the 1700s. The extensive pedigree and phenotypic information available for the Thoroughbred horse population provides a unique opportunity to examine the effects of 300 years of selective breeding on genetic load. By analysing the relationship between inbreeding and racing performance of 135,572 individuals, we found that selective breeding has not efficiently alleviated the Australian Thoroughbred population of its genetic load. However, we found evidence for purging in the population that might have improved racing performance over time. Over 80% of inbreeding in the contemporary population is accounted for by a small number of ancestors from the foundation of the breed. Inbreeding to these ancestors has variable effects on fitness, demonstrating that an understanding of the distribution of genetic load is important in improving the phenotypic value of a population in the future. Our findings hold value not only for Thoroughbred and other domestic breeds, but also for small and endangered populations where such comprehensive information is not available.

The Thoroughbred horse population is one of the largest closed populations of animals in the world. Thoroughbreds are extremely valuable because of the large amount of prizemoney on offer and the high residual value of superior athletes. All Thoroughbred horses trace their ancestry back to three paternal lines, due to the narrow bottleneck at the foundation of the population[1–3]. More than 300 years of breeding practices have produced signatures of selection in the 21st century Thoroughbred population, contributing to the superior athleticism of the breed[4,5]. At the same time, these practices have increased levels of inbreeding and reduced the genetic diversity of Thoroughbreds compared with other domestic horse breeds[3,6,7].

To our knowledge, there has been no detailed examination of the effects of inbreeding on the racing performance of Thoroughbred horses and the genetic load of the population. Genetic load, the presence of unfavourable genetic material, is a reflection of a population's fitness because a higher genetic load leads to a lower mean fitness level[8]. A large proportion of genetic load consists of recessive deleterious mutations, known as mutational load. Inbreeding can expose mutational load because it increases an individual's chance of inheriting two copies of recessive deleterious alleles from a common ancestor[8,9]. The subsequent decrease in fitness caused by these expressed recessive deleterious mutations is thought to be a major cause of inbreeding depression[10]. Other mechanisms believed to contribute to inbreeding depression include epistatic interactions and reductions in favourable heterozygosity[10,11].

The inevitable effect of selection in a closed population is an increase in the level of inbreeding[12,13]. There is some evidence that continued inbreeding for selection can purge a population of some or all of its genetic load, such that new inbreeding events have negligible or even positive effects on phenotype[9]. Although some domestic and wild populations show signs of purging[14–16], others still show strong signs of inbreeding depression even after multiple population bottlenecks and inbreeding events[17–19]. Purging is most likely to occur in populations under strong selection and slow rates of inbreeding, allowing deleterious alleles to be effectively eliminated rather than

[1]School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia. [2]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, 75007, Uppsala, Sweden. Correspondence and requests for materials should be addressed to E.T.T. (email: evelyn.todd@sydney.edu.au)

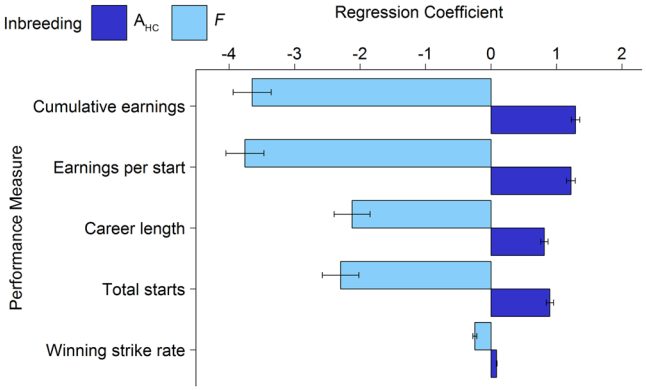

Figure 1. Regression coefficients showing the relationship between measures of racing performance and inbreeding in Thoroughbred horses ($n = 135{,}572$). All measures of racing performance have a negative relationship with F but a positive association with A_{HC}. Error bars represent 1 standard error around the mean. Regression coefficients and standard errors were divided by the standard error of their respective traits. The relationship between each measure of inbreeding and racing performance was highly significant ($P < 0.001$).

fixed by genetic drift[11,20]. Additionally, inbreeding for favourable phenotypic characteristics can have unexpected negative implications through deleterious alleles hitchhiking on regions of the genome under positive selection, thereby increasing their frequency in the population[21–23].

Understanding the effects of selection is further complicated by the uneven distribution of genetic load in a population. Inbreeding to different ancestors can have varying effects on fitness, such that the total proportion of alleles identical by descent (IBD) might not be an accurate reflection of mutational load[24–26]. This raises the possibility that inbreeding in different pedigree lines has variable effects on genetic load in the Thoroughbred population.

The availability of extensive phenotypic and pedigree records, dating back to the late 18th century, makes the Thoroughbred population ideal for studying the long-term, population-wide effects of selection on performance and genetic load. Here, we examine the effects of inbreeding on racing performance and mutational load in the Australian Thoroughbred population. Australia has the second-largest racing and breeding population in the world, containing approximately 15% of all Thoroughbreds[27].

We analyse a sample of 135,572 individuals, representing all Thoroughbred horses that had one or more race starts in Australia between 2000 and 2011. A genealogy of these individuals, dating back to the founders of the population ($n = 257{,}249$), is also included in our analyses. Although some lines of pedigree are incomplete, we have comprehensive pedigree information for all individuals in the racing performance data set, making our inbreeding estimates highly accurate. The availability of extensive pedigree records not only allows us to study broad population trends over time, but also to determine whether the selection for optimal racing performance has alleviated mutational load. We use these data to measure inbreeding and ancestral coefficients for all individuals. We also identify the ancestors that have made the greatest genetic contributions, in order to understand better the distribution of mutational load in the population. For a representative subset of individuals, we perform high-density genotyping to determine whether inbreeding load is reflected at the genomic level.

Results and Discussion

The effects of inbreeding and purging on racing performance. Our analysis of data from 135,572 Thoroughbred horses revealed a strong negative relationship (all $P < 0.001$, Fig. 1) between Wright's inbreeding coefficient, F, and five measures of racing performance that encompass a range of factors that contribute to exercise performance[28,29]. These included two measures that are based on the assumption that more successful individuals earn more prizemoney: cumulative prizemoney earnings and prizemoney earnings per start. We also included two measures of constitutional soundness: total number of race starts and career length. Finally, we accounted for consistency of performance with the measure winning strike rate.

The negative relationship between F and performance can be explained by a genetic load of partially deleterious alleles still being carried by the population. We expect that the alleles causing the observed inbreeding depression are more difficult to select out of the population than those with lethal or debilitating effects on juvenile or embryonic survival[10,21,30–32]. Population bottlenecks that occurred during the ancestry of the Thoroughbred, including the domestication of the horse[33], and the foundation of the breed[2,3], might have increased the frequency of deleterious alleles through genetic drift. It is also possible that continued inbreeding of the Thoroughbred population over the past 300 years has inadvertently increased the frequency of deleterious variants in the population, potentially through hitchhiking on selective sweep regions[13,21,23]. As a result of many generations of inbreeding, the average F of the 21st century Thoroughbred population is 0.139 ($s = 0.011$).

In contrast with the results from Wright's inbreeding coefficient, the ancestral history coefficient, A_{HC}, showed a strong positive association with racing performance (all $P < 0.001$, Fig. 1). This statistic, described by Baumung, *et al.*[34], counts the number of times that an allele has been IBD in an individual's pedigree, thus providing a comprehensive reflection of selection for favourable traits over time. The A_{HC} statistic is based on the assumption that an allele that has been IBD multiple times in an individual's pedigree is likely to have a neutral or positive effect

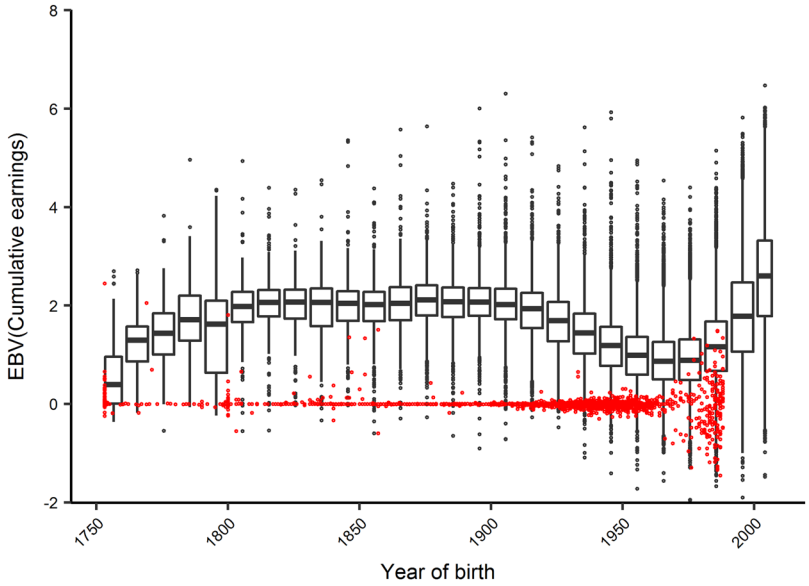

Figure 2. The distribution of estimated breeding values (EBVs) over time for Australian Thoroughbred horses ($n = 257,249$), based on the cumulative earnings of 135,572 individuals that raced between 2000 and 2010. Bins were calculated over intervals of 0.2, with each bin representing a 10-year period. Individuals with unknown parents are shown in red. The EBV results for the other measures of racing performance follow the same trends and are included in the Appendix.

on fitness. In contrast, an allele that is IBD for the first time is more likely to have a negative effect on fitness. Therefore, individuals with higher A_{HC} are more likely to contain larger proportions of alleles in their genomes that have been positively selected over many generations. It is possible for an individual with a comprehensive pedigree to have an A_{HC} greater than 1. As a consequence of the comprehensive and inbred pedigree, the reference population had average A_{HC} of 1.973 ($s = 0.089$).

The positive relationship between A_{HC} and all measures of racing performance is possibly due to the many generations of selective breeding that have increased the frequency of alleles associated with positive improvements with exercise physiology. These alleles will appear IBD more times in the pedigrees of each subsequent generation, thus driving up A_{HC} (Appendix S2). Our results indicate that inbreeding for selection has effectively increased the frequencies of favourable alleles, but has not completely eliminated genetic load from the population. Considering this finding, it is unsurprising that parts of the Thoroughbred genome show signatures of selective sweeps linked to genes related to athletic performance, including formation of muscular fibres, upregulation of mitochondrial activity, angiogenesis, brown adipose tissue formation, and lipid metabolism[5,35]. In agreement with our results, there is some evidence for selection improving racing performance in another horse breed, the Norwegian cold-blooded trotter[32].

Both F and A_{HC} showed the strongest associations with cumulative earnings and earnings per start (Fig. 1). We expect that this is because these measures reflect not only talent, but also good constitution because horses that race more are more likely to win more prizemoney. The smallest regression coefficient was for winning strike rate, probably because this measure is a crude estimate of consistency and does not reflect the race class, or the finishing order of a horse on non-winning occasions.

The estimated breeding values of the population over time. We found that selective breeding practices have not increased the overall performance levels of the population over time. We implemented a numerator relationship matrix in conjunction with a linear mixed model to account for additive genetic relationships between animals in the pedigree (Materials and Methods). Based on the racing performance of contemporary individuals ($n = 135,572$), we used this relationship matrix to calculate the estimated breeding values (EBVs) of all individuals in their pedigree ($n = 257,249$)[36,37]. The large increase in EBVs at the foundation of the population indicates that early selection events resulted in an initial jump in the frequency of favourable alleles (Fig. 2). After this initial increase, the distribution of EBVs remains constant; demonstrating that selective breeding from the early 19th century was not effective in improving the racing performance of the population. The level of F has increased constantly during this time (Fig. S9), so we conclude that inbreeding has not effectively removed mutational load from the population. This explains why we observe strong inbreeding depression persisting in the contemporary population. We expect that this is due in part to a change in racing and training regimes over time that, in turn, has changed selection pressures on the population[38]. In the 18th and early 19th century, Thoroughbred races were held over a distance of several miles, with each horse participating in multiple heats on the same day. In the 20th century, focus shifted to breeding sprinters and early developers for two-year-old racing[39]. Similarly, there was very little increase over time in the EBVs of Polish Warmblood horses despite selection

Ancestor name	Year of birth	Percentage contribution by each ancestor	
		*p*F	pA_{HC}
Herod	1758	19.87	25.13
Eclipse	1764	11.5	12.97
St Simon	1881	8.74	4.58
Godolphin Arabian	1724	8.34	10.34
Touchstone	1831	7.73	5.79
Stockwell	1849	7.15	4.76
Rachel	1763	5.75	6.32
Snap	1750	5.41	5.77
Partner	1718	3.62	12.97
Roxana	1718	2.28	2.49
Total contribution		80.40	82.18

Table 1. The average partial F (pF_i) and A_{HC} (pA_{HCi}) coefficients of the contemporary population for the 10 ancestors with the greatest marginal contributions to the modern Australian Thoroughbred population ($n = 135{,}572$). The final pair of columns shows the total average contribution of all 10 ancestors to the F and A_{HC} coefficients. All values are expressed as a percentage of the total F or A_{HC} value.

for performance, indicating that intensive selection might be necessary to improve the mean value of complex quantitative traits in a population[40].

The dip in EBVs between 1930 and 1980 can also be partly attributed to an increased number of individuals with unknown pedigree information, as shown in red on Fig. 2. This, together with the increased variability of EBVs during this period, could also be due to the presence of less successful pedigree lines that have not been purged from the modern population. We expect that the increase in the average EBV from 1980 onwards is partly due to the introduction of parental testing in the 1980s, leading to complete pedigrees for all registered individuals. The increasing trend in EBVs over recent generations indicates a possibility for future improvement in the population's overall phenotypic quality.

The uneven ancestral genetic contribution in the contemporary Thoroughbred population. Selective breeding practices are likely to result in uneven ancestral genetic contributions, favouring ancestors carrying beneficial alleles and leading to the extinction of less successful ancestral lines[25,41,42]. We found that a small number of ancestors in the early years of the breed formation accounted for much of the inbreeding coefficient in the modern Australian Thoroughbred population.

We found that 10 ancestors accounted for, on average, over 80% of the IBD alleles in the modern Australian Thoroughbreds (Table 1). Almost 20% of the IBD alleles in the contemporary population were attributed to a single individual, Herod. We selected these 10 ancestors because they provided the greatest marginal contributions to the individuals in our racing performance data (Appendix S4). The greatest marginal contributors are selected by first identifying the single ancestor with the greatest contribution to the population, and then subsequently finding the other ancestors that provide the greatest genetic contributions not accounted for by previously selected ancestors[43] (Appendix S4). We then estimated the proportion of F (pF_i) and A_{HC} (pA_{HCi}) for each individual in our data set that is attributed to each of these ancestors[34,44].

We identified these individuals as superior athletes that were also highly successful at stud. Historical records show that most of these individuals are closely related to each other (Fig. S8). One of them, Godolphin Barb, was one of the three foundation stallions of the breed in the early 18th century[1]. He has been reported to contribute to 13.8% of the genetic makeup of British Thoroughbred horses[3]. Another of the foundation stallions, Eclipse, was identified as the source of a Y chromosome mutation that is near fixation in the modern Thoroughbred population[45].

The 10 notable ancestors accounted for over 82% of the A_{HC} coefficient in their modern descendants (Table 1). We expected this relationship because these individuals appear many generations back in the pedigree of modern horses. For alleles inherited from them to have such a large contribution to F, they must appear IBD many times in the pedigrees of their descendants. In concordance with the principle of the A_{HC} coefficient, alleles that are found IBD multiple times in the pedigree are likely to have neutral or beneficial effects on fitness. These findings are reflected in the positive trends in F and A_{HC} over time in the population (Fig. S9).

Uneven distribution of genetic load between different ancestors. We found evidence that founder-specific inbreeding depression differentially affects racing performance in the Australian Thoroughbred population (Fig. 3). We determined the distribution of genetic load between the 10 dominant ancestors by using linear mixed models to examine the relationship between partial inbreeding coefficients and racing performance. Genetic load may be unevenly distributed between different ancestors, such that inbreeding to different individuals can have a variable effect on fitness[24–26]. If inbreeding to a particular ancestor results in a reduction in the racing performance of their descendants, a higher proportion of the genetic load in the population can be attributed to them. The variation in genetic load between different ancestors indicates that inbreeding depression in Thoroughbreds is due to a small number of loci that have large effects on performance[24,25,46].

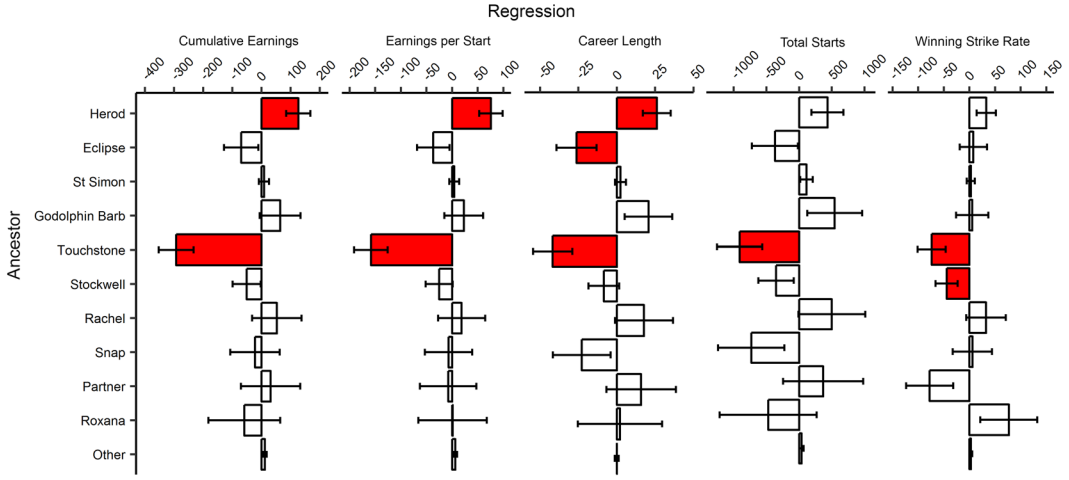

Figure 3. Inbreeding to different ancestors has variable effects on five measures of racing performance in modern Australian Thoroughbred horses. Partial inbreeding coefficients were calculated for the 10 ancestors with the greatest marginal contributions to the contemporary Australian Thoroughbred population. The relationship between each partial coefficient and inbreeding was analysed using regression coefficients from restricted maximum likelihood models. Error bars represent 1 standard error from the mean. This plot uses the same data set as in Fig. 1, but with each inbreeding coefficient split into partials. Red bars denote significant relationships.

We found that inbreeding resulting from four ancestors had significant effects on racing performance. Individuals with more IBD alleles attributed to Herod had greater cumulative earnings, earnings per start, and career length. This does not mean that increased inbreeding to Herod has had no negative effects on the phenotypic value of his descendants, but that overall they exhibit less inbreeding depression than other, equally inbred individuals[25]. Conversely, inbreeding to Eclipse, Stockwell, and Touchstone had negative effects on the racing performance of their descendants. We propose that these negative effects are partly due to the "cost of domestication"[23], whereby inbreeding these individuals has inadvertently selected for deleterious alleles linked to sites that have undergone selective sweeps[21,22].

Additionally, historical reports describe these stallions to be potential carriers of disease alleles, which may have predisposed their descendants to common conditions that reduce racing performance. Touchstone was reported by his contemporaries to have a number of conformational and behavioural issues[47], which might also have contributed to the reduced level of performance in his descendants. Although Eclipse was a superior racehorse, his grandsire suffered from exercise-induced pulmonary haemorrhage (bleeding from the lungs)[48]. This hereditary condition reduces racing success[49,50], and recurrent episodes result in a horse's permanent ban from racing in Australia. Inheritance of this condition might be a contributing factor to the reduced career lengths of Eclipse's descendants. Individuals with higher levels of inbreeding to Stockwell show reduced winning strike rates, although this might be a statistical abnormality because $P = 0.04$. However, Stockwell's mother suffered from the congenital condition of laryngeal neuropathy (paralysis of the larynx)[51–53], which may partly explain the observed reduction in performance.

We expect that most of these ancestors have passed on a mix of alleles with both positive and negative effects, such that inbreeding to them has the same effect as inbreeding to other individuals in the pedigree. However, it is also possible that two individuals inbred to the same ancestor could have inherited different sets of loci from different ancestral paths, making this ancestor's effect on fitness variable between different descendants[46]. An instance of this has been found in cattle, where the occurrence of ectodermal dysplasia in a number of calves from unaffected parents was traced back to a *de novo* mutation in one bull[54]. The condition was only revealed through inbreeding of his descendants, when some of their progeny inherited two copies of the disease allele. This example demonstrates that inbreeding to a particular individual can have highly variable effects on fitness levels between their different descendants.

Considering the strong evidence for an uneven distribution in genetic load, we conclude that the majority of inbreeding depression is only due to small proportion of IBD alleles[25,42]. Consequently, we suggest that simply measuring the proportion of IBD alleles in the genome does not provide a comprehensive reflection of a population's genetic load. Understanding the heterogeneous distribution of genetic load is important in assisting breeding decisions to minimize inbreeding to ancestors that negatively affect fitness[42].

Relationship between genome-based inbreeding coefficients and racing performance. In contrast with the pedigree-based estimates of inbreeding, we found that genomic measures of inbreeding showed no overall relationship with any measure of racing performance (Table 2). For a representative subset of the population ($n = 122$), we estimated genomic inbreeding levels as the proportion of the genome consisting of runs of homozygosity (F_{ROH}). This method reflects inbreeding levels by capturing long, homologous tracts of DNA inherited from a common ancestor that have not been broken by recombination[55–59]. For our analyses, we selected

	F_{ROH_5}	F_{ROH_12}	F	A_{HC}
Cumulative earnings	1.95 (11.62)	5.05 (14.35)	−10.56 (19.74)	3.04 (3.51)
Earnings per start	0.63 (8.53)	2.30 (10.54)	−10.10 (14.55)	1.61 (2.58)
Career length	−0.69 (2.32)	−0.84 (2.86)	−3.26 (3.64)	0.54 (0.69)
Total starts	−10.61 (87.49)	−24.01 (107.94)	16.32 (142.13)	38.81 (25.81)
Winning strike rate	−2.17 (4.16)	−5.08 (5.14)	−17.37 (6.61)*	−0.43 (1.15)

Table 2. Regression coefficients of linear mixed estimating the association between five measures of racing performance and pedigree-based and genomic coefficients ($n = 122$). Sex and year of birth were added as fixed effects and a numerator relationship matrix as a random effect in each model. Cumulative earning, earnings per start, and career length were log transformed for a normal distribution and analysed with a linear mixed model. Total starts was analysed using a Poisson generalized linear mixed model and winning strike rate using a binomial generalized linear mixed model. Inbreeding was measured using the pedigree measures of: Wright's inbreeding coefficient (F) and the ancestral history coefficient (A_{HC}). Genealogical inbreeding was measured as the proportion of runs of homozygosity (ROH) in the genome with the minimal lengths of 5MB (F_{ROH_5}) and 12MB (F_{ROH_12}). Standard errors are shown in parentheses. $^*P < 0.05$; $^{**}P < 0.001$.

minimal length thresholds of 5 Mb (F_{ROH_5}) and 12 Mb (F_{ROH_12}) to correspond to old and new inbreeding, respectively (Appendix S2).

For this smaller data set, however, we also found that the F and A_{HC} coefficients for these individuals also showed no relationship with performance (Table 2). Considering that this relationship was significant for a larger sample size, we conclude that a sample size of 122 was not sufficient to capture the relationship between inbreeding and performance. Our models were unable to account for a number of confounding environmental factors that could affect racing performance (such as training regime, jockey success, and foal-rearing process), so a large sample size is needed to tease out the underlying relationship between inbreeding and performance. There is also a large continuum between the best- and worst-performing individuals in such a large population that might not be captured by a small subset of individuals. Our findings indicate that caution should be exercised in studies of smaller populations.

Molecular estimates of inbreeding are often considered to be superior to genealogical measures because they account for the unpredictable nature of recombination and inaccurate pedigree-recording information. However, the parameters of F_{ROH} measurements should also be chosen carefully, so that they accurately reflect inbreeding levels. The accuracy of these estimations might be affected by inadequate SNP density[56,60] and long tracts of ROH persisting in areas of low recombination[61,62] (Appendix S2). Many studies use different parameters for genotyping densities, data trimming, and ROH, making comparisons between them difficult to draw.

We found that the correlation between F_{ROH} and F in our data set (Fig. S3) was lower than that reported in other domestic species[63,64], which may partly explain the contrasting results. We found that a large proportion of the inbreeding coefficient in the Australian Thoroughbred population was accounted for by ancestors many generations back in the pedigree. Inbreeding to distant ancestors results in shorter ROH regions that might not be captured by the SNP density used in our analysis (Appendix S2).

For these reasons, we believe that for large populations with comprehensive pedigrees, genealogical measures of inbreeding can provide important inferences if the size of the pedigree is much larger than the number of individuals genotyped. The use of pedigree data allows inferences to be made for deceased individuals, for which genotyping might not be possible. Additionally, using a pedigree to analyse trends over time can be advantageous because it might not be possible to obtain molecular data for deceased individuals (such as the founders of the population). Pedigrees also provide the opportunity to estimate the effects of specific individuals over time on the fitness of their descendants.

Conclusions

In this study, we have presented the effects of inbreeding and selection in a very large population with extensive phenotypic and pedigree records. Our analyses have shown that genetic load can still persist in a population even after many generations of inbreeding. However, we have also found evidence that multiple generations of inbreeding for selection can have positive effects on the overall genetic value of a population. We suggest that using EBVs whilst managing inbreeding levels will increase the efficiency of selection to reduce inbreeding depression in subsequent generations. Further, our findings highlight the need for caution in studies with small sample sizes because they can lead to inaccurate inferences about the effects of inbreeding.

We have also found evidence that the genetic load is unevenly distributed in the Thoroughbred population. This indicates that studies of inbreeding need to account for heterogeneity between different ancestors, because the total proportion of IBD alleles might not accurately reflect genetic load. Understanding the distribution of genetic load in the population will assist in breeding decisions to reduce disease alleles and improve the overall fitness of the population in future generations. Our findings open the possibility of evaluating the effects of particular individuals on the fitness of the population in order to improve phenotypic quality and reduce genetic load in the future.

Materials and Methods

Calculating pedigree-based inbreeding coefficients. Racing Australia provided race records for all individuals that had participated in a race start in Australia between 2000 and 2010 ($n = 135,572$). A genealogy of all horses born after 1970, dating back to the founders of the population, was provided by the Australian Stud

Book ($n = 500,477$) (Appendix S1). We trimmed the pedigree file so that it only included the ancestors of the individuals in our data set, leaving a pedigree size of 257,249. We found that all individuals included in our analysis had a comprehensively recorded pedigree (an average of 24.60 discrete generational equivalents of known pedigree[65,66]). Before 1980, however, a small number of individuals appear in the stud book with no recorded pedigree[67,68] (Appendix S6). These individuals accounted for 1.4% of the total ancestors included in our genealogy file, and mostly appear more than 6 generations back in the pedigree.

We estimated inbreeding levels for all individuals in the data set using Wright's inbreeding coefficient (F)[69]. We used this traditional measure of quantifying inbreeding to allow our results to be compared with those from previous studies. We also used the pedigree data to estimate several ancestral inbreeding coefficients that account for genetic load (SI Materials and Methods, Appendix S1, S2)[18,34,70]. We selected the ancestral history coefficient (A_{HC}) for further analysis because this measure counts the number of times that an allele has been IBD in an individual's pedigree, thus providing a comprehensive reflection of selection for favourable traits over time[34].

We calculated F and A_{HC} for all individuals in the pedigree using 10^6 replications of simulated gene drops in GRain 1.0[34] (Appendix S1). This method uses Mendelian segregation rules to simulate gene flow through a population by flagging each allele as it runs through the pedigree. These data are then used to estimate the probability-based inbreeding coefficients[71]. The accuracy of the results depends on the number of replications performed, which is proportional to the number of unlinked loci calculated in the analysis[34]. We checked the accuracy of our output by comparing F estimations using GRain with a deterministic approach as implemented by PEDIG[66]. Estimates from the two methods had a correlation coefficient of 0.99, indicating high accuracy of the inbreeding estimations by GRain.

We identified the 20 ancestors that provided the greatest marginal contributions to the population of 135,572 individuals by using iterations in the *prog_orig.f* program in PEDIG[43,66]. We then used GRain to calculate pF and pA_{HC} of each ancestor for each individual in our data set. Ten ancestors were chosen for further analysis, and their identities were determined using the Australian Stud Book and the online pedigree database (pedigreequery.com).

Estimating inbreeding from genomic data. We selected a representative subset of individuals for high-density genotyping ($n = 128$). These individuals were selected to provide a reflection of different bloodlines in the population and a continuum of racing successes. We used these data to estimate the proportion of the genome consisting of runs of homozygosity (F_{ROH}).

To estimate genome-based levels of inbreeding, we first extracted DNA from hair samples (collected under approval from University of Sydney Ethics Committee N00-2009-3-5109) using the Qiagen Gentra® Puregene® Tissue Kit (Qiagen, Redwood City, CA, USA). We genotyped 105 individuals using the Equine SNP70 BeadChip (Illumina, San Diego, CA, USA), which consists of 65,102 SNPs evenly distributed throughout the equine genome. Additionally, we typed 23 individuals on the Axiom Affymetrix SNP Chip (670,671 SNPs), when this higher density array became available at a later date. We used custom Perl scripts to extract only the SNPs that were common to these two panels, which we then used in further analyses.

SNP data were edited and analysed using PLINK 1.07[72]. The data were trimmed to be in concordance using the following parameters: minor allele frequency > 0.01; individual call rate > 0.9; and SNP call rate > 0.9[6,73]. This process yielded a final data set comprising 45,451 SNPs for each of 122 individuals. Additionally, we only analysed autosomal SNPs in order to exclude any bias between male and female.

We used these data to estimate the proportion of the genome consisting of runs of homozygosity (F_{ROH}). To define the parameters of an ROH, we set the minimum density to 0.05 Mb/SNP and the largest gap to 1 Mb, in accordance with the settings used by Goddard, *et al.*[74] and Silió, *et al.*[63]. We set the minimal number of SNPs in each ROH to 20, because our SNP coverage was approximately 1 SNP every 50 Mb, making this sufficient to distinguish an ROH of 1 Mb. ROH lengths were calculated as a proportion of total ROH length in relation to the total equine autosome size of 2,242,879,462 bp.

Measuring racing performance. We selected five different measures of racing performance that account for talent, consistency, and constitutional soundness[28,29,75]. These measures were: cumulative earnings ($AU), earnings per start ($AU), career length (months), total number of race starts, and winning strike rate. Cumulative earnings and earnings per start favour talented individuals, because more prestigious races carry larger prizemoney purses. Career length and total starts favour individuals with good constitutions; individuals with health and conformational defects are unable to race for extended periods. Winning strike rate accounts for consistency in horses, because more talented horses are expected to win a higher proportion of their race starts (Appendix S3).

Statistical analysis. The relationship between each measure of inbreeding and racing performance was analysed using (generalized) linear mixed models in ASReml-R 3.0[76]. We used the five measures of racing performance as outcome variables. Cumulative earnings, earnings per start and career length were analysed with linear mixed models. These variables were log-transformed; to accommodate zero-value in these measure, $100 was added to all career earnings and earnings per starts, and 1 month to all career length values. Total starts was analysed using a Poisson generalized linear mixed model and winning strike rate using a binomial generalized linear mixed model.

Each model included a predictor variable of either F, A_{HC}, or F partitioned into partial coefficients for each of the 10 important ancestors, making a total of 15 models. Sex and year of birth were also included as predictor variables in each model. We also included a random animal effect that was associated with the numerator relationship matrix derived from the pedigree ($n = 257, 249$).

The significance of fixed effects was assessed using Wald tests. To allow comparisons of regression coefficients across different traits, the regression coefficients were divided by the standard deviation of their respective traits. EBVs were obtained from the fitted models. We summarised the EBV distributions over time in 10-year bins, which approximately represents one generation interval. We calculated the average generational interval to be 10.5 years using the *intgen.f* program from PEDIG[66].

Data Availability Statement. The data that support the findings of this study are available from Racing Australia and the Australian Stud Book. However, restrictions apply to the availability of these data, which were used under license for the current study, and so they are not publicly available. Data are, however, available from the authors upon reasonable request and with permission of Racing Australia. The data set can also be accessed from the public repositories of www.racingaustralia.horse and www.studbook.org.au.

References

1. Weatherby, J. *An Introduction to a General Stud Book*. (Weatherby and Sons, 1791).
2. Corbin, L. J. *et al*. Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Anim. Genet.* **41**, 8–15 (2010).
3. Cunningham, E. P., Dooley, J. J., Splan, R. K. & Bradley, D. G. Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to Throughbred horses. *Anim. Genet.* **32**, 360 (2001).
4. Petersen, J. L. *et al*. Genome-Wide Analysis Reveals Selection for Important Traits in Domestic Horse Breeds. *PLoS Genet*. **9**, e1003211 (2013).
5. Gu, J. *et al*. A genome scan for positive selection in Thoroughbred horses. *PLoS ONE* **4**, e5767 (2009).
6. Petersen, J. L. *et al*. Genetic Diversity in the Modern Horse Illustrated from Genome-Wide SNP Data. *PLoS ONE* **8**, e54997 (2013).
7. Metzger, J. *et al*. Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genomics* **16**, 764 (2015).
8. Hedrick, P. W. & Garcia-Dorado, A. Understanding InbreedingDepression, Purging, and Genetic Rescue. *Trends in Ecol. Evol.* **31**, 940–952 (2016).
9. Crnokrak, P. & Spencer, C. H. B. Perspective: purging the genetic load: a review of the experimental evidence. *Evolution* **56**, 2347–2358 (2002).
10. Charlesworth, D. & Willis, J. H. The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**, 783–796 (2009).
11. Leroy, G. Inbreeding depression in livestock species: review and meta-analysis. *Anim. Genet.* **45**, 618–628 (2014).
12. Woolliams, J. A., Berg, P., Dagnachew, B. S. & Meuwissen, T. H. E. Genetic contributions and their optimization. *J. Anim. Breed. Genet.* **132**, 89–99 (2015).
13. Zhang, Q., Guldbrandtsen, B., Bosse, M., Lund, M. S. & Sahana, G. Runs of homozygosity and distribution of functional variants in the cattle genome. *BMC Genomics* **16**, 542 (2015).
14. Mc Parland, S., Kearney, F. & Berry, D. P. Purging of inbreeding depression within the Irish Holstein-Friesian population. *Genet. Select. Evol.* **41**, 16 (2009).
15. Moreno, E., Pérez-González, J., Carranza, J. & Moya-Laraño, J. Better fitness in the Capture Cuvier's Gazelle despite inbreeding increase: evidence for purging? *PLoS ONE* **10** (2015).
16. Laws, R. J. & Jamieson, I. G. Is lack of evidence of inbreeding depression in a threatened New Zealand robin indicative of reduced genetic load? *Anim. Conserv.* **14**, 47–55 (2011).
17. Kennedy, E. S., Grueber, C. E., Duncan, R. P. & Jamieson, I. G. Severe inbreeding depression and no evidence of purging in an extremely inbred wild species- the Chatham Island black robin. *Evolution* **68**, 987–995 (2014).
18. Kalinowski, S. T., Hedrick, P. W. & Miller, P. S. Inbreeding depression in the Speke's gazelle captive breeding program. *Conserv. Biol.* **14**, 1375–1384 (2000).
19. Boakes, E. H., Wang, J. & Amos, W. An investigation of inbreeding depression and purging in captive pedigreed populations. *Heredity* **98**, 172–182 (2007).
20. Kristensen, T. N. & Sørensen, A. C. Inbreeding - lessons from animal breeding, evolutionary biology and conservation genetics. *Anim. Sci.* **80**, 121–133 (2005).
21. Marsden, C. D. *et al*. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc. Natl. Acad. Sci. USA* **113**, 152–157 (2016).
22. Cruz, F. *et al*. The Legacy of domestication: accumulation of deleterious mutations in the dog genome. *Mol. Biol. Evol.* **25**, 2331–2336 (2008).
23. Lu, J. *et al*. The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends Genet.* **22**, 126–131 (2006).
24. Lacy, R. C., Alaks, G. & Walsh, A. Hierarchical Analysis of Inbreeding Depression in Peromyscus polionotus. *Evolution* **50**, 2187–2200 (1996).
25. Gulisija, D., Gianola, D., Weigel, K. A. & Toro, M. A. Between-founder heterogeneity in inbreeding depression for production in Jersey cows. *Livest Sci* **104**, 244–253 (2006).
26. Casellas, J., Varona, L., Ibáñez-Escriche, N., Quintanilla, R. & Noguera, J. L. Skew distribution of founder-specific inbreeding depression effects on the longevity of Landrace sows. *Genet Res (Camb)* **90**, 499–508 (2008).
27. Australian Racing Board. 2015/2016 Australian Racing Fact Book. (Australian Racing Board, Sydney, Australia, 2016).
28. Velie, B. D., Hamilton, N. A. & Wade, C. M. Heritability of racing performance in the Australian Thoroughbred racing population. *Anim. Genet.* **46**, 23–29 (2015).
29. Velie, B. D., Hamilton, N. A. & Wade, C. M. Performance selection for Thoroughbreds racing in Hong Kong. *Equine Vet. J.* **47**, 43–47 (2015).
30. Charlesworth, B. & Charlesworth, D. The genetic basis of inbreeding depression. *Genet. Res.* **74**, 329–340 (1999).
31. Hinrichs, D. *et al*. Analysis of inbreeding depression in the first litter size of mice in a long-term selection experiment with respect to the age of the inbreeding. *Heredity* **99**, 81–88 (2007).
32. Klemetsdal, G. The effect of inbreeding on racing performance in Norwegian cold- blooded trotters. *Genet. Select. Evol.* **30**, 351–366 (1998).
33. Schubert, M. *et al*. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci. USA* **111**, E5661–E5669 (2014).
34. Baumung, R. *et al*. GRain: a computer program to calculate ancestral and partial inbreeding coefficients using a gene dropping approach. *J. Anim. Breed. Genet.* **132**, 100–108 (2015).
35. Hill, E. W., Gu, J., McGivney, B. A. & MacHugh, D. E. Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. *Anim. Genet.* **41**, 56–63 (2010).
36. Henderson, C. R. Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423–447 (1975).
37. Bijma, P. & Woolliams, J. A. Prediction of rates of inbreeding in populations selected on best linear unbiased prediction of breeding value. *Genetics* **156**, 361A–3373 (2000).

38. Cassidy, R. Blood will tell In *The sport of kings: kinship, class and thoroughbred breeding in Newmarket* 140–157 (Cambridge University Press, 2002).
39. Vamplew, W. Development of racing in *The turf: a social and economic history of horse racing* Ch. 1, 1–28 (Allen Lane, 1976).
40. Borowska, A., Wolc, A. & Szwaczkowski, T. Genetic variability of traits recorded during 100-day stationary performance test and inbreeding level in Polish warmblood stallions. *Archiv fur Tierzucht* **54**, 327–337 (2011).
41. Dadar, M., Mahyari, S. A., Rokouei, M. & Edriss, M. A. Rates of inbreeding and genetic diversity in Iranian Holstein Cattle: Inbreeding, Genetic Diversity. *Anim. Sci. J.* **85**, 888–894 (2014).
42. Rodrigáñez, J., Toro, M. A., Rodriguez, M. C. & Silió, L. Effect of founder allele survival and inbreeding depression on litter size in a closed line of Large White pigs. *Anim. Sci.* **67**, 573–582 (1998).
43. Boichard, D., Maignel, L. & Verrier, É. The value of using probabilities of gene origin to measure genetic variability in a population. *Genet. Select. Evol.* **29**, 5–23 (1997).
44. Nagy, I. *et al*. Genetic diversity and population structure of the synthetic Pannon White rabbit revealed by pedigree analyses. *J. Anim. Sci.* **88**, 1267–1275 (2010).
45. Wallner, B. *et al*. Identification of genetic variation on the horse Y chromosome and the tracing of male founder lineages in modern breeds. *PLoS ONE* **8**, e60015 (2013).
46. Casellas, J., Piedrafita, J., Caja, G. & Varona, L. Analysis of founder-specific inbreeding depression on birth weight in Ripollesa lambs. *J. Anim. Sci.* **87**, 72–79 (2009).
47. Fairfax-Blakeborough, J. *Northern turf history, vol. III* (J.A. Allen & Co, 1970).
48. Erickson, H. H., Kindig, C. A. & Poole, D. C. Exercise-induced pulmonary hemorrhage: A new concept for prevention. *J Equine Vet Sci.* **20**, 164–167 (2000).
49. Morley, P. S., Bromberek, J. L., Saulez, M. N., Hinchcliff, K. W. & Guthrie, A. J. Exercise-induced pulmonary haemorrhage impairs racing performance in Thoroughbred racehorses. *Equine Vet. J.* **47**, 358–365 (2015).
50. Velie, B. D., Raadsma, H. W., Wade, C. M., Knight, P. K. & Hamilton, N. A. Heritability of epistaxis in the Australian Thoroughbred racehorse population. *Vet. J.* **202**, 274–278 (2014).
51. Ewart, J. C. Horse-breeding and horse-racing. *Nature* **99**, 346–347 (1917).
52. Boyko, A. R. *et al*. Genomic analysis establishes correlation between growth and laryngeal neuropathy in Thoroughbreds. *BMC Genomics* **15**, 259–259 (2014).
53. Dupuis, M. C. *et al*. Detection of copy number variants in the horse genome and examination of their association with recurrent laryngeal neuropathy. *Anim. Genet.* **44**, 206–208 (2013).
54. Bourneuf, E. *et al*. Rapid discovery of *de novo* deleterious mutations in cattle enhances the value of livestock as model species. *Sci. Rep.* **7**, 11466 (2017).
55. Keller, M. C., Visscher, P. M. & Goddard, M. E. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* **189**, 237–249 (2011).
56. Purfield, D. C., Berry, D. P., McParland, S. & Bradley, D. G. Runs of homozygosity and population history in cattle. *BMC Genet.* **13**, 70 (2012).
57. Leutenegger, A.-L. *et al*. Estimation of the inbreeding coefficient through use of genomic data. *Am. J. Hum. Genet.* **73**, 516–523 (2003).
58. McQuillan, R. *et al*. Runs of homozygosity in European populations. *Am. J. Hum. Genet.* **83**, 359–372 (2008).
59. Kristensen, T. N., Pedersen, K. S., Vermeulen, C. J. & Loeschcke, V. Research on inbreeding in the 'omic' era. *Trends Ecol. Evol.* **25**, 44–52 (2010).
60. Marras, G. *et al*. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Anim. Genet.* **46**, 110–121 (2015).
61. Gibson, J., Morton, N. E. & Collins, A. Extended tracts of homozygosity in outbred human populations. *Hum. Mol. Genet.* **15**, 789–795 (2006).
62. Pemberton, T. J. *et al*. Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.* **91**, 275–292 (2012).
63. Silió, L. *et al*. Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. *J. Anim. Breed. Genet.* **130**, 349–360 (2013).
64. Kim, E.-S. *et al*. Effect of artificial selection on runs of homozygosity in U.S Holstein cattle. *PloS ONE* **8**, e80813 (2013).
65. Woolliams, J. A. & Mäntysaari, E. A. Genetic contributions of Finnish Ayrshire bulls over four generations. *Anim. Sci.* **61**, 177–187 (1995).
66. Boichard, D. In *7th World Congress on Genetics Applied to Livestock Production* Vol. paper 28–13 (Montpellier, France, 2002).
67. O'Loghlen, F. *Champions of the Turf: By 'Eurythmic' (Frank O'Loghlen)*. (F. H. Johnston Publishing Company, 1945).
68. Ford, M. *The Australian Stud Book History*. Vol. 1 (The Australian Stud Book, 2006).
69. Wright, S. Coefficients of inbreeding and relationship. *Am. Nat.* **56**, 330–338 (1922).
70. Ballou, J. D. Ancestral inbreeding only minimally affects inbreeding depression in mammalian populations. *J. Hered.* **88**, 169–178 (1997).
71. Suwanlee, S., Baumung, R., Sölkner, J. & Curik, I. Evaluation of ancestral inbreeding coefficients: Ballou's formula versus gene dropping. *Conserv. Genet.* **8**, 489–495 (2007).
72. Purcell, S. *et al*. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
73. Ferencakovic, M., Solkner, J. & Curik, I. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet. Select. Evol.* **45**, 42–42 (2013).
74. Goddard, M. E., Powell, J. E. & Visscher, P. M. Reconciling the analysis of IBD and IBS in complex trait studies. *Nat. Rev. Genet.* **11**, 800–805 (2010).
75. Velie, B. D., Wade, C. M. & Hamilton, N. A. Profiling the careers of Thoroughbred horses racing in Australia between 2000 and 2010. *Equine Vet. J.* **45**, 182–186 (2013).
76. Gilmour, A. R., Gogel, B. J., Cullis, B. R. & Thompson, R. *ASReml User Guide Release 3.0*. (VSN International Ltd, 2009).

Acknowledgements

We acknowledge Racing Australia and The Australian Stud Book for their support and providing data for this study. We are grateful to the anonymous donors of DNA samples. We thank Emeritus Professor Frank Nicholas for his comments on the manuscript. SYWH was supported by the Australian Research Council. This research was supported by Sydney Informatics Hub, funded by the University of Sydney.

Author Contributions

E.T.T., S.Y.W.H., P.C.T. and N.A.H. designed research, E.T.T., B.D.V. and R.A.A. performed research, E.T.T. and P.C.T. analysed data; and E.T.T., S.Y.W.H., P.C.T. and N.A.H. wrote the paper.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-24663-x.

Competing Interests: Data was provided by Racing Australia and The Australian Stud Book.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Todd et al. BMC Genetics (2020) 21:41
https://doi.org/10.1186/s12863-020-00847-1

Appendix 3

BMC Genetics

RESEARCH ARTICLE Open Access

The effects of inbreeding on covering success, gestation length and foal sex ratio in Australian thoroughbred horses

Evelyn T. Todd[1*], Natasha A. Hamilton[2], Brandon D. Velie[1] and Peter C. Thomson[1]

Abstract

Background: Horses produce only one foal from an eleven-month gestation period, making the maintenance of high reproductive rates essential. Genetic bottlenecks and inbreeding can increase the frequency of deleterious variants, resulting in reduced reproductive levels in a population. In this study we examined the influence of inbreeding levels on foaling rate, gestation length and secondary sex ratio in Australian Thoroughbred mares. We also investigated the genetic change in these traits throughout the history of the breed. Phenotypic data were obtained from 27,262 breeding records of Thoroughbred mares provided by three Australian stud farms. Inbreeding was estimated using the pedigree of each individual dating back to the foundation of the breed in the eighteenth century.

Results: While both gestation length and foaling rate were heritable, no measurable effect of inbreeding on either trait was found. However, we did find that the genetic value for both traits had decreased within recent generations. A number of environmental factors also had significant effects on foaling rate and gestation length. Secondary sex ratio had only an extremely small paternal heritable effect and was not susceptible to environmental influences.

Conclusions: In contrast to racing performance, inbreeding had no measurable effect on foaling rate or gestation length in Australian Thoroughbred horses. This could be because the level of inbreeding in the population examined is not high enough to show a discernible effect on reproductive traits. Populations that experience higher levels of inbreeding due to use of artificial reproductive technologies or extremely small population sizes may show a more pronounced reduction in natural foaling rate or gestation length. It is also possible that the intensive management techniques used in the Thoroughbred population masks any negative effects of inbreeding. The decrease in the genetic value of foaling rate is likely to be because horses with unfavourable genetic potential have not yet been selected out of the population. The change in genetic value of gestation length may be due to selective breeding favouring horses with shorter pregnancies. We also found that prioritising the mating of older mares, and avoiding out of season mating could lead to an increased breeding success.

Keywords: Inbreeding, Genetic diversity, Thoroughbred horse, Fertility

* Correspondence: evelyn.todd@sydney.edu.au
[1]School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia
Full list of author information is available at the end of the article

Background

Horses have an 11 month gestation period, both conceiving and giving birth in the spring and summer [1]. Like most spring seasonal-breeding animals, increased photoperiod induces ovulation cycles in mares. Survival of twins is rare, so mares can only produce one foal a year. The long gestation period and short breeding season make the maintenance of good fertility rates in horse populations imperative to provide commercial returns for domestic breeds, and to increase the size of endangered populations [2, 3].

Deleterious genetic variants have accumulated in the genomes of modern horses as a result of population bottlenecks during domestication and breed foundation events [4, 5]. In this process, also known as the "cost of domestication", deleterious mutations increase in frequency by "hitchhiking" on selective sweep regions [6]. These mutations can also increase in frequency through inbreeding, selective breeding and genetic drift in a population [7]. The presence of these variants in a population can have negative consequences for overall fitness, including a decrease in fertility rates. However, it is also possible that selective breeding over a number of generations may have removed some or all of these deleterious variants from contemporary horse populations. Evidence of positive selection in regions harbouring genes related to conceptus development have been found in domestic horse breeds [8, 9], indicating that fertility rates may have been targeted and improved by breeding practices.

The effects of inbreeding on reproductive traits vary between studies. Increased inbreeding levels were associated with reduced fertility in some domestic and wild horse populations [2, 3, 10]. Impaired ovarian function resulting from high levels of inbreeding was reported in the Przewalski's horse, the most closely related species to the domestic horse [3]. Conversely, a number of studies in other horse breeds have shown no relationship between inbreeding levels and reproductive traits [11–13]. Varying relationships between inbreeding and reproductive performance also exist for a number of other domestic animal populations [14–18]. It is possible that the effects of inbreeding on fertility may vary between different populations depending on the rate of increase in inbreeding, selective pressures and genetic diversity.

As well as affecting fertility, there is some evidence that increased inbreeding can skew secondary sex ratios (the sex ratio at birth) in animal populations [19]. Variations in sex ratio exist due to an increased chance of early conceptus loss of one sex under different conditions [20]. As maternal condition declines due to environmental stresses or inbreeding, the chance of producing a viable male conceptus may also decrease [19, 21]. Early female horse conceptuses produce more insulin like growth factor-1 than males, which may promote their survival in adverse conditions [20]. It is possible that the environment at the time of conception or levels of inbreeding in horses may favour the survival of one sex.

Maintaining high reproductive rates is particularly important for the Thoroughbred horse breed. The Thoroughbred population has been closed since the eighteenth century, resulting in reduced levels of genetic diversity in current individuals [22–24]. Although selective breeding in Thoroughbred horses focusses on improving racetrack performance, the prohibition on reproductive technologies also makes it essential to maintain good fertility rates in the population. Thoroughbred foals that are born early in the spring season are assumed to have a size and maturing advantage over their peers. Mares that can conceive within 30 days after parturition will give birth at the same time next year. However, mares often take more than one covering each season, leading to their parturition date being delayed until later in the spring every year [25]. When a mare's parturition date reaches the end of spring, she cannot be covered until the next year. This will result in a two-year period in which she does not produce a foal.

Despite many generations of selective breeding for athletic ability, increased inbreeding is associated with reduced racing performance in Thoroughbred horses [24]. In this study we examine the effects of inbreeding levels on foaling rate, gestation length and secondary sex ratio in Thoroughbred mares. We use the pedigree data of twenty-first century Thoroughbred horses to estimate the heritability and the effects of inbreeding on these three reproductive traits. We also evaluate the environmental effects on these reproductive traits in Australian Thoroughbred mares. Additionally, we used estimated breeding values to measure the genetic change in foaling rate and gestation length since the foundation of the Thoroughbred population in the eighteenth century. Estimated breeding values (or genetic values) can measure the genetic potential of an animal for a trait based on the phenotypic information of themselves and their relatives. Genetic values are not commonly used for Thoroughbred horses but are utilized to assist in breeding management for other horse and livestock breeds. Although phenotypic data are not available for previous generations, we utilize the comprehensive pedigree information available for Thoroughbred horses to calculate the genetic values for the ancestors in the pedigree, based on the reproductive trait data available for their modern day descendants.

Results

Data were available for 27,296 coverings of 12,922 mares bred to 131 stallions between 2000 and 2017. The pedigree of these individuals dating back to the founders of the population consisted of 92,852 individuals. The

conceptuses, mares and stallions in the dataset had an average pedigree depth of 29 generations and a mean level of inbreeding of 0.156 ± 0.012 (mean ± SD).

The overall proportion of mares with a positive 15 day scan each season was 81.65% (22,287 out of 27,296). Based on the estimated variances from the mixed models, the maternal heritability of foaling rate was 0.058 (± 0.015), and 0.00 (± 0.00) for paternal heritability. Estimated breeding values for all individuals in the pedigree based on the foaling rate of their modern descendants showed a decrease in recent generations from a mean of − 0.002 (± 0.092) in 1990 to − 0.133 (± 0.138) by 2017 (Fig. 1). Variation in the estimated breeding values of foaling rate was also estimated to increase dramatically from 1990 onwards (between − 1.05 and 0.502) from previous years (between − 0.468 to 0.264) (Fig. 1). Mares that were covered later in the season showed a significant reduction in foaling success ($P < 0.001$) (Fig. 2). There was an overall decrease in foaling rate with increasing mare age ($P < 0.001$) (Fig. 3). Foaling rate had no relationship with the sire, dam or conceptus inbreeding level ($P = 0.142$, 0.788 and 0.701 respectively).

The sex of 7578 live foal births were recorded in the dataset, 3785 of which were colts (49.95%) and 3793 were fillies (50.05%). Secondary sex ratio did not have an estimated heritable component for maternal genetic estimates (0.005 (± 0.026)), but had a small paternal heritability estimate of 0.011 (± 0.005). The sex ratio was not influenced by the sire, dam or foal inbreeding level ($P = 0.637$, 0.746 and 0.899, respectively). Environmental variables of mare age and month of birth also had no significant relationship with sex ratio ($P = 0.495$ and 0.337, respectively).

Gestation length data were available for 764 foals from 152 mares covered by 89 stallions. Gestation length was normally distributed, with a mean of 341 days (± 8.633), a minimum of 311 and a maximum of 376. Based on the estimated variances from the mixed model, the maternal heritability of gestation length was found to be 0.562 (± 0.042) and the paternal heritability was 0.004 (± 0.001). The average estimated breeding value in the pedigree remained at a mean of 0.03 (± 0.241) and showed little variation between − 4.695 and 3.959 from the foundation of the population until 1990 (Fig. 4). Since 2000, the average estimated breeding value for gestation length has decreased to − 0.7 (± 1.91) and variation has increased, with a minimum of − 11.82 and a maximum of 15.41. There was a significant decrease in gestation length in the later months of the season ($P < 0.001$) (Fig. 5). Gestation length increased linearly with mare age ($P < 0.001$), going from a mean of 342 days at 2 years old to a mean of over 354 days by 24 years old (Fig. 6). Male foals had a significantly longer predicted gestation length (349 days) than female foals (346 days) ($P < 0.001$). Gestation length had no significant association with the sire ($P = 0.087$), dam ($P = 0.419$) or foal ($P = 0.062$) inbreeding level.

Discussion

We found that inbreeding had no measurable effect on foaling rate, sex ratio or gestation length in Australian Thoroughbred horses. This may be because intensive management techniques used on commercial Thoroughbred stud farms have masked any negative effects of inbreeding on these traits. Some studies have similarly shown that inbreeding has no effect on gestation length [12, 13], or foaling rate [11] in domestic horses. However, high levels of inbreeding in the endangered Przewalski and Sorraia horse breeds are associated with decreased birthing rates [2, 3]. It is possible that

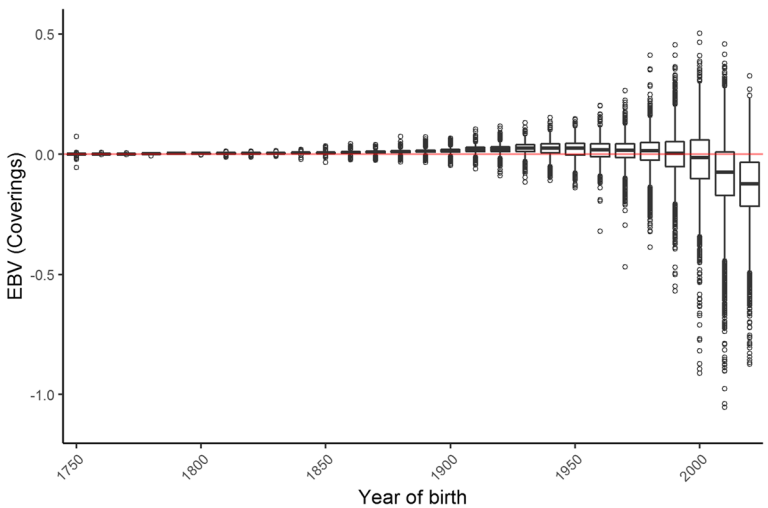

Fig. 1 Boxplot of the distribution of estimated breeding values (EBVs) over time for Thoroughbred horses ($n = 95,663$), based on the foaling rate of 27,962 individuals bred between 2000 and 2017. Each bin represents a 10-year period

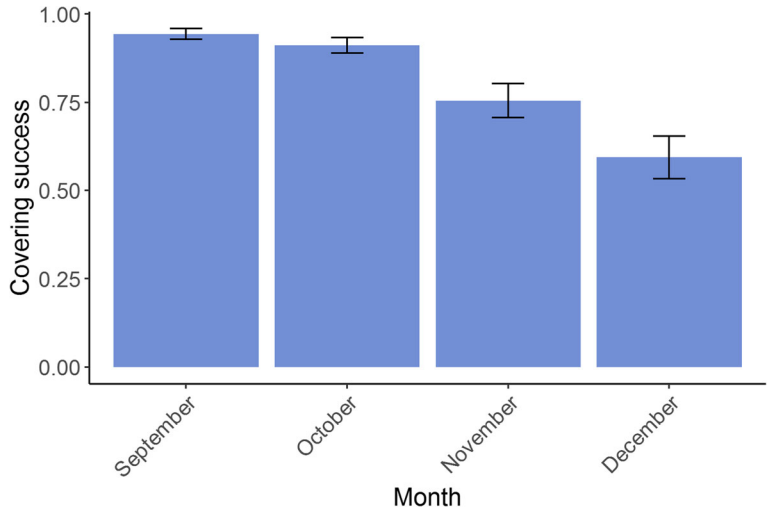

Fig. 2 The relationship between the predicted values of foaling rate by month of covering for Australian Thoroughbred horses between 2000 and 2017 ($n = 27{,}962$). The error bars represents ±1 standard error of the predicted value

inbreeding has no measurable effects on reproductive traits until it reaches a very high level. Both the Sorraia and the Przewalski's horse population have extremely high inbreeding levels (0.21 and 0.38, respectively) due to small effective population sizes and recent severe population bottlenecks [2, 3, 26, 27]. In contrast, the relatively high average inbreeding coefficient (0.156) found for Thoroughbreds in this study is due to many generations of slow inbreeding. A large increase in the rate of inbreeding may lead to more noticeable effects on the reproductive traits of Thoroughbred horses in the future. The prohibition of artificial reproductive technologies (e.g. artificial insemination, cloning and embryo transfer) likely limits the rate of increase in inbreeding in the Thoroughbred population. The use of these technologies in other horse populations has been associated with increases in the rate of inbreeding [28, 29], which could show more measurable effects on reproductive traits. It is also important for Thoroughbred breeders to be vigilant in their selection of sires and dams to avoid overbreeding to successful families as it may increase inbreeding rates in future generations and have unexpected negative effects on the population.

The maternal heritability estimate for gestation length based on our models was 0.562, higher than previous estimates in horses of 0.18–0.39 [12, 30, 31]. This increased heritability estimate may be because all mares in this study have been intensively managed in the same

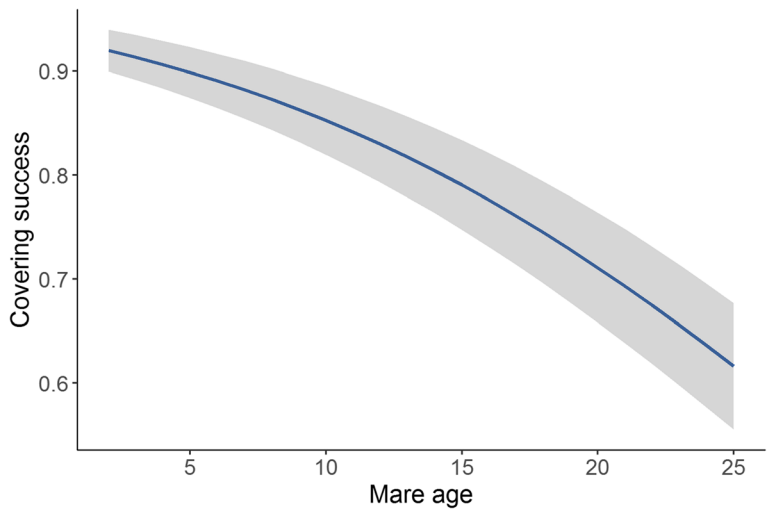

Fig. 3 The relationship between the predicted values of foaling rate and mare age for Australian Thoroughbred horses between 2000 and 2017 ($n = 27{,}962$). The grey band represents ± standard error of the predicted value

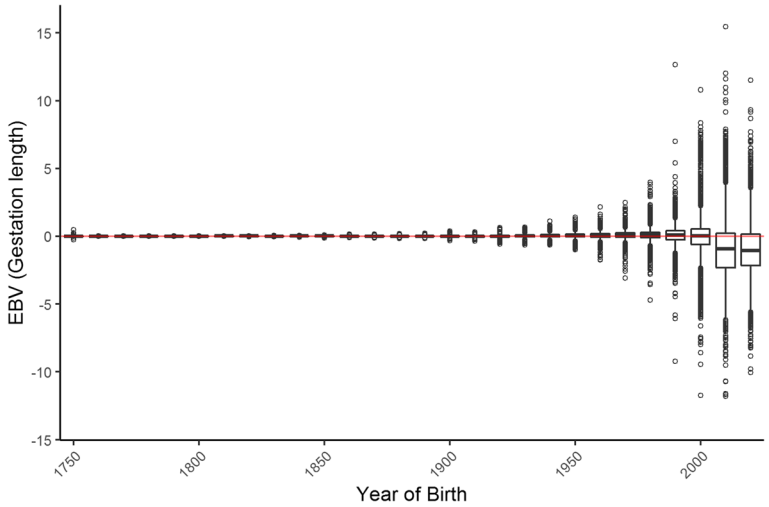

Fig. 4 Boxplot of the distribution of estimated breeding values (EBVs) over time for Thoroughbred horses ($n = 95{,}663$), based on the gestation length of 764 individuals bred between 2000 and 2017. Each bin represents a 10-year period

way, minimising the amount of environmental variation that can reduce such estimates. Sires were also estimated to have a smaller, but still significant genetic effect on the gestation length of the foal. Stallions have similarly been found to influence gestation length in other horse breeds [30, 31], and such knowledge could assist in breeding management decisions. For example, avoiding the mating of mares and stallions genetically predisposed to longer gestation lengths, particularly towards the end of the season, could aid in avoiding delayed parturition dates.

Foaling rate had a lower but still significant maternal heritability estimate of 0.058, which similarly to gestation length was slightly higher than previous estimates (0.013–0.024) [32]. Unexpectedly, paternal heritability estimates from our models found no genetic influence of the sire on foaling rate. Only 131 sires were included in this study, so it is possible that a larger sample size would reveal significant heritable effects. Additionally, Thoroughbred stallions that show suboptimal fertility may be gelded and returned to racing. Analyses that include these individuals may reveal more measurable paternal genetic effects on covering success.

In contrast to gestation length and covering success, we found that secondary sex ratio had a negligible maternal heritable component (0.005) with a high standard error, indicating that genetic variation in the mare has no influence on the sex of the foal. However, paternal heritability estimates revealed an extremely small but

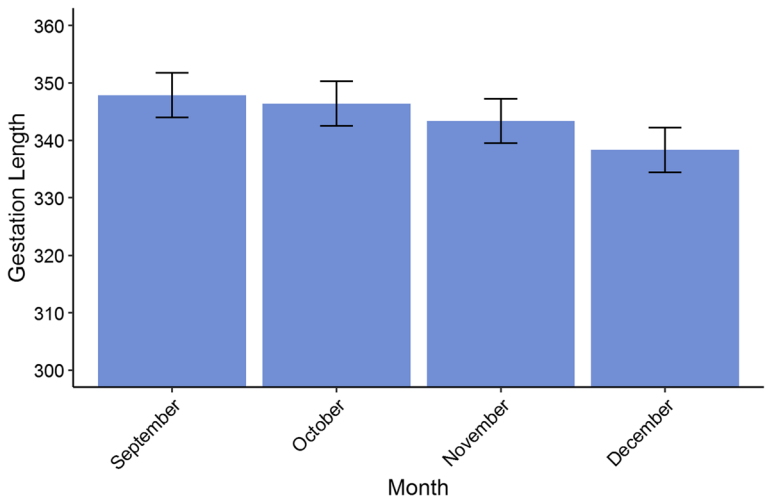

Fig. 5 The relationship between the predicted values of gestation length by month of covering for Australian Thoroughbred horses between 2000 and 2017 ($n = 764$). The error bars represents ±1 standard error of the predicted value

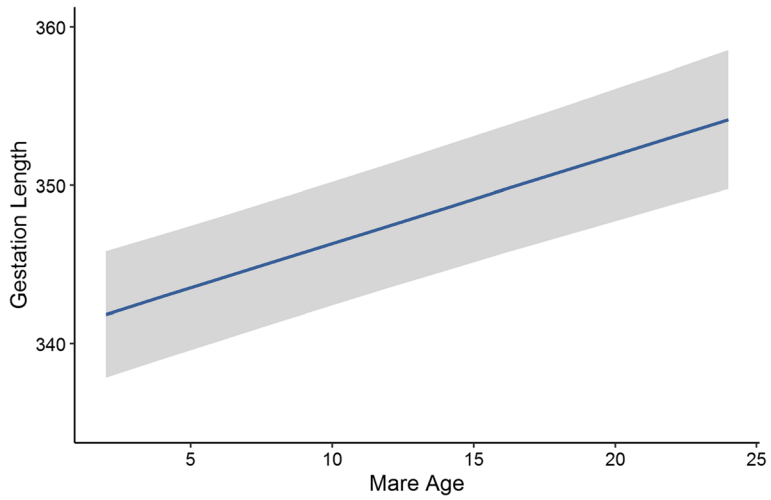

Fig. 6 The relationship between the predicted values of gestation length by mare age for Australian Thoroughbred horses between 2000 and 2017 ($n = 764$). The grey band represents ± standard error of the predicted value

nonetheless significant effect of sire genetic variation on foal sex ratios (0.011). Other studies in mammals have also found evidence of male-driven sex ratio bias [33, 34]. However, in contrast to these studies, we found no association between the inbreeding level of the sire and offspring sex ratio [33, 34]. It is postulated that higher quality, less inbred fathers produce more male offspring [35]. The artificial breeding practices used in the Thoroughbred population that only allow a small proportion of high-quality males to breed may mitigate any effects of inbreeding on offspring sex ratios, and result only a small measurable heritable effect.

In this study there was found to be little improvement in the estimated breeding value (genetic value) for reproductive traits in Thoroughbred horses from the eighteenth century until present day. The horses for which these values are obtained only include the direct ancestors to the horses with reproductive trait records in our study. Some individuals in previous generations may have had greater variation in genetic value but do not appear in the pedigree of modern Thoroughbreds. We found no change in the mean genetic value for foaling rate until the most recent three generations (Fig. 1). Selection may favour mares that produce many offspring, making them more likely to be present in the pedigree of future generations. On the other hand, female families with low fertility have less chance of appearing in the pedigree of Thoroughbreds in future generations. It is possible that if we had been able to evaluate the genetic value for all horses in the studbook from previous generations, we would have found a greater spread of values. Another reason for the minimal variation in the genetic value of earlier generations may be due to the lack of information conveyed in the binary trait of foaling rate. The near zero values may represent regression to the mean when there is limited information from previous generations.

In the past 30 years (approximately three generations), the average genetic value for foaling rate has decreased, and variation has increased. The decreasing mean may be because horses with a lower genetic value have not yet been selected out of the current Thoroughbred population. These horses may not appear in the pedigree of future Thoroughbreds, such that the values from their generation will show minimal variation. The decreasing average genetic value conflicts with an increased foaling rate reported in recent years [36]. It is likely that the improving fertility rates in the population are due to better management techniques rather than genetic gain. Recent advances in veterinary treatments has led to widespread use of hormonal therapies to increase foaling rate [36]. Increasing commercial demand for mares with good fertility may explain the outlying individuals with high genetic potential in recent years. However, widespread use of intensive veterinary treatments could make selection against individuals with lower genetic value less efficient, resulting in long term reductions in natural reproductive levels of the Thoroughbred population.

Similarly to foaling rate, the genetic value for gestation length has shown little variation since the foundation of the breed (Fig. 4). However, unlike foaling rate, gestation length is not a binary trait. Selection against horses with longer gestation lengths (corresponding to higher genetic values) is likely to occur because they can produce fewer foals throughout their lifetimes. Extremely short gestation lengths may also be selected against because foals born prematurely will have a lower survival rate. These factors will remove individuals with high and low

values, resulting in minimal variation and a mean of zero in more distant ancestral generations. It is also possible that the low genetic diversity in previous generations of the pedigree has resulted in reduced variation of the genetic value of these individuals based on breeding records from the current population. In the past 20 years, the average genetic value has reduced, showing that gestation length in Thoroughbred mares has, on average, become shorter. The increase in commercial Thoroughbred breeding during this time may have favoured mares with slightly shorter gestation lengths, as they are more likely to be successfully covered again in the same season. Foals that are born earlier in the season can have a competitive advantage over their peers when racing at a young age, which also favours mares with a decreased gestation length. Variation in genetic values of gestation length has also increased in the recent generations of the pedigree, indicating that there are increasing numbers of mares with genetic potential for very long and short gestation lengths. These individuals may be selected out of the population in future generations.

In contrast to foaling rate and gestation length, the genetic values for racing performance in the Thoroughbred population have increased in recent generations [24]. This may be because Thoroughbred horses are directly selected for good racing success rather than fertility. If racing performance traits are driven primarily by positive selection (selection for advantageous alleles), this would explain the increase in genetic values for these traits over the past few generations. On the other hand, reproductive traits may be driven by mostly negative selection against disadvantageous alleles, explaining the decrease in genetic values of foaling rate in recent generations because there has not been an opportunity for the population to be purged of these genes. Thoroughbred horses have been selectively bred for racing performance since the start of the eighteenth century. However, it is likely that domestic horses have been selected for reproductive traits for many generations prior to the foundation of the Thoroughbred breed. Common signatures of selection for fertility have been found in domestic horse breeds including the Thoroughbred [8, 9]. This could explain why an increase in genetic values for racing performance traits, but not fertility measures, is seen at the foundation of the Thoroughbred breed.

We also found that a number of environmental effects have significant influences on both gestation length and foaling rate. Our results showed that mares foaling down later in the season had significantly shorter gestation lengths (Fig. 5). Our findings agree with previous studies [12, 30, 32] and highlight the importance of photoperiod length for inducing parturition in horses [37]. This pattern may strongly depend on the mare's location, as photoperiod variation is highly dependent on the proximity to the equator.

We also found that mares who produced male foals had significantly longer gestation lengths than those who produced female foals, possibly because of differences in maternal-foetal hormonal interactions [12, 13, 32, 38]. However, an average difference of 3 days between the gestation lengths of colts and fillies is unlikely to impact breeding management decisions. Gestation length also increased with mare age, which could be explained by changes in hormonal, nutritional and uterine changes as a female ages [13] (Fig. 6). Some studies have found a similar linear increase with age [32], whereas others have found that gestation length is longer in both younger and older mares [12]. This pattern may be dependent on the veterinary management provided to maiden mares. Foaling rate declined with increasing mare age, with mares over the age of 20 having less than 70% success (Fig. 3). Foaling rate reduced dramatically in November and December, most likely due to the accumulation of less fertile mares at the end of the season (Fig. 2). To optimise foaling rate, the breeding of older mares would need to be prioritized because they tend to have longer gestation lengths and lower foaling rates.

In contrast to foaling rate and gestation length, secondary sex ratio was not significantly influenced by any environmental effects included in these models. Mares in a poor nutritional condition at conception have been reported to have an increased chance of successfully carrying a female foetus, with reports of female foal ratios up to 80% [39, 40]. In the Mangalarga Marchador horse breed, higher ratios of female foals were found in older mares [21]. Additionally, increased dam inbreeding in cattle is associated with higher female birth rate [19]. However, we postulate that the high level of veterinary management and care provided to Thoroughbred horses in commercial stud farms examined in this study has resulted in no environmental or inbreeding factors having a measurable influence on foal sex ratios. Differing management of other horse populations (e.g. feed and hormonal supplements) may show more measurable effects on sex ratio. Wild animal populations with more variable environmental conditions and rates of inbreeding may also show different trends.

Conclusions

In this study we found that inbreeding had no measureable effect on reproductive traits in Australian Thoroughbred horses. Although this contrasts with previous findings in racing performance traits, we postulate that inbreeding levels are not yet high enough to have a measurable effect on the Thoroughbred reproduction traits examined in the current study. The effects of artificial reproductive technologies on natural fertility rates in other horse populations should be examined to ensure that such practices do not result in long-term reductions

Table 1 Mixed models of each fertility trait for Australian Thoroughbred horses. Foal, sire and dam inbreeding were estimated in separate models. Model descriptions are in R syntax, not full mathematical notation

Outcome variable	Model type	Models
Gestation length	LMM	GL ~ mare.age + month + foal.sex + season + *F* + ped (mare) + ide (mare) GL ~ mare.age + month + foal.sex + season + *F* + ped (sire) + ide (sire)
Foaling rate	Binomial GLMM	CS ~ mare.age + month + season + *F* + ped (mare) + ide (mare) CS ~ mare.age + month + season + *F* + ped (sire) + ide (sire)
Foal sex ratio	Binomial GLMM	SR ~ mare.age + month + season + *F* + ped (mare) + ide (mare) SR ~ mare.age + month + season + *F* + ped (sire) + ide (sire)

LMM Linear mixed model, *GLMM* Generalised linear mixed model, *GL* Gestation length, *CS* Foaling rate, *SR* Foal sex ratio, mare.age = age of the mare at the time of covering, month = month of covering, *F* = inbreeding coefficient of the sire, dam or foal (models were run separately for each), ped = pedigree of the mare, ide = permanent environmental effect of the mare

in natural fertility levels. Genomic scans for reproduction traits in Thoroughbred and other breeds may assist in understanding genetic variation that influences fertility. We also found that unlike racing performance, there has been little increase in the breeding value of reproductive traits in Thoroughbred horses. Breeding values of foaling rates have decreased in recent generations, possibly because these traits are primarily governed by negative rather than positive selection. Further monitoring of these traits in future generations would assist in understanding the selective forces influencing these traits.

Methods

Reproductive trait data were provided by three large Australian Thoroughbred stud farms that provide a representative sample of the population as a whole. These data included the mating records of 12,922 mares bred to 131 stallions between 2000 and 2017. The scan status of each mare covering (negative or positive at the first scan 15 days after covering) was transformed into a binary trait. The sex of each live foal recorded in the dataset ($n = 7578$) was also transformed into a binary trait (female = 1, male = 0). Additionally, more detailed reproductive trait data were available for 152 mares mated to 89 stallions over multiple seasons ($n = 764$ foals), including the date of the foal birth. The gestation length of each live foal birth was calculated from these data.

The pedigree for all the mares, stallions and conceptuses included in the study dating back to the founders of the population consisted of 92,852 records. We used the CFC program (version 1) to reorder the pedigree to ensure that each individual was listed after their parents and to estimate Wright's inbreeding coefficient for all individuals [41]. We also used CFC to estimate the overall average of the average number of generations in the pedigree for each mare and stallion [41].

The genetic and environmental influences on the foaling rate, gestation length and sex ratio were estimated in ASReml-R [42], using a linear mixed model for gestation length, and a generalised liner mixed model for the binary traits of foaling rate and foal sex. Details of the fixed and random effects in each model are included in Table 1. The outcome variables were foaling rate, the sex of the foal and the gestation length. The fixed effects included in each model were: inbreeding coefficient (of the mare, the stallion and the foal), the month of covering, the year of covering, the age of the mare and the stud farm. An animal model was implemented through an inverse relationship matrix using the pedigree in ASReml-R [42]. Multiple coverings of the same mare in the same season were accounted for in the models by inclusion of a permanent environment effect of the mare, as a random effect.

The heritability of each outcome variable was estimated using the variance component estimates of the fitted models. Significance of model terms was evaluated through Wald statistics and the estimated value of each fixed effect tabulated. Estimated breeding values (i.e. genetic values) for foaling rate and gestation length were calculated as best linear unbiased predictions for each individual in the pedigree using the models fitted in ASReml-R. This method uses the available phenotypic data and the associated pedigree structure in the models to provide genetic value estimates for all animals in the pedigree.

Acknowledgements

The authors acknowledge the technical assistance provided by the Sydney Informatics Hub, a Core Research Facility of the University of Sydney.

Authors' contributions

ETT collected and collated the data. ETT and PCT analysed the data. ETT wrote the manuscript. ETT, PCT, BDV and NAH all edited, read and approved the final manuscript.

Funding

No funding was provided.

Availability of data and materials

The data that support the findings of this study are available upon reasonable request but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

Ethics approval and consent to participate

DNA samples were collected under approval from University of Sydney Ethics Committee N00–2009–3-5109.

Consent for publication

Not applicable.

Competing interests
NAH is supported by Racing Australia in the form of salary. All other authors declare that they have no competing interests.

Author details
[1]School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia. [2]Racing Australia Equine Genetics Research Centre, Racing Australia, Sydney, NSW 2000, Australia.

Received: 5 September 2019 Accepted: 24 March 2020
Published online: 08 April 2020

References

1. Aurich C. Reproductive cycles of horses. Anim Reprod Sci. 2011;124(3):220–8.
2. Kjöllerström HJ, Gama LT, Oom MM. Impact of inbreeding on fitness-related traits in the highly threatened Sorraia horse breed. Livest Sci. 2015;180:84–9.
3. Collins CW, Songsasen NS, Vick MM, Wolfe BA, Weiss RB, Keefer CL, et al. Abnormal reproductive patterns in Przewalski's mares are associated with a loss in gene diversity. Biol Reprod. 2012;86:2.
4. Schubert M, Jónsson H, Chang D, Sarkissian CD, Ermini L, Ginolhac A, et al. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. Proc Natl Acad Sci U S A. 2014;111(52):E5661–E9.
5. Librado P, Fages A, Gaunitz C, Leonardi M, Wagner S, Khan N, et al. The evolutionary origin and genetic makeup of domestic horses. Genetics. 2016; 204(2):423–34.
6. Marsden CD, Vecchyo DO-D, O'Brien DP, Taylor JF, Ramirez O, Vilà C, et al. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proc Natl Acad Sci U S A. 2016;113(1): 152–7.
7. Moyers BT, Morrell PL, McKay JK. Genetic costs of domestication and improvement. J Hered. 2017;109(2):103–16.
8. Metzger J, Karwath M, Tonda R, Beltran S, Agueda L, Gut M, et al. Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. BMC Genomics. 2015;16(1):764.
9. Gu J, Orr N, Park SD, Katz LM, Sulimova G, MacHugh DE, et al. A genome scan for positive selection in thoroughbred horses. PLoS One. 2009;4(6): e5767.
10. Klemetsdal G, Johnson M. Effect of inbreeding on fertility in Norwegian trotter. Livest Prod Sci. 1989;21(3):263–72.
11. Muller-Unterberg M, Wallmann S, Distl O. Effects of inbreeding and other systematic effects on fertility of Black Forest draught horses in Germany. Acta Vet Scand. 2017;59:70.
12. Valera M, Blesa F, Santos RD, Molina A. Genetic study of gestation length in Andalusian and Arabian mares. Anim Reprod Sci. 2006;95(1/2):75–96.
13. Ewert M, Luders I, Borocz J, Uphaus H, Distl O, Sieme H. Determinants of gestation length in thoroughbred mares on German stud farms. Anim Reprod Sci. 2018;191:22–33.
14. Lazzari G, Colleoni S, Duchi R, Galli A, Houghton FD, Galli C. Embryonic genotype and inbreeding affect preimplantation development in cattle. Reproduction. 2011;141(5):625–32.
15. Leroy G, Phocas F, Hedan B, Verrier E, Rognon X. Inbreeding impact on litter size and survival in selected canine breeds. Vet J. 2015;203(1):74–8.
16. Martikainen K, Tyrisevä AM, Matilainen K, Pösö J, Uimari P. Estimation of inbreeding depression on female fertility in the Finnish Ayrshire population. J Anim Breed Genet. 2017;134(5):383–92.
17. Saura M, Fernández A, Varona L, Fernández AI, de Cara MÁR, Barragán C, et al. Detecting inbreeding depression for reproductive traits in Iberian pigs using genome-wide data. Genet Sel Evol. 2015;47(1):1.
18. Zhang C, MacNeil MD, Kemp RA, Dyck MK, Plastow GS. Putative loci causing early embryonic mortality in Duroc swine. Front Genet. 2018;17(9):655.
19. Mc Parland S, Kearney JF, Rath M, Berry DP. Inbreeding effects on milk production, calving performance, fertility, and conformation in Irish Holstein-Friesians. J Dairy Sci. 2007;90(9):4411–9.
20. Beckelmann J, Budik S, Helmreich M, Palm F, Walter I, Aurich C. Sex-dependent insulin like growth factor-1 expression in preattachment equine embryos. Theriogenology. 2013;79(1):193–9.
21. Santos MM, Maia LL, Nobre DM, Oliveira Neto JF, Garcia TR, Lage MCGR, et al. Sex ratio of equine offspring is affected by the ages of the mare and stallion. Theriogenology. 2015;84(7):1238–45.
22. Cunningham EP, Dooley JJ, Splan RK, Bradley DG. Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to Throughbred horses. Anim Genet. 2001;32(6):360.
23. Corbin LJ, Blott SC, Swinburne JE, Vaudin M, Bishop SC, Woolliams JA. Linkage disequilibrium and historical effective population size in the thoroughbred horse. Anim Genet. 2010;41(2):8–15.
24. Todd ET, Ho SYW, Thomson PC, Ang RA, Velie BD, Hamilton NA. Founder-specific inbreeding depression affects racing performance in thoroughbred horses. Sci Rep. 2018;8(1):6167.
25. Kuhl J, Stock KF, Wulf M, Aurich C. Maternal lineage of warmblood mares contributes to variation of gestation length and bias of foal sex ratio. PLoS One. 2015;10(10):e0139358.
26. Der Sarkissian C, Ermini L, Schubert M, Yang Melinda A, Librado P, Fumagalli M, et al. Evolutionary genomics and conservation of the endangered Przewalski's horse. Curr Biol. 2015;25(19):2577–83.
27. Liu G, Shafer ABA, Zimmermann W, Hu D, Wang W, Chu H, et al. Evaluating the reintroduction project of Przewalski's horse in China using genetic and pedigree data. Biol Conserv. 2014;171:288–98.
28. Teegen R, Edel C, Thaller G. Population structure of the Trakehner hrse breed. Animal. 2009;3(1):6–15.
29. Roos L, Hinrichs D, Nissen T, Krieter J. Investigations into genetic variability in Holstein horse breed using pedigree data. Livest Sci. 2015;177:25.
30. Bene S, Benedek Z, Nagy S, Szabo F, Polgar PJ. Some effects on gestation length of traditional horse breeds in Hungary. J Cent Eur Agric. 2014;15(1):1.
31. Rodrigues JA, Gonçalves AR, Antunes L, Bettencourt EV, Gama LT. Genetic and environmental factors influencing gestation length in Lusitano horses. J Equine Vet Sci. 2020;84:102850.
32. Christmann A, Sieme H, Martinsson G, Distl O. Genetic and environmental factors influencing gestation length and parturition conception interval in Hanoverian Warmblood. Livest Sci. 2017;199:63–8.
33. Malo AF, Gilbert TC, Riordan P. Drivers of sex ratio bias in the eastern bongo: lower inbreeding increases the probability of being born male. Proc Biol Sci. 2019;286(1902):20190345.
34. Malo AF, Martinez-Pastor F, Garcia-Gonzalez F, Garde J, Ballou JD, Lacy RC. A father effect explains sex-ratio bias. Proc Biol Sci. 2017;284:1861.
35. Douhard M, Festa-Bianchet M, Coltman D, Pelletier F. Paternal reproductive success drives sex allocation in a wild mammal. Evolution. 2016;70(2):358–68.
36. Allen WR, Wilsher S. Half a century of equine reproduction research and application: a veterinary tour de force. Equine Vet J. 2018;50(1):10–21.
37. Nolan MB, Walsh CM, Duff N, McCrarren C, Prendergast RL, Murphy BA. Artificially extended photoperiod administered to pre-partum mares via blue light to a single eye: observations on gestation length, foal birth weight and foal hair coat at birth. Theriogenology. 2017;100:126–33.
38. Wulf M, Erber R, Ille N, Beythien E, Aurich J, Aurich C. Effects of foal sex on some perinatal characteristics in the immediate neonatal period in the horse. J Vet Behav. 2017;18:37–42.
39. Aurich C, Schneider J. Sex determination in horses-current status and future perspectives. Anim Reprod Sci. 2014;146(1–2):34–41.
40. Cameron EZ, Linklater WL, Stafford KJ, Veltman CJ. Birth sex ratios relate to mare condition at conception in Kaimanawa horses. Behav Ecol. 1999;10(5): 472–5.
41. Sargolzaei M, Iwaisaki H, Colleau JJ. CFC: a tool for monitoring genetic diversity. In: Proceedings of the 8th world congress on genetics applied to livestock production; 2006.
42. Gilmour AR, Gogel BJ, Cullis BR, Thompson R. ASReml user guide release 3.0. Hemel Hempstead: VSN International Ltd; 2009.

Publisher's Note

Appendix 4

SCIENTIFIC REPORTS
nature research

Check for updates

OPEN

A genome-wide scan for candidate lethal variants in Thoroughbred horses

Evelyn T. Todd[1]✉, Peter C. Thomson[1], Natasha A. Hamilton[2], Rachel A. Ang[1], Gabriella Lindgren[3,4], Åsa Viklund[3], Susanne Eriksson[3], Sofia Mikko[3], Eric Strand[5] & Brandon D. Velie[1]

Domestic animal populations are often characterised by high rates of inbreeding and low effective population sizes due to selective breeding practices. These practices can result in otherwise rare recessive deleterious alleles drifting to high frequencies, resulting in reduced fertility rates. This study aimed to identify potential recessive lethal haplotypes in the Thoroughbred horse breed, a closed population that has been selectively bred for racing performance. In this study, we identified a haplotype in the *LY49B* gene that shows strong evidence of being homozygous lethal, despite having high frequencies of heterozygotes in Thoroughbreds and other domestic horse breeds. Variant analysis of whole-genome sequence data identified two SNPs in the 3′UTR of the *LY49B* gene that may result in loss of function. Analysis of transcriptomic data from equine embryonic tissue revealed that *LY49B* is expressed in the trophoblast during placentation stage of development. These findings suggest that *LY49B* may have an essential, but as yet unknown function in the implantation stage of equine development. Further investigation of this region may allow for the development of a genetic test to improve fertility rates in horse populations. Identification of other lethal variants could assist in improving natural levels of fertility in horse populations.

There is estimated to be a high rate of natural embryonic mortality in mammals. A large proportion of these embryonic losses occur soon after fertilisation, such that pregnancies often go undetected, with the only sign being reduced fertility[1]. Mutation screens in mice reveal that many genes are essential for development, with knockout of 29% of genes tested resulting in embryonic death by day 14[2,3]. Although mutations in these genes are expected to be under strong negative selection due to being completely deleterious, many species are estimated to carry between one and two recessive lethal mutations per genome[4]. However, single mutations are often uncommon in a population, such that unrelated individuals are unlikely to carry the same recessive lethal mutations[5–7]. The likelihood of an individual inheriting two copies of the same lethal mutation is dramatically increased by inbreeding events, whereby alleles that are identical by descent are inherited from a common ancestor[8–10].

In recent years, a number of studies in livestock have identified embryonic lethal mutations at high frequencies due to intensive selective breeding practices[11–18]. This is often due to a limited number of sires with desirable characteristics making large genetic contributions to the population[11,19]. Moreover, population bottlenecks due to domestication and breed formation have also resulted in increased deleterious mutation loads and diminished gene pools in many domestic breeds[20–23]. These processes lead to a reduction in genetic diversity indices including effective population size, which is defined as the theoretical population size that shows the same rate of loss in genetic diversity as the study population (N_e)[24,25]. Reduced genetic diversity can increase the risk of drift and inbreeding events in future generations of a population. Lethal mutations that have reached high frequencies are often detected by deviations from the Hardy–Weinberg equilibrium with a lack of homozygotes for one allele[12]. Characterisation of such mutations can assist in improving breeding decisions to increase fertility rates in these populations and prevent these mutations from drifting to higher frequencies[26,27].

[1]School of Life and Environmental Sciences, The University of Sydney, Sydney, Australia. [2]Racing Australia Equine Genetics Research Centre, Sydney, Australia. [3]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden. [4]Livestock Genetics, Department of Biosystems, KU Leuven, Leuven, Belgium. [5]Department of Companion Animal Clinical Sciences, Faculty of Veterinary Medicine, Norwegian University of Life Sciences, Oslo, Norway. ✉email: evelyn.todd@sydney.edu.au

Population	Sample size	Reference	6:38278097				6:38278874			
			Expected GG	GG	AG	AA	Expected CC	CC	AC	AA
Australian Thoroughbreds	156	Own data	7*	0	66	90	7*	0	66	90
Japanese Thoroughbreds	370	Fawcett et al.[69]	9*	0	117	253	9*	0	117	253
Swedish Warmblood	380	Privately provided, Ablondi et al.[70]	4*	0	75	304	4*	0	74	306
Coldblooded Trotter	646	Privately provided, Velie et al.[84]	26*	0	258	388	28	22	226	393
Quarter Horse	137	Petersen et al.[65]	17*	0	97	40	17*	0	97	40
Exmoor Pony	285	Velie et al.[71]	0	0	1	279	0	0	1	282
Various breeds	582	Petersen et al.[30]	15	0	85	497	15	0	85	497

Table 1. The allele frequencies of two adjacent SNPs with an absence of minor homozygotes in genotype data from two Thoroughbred horse datasets. The expected number of minor homozygotes in each population was calculated under Hardy–Weinberg equilibrium. Observed genotype frequencies that significant deviate from Hardy–Weinberg equilibrium frequency expectations ($p < 0.05$) are denoted with an asterisk.

The identification of high frequency lethal variants is of particular interest in domestic horse populations. Although a recent study has identified some candidate mutations[28], to date there has been no published comprehensive characterisation of common embryonic lethal alleles in horse populations. Despite the large variety of domestic horse breeds found throughout the world, many breeds suffer from low within-breed diversity and small N_e[29–31]. Some horse breeds with large census population sizes also experience low genetic diversity due to intense artificial selective breeding practices and closed population structures[30,31]. Maintaining good fertility rates is particularly important for horse populations due to the seasonal nature of breeding and the low individual fertility output, as mares produce only one foal from an eleven month gestation period[32]. Despite the extensive use of hormonal therapies to increase covering success in many domestic horse populations, per cycle pregnancy rates in some breeds only average around 65%, suggesting the presence of unknown variables that may reduce fertility[33].

In this study, we aimed to characterise variants at high frequencies that may cause lethality in the Thoroughbred horse population. The Thoroughbred breed is of particular interest due to the closed population structure since the foundation of the studbook in the eighteenth century[34]. The population has since been intensely selected for the improvement of athletic abilities[35,36], resulting in contemporary Thoroughbred horses being characterised by high levels of inbreeding and a small N_e[31,37–39]. Due to selective breeding practices, all Thoroughbred horses can trace their ancestry back to a small number of individuals from the foundation of the breed[37,38]. Genetic diversity in the Thoroughbred breed has been reduced in recent decades due to the increased commercialisation of popular stallions providing large genetic contributions to the population[40]. Although such practices are in line with selective breeding principles[41], they could also inadvertently increase the frequency of embryonic lethal variants in the population. Reproductive technologies such as artificial insemination are banned in the Thoroughbred population, making the maintenance of high levels of natural fertility imperative. Additionally, Thoroughbred horses have been used as foundation stock for other popular horse breeds including The Quarter Horse, Standardbred, and many Warmblood breeds[30]. Therefore, identification of lethal variants in Thoroughbreds is also likely to assist in the breeding management of these populations. We also aimed to determine the frequency of any potentially lethal variants identified in the Thoroughbred population in other horse breeds and examine their transcriptomic profile in embryonic tissue.

Results

Identifying candidate lethal SNPs at high frequencies in Thoroughbred horses. Analysis of genotype data from Thoroughbred horses ($n = 156$) identified only two adjacent, linked SNPs that significantly deviated from the Hardy–Weinberg equilibrium with an absence of homozygotes (Table 1). Under Hardy–Weinberg equilibrium, seven minor allele homozygotes were expected for both of these SNPs in the dataset. Genotype data from Japanese Thoroughbred horses ($n = 370$) also showed an absence of homozygotes for these SNPs (Table 1). In this dataset, the expected number of minor homozygotes for these SNPs under Hardy–Weinberg equilibrium was nine. Despite a complete absence of homozygotes, almost 35% of Thoroughbreds across both of these datasets were heterozygous for this two-SNP haplotype. These SNPs also showed an absence or reduction of minor homozygotes in genotype data for other domestic horse breeds ($n = 2030$, Table 1).

These two candidate SNPs mapped to the coordinates of 6:38278097 (rs68661802) and 6:38278874 (rs68663106), which are found in an intronic region of the *LY49B* gene on chromosome 6. This gene is part of the *LY49* gene family, which plays an important role in innate immunity. There are five functional members of the *LY49* gene family in *Equus caballus*, all of which closely grouped together on chromosome 6. Since both of the SNPs mapped to a non-coding region of the *LY49B* gene, the likelihood of either being a causal variant for lethality is low.

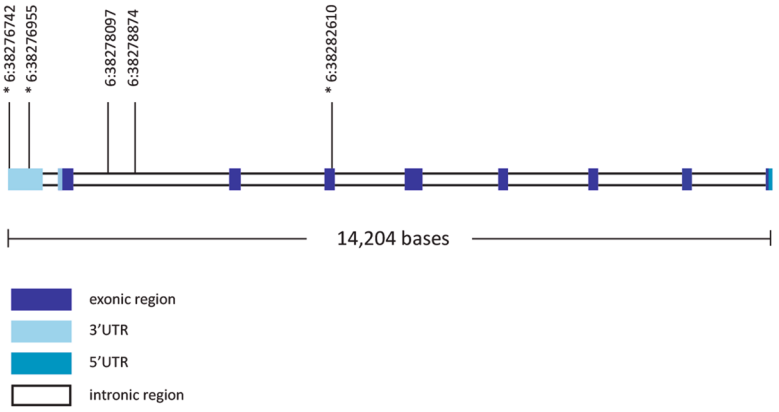

Figure 1. The equine *LY49B* gene structure and SNP positions. The two variants in the intronic region (6:38278097 and 6:38278874) were identified in preliminary analysis as showing a significant absence of homozygotes for one allele. The three variants marked with a * are in linkage disequilibrium to these SNPs and may cause a loss of function in homozygous state. The structure of the gene is based on the EquCab 2.0 reference genome where the *LY49B* gene is on the reverse strand.

Gene	Sequence
LY49B	AAAGACTTTCTCAG**G**GCCATTAAAGAGATGGGAAACTGC TTTCCAAAGAC
LY49C	AGAGAATTTCCCAG**G**GCCATTAAAGAGAAGAGCAACTGA TTTCCAAAGAC
LY49D	AGAGAATTTCTCAG**G**GCCATTAAAGAGAAGGGCAACTGA TTTCCAAAGAC
LY49E	AGAGAATTTCTCAG**G**GCCATTAAAGAGAAGGGCAACTGA TTTCCAAAGAC
LY49F	AGAGAATTTTGCAG**G**GTCATTAAAGAGAGGGGTAACTGC TTTCCAAAGAC

Table 2. Amino acid residue sequence in a conserved area of the 3′UTR found in all *Equus caballus LY49* genes as mapped in the EquCab2.0 assembly. The SNP position of 6:38276955G > A is highlighted in bold.

Phylogenetic origin of the candidate SNPs. According to the phylogenetic tree generated by Petersen et al.[29,30], and their associated SNP data, the SNPs of interest were present in heterozygous state across most phylogenetic branches of domestic horse breeds. Of the 32 breeds in this dataset, 23 had at least one heterozygote for both SNPs of interest. Notably, this two-SNP haplotype was not found in genotype data from one branch of the tree which contains the North Swedish Horse ($n = 19$), Norwegian Fjord Horse ($n = 21$) and Exmoor Pony ($n = 24$) (Table S1). A larger sample of Exmoor Pony data ($n = 274$, Table 1) found only one heterozygote for this haplotype.

Frequency of the candidate SNPs in other breeds. Analysis of SNP data from other domestic breeds showed that heterozygotes for the SNPs of interest were at a particularly high frequency in the Quarter Horse population (71%, $n = 137$) (Table 1, Table S2). The proportion of heterozygotes was also high in Swedish Warmbloods ($n = 380$) and Norwegian-Swedish Coldblooded Trotters ($n = 641$), being 20% and 40% respectively (Table 1). Smaller datasets also revealed that Belgian Draft ($n = 19$), French Trotter ($n = 17$), Paint ($n = 15$), Morgan ($n = 19$), Mongolian Paulista ($n = 19$) and Tuva ($n = 15$) breeds may also have a high proportion of heterozygotes for this haplotype in their populations (Table S1).

Identifying candidate causal variants using whole genome sequence data. To further investigate SNP frequencies in this region, variants were called from whole-genome sequence data of 90 domestic horses. The two SNPs identified in the preliminary analysis showed a complete absence of homozygotes for their minor alleles in these individuals (Table S3). Additionally, a number of variants closely linked to these SNPs were identified (Table S3, Figure S1). Annotation of these loci using SIFT[42] identified three variants that may result in changes to protein structure or expression, so these represent the most likely candidates to cause lethality in homozygous state (Fig. 1).

The first of these variants, 6:38282610G > A (rs68663123), was located in an exonic region of the *LY49B* gene and resulted in an amino acid change from a phenylalanine to a serine residue. This substitution is located next to a tryptophan residue that appears to be highly conserved across members of the *LY49* family and across species.

Tissue	Gene count (FPKM)		
Trophectoderm	Day 15	Day 22	Day 25
	0.000	0.031	0.024
Inner cell mass	Day 16	Day 23	Day 24
	0.00	0.00	0.00

Table 3. Gene counts from RNA sequence data of three trophectoderm and three inner cell mass tissue samples from equine embryos. Transcript counts are in fragments per kilobase/million (FPKM).

However, there is little conservation of the phenylalanine residue across taxa; some species have a phenylalanine and others a serine at this position. This SNP is annotated as being "tolerated" in SIFT.

Two other variants that were closely linked to the candidate SNPs (6:38276742A > T and 6:38276955G > A) were found within the 3′UTR (3′untranslated region) of the *LY49B* gene. Alignment of the 3′UTR of the five functional *LY49* genes in *Equus caballus* revealed that the region containing the SNP 6:38276955G > A (rs1139567427) is highly conserved in all members of the *LY49* gene family (Table 2). This region may be important for mRNA stability and translation into a functional protein. The other variant 6:38276742A > T (rs1137325172) was found in an AU-rich region at the end of the *LY49B* mRNA transcript, which is often associated with polyadenylation and post translation stability.

Transcriptomic analysis of RNA sequence data. Measurable levels of *LY49B* mRNA were not detected in equine trophectoderm tissue collected on day 16 of development. However, *LY49B* mRNA was observed in trophectoderm tissue collected on days 23 and 24 of development (Table 3). Additionally, *LY49B* mRNA transcripts were detected in microarray data from equine chorion and chorionic girdle tissue between days 27 and 34 of development (Table S4). Inner cell mass tissue collected on days 15, 22 and 25 of development did not show any measurable transcription of *LY49B* (Table 3). The genotypes of the candidate SNPs in the mRNA samples analysed were unknown.

Discussion

Analysis of genotype data identified a two-SNP haplotype as a strong candidate for harbouring a variant that causes lethality in homozygous state. The SNPs identified in this preliminary analysis mapped to an intronic region in the *LY49B* gene on ECA6 (Table 1). The *LY49B* gene belongs to the *LY49* (Killer cell lectin-like receptor subfamily A) family of receptors, which consists of five functional members in *Equus caballus*[43]. Other species (including humans) have a functionally similar, but structurally different gene family called *KIR* (Killer cell immunoglobin line receptors)[44]. The *LY49*/*KIR* gene family are expressed across various types of immune cells, and mediate their function through bindings to MHC-1[45]. The *LY49B* gene is expressed in myeloid cells where it regulates their activity through an inhibitory effect, possibly to prevent their spontaneous activation[46]. Despite the important role that they play in immunity, the function of *LY49* genes in development is currently unknown. In humans, incompatibilities between foetal *KIR* and maternal *MHC* (*HLA*) genotypes are associated with an increased risk of miscarriage and preeclampsia[47–49]. Additionally, knockdown of *LY49* in mice showed a high rate of implantation failure[50,51]. These findings indicate that *LY49B* may play an important role in maternal/foetal compatibility and implantation success in horses.

Analysis of transcriptomic data found that *LY49B* was first expressed in equine trophoblast tissue during the placental development stage. The first evidence of *LY49B* expression was found on day 23–24 of development (Table 3), during which the glycoprotein capsule surrounding the embryo is broken down and placental tissue starts to develop[52]. Measurable expression of *LY49B* was also found in chorion and chorionic girdle tissues between days 27 and 34 of development (Table S4). During this time, trophoblast cells rapidly proliferate to form the chorionic girdle, which then invades the endometrium to form epithelial cups[53]. It is possible that LY49B is important for successful implantation of the embryo by mediating the action of MHC-1 which is expressed during this time[54,55]. Further investigations into the role of *LY49B* in equine development would confirm whether impaired function causes lethality and the stage of development at which this occurs.

Variant calling in whole-genome sequence data from 90 domestic horses further confirmed an absence of minor homozygotes for the two SNPs of interest. Three variants closely linked to these SNPs were also identified in these data as the most likely candidates to cause loss of function in the *LY49B* gene and result in lethality in homozygous state (Table S3). One SNP was a missense variant in the coding region of the *LY49B* gene that results in the substitution of a negatively charged serine for an aromatic phenylalanine residue. However, lack of conservation of this SNP in *LY49* genes across taxa makes it seem unlikely to be a causative variant for embryonic lethality. Two other variants found in the 3′UTR of the *LY49B* gene were also closely linked to the SNPs identified in the preliminary analysis, and seemed more likely candidates to cause embryonic lethality in homozygous state.

The 3′UTR of a gene is responsible for transcriptional stability through the binding of miRNAs and RNA binding proteins[56]. The addition of the polyadenylation tail to the 3′UTR is also essential to ensure proper processing and translation of the mRNA strand[57]. Mutations in the 3′UTR can lead to degradation of the mRNA, resulting in reduced or inhibited translation even when the gene is transcribed[58]. Variation in the 3′UTR of genes are associated with a number of diseases including Huntington's and breast cancer in humans[59,60]. Additionally, SNPs in the 3′UTR are associated with traits in livestock including milk production in cows, muscularity in sheep and obesity in horses[58,61,62].

Despite the importance of the 3′UTR for the mRNA stability and normal expression of a gene, little is known about how specific polymorphisms can affect post-transcriptional processing. This makes it difficult to identify how the 3′UTR variants identified in this study could affect the translation of *LY49B* mRNA into a functional protein. The 3′UTR variant 6:38276955G > A was identified as a possible candidate for embryonic lethality because it is highly conserved between all members of the *LY49B* family (Table 2) in horses, so may play an important role in mRNA stability. The other 3′UTR variant (6:38276742A > T) is found in an AU-rich region at the end of the transcript, so may be important for the addition of the polyadenylation tail. Further examination of the effects that these variants have on post-transcriptional processing would determine if they impact the normal expression of *LY49B* in horses.

Despite an absence of homozygotes, the two intronic SNPs identified in this study were found at high heterozygote frequencies in the Thoroughbred population. Mares are often covered multiple times in a season, which may explain why a more discernible reduction in fertility has not been observed as a result of the high frequency of this variant. However, the presence of lethal variants at high frequencies may result in more coverings being required for each mare in a season. Currently, there is no evidence that variation in the *LY49B* gene is associated with phenotypic advantages in horses. However, it is possible that one of the variants linked to these SNPs results in a phenotypic advantage in heterozygotes, which could explain why they have reached such high frequencies in the breed. It is also possible that selective breeding practices favouring a limited number of stallion bloodlines are responsible for this potentially lethal haplotype drifting to high frequencies in the Thoroughbred population. This would be most likely to occur if a stallion that made a large genetic contribution to the population was a carrier. A similar instance has recently been documented in cattle, where a lethal variant at a high frequency was traced back to a sire with an extensive genetic influence on the population[19].

The presence of this potentially lethal haplotype across many diverse breeds of domestic horses indicates that it may not be the result of a recent mutation present only in the Thoroughbred population. Rather, heterozygotes for this haplotype may have been present in pre-domesticated horses as a rare variant, and have become more frequent in some domestic breeds as the result of population bottlenecks due to breed formations and selective breeding practices. Domestication and breed formation events have been well documented to result in increased deleterious mutation loads in horses and other domestic species[22,31,63,64]. A high proportion of heterozygotes for this haplotype were found in some breeds closely related to the Thoroughbred including the Paint, French Trotter, Morgan and Quarter Horse. Notably, over 70% of Quarter Horse samples included in this study were heterozygous for these SNPs (Table 1). The Quarter Horse has an open stud book, and higher genetic diversity than the Thoroughbred population[65], making the high frequency of a potentially lethal haplotype at first surprising. The Quarter Horse dataset reportedly did not contain full or half siblings[65], but the collection of samples from one geographical area may not fully reflect the diversity of the worldwide population. An average relatedness analysis of these samples noted the large genetic influence of one particular Thoroughbred stallion[65], which may explain the high frequency of heterozygotes observed in this population. However, the extremely high frequency of heterozygotes in this breed may be due to selective breeding favouring these individuals.

The Belgian Draft, Mangalarga Paulista and Tuva breeds also show a high proportion of heterozygotes, but are more distantly related to the Thoroughbred and to each other. Therefore, the high frequency of heterozygotes in these breeds may be due to independent genetic drift events. Heterozygotes for this haplotype were notably absent from one branch of the tree containing small heavy horses from Northern Europe, which are more distantly related to the Thoroughbred. A larger dataset of Exmoor Pony samples from this phylogenetic branch revealed one heterozygote for this haplotype (Table 1). This could be due to a calling error, but it is also possible that these SNPs exist at very low frequencies in these breeds. The small sample size of the genotype data for many individual breeds in this study means that heterozygote frequencies across all subpopulations found throughout the world may deviate from that reported here. However, these data provide an indication of breeds with high proportions of heterozygotes for this region. Analysis of SNP data from Northern-Swedish Coldblooded Trotters identified 22 homozygotes for the SNP at position 6:38278874. It is likely that there has been recombination between this SNP and the causal variant, and may appear more frequent in this population due to differences in breed history and recombination patterns[66]. However, additional analyses are required to explore this further. Overall, our findings suggest that this region shows evidence harbouring a homozygous lethal variant, yet a high proportion of heterozygotes are found across many domestic horse breeds.

In this study, we identified a haplotype at high heterozygote frequencies in the Thoroughbred horse population that is a strong candidate for harbouring a variant causing lethality in homozygous state. Similar analyses on larger datasets in other livestock populations have identified multiple lethal haplotypes, so it is likely that other such variants are present at high frequencies in the Thoroughbred population but were not captured in this study. Additionally, the use of commercial SNP arrays only allows for the identification of variants with high minor allele frequencies in populations. Analysis of larger sample sizes, and using higher density genotype data could allow for identification of other variants associated with lethality in domestic horses. The identification of this potentially lethal haplotype demonstrates the potential implications of heavily favouring a limited number of bloodlines in selective breeding practices. Further characterisation of lethal haplotypes in other breeds would also assist in breeding management to increase per covering fertility rates in domestic horse populations.

Methods

DNA extractions. DNA was extracted from the hair samples of Australian Thoroughbred horses using the Qiagen Gentra Puregene Tissue Kit (Qiagen, Redwood City, CA, USA). DNA was extracted from the hair samples of Norwegian-Swedish Coldblooded Trotters and Swedish Warmbloods by incubating the samples for 2 h at 56 °C with Chelex 100 Resin (Bio-Rad Laboratories, Hercules, CA) and Proteinase K (20 mg/mL; Merck KgaA, Darmstadt, Germany). The Proteinase K was then inactivated by incubating for 10 min at 95 °C and DNA resus-

pended in low TE (1 mM Tris, 0.1 mM EDTA). DNA was extracted from blood samples using the Qiasymphony DSP DNA mini kit (Qiagen, Hilden, Germany).

Initial genotyping. Genotype data from a representative sample of Thoroughbreds were used to identify SNPs with a high proportion of heterozygotes, but an absence of homozygotes for one allele. Genome-wide SNP data were generated for 156 Australian Thoroughbred horses by genotyping samples on either the Illumina 70 K Chip (65,102 SNPs) ($n=102$) or the Affymetrix 670 K Chip (670,796 SNPs) ($n=54$). Common genotyped SNPs between the two arrays were scanned for deviations from the Hardy–Weinberg equilibrium with an absence of homozygotes for one allele using PLINK (version 1.9)[67]. The p values were adjusted using a false discovery rate correction with the R package "qvalue"[68]. Since SNPs with an absence of homozygotes could indicate a calling error, the search was narrowed to only include adjacent SNPs that fit such criteria.

The frequencies of the candidate SNPs were then examined in publicly available genotype data from Japanese Thoroughbreds ($n=370$) typed on the Affymetrix 670 K Chip[69] and these were added the Thoroughbred sample. The SNP frequencies were then characterised from genotype data from Swedish Warmbloods ($n=380$)[70] and Norwegian-Swedish Coldblooded Trotters ($n=646$)[71] typed on the Affymetrix 670 K Chip. Publicly available data from Exmoor Ponies ($n=285$, typed on the Affymetrix 670 K Chip)[71], Quarter Horses ($n=137$, typed on the Illumina 70 K Chip)[72] and horses of 32 different domestic breeds ($n=582$, typed on the Illumina 50 K Chip)[29,30] were also included in this preliminary scan for SNP frequencies. In these data, raw intensities were plotted to check for calling errors. If potential calling errors were detected, SNPs were recalled using a mixture model fitted with an expectation–maximization algorithm in R.

Variant discovery and mapping. Publicly available whole-genome sequence data were used to further examine the frequencies of the candidate SNPs identified in the initial genotype analysis, and to identify linked variants. Paired end whole-genome sequence data from 90 horses of different domestic breeds were used in this analysis (Table S5). The whole genome datasets were downloaded from the European Nucleotide Archives (ENA, https://www.ebi.ac.uk/ena) which included horses of different domestic breeds (PRJEB14779, $n=70$) and additional Thoroughbred samples (PRJNA168142, $n=16$ and PRJNA184688, $n=4$) (Table S5).

The SNP array used in the initial genotyping analysis was developed based on coordinates of the EquCab2.0 reference genome. For consistency, we used the EquCab2.0 assembly as a reference for the whole-genome sequence analysis. The EquCab2.0 assembly was also used because of an issue with resolution in the area of interest on the newer EquCab3.0 assembly. The raw reads were mapped to the EquCab2.0 reference genome using BWA-MEM algorithm from Burrows-Wheeler Alignment Tool (version 0.7.17)[73]. Duplicate reads were flagged using Samblaster (version 0.1.22)[74], and base recalibration was performed using Genome Analysis Toolkit (GATK) (version 4.0.8.1)[75]. Variants (SNPs and INDELs (insertions and deletions)) were called using Haplotype Caller and then filtered using the standard hard filtering recommendations in GATK[76]. The individual SNPs were then filtered to only include high quality allele calls with an average filtered depth over 10 and a Phred score over 20.

Variants that were linked to the SNPs identified from the genotype data were produced using the LD function in PLINK (version 1.9) with a window size of 5 Mb[67]. Only SNPs with an r^2 value of over 0.8 and a D' value >0.9 were shortlisted. The effects of each SNP on gene structure and function was characterised using SIFT (version 4G)[42]. The conservation of variants across taxa was analysed using the NCBI Conserved Domain Database Search[77].

Transcriptomic analysis. Publicly available RNA sequence data were used to examine expression levels of the genes of interest in embryonic tissue. The data included equine inner cell mass tissue (collected at day 15, 22 and 25, $n=3$) and trophectoderm tissue (collected at day 16, 23 and 24, $n=3$) from the Functional Annotation of ANimal Genomes (FAANG) equine biobank (available from ENA under the project name PRJNA223157)[78]. Adaptors were trimmed using bbduk from BBtools (version 37.98)[79]. Reads were aligned to the EquCab 2.0 genome using STAR (version 2.7.2b)[80]. Counts were generated using featurecounts from Subread package(version 1.5.1)[81], then quantified in fragments per kilobase/million (FPKM) using the R package "edgeR"[82] with the Equus_caballus_Ensembl_94 file used for annotation. Microarray data for chorion ($n=19$) and chorionic girdle ($n=19$) tissue collected from horse embryos between days 27–34 of development[83] were also examined for gene expression levels.

Ethics statement. Hair samples from Australian Thoroughbred horses were collected under approval from University of Sydney Ethics Committee (Number: N00-2009-3-5109.) Written informed consent to use the animals in this study was obtained from the owners of the animals. The hair samples from Swedish Warmblood horses were originally collected for parentage testing and stored in the biobank at the Animal Genetics Laboratory, SLU so ethics approval was not applicable. Hair and blood samples of Norwegian-Swedish Coldblooded Trotters were collected under approval from the Ethics Committee for Animal Experiments in Uppsala, Sweden (Number: C 121/14). All the methods were performed in accordance with the guidelines set out by the respective Animal Ethics Committees and the guidelines contained in the Guide for the Care and Use of Laboratory Animals. No experimental procedure was performed on live animals. All other data was downloaded from publicly available repositories.

Data availability

The whole-genome sequence data used in this study is publicly available for The European Nucleotide Archive. Genotype data for Japanese Thoroughbreds, Exmoor Ponies, Quarter Horses and other domestic horse breeds can be found in the supplementary information of their respective papers (https://doi.org/10.1371/journ

al.pone.0218407, 10.1371/journal.pone.0152966, 10.1093/jhered/est079 and 10.1371/journal.pone.0054997). The exceptions are genotype data from Australian Thoroughbred, Swedish Warmbloods and North Swedish Coldblooded Trotters, which are available on request but restrictions apply to the availability of these data which were used under license for the current study and so are not publicly available.

Received: 13 May 2020; Accepted: 26 June 2020
Published online: 04 August 2020

References

1. Spencer, T. E. Early pregnancy: concepts, challenges, and potential solutions. *Anim. Front.* **3**, 48–55. https://doi.org/10.2527/af.2013-0033 (2013).
2. White, J. K. *et al.* Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell* **154**, 452–464. https://doi.org/10.1016/j.cell.2013.06.022 (2013).
3. Mary, E. D. *et al.* High-throughput discovery of novel developmental phenotypes. *Nature* https://doi.org/10.1038/nature19356 (2016).
4. Gao, Z., Waggoner, D., Stephens, M., Ober, C. & Przeworski, M. An estimate of the average number of recessive lethal mutations carried by humans. *Genetics* **199**, 1243–1254. https://doi.org/10.1534/genetics.114.173351 (2015).
5. Ballinger, M. A. & Noor, M. A. F. Are lethal alleles too abundant in humans?. *Trends Genet.* **34**, 87–89. https://doi.org/10.1016/j.tig.2017.12.013 (2018).
6. Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**, 335 (1972).
7. Sulem, P. *et al.* Identification of a large set of rare complete human knockouts. *Nat. Genet.* **47**, 448–452. https://doi.org/10.1038/ng.3243 (2015).
8. Curik, I., Ferencakovic, M. & Solkner, J. Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest. Sci.* **166**, 26–34. https://doi.org/10.1016/j.livsci.2014.05.034 (2014).
9. Keller, L. F. & Waller, D. M. Inbreeding effects in wild populations. *Trends Ecol. Evol.* **17**, 230–241. https://doi.org/10.1016/S0169-5347(02)02489-8 (2002).
10. Charlesworth, D. & Willis, J. H. The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**, 783–796. https://doi.org/10.1038/nrg2664 (2009).
11. Hoff, J. L., Decker, J. E., Schnabel, R. D. & Taylor, J. F. Candidate lethal haplotypes and causal mutations in Angus cattle. *BMC Genom.* https://doi.org/10.1186/s12864-017-4196-2 (2017).
12. VanRaden, P. M., Olson, K. M., Null, D. J. & Hutchison, J. L. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *J. Dairy Sci.* **94**, 6153–6161. https://doi.org/10.3168/jds.2011-4624 (2011).
13. Fasquelle, C. *et al.* Balancing selection of a frame-shift mutation in the MRC2 gene accounts for the outbreak of the crooked tail syndrome in Belgian Blue cattle. *PLoS Genet.* **5**, e1000666. https://doi.org/10.1371/journal.pgen.1000666 (2009).
14. Zhang, C., MacNeil, M. D., Kemp, R. A., Dyck, M. K. & Plastow, G. S. Putative loci causing early embryonic mortality in Duroc swine. *Front. Genet.* **17**, 655. https://doi.org/10.3389/fgene.2018.00655 (2018).
15. Charlier, C. *et al.* NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. *Genome Res.* **26**, 1333–1341. https://doi.org/10.1101/gr.207076.116 (2016).
16. Sonstegard, T. S. *et al.* Identification of a nonsense mutation in CWC15 associated with decreased reproductive efficiency in Jersey cattle. *PLoS ONE* **8**, e54872. https://doi.org/10.1371/journal.pone.0054872 (2013).
17. Derks, M. F. L. *et al.* A systematic survey to identify lethal recessive variation in highly managed pig populations. *BMC Genom.* https://doi.org/10.1186/s12864-017-4278-1 (2017).
18. Derks, M. F. L. *et al.* Loss of function mutations in essential genes cause embryonic lethality in pigs. *PLoS Genet.* **15**, e1008055. https://doi.org/10.1371/journal.pgen.1008055 (2019).
19. Bourneuf, E. *et al.* Rapid discovery of de novo deleterious mutations in cattle enhances the value of livestock as model species. *Sci. Rep.* **7**, 11466. https://doi.org/10.1038/s41598-017-11523-3 (2017).
20. Schubert, M. *et al.* Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc. Natl. Acad. Sci. USA* **111**, E5661–E5669. https://doi.org/10.1073/pnas.1416991111 (2014).
21. Cruz, F. *et al.* The legacy of domestication: accumulation of deleterious mutations in the dog genome. *Mol. Biol. Evol.* **25**, 2331–2336. https://doi.org/10.1093/molbev/msn177 (2008).
22. Marsden, C. D. *et al.* Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc. Natl. Acad. Sci. USA* **113**, 152–157. https://doi.org/10.1073/pnas.1512501113 (2016).
23. Bosse, M., Derks, M. & Groenen, M. Deleterious alleles in the context of domestication, inbreeding, and selection. *Evol. Appl.* **12**, 6–17. https://doi.org/10.1111/eva.12691 (2019).
24. Wright, S. Evolution in Mendelian populations. *Genetics* **16**, 97–159 (1931).
25. Husemann, M., Zachos, F. E., Paxton, R. J. & Habel, J. C. Effective population size in ecology and evolution. *Heredity* https://doi.org/10.1038/hdy.2016.75 (2016).
26. García-Dorado, A. Understanding and predicting the fitness decline of shrunk populations: inbreeding, purging, mutation, and standard selection. *Genetics* **190**, 1461–1476. https://doi.org/10.1534/genetics.111.135541 (2012).
27. Casas, E. & Kehrli, M. E. A review of selected genes with known effects on performance and health of cattle. *Front. Vet. Sci.* https://doi.org/10.3389/fvets.2016.00113 (2016).
28. Jagannathan, V. *et al.* Comprehensive characterization of horse genome variation by whole-genome sequencing of 88 horses. *Anim. Genet.* **50**, 74–77. https://doi.org/10.1111/age.12753 (2019).
29. Petersen, J. L. *et al.* Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genet.* **9**, e1003211. https://doi.org/10.1371/journal.pgen.1003211 (2013).
30. Petersen, J. L. *et al.* Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLOS ONE* **8**, e54997. https://doi.org/10.1371/journal.pone.0054997 (2013).
31. Orlando, L. & Librado, P. Origin and evolution of deleterious mutations in horses. *Genes* **10**, 649. https://doi.org/10.3390/genes10090649 (2019).
32. Aurich, C. Reproductive cycles of horses. *Anim. Reprod. Sci.* **124**, 220–228. https://doi.org/10.1016/j.anireprosci.2011.02.005 (2011).
33. Allen, W. R. & Wilsher, S. Half a century of equine reproduction research and application: a veterinary tour de force. *Equine Vet. J.* **50**, 10–21. https://doi.org/10.1111/evj.12762 (2018).
34. Weatherby and Sons. *An Introduction to the General Stud Book*. (Weatherby and Sons, 1791).
35. Hill, E. W., Gu, J., McGivney, B. A. & MacHugh, D. E. Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. *Anim. Genet.* **41**, 56–63. https://doi.org/10.1111/j.1365-2052.2010.02104.x (2010).

36. Gu, J. *et al.* A genome scan for positive selection in Thoroughbred horses. *PLoS ONE* **4**, e5767. https://doi.org/10.1371/journal.pone.0005767 (2009).
37. Cunningham, E. P., Dooley, J. J., Splan, R. K. & Bradley, D. G. Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to Throughbred horses. *Anim. Genet.* **32**, 360. https://doi.org/10.1046/j.1365-2052.2001.00785.x (2001).
38. Todd, E. T. *et al.* Founder-specific inbreeding depression affects racing performance in Thoroughbred horses. *Sci. Rep.* **8**, 6167. https://doi.org/10.1038/s41598-018-24663-x (2018).
39. Corbin, L. J. *et al.* Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Anim. Genet.* **41**, 8–15. https://doi.org/10.1111/j.1365-2052.2010.02092.x (2010).
40. McGivney, B. A. *et al.* Genomic inbreeding trends, influential sire lines and selection in the global Thoroughbred horse population. *Sci. Rep.* **10**, 466. https://doi.org/10.1038/s41598-019-57389-5 (2020).
41. Woolliams, J. A., Berg, P., Dagnachew, B. S. & Meuwissen, T. H. E. Genetic contributions and their optimization. *J. Anim. Breed. Genet.* **132**, 89–99. https://doi.org/10.1111/jbg.12148 (2015).
42. Ng, P. C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814. https://doi.org/10.1093/nar/gkg509 (2003).
43. Futas, J. & Horin, P. Natural killer cell receptor genes in the family Equidae: not only Ly49. *PLoS ONE* **8**, e64736. https://doi.org/10.1371/journal.pone.0064736 (2013).
44. Kelley, J., Walter, L. & Trowsdale, J. Comparative genomics of natural killer cell receptor gene clusters. *PLoS Genet.* **1**, 129–139. https://doi.org/10.1371/journal.pgen.0010027 (2005).
45. Rahim, M. M. A. *et al.* Ly49 receptors: innate and adaptive immune paradigms. *Front. Immunol.* **5**, 145. https://doi.org/10.3389/fimmu.2014.00145 (2014).
46. Gays, F. *et al.* Ly49B is expressed on multiple subpopulations of myeloid cells. *J. Immunol. Res.* **177**, 5840–5851. https://doi.org/10.4049/jimmunol.177.9.5840 (2006).
47. Hiby, S. *et al.* Association of maternal killer-cell immunoglobulin-like receptors and parental HLA-C genotypes with recurrent miscarriage. *Hum. Reprod.* **23**, 972–976. https://doi.org/10.1093/humrep/den011 (2008).
48. Hiby, S. E. *et al.* Materal activating KIRs protect against human reproductive failure mediated by fetal HLA-C2. *J. Clin. Investig.* **120**, 4102 (2010).
49. Long, W. *et al.* Association of maternal KIR and fetal HLA-C genes with the risk of preeclampsia in the Chinese Han population. *Placenta* **36**, 433–437. https://doi.org/10.1016/j.placenta.2014.05.008 (2015).
50. Leon, L. *et al.* Ly49 knockdown in mice results in aberrant uterine crypt formation and impaired blastocyst implantation. *Placenta* **39**, 147–150. https://doi.org/10.1016/j.placenta.2016.01.013 (2016).
51. Lima, P. D. *et al.* Ly49 receptors activate angiogenic mouse DBA+ uterine natural killer cells. *Cell. Mol. Immunol.* **11**, 467–476. https://doi.org/10.1038/cmi.2014.44 (2014).
52. Allen, W. R. & Wilsher, S. A review of implantation and early placentation in the mare. *Placenta* **30**, 1005–1015. https://doi.org/10.1016/j.placenta.2009.09.007 (2009).
53. Noronha, L. E. & Antczak, D. F. Maternal immune responses to trophoblast: the contribution of the horse to pregnancy immunology. *Am. J. Reprod. Immunol.* **64**, 231–244. https://doi.org/10.1111/j.1600-0897.2010.00895.x (2010).
54. Donaldson, W. L., Oriol, J. G., Pelkaus, C. L. & Antczak, D. F. Paternal and maternal major histocompatibility complex class I antigens are expressed co-dominantly by equine trophoblast. *Placenta* **15**, 123–135. https://doi.org/10.1016/S0143-4004(05)80449-7 (1994).
55. Bacon, S. J., Ellis, S. A. & Antczak, D. F. Control of expression of major histocompatibility complex genes in horse trophoblast. *Biol. Reprod.* **66**, 1612–1620. https://doi.org/10.1095/biolreprod66.6.1612 (2002).
56. Steri, M., Idda, M. L., Whalen, M. B. & Orrù, V. Genetic variants in mRNA untranslated regions. *Wiley Interdiscip. Rev. RNA* https://doi.org/10.1002/wrna.1474 (2018).
57. Di Giammartino, D. C., Nishida, K. & Manley, J. L. Mechanisms and consequences of alternative polyadenylation. *Mol. Cell* **43**, 853–866. https://doi.org/10.1016/j.molcel.2011.08.017 (2011).
58. Lewis, S. L. *et al.* Genomewide association study reveals a risk locus for equine metabolic syndrome in the Arabian horse 1. *J. Anim. Sci.* **95**, 1071–1079. https://doi.org/10.2527/jas.2016.1221 (2017).
59. Dorairaj, J. *et al.* A germline mutation in the BRCA1 3'UTR predicts Stage IV breast cancer. *BMC Cancer* **14**, 421. https://doi.org/10.1186/1471-2407-14-421 (2014).
60. Kim, K. H. *et al.* Full sequence of mutant huntingtin 3′-untranslated region and modulation of its gene regulatory activity by endogenous microRNA. *J. Hum. Genet.* **64**, 995–1004. https://doi.org/10.1038/s10038-019-0639-8 (2019).
61. Hou, J. *et al.* Two mutations in the caprine MTHFR 3'UTR regulated by microRNAs are associated with milk production traits. *PLoS ONE* **10**, e0133015–e0133015. https://doi.org/10.1371/journal.pone.0133015 (2015).
62. Alex, C. *et al.* A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat. Genet.* **38**, 813. https://doi.org/10.1038/ng1810 (2006).
63. Librado, P. *et al.* Ancient genomic changes associated with domestication of the horse. *Science* **356**, 442–445. https://doi.org/10.1126/science.aam5298 (2017).
64. Fages, A. *et al.* Tracking five millennia of horse management with extensive ancient genome time series. *Cell* **177**, 1419-1435.e1431. https://doi.org/10.1016/j.cell.2019.03.049 (2019).
65. Petersen, J. L., Mickelson, J. R., Cleary, K. D. & McCue, M. E. The American Quarter Horse: population structure and relationship to the Thoroughbred. *J. Hered.* **105**, 148–162. https://doi.org/10.1093/jhered/est079 (2014).
66. Beeson, S. K., Mickelson, J. R. & McCue, M. E. Exploration of fine-scale recombination rate variation in the domestic horse. *Genome Res.* **29**, 1744. https://doi.org/10.1101/gr.243311.118 (2019).
67. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575. https://doi.org/10.1086/519795 (2007).
68. Storey, J., Bass, A., Dabney, A. & Robinson, D. Qvalue: q-value estimation for false discovery rate control v. R package version 2.18.0. (Accessed 6 September 2018); https://github.com/jdstorey/qvalue.
69. Fawcett, J. A. *et al.* Genome-wide SNP analysis of Japanese Thoroughbred racehorses. *PLoS ONE* https://doi.org/10.1371/journal.pone.0218407 (2019).
70. Ablondi, M., Viklund, A., Lindgren, G., Eriksson, S. & Mikko, S. Signatures of selection in the genome of Swedish Warmblood horses selected for sport performance. *BMC Genom.* https://doi.org/10.1186/s12864-019-6079-1 (2019).
71. Velie, B. D. *et al.* Using an inbred horse breed in a high density genome-wide scan for genetic risk factors of insect bite hypersensitivity (IBH). *PLoS ONE* https://doi.org/10.1371/journal.pone.0152966 (2016).
72. Petersen, J. L., Valberg, S. J., Mickelson, J. R. & McCue, M. E. Haplotype diversity in the equine myostatin gene with focus on variants associated with race distance propensity and muscle fibre type proportions. *Anim. Genet.* **45**, 827–835. https://doi.org/10.1111/age.12205 (2014).
73. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324 (2009).
74. Faust, G. & Hall, I. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505. https://doi.org/10.1093/bioinformatics/btu314 (2014).

75. McKenna, A. *et al.* The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data (Report). *Genome Res.* **20**, 1297–1303. https://doi.org/10.1101/gr.107524.110 (2010).
76. Depristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491. https://doi.org/10.1038/ng.806 (2011).
77. Lu, S. *et al.* CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkz991 (2019).
78. Iqbal, K., Chitwood, J. L., Meyers-Brown, G. A., Roser, J. F. & Ross, P. J. RNA-Seq transcriptome profiling of equine inner cell mass and trophectoderm. *Biol. Reprod.* https://doi.org/10.1095/biolreprod.113.113928 (2014).
79. Bushnell, B. BBMap short read aligner. (Accessed 29 August 2019); https://sourceforge.net/projects/bbmap.
80. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21. https://doi.org/10.1093/bioinformatics/bts635 (2013).
81. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930. https://doi.org/10.1093/bioinformatics/btt656 (2014).
82. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140. https://doi.org/10.1093/bioinformatics/btp616 (2010).
83. Read, J. E. *et al.* Dynamic changes in gene expression and signalling during trophoblast development in the horse. *Reproduction* https://doi.org/10.1530/REP-18-0270 (2018).
84. Velie, B. D. *et al.* A genome-wide association study for harness racing success in the Norwegian-Swedish coldblooded trotter reveals genes for learning and energy metabolism. *BMC Genet.* **19**, 1–13. https://doi.org/10.1186/s12863-018-0670-3 (2018).

Acknowledgements
The authors acknowledge the technical assistance provided by the Sydney Informatics Hub, a Core Research Facility of the University of Sydney.

Author contributions
N.A.H., B.D.V., R.A.A., G.L., A.V., S.E., S.M. and E.S. assisted in the collection and genotyping of the DNA samples. E.T.T., B.D.V., P.C.T. and N.A.H. designed the project. E.T.T. analysed the data and wrote the manuscript. All authors edited, read and approved the final manuscript.

Competing interests
N.A.H. is supported by Racing Australia in the form of salary. All other authors declare that they have no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-68946-8.

Correspondence and requests for materials should be addressed to E.T.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.