ANNUAL REVIEWS

Annual Review of Animal Biosciences

Genetics of Thoroughbred Racehorse Performance

Ernest Bailey,[1] Jessica L. Petersen,[2] and Theodore S. Kalbfleisch[1]

[1]MH Gluck Equine Research Center, University of Kentucky, Lexington, Kentucky, USA; email: ebailey@uky.edu, ted.kalbfleisch@uky.edu

[2]Department of Animal Science, University of Nebraska, Lincoln, Nebraska, USA; email: jessica.petersen@unl.edu

Annu. Rev. Anim. Biosci. 2022. 10:131–50

First published as a Review in Advance on November 15, 2021

The *Annual Review of Animal Biosciences* is online at animal.annualreviews.org

https://doi.org/10.1146/annurev-animal-020420-035235

Copyright © 2022 by Annual Reviews. All rights reserved

ANNUAL REVIEWS CONNECT

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Keywords

heritability, GWAS, genome-wide association study, genomic, equine, racing, athlete, artificial selection

Abstract

Thoroughbred horses have been selected for racing performance for more than 400 years. Despite continued selection, race times have not improved significantly during the past 60 years, raising the question of whether genetic variation for racing performance still exists. Studies using phenotypes such as race time, money earned, and handicapping, however, demonstrate that there is extensive variation within these traits and that they are heritable. Even so, these are poor measures of racing success since Thoroughbreds race at different ages and distances and on different types of tracks, and some may not race at all. With the advent of genomic tools, DNA variants are being identified that contribute to racing success. Aside from strong associations for myostatin variants with best racing distance, weak to modest associations with racing phenotypes are reported for other genomic regions. These data suggest that diverse genetic strategies have contributed to producing a successful racehorse, and genetic variation contributing to athleticism remains important.

Downloaded from www.AnnualReviews.org
Guest (guest)
IP: 18.220.137.164
On: Fri, 03 May 2024 20:09:24

INTRODUCTION

Thoroughbred horse racing traces its origins to the 1660 restoration of the monarchy in England. Charles II had a particular interest in horse racing and established a horse breeding and racing center at Newmarket, near London. Horse racing had been popular in England during preceding centuries; however, Charles II and his fellow enthusiasts began the methodical breeding of superior racehorses and raised the profile of the sport (1, 2). Thus, enthusiasm for Thoroughbred racing grew, and subsequently, the sport and its bloodstock were exported worldwide, to France, the Americas, South Africa, Australia, Japan, and elsewhere (1, 2). Although each region developed and celebrated its own champions, the international appeal of Thoroughbred racing resulted in racing and breeding stock becoming global travelers. Races of two-, three-, and four-mile heats were still run at the end of the eighteenth century, but prestige went to shorter races, less than two miles, and was generally restricted to three-year-old horses. As a result, since the late 1700s, selection for racing performance in Thoroughbreds has focused on performance at the ages of two and three years for these shorter distances.

A Thoroughbred's value depends upon its ability to win races. Success for the Thoroughbred breeder is not measured by the number of horses produced or growth rates of foals. Success lies, instead, in buyers' confidence that a horse will have a productive racing career. Gambling on the outcome of horse races is another important aspect of the sport. Indeed, gambling propelled Thoroughbred racing from being the "Sport of Kings" to a sport widely popular with people from all walks of life. For $2, anyone can proverbially "own a piece" of a million-dollar racehorse for two minutes (**Figure 1**).

The number of horses founding the Thoroughbred breed has been a subject of controversy; however, it is generally acknowledged that a relatively small number of horses were responsible. Many founders of the breed are known from historical records and include Oriental (Arabian, Middle Eastern, and North African) horses, as well as horses native to the British Isles (1, 2). Cunningham et al. (3) studied a random sample of 200 British Thoroughbreds, born between 1770 and 1990, and traced their pedigrees to estimate the foundation stock of the British Thoroughbred.

Figure 1

Thoroughbred racing at the Keeneland Race Track in Lexington, Kentucky. Photograph by Allen Page.

Downloaded from www.AnnualReviews.org
Guest (guest)
IP: 18.220.137.164
On: Fri, 03 May 2024 20:09:24

Using this method, they identified 158 foundation horses, born between 1665 and 1860. The contribution of those 158 founders accounted for 81% of the genetics of the sample. A large number of these founders were identified as being of Oriental origin, variously identified as Arabs, Turks, and Barbs. Wallner and coworkers (4, 5) studied DNA variation on the Y chromosome and deduced that the haplotype for the Y chromosome of most European horse breeds, including the Thoroughbred, originated from Oriental stallions dating to approximately 700 years ago and possibly earlier. Therefore, the import and use of Oriental stallions in development of the Thoroughbred breed during the sixteenth to eighteenth centuries was part of this ongoing trend. With respect to mares, Bower et al. (6) compared maternally inherited mitochondrial DNA sequences among different breeds and found some mitochondria genotypes among modern Thoroughbred mares distinct from those found among modern Arabian horses. They concluded that English mares (for example, Irish Hobby horses and Galloways) as well as other European mares were significant contributors to the foundation of the Thoroughbred breed. Finally, although two of the three stallions often noted as having the greatest impact on the breed are sometimes identified as Arabian, a comparison of genetic variants found among Thoroughbred and Arabian horses demonstrated that today's Thoroughbred has a genomic identity distinct from modern Arabian horse populations (7).

In summary, the Thoroughbred racehorse is not a type of Arabian horse, modified by selection, but rather is derived from crosses of Oriental horses and horses indigenous to the British Isles followed by selection for size, strength, and stamina. The health, athleticism, and size of the modern Thoroughbred led to its subsequent use as foundation stock for many breeds developed during the past two centuries. In addition, Thoroughbred horses have often been crossed with horses in other populations to increase athleticism and genetic diversity. Consequently, the genetics of Thoroughbreds are shared with most modern horse breeds (8).

The fundamental principles of genetics were not developed fully until the twentieth century, and consequently, breeders made some early missteps when seeking to identify successful breeding patterns [reviewed by Binns & Morris (2)]. One popular, but errant, pattern included the Figure System developed by Bruce Lowe in the late 1800s, which encouraged selection based on prominent dam families. Another system, called Dosage, developed by Jean-Joseph Vuillier in the early 1900s, emphasized selection based on the presence of prominent stallions, deep in the pedigree and designated Chef-de-Race. However, modern Thoroughbred breeding has been guided by an understanding of Mendelian genetics and the observation articulated by Joe Estes (9, p. 101), the long-time editor of *The Blood Horse*, that "Pedigrees are useful only when we are ignorant of the merit of the individual, and not very useful then." The challenge for breeders has been determining how to best identify merit in race horses.

DECLINING RATE FOR RECORD TIMES IN PROMINENT RACES

Race time (time to complete a race) seems an obvious and potentially objective measure of racing performance, because the winner of any race will clearly have completed the distance in the shortest time. To assess how breeding and training practices have improved racing performance, scientists compared the winning race-time records for prominent races across several centuries (10–13). These prominent races were established by each country conducting Thoroughbred racing for the best three-year-old horses on an annual basis. For example, in the United Kingdom, the Epsom Derby (since 1780), the Oaks (since 1779), and the St. Leger (since 1776) have each been run annually since their inception. In the United States, the Triple Crown is a series of three races for three-year-old horses: the Kentucky Derby (since 1875), the Preakness (since 1873), and the Belmont (since 1867). Hámori & Halász (14) compared times for other prominent races in Europe, noting that the improvement of times had been limited during the previous 60 years

Downloaded from www.AnnualReviews.org
Guest (guest)
IP: 18.222.137.xxx
On: Fri, 03 May 2024 20:09:24

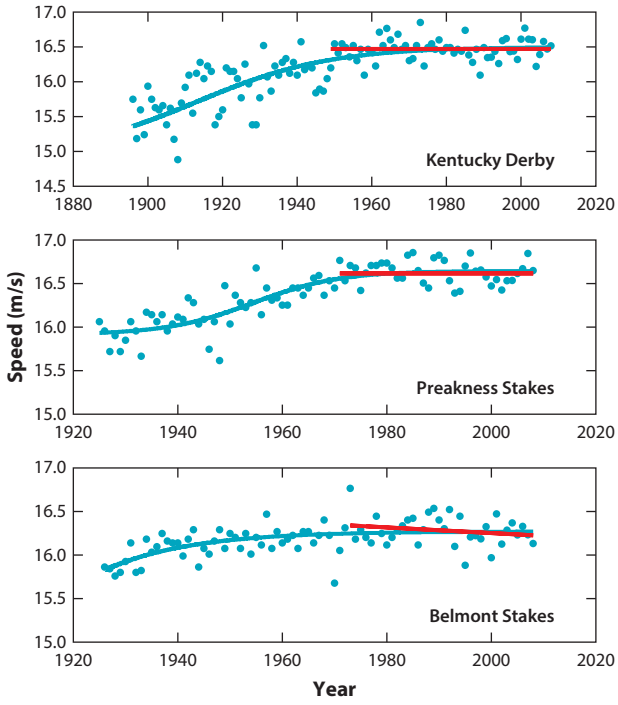

Figure 2

Temporal patterns of winning speeds in the Triple Crown races. Blue dots are winning speeds in the years shown. Red lines are regressions for data in the plateau of each record; any slope of these regression lines is statistically insignificant. Blue lines are the best-fit logistic models. Figure adapted with permission from Denny (13); copyright 2008 Company of Biologists.

as compared to the century prior. More recently, scientists reported that the winning times in some prominent Thoroughbred horse races have not improved significantly after 1910 (10–13) (**Figure 2**). This raised the question: After 400 years of selection, has genetic improvement reached the point at which further improvement is not possible using existing breeding stock? Denny (13) argued persuasively that a limit to maximum accomplishment may have been reached, or nearly reached, for racehorse performance. At the same time, Cunningham (15) used heritability estimates and, noting considerable variation within the Thoroughbred population, suggested that, whereas rare horses may be capable of record-setting performances, this is not true of most horses in the breeding population. Genetics continues to be a useful tool for Thoroughbred breeders. This review examines what is known about the genetics of performance and projects how these tools can best be used to improve the sport.

GENETIC IMPROVEMENT OF THOROUGHBRED RACEHORSES

The Extent of Variation

During the past century, quantitative genetic methods have been developed to assess genetic variation in livestock. When genetic variation contributes to a particular trait, that trait is said to be heritable. Heritability is a measure of the proportion of variation in any particular trait that is due to genetics (the rest is then attributed to the environment, including factors such as training,

Downloaded from www.annualreviews.org
Guest (guest)
IP: 18.220.137.164
On: Fri, 03 May 2024 20:09:24

nutrition, and track surface). Indeed, the identification of estimated breeding values and rigorous application of selection toward heritable traits led to dramatic improvement of production phenotypes in cattle, sheep, pigs, and chickens. Methods of quantitative genetics were also applied to Thoroughbreds, demonstrating that racing performance is heritable (16–18). However, an objective measure to estimate heritability has been elusive.

Heritability of Racing Performance

Racing performance is complex, with horses competing over different distances (1,000 m to 2,400 m), at different ages (often beginning at two years old), on different surfaces (dirt, turf, circular, open), and in diverse weather (fast tracks, sloppy tracks), among other variables. Oki and coworkers (19, 20) reported significant effects of age, jockey, weight, and even racing season on race times. Scientists have used diverse measures of performance, including best time, average time, cumulative earnings, log of cumulative earnings, earnings per start, log of earnings per start, average handicap weight, best handicap weight, and success measured in terms of wins, places, and shows. Estimates of heritability varied among studies depending upon the choice of phenotype, method of analyses, quality of data, and size of the data set. These heritability estimates serve to guide breeders as to what proportion of performance (e.g., race time, winnings) is due to variation that the individual can pass to his/her progeny; traits with high heritability will respond more favorably to selection than those with low heritability. As with any species, there are challenges to arriving at sound estimates of heritability, especially given the diversity of management and among racing programs across the world. Excellent reviews have been published that discuss the problems with and productive approaches to estimating heritability (16, 21–25). For the purpose of this review, the conclusions from various studies of heritability are summarized below and limited to the traits race time, money earned, and handicapping. It is also important to acknowledge that for lowly heritable traits, improvements in performance may be derived most easily from managing environmental factors rather than genetics. For traits with low heritability but high value, selection to improve the genetic merit of individuals, while slow, may still offer substantial benefit.

Heritability for Measures of Race Time

The phenotype of race time (time to complete a particular race distance) seems straightforward. Furthermore, horses race on multiple occasions, so it is possible to get repeated records. Unfortunately, horses do not always race the same distance, nor do all horses have records. Only 40–70% of Thoroughbred horses enter races, and even then, race time is recorded only for the winning horse. For example, Jockey Club records for 2016 from the United States show that only 70% of the horses born and registered competed by the time they were three years old, and of those, only 50% had won at least one race. As a consequence, heritability estimates varied widely among the different studies, from 0.0 to 0.60. Most measures of race time had low heritability, with some reviewers concluding that the most appropriate values probably lie between 0.10 and 0.20 (16, 21–25).

Oki and coworkers (19, 26) noted that heritability estimates varied by distance and suggested that different racing distances should be regarded as different traits. Heritability estimates were different when race times for each distance were considered separately. **Table 1** compares heritability estimates for finish time at different distances from several studies (25, 27–30). Velie et al. (25) based their calculations on winning times, whereas the other three studies estimated race times for all horses based on their distance from the winner at the finish. Heritability for race time was higher for shorter distances in three of the studies (28–30). Oki et al. (26) and Mota et al. (29) also compared times for horses racing multiple distances and reported that correlations in race times decreased consistently as the differences in distance increased.

Downloaded from www.AnnualReviews.org
Guest (guest)
IP: 18.222.x.x
On: Fri, 03 May 2024 20:09:24

Table 1 Heritability (h^2) and repeatability (r) for finish times at different distances on dirt tracks

Study	Moritsu et al. (27)		Oki et al. (28)		Mota et al. (29)		Ekız & Koçak (30)		Velie et al. (25)	
Venue	Japan		Japan		Spain		Turkey		Australia	
Method	Best time		Estimated time		Estimated time		Estimated time		Winning time	
Distance (m)	h^2	r	h^2	r	h^2	r	h^2	r	h^2	r
1,000	-	-	0.191	0.669	0.29	0.63	-	-	0.01	0.02
1,100	-	-	-	-	0.21	0.47	-	-	<0.01	0.21
1,200	0.11	-	0.217	0.645	0.15	0.38	0.353	0.364	0.05	0.13
1,300	-	-	-	-	0.10	0.24	0.309	0.380	-	-
1,400	-	-	0.121	0.509	0.06	0.20	0.265	0.363	<0.01	0.15
1,500	-	-	-	-	0.04	0.15	0.228	0.363	<0.01	0.15
1,600	-	-	0.086	0.556	0.05	0.19	0.248	0.373	0.03	0.12
1,800	0.09	-	0.165	0.671	-	-	0.214	0.289	<0.01	0.25
2,000	-	-	-	-	-	-	0.227	0.404	0.01	0.15

Clearly, measures of time are complex and fraught with confounding variables that limit their efficacy to estimate heritability. Langlois (21, p. 44) suggested that it is not surprising that time performance is improving very little, because "Thoroughbreds do not race against the clock but against each other." He suggested that strategic and psychological aspects associated with racing may contribute more to competition. A horse that is prevented from passing will not have an optimum time. A horse winning by several lengths may not be pushed to finish strong. A horse that passes on a turn will travel farther than a horse running on the inside. And, as noted above, weather, track surface, and other factors can affect race time to a greater extent than genetics, leading to low repeatability for this trait (19, 20, 25).

Heritability for Measures of Money Earned

Success also can be measured with prize money. Modern Thoroughbred breeders recognize genetic merit by comparing racetrack earnings among horses. In 1948, the editor of *The Blood Horse*, Joe Estes (31, 32), developed and published the Average Earning Index, which identified and ranked Thoroughbred stallions based on relative earnings of their offspring by year. In any race, prize money is usually awarded to the first three to five finishers. Races are also classified by the quality of the field of horses; the prize money increases with the quality of the race. Prize money can range from thousands to millions of US dollars. The Average Earning Index has become a major tool for breeders to evaluate breeding stock (33, 34). Because the distribution of prize money in races is not normally distributed, e.g., first place earns significantly more than second, and third place earns even less, transformation of money earned to log of money earned, or even log of money earned per start, has been found to produce higher, and more consistent, heritability estimates than cumulative winnings (17).

The use of money earned has been well studied and described in several very good reviews (16, 21–25). Age was not always considered when making the calculations, and Langlois & Blouin (18) suggested using an index to account for earnings at different ages. One review suggested that moderate heritability of 0.30–0.40 is reasonable for log of annual earnings in Thoroughbreds (23). However, as with measures of race time, the heritability values reported for earnings vary widely depending on which aspect of earnings was measured (cumulative earnings, log of cumulative earnings, earnings per start, log of earnings per start) and are complicated by age and the fact that not all horses in the breeding population have earnings. In fact, using a simulated population and various

Downloaded from www.annualreviews.org
Guest (guest)
IP: 18.220.137.164
On: Fri, 03 May 2024 20:09:24

scenarios of record availability, Burns et al. (35) demonstrated that a lack of performance records from many horses in the breeding population resulted in decreased estimates of heritability.

Heritability for Measures of Handicap Rating

A horse cannot run as fast when it carries more weight. For this reason, weight carried by a horse is standardized for many races to directly compare the ability of those horses. For example, in the Kentucky Derby, the weight of the jockey, saddle, and bridle must be 126 lb for colts and 121 lb for fillies. Metal bars are added to saddle pockets to bring the weight up, if necessary. However, in other races, the goal is to equalize the horses' ability in order to measure the skills of the jockeys and competitive performance of the horses. Weight is added to the saddle of better horses to compensate for differences in their athletic abilities. An expert, usually the secretary of the racetrack offering the race, evaluates the distance (in body lengths) between horses in previous races and then assigns weights to each horse with the objective of allowing, in principle, each horse to be competitive (further complicating the use of time to measure performance!). The added weight is called the handicap. For example, during a 1920 handicap race, the great American racehorse Man O'War carried 138 lb in one race against rivals carrying as little as 118 lb and still won. The term handicap has been extended to encompass the subjective evaluation of a horse's capability.

Moritsu and coworkers (27, 36) determined heritability estimates in Japan were higher using handicap values (0.18) than in their previous studies for race time (0.11 for 1,200 m and 0.09 for 1,800 m). Velie et al. (25) reported heritability values in Australia of only 0.06 and 0.03 for average and highest handicap, respectively, and <0.01–0.05 for time. In general, heritability values for handicapping values vary widely between venues and appear to depend on the methods used for determining handicaps. Although the results are not universally applicable, they can be informative within a particular venue.

Timeform is a handicap measure commonly used in Great Britain and Ireland, developed by Phil Bull in 1948 for his publication of the same name. Bull evaluated past racing performances of horses and assigned handicap ratings to predict future performances. Gaffney & Cunningham (10) used Timeform ratings generated at the end of the racing season for three-year-olds to estimate heritability and measure changes in performance for 31,263 horses racing between 1952 and 1977. They concluded that a reasonable heritability estimate for Timeform was approximately 0.35. They also compared the Timeform scores for sires of these horses and determined an increase in Timeform rating of ~1% per year. While discussing the data, they noted the scores were higher when calculated based on the sires alone and suggested the existence of a bias in how horses are managed depending on the horses' perceived values.

Summary Regarding Heritability

The consensus from these studies is that genetic variation exists in the Thoroughbred horse population for traits associated with racing performance. From the foregoing material, it should be clear that the phenotype encompassed by the term racing performance is more complex than simply measuring record times of the best three-year-old horses in prominent races. Estimates of heritability fell into the low to moderate range for most traits. Consequently, phenotypic selection for these traits is predicted to be effective but with slow improvement. To contribute to further improvement in these phenotypes, the combination of phenotypic selection with improved management, and a better understanding of dominance, epistasis, and how genetics and the environment interact, is likely necessary. A new approach for the application of genetics to select and breed better racehorses became available when the horse genome was sequenced in 2007.

Downloaded from www.annualreviews.org
Guest (guest)
IP: 18.222.137.xxx
On: Fri, 03 May 2024 20:09:24

Whole-Genome Sequencing of the Horse and Candidate Genes

With the advent of whole-genome sequencing and genome annotation for the horse, it became possible to identify and investigate the actual genes contributing to racing performance. The equine genome sequence was annotated first with reference to genes in other species and made available in 2007 (on three websites: the National Center for Biotechnology Information, the University of California, Santa Cruz Genome Browser, and Ensembl) (37, 38). The genome is large, with 2.5 million base pairs and more than 20,000 protein-coding genes identified. At the same time, more than 450,000 genetic elements are thought to regulate these genes, and work is ongoing to identify them (39–41). Among diverse horse breeds, more than 25 million genetic variants, called single-nucleotide polymorphisms (SNPs), have been identified (42). Subsequently, several tools were produced allowing genotyping of large numbers (50,000–670,000) of SNPs, enabling studies of the relationships among individuals based on genetic variation (42, 43). These genotyping tools are commonly referred to as SNP chips and are used to study genetic diversity among 36 breeds (8). The range for expected heterozygosity, a standard measure of diversity, ranged from 0.232 to 0.311. Ranked from most to least diverse, the Thoroughbred ranked 31 out of the 36 breeds evaluated. The breeds with lower heterozygosity were usually breeds with smaller populations, some of which had experienced recent bottlenecks that reduced diversity. In contrast, the Thoroughbred horse has a large population with a worldwide distribution. Consequently, the lower expected heterozygosity is likely the result of several centuries of selection for racing performance, effectively removing much of the genetic diversity that contributes to poor performance. To that point, these genomic tools were applied to discover genes that impact racing ability.

Candidate Genes for Athletic Performance

For the first time, it became possible to study every annotated gene in the horse and ask whether variation in each particular gene has an impact on racing success. Genes were identified as candidates through several processes. Many genes were identified as impacting athletic performance in other species. Studies in humans identified 221 genes suspected to affect athleticism (44). The gene *myostatin* (*MSTN*) was a candidate based on observations of its effects on racing performance in whippet dogs (45) and extreme musculature in cattle (46). Schröder et al. (47) reviewed the nature of equine athletic performance and identified candidate genes based on those known to be involved in musculoskeletal, hemodynamic, mitochondrial, and health-related aspects of athleticism.

Signatures of Selection

Comparisons of Thoroughbreds to non-Thoroughbreds, called investigations of signatures of selection, identified chromosomal regions for which variants are at a higher frequency among Thoroughbred horses than they are among individuals within other breeds. The premise for these differences is that these regions have been selected in Thoroughbred horses because they have an impact on racing and were not selected in other horse breeds. Two studies, described below, identified chromosome regions that differed when comparing Thoroughbred to non-Thoroughbred horses.

The first study for signatures of selection was conducted before the availability of the SNP chip with other genetic markers dispersed across the genome, namely, microsatellites. Gu et al. (48) did the first genome-wide scan for performance-linked markers using 394 microsatellite markers to compare the distribution of variants between 112 Thoroughbred and 52 non-Thoroughbred horses. Regions that were similar among Thoroughbreds and different among non-Thoroughbreds were then hypothesized to harbor genes and genetic variation that contribute

Downloaded from www.annualreviews.org
Guest (guest)
IP: 18.220.137.164
On: Fri, 03 May 2024 20:09:24

Table 2 Results from GWAS for measures of racing success in Thoroughbred racehorses

Study	Phenotype and approach	Chromosome: approximate region (Mb) (EquCab 2.0 unless noted otherwise)	Candidate genes nominated by author
Gu et al. (48)	Signatures of selection using 394 microsatellite loci on 112 Thoroughbreds and 52 non-Thoroughbreds	ECA1: ~71.2	*ACTA1*, *ACTN2*, *AGT*, *GGPS1*, *TOMM20*
		ECA4: ~38.6	*ACN9*
		ECA9: ~18.9	*ADHFE1*, *MTFR1*
		ECA21: ~46.7	*DNAH5*
		ECA22: ~0.2	*ACSS1*
		ECA25: ~25.7	*ATP6V1G1*, *GSN*, *HSDL2*, *LTB4DH*, *MUSK*, *NDUFA8*, *PTGs1*, *SVEP1*, *TNC*, *UGCG*
Petersen et al. (8)	GWAS with 23,401 SNPs on 744 horses of 33 breeds for signatures of selection	ECA17: 20.7–23.2	*KCNRG*, *TRIM3*, *SPRYD7*, *KONA3*, *EBPL*, *ARL***, *RCBTB1*, *PHF11*, *SETDB2*, *CAB39L*, *CDADC1*, *FNDC3A*, *CYSLTR2*, *RCBTB2*, *RB1*, *LPAR6**, *ITM2B*, *MED4*, *NUDT15*, *SUCLA2**
		ECA14:41–42.5	10 genes, but not identified
Shin et al. (52)	Race times 240 horses/retest 190 SNPs with 916 horses	ECA5: ~40.7	None specified
		ECA8: 59.0–61.2, 115.9	*INPP5J*, *OSP2*, *MVK*
		ECA16: 14.2–17.9	*CNTN3*, *PDZRN3*, *PPP4R2*, *GXYLT2*, *SHQ1*
		ECA20: 30.1–32.1	*VARS2*,
		ECA21: 1.1–18.3, 47.5–49.9	*ARL15*, *CCT5*, *TAS2R1*
		ECA28: ~17.4	None specified
		ECA 30: ~38.7	*RGS7*
Hill et al. (49)	Best distance: Elite racehorses by distance; GWAS with SNP chip	ECA18: 65.8–67.5	*MSTN*
Binns et al. (50)	Best distance: Elite winners by distance; GWAS with SNP chip	ECA18: 65.8–67.2	*MSTN*
Tozaki et al. (51)	Lifetime earnings and rank; 1,440 microsatellites followed by fine-mapping of SNPs	ECA18: 62.6–66.8	*MSTN*
Farries et al. (53)	Precocity: age of work out at first race, best race, age at best race, best win distance	ECA1: ~37.1	*HTR7*
		ECA18: ~66.4	*MSTN*
Farries et al. (54)	Speed measures: PCA of speed indices during training; compare Elite, non-Elite, and not-raced among two-year-old Thoroughbreds	ECA8: ~90.2, ~94.6, ~97.4, ~101.1	*MN1*, *CRYBA4*, *TPST2*, *MYO18B*
		ECA2: ~100.2	*SCLT1*
		ECA11: ~42.4, ~42.9, ~44.3	*FOXN1*, *MYO18A*, *ABR*
McGivney et al. (55)	Durability: lifetime starts, starts at 2 and 3, longevity, raced versus not-raced	ECA7: 39.2–44.8	*NTM*, *OPCML*, others
		ECA7: 62.1–62.3	*PRCP*, others
Han et al. (57)	CSS for 99 horses from Australia versus rest of world; corroborate with GWAS for ECA14	(EquCab3.0) ECA14: 32.2–37.5	*PCDHGC5*, *APBB3*, *HBEGF*, *NDUFA2*, *SRA1*, *NR3C1*, *ARHGAP26*
		(EquCab3.0) ECA6: 34.4–35.3	*CLSTN3*
		(EquCab3.0) ECA16: 24.3–26.5	*ATXN7*

Abbreviations: CSS, composite selection signals; GWAS, genome-wide association studies; PCA, principal component analysis; SNP, single-nucleotide polymorphism.

to racing performance. They reported positive signals on horse chromosomes 1, 4, 9, 21, 22, and 25 and suggested some genes underlying those regions as candidate genes that contribute to racing ability (**Table 2**). After SNP genotyping arrays became available, Petersen et al. (8) used the SNP chip to compare the distribution of variants among 33 modern horse breeds to identify signatures

Downloaded from www.annualreviews.org
Guest (guest)
IP: 18.222.157.xxx
On: Fri, 03 May 2024 20:09:24

of selection in diverse breeds. The study suggested associations with regions for traits such as color, gait, size, and performance in breeds consistent with common uses and selection on those breeds. They reported variants in two regions that denoted possible selection in the Thoroughbred, one on ECA2 and another on ECA14 (**Table 2**).

Genome-Wide Association Studies (GWAS) Comparing Thoroughbreds

Since 2010, genome-wide association studies (GWAS) became the predominant approach to discover genetic influences on racing performance. Groups of horses with phenotypic differences were genotyped using SNP chips, and the distributions of SNP variants among the groups were compared. If there were statistically significant differences in the distribution of SNPs among horses based on the phenotypic differences, it was hypothesized that genes underlying that chromosome region represented by those SNPs may have a genetic influence on the trait. Annotation of the genome, as well as knowledge of gene function, allowed nomination of candidate genes within these regions. The most common GWAS approach has been to compare Thoroughbreds based on performance. One of the most common comparisons was of horses that had won prestigious races, for example, the Stakes Races in the United States, to horses that had race records but had not achieved a win in a Stakes race. Winners of prestigious races are referred to as Elite, whereas horses that race without winning in prestigious races are referred to as non-Elite. The approach initially used combined GWAS to identify chromosome regions associated with performance and then investigated the candidate genes that may lie within the region. Horses were categorized by phenotype and tested for SNPs using one of the SNP chips. The results of some of those studies are described below and in **Table 2**. Also using GWAS, a series of publications demonstrated the effect of a chromosome region on best race distance for a horse (49–51). The premise for those studies was to investigate the candidate gene *MSTN* and is discussed below in greater detail.

Shin et al. (52) used race time as a phenotype and conducted a two-stage GWAS in which they identified candidate genomic regions for performance among 240 Thoroughbreds, followed by an assessment of the 190 most significant SNPs among 916 additional horses. They reported statistically significant associations on horse chromosomes 5, 8, 16, 20, 21, 28, and 30, which included 17 candidate genes related to muscle function.

Farries et al. (53) conducted a GWAS of Thoroughbred horses from five continents and compared traits equated with performance at an early age (precocity), finding associations on horse chromosomes 18 and 1. The strongest signal was associated with *MSTN* on ECA18.

Farries et al. (54) also measured velocity, acceleration, and distance covered for 294 Thoroughbreds in training and compared the principal component analyses for these traits, called race indices, among horses with Elite, non-Elite, and no-racing careers in a GWAS using SNP variants. Although highly significant results were not achieved, SNPs on chromosomes 2, 8, and 11 reached levels suggesting an influence, and candidates were identified in those regions.

McGivney et al. (55) conducted GWAS to identify genetic variants associated with racecourse starts in Irish Thoroughbreds. They determined a heritability of 0.11 for lifetime starts and starts at two and three years old. This heritability score corroborated the results of Velie et al. (56), who estimated heritability values of 0.12 for longevity (elapsed time from first to last race) and 0.11 for persistence (number of races during career) in Australian Thoroughbreds. A GWAS identified two significant regions on ECA7. A validation set of 528 horses confirmed the association, and a set of SNPs was identified for use in a genomic prediction model that explained 24.7% of the phenotypic variation in lifetime starts. Genes in the region that may function to alter this phenotype included several that, in other species, were thought to influence behavior.

Downloaded from www.AnnualReviews.org
Guest (guest)
IP: 18.220.137.164
On: Fri, 03 May 2024 20:09:24

Han et al. (57) noted that GWAS could miss some significant markers based on signals of selection and used an approach called composite signal selection to compare Australian Thoroughbreds to Thoroughbreds from elsewhere in the world. In addition, they performed a GWAS for principal component analysis scores for exercise measures and identified a region on ECA14 that corroborated the results of the composite signal selection study and implicated *PCDHGC5*, finding a significant difference for linked SNPs comparing Elite versus non-Elite Australian horses ($P = 0.01$) but not in European horses.

Summary from GWAS

These studies provide statistical evidence that specific regions of the genome may contain variation impacting racing performance. In principle, one could select horses based on genetic signatures from these regions to obtain a genetic advantage. However, with the exception of ECA18, these studies tend to implicate different chromosome regions. This suggests that many and different genetic factors influence racing performance.

Evaluation of Candidate Genes

Genetic selection using genetic markers associated with a particular phenotype has been applied for livestock improvement with other species, notably cattle. Success depends upon the association strength and selection intensity. However, these association studies and advances in molecular genetics make it possible to dig down and seek the underlying basis for the association by identifying the precise variant responsible and its physiological effect. So far, only a few candidate genes have been investigated.

Myostatin and Best Distance

Several of the GWAS in **Table 2** implicated *MSTN* as having an impact on various aspects of racing performance. This discovery was made within a year of publication of the whole-genome sequence for the horse. Hill et al. (49) reported an association between an intron variant in the gene *MSTN* and performance at short and long distances. The impetus for the work was the observation that whippet dogs that were heterozygous for a *MSTN* polymorphism were more competitive than homozygotes for either variant (45). *MSTN* inhibits development of glycolytic type IIB muscle fibers, and individuals with decreased *MSTN* expression have a more developed musculature. The intronic variant of the *MSTN* gene was found at EquCab 2.0: ECA18: g.66493737T>C, and the authors noted the C allele was associated with sprinting ability, whereas the T allele was associated with endurance. **Figure 3**, from that report (49), shows the distribution of genotypes found among winners at different distances. Subsequently, Hill et al. (58) conducted a GWAS using 40,977 SNPs comparing horses that had racing successes at short, medium, or long distances and demonstrated an association with the region harboring *MSTN*. During that study, they resequenced *MSTN*, leading to the identification of three additional SNPs in the 3′ untranslated region, as well as a polymorphic insertion of a repetitive element (ERE-1) in the gene promoter. The observation was confirmed in two GWAS showing an association of this region with lifetime earnings (51) and with optimum race distance (50, 51). Petersen et al. (59) investigated an association of *MSTN* with muscle fiber types and suggested that the ERE-1 element was the cause of the phenotype and not the T>C SNP. Subsequently, investigations of the ERE-1 variant in the *MSTN* promoter demonstrated that ERE-1 downregulated the expression of the gene in horses with superior sprinting ability (60, 61). Both intronic variants in *MSTN* were found in other breeds, and their distributions were consistent with the athletic character of other breeds. Specifically,

Downloaded from www.annualreviews.org
Guest (guest)
IP: 18.222.x.x
On: Fri, 03 May 2024 20:09:24

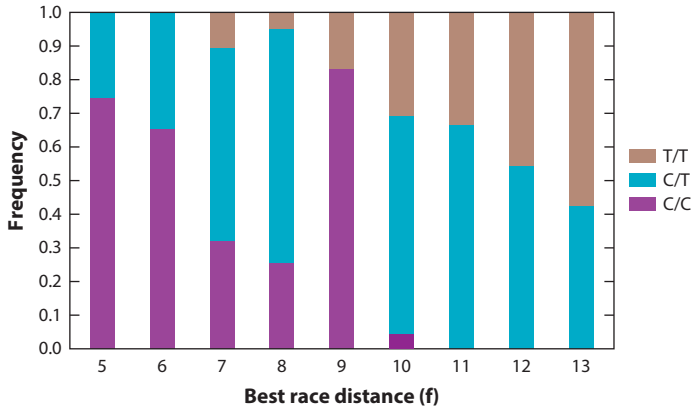

Figure 3

Optimal *MSTN* genotype for racing distance. Distribution of C/C, C/T, and T/T genotypes among $n = 179$ Elite Thoroughbreds. Figure modified from Hill and coworkers (49) with permission from authors.

Arabian horses, renowned for endurance racing, were almost fixed for the T variant, whereas American Quarter horses, highly selected for sprinting, were almost fixed for the C variant, which was nearly always inherited together with the ERE-1 element (59, 62). In fact, in the Quarter Horse and Thoroughbred breeds, the intronic, C allele, and ERE-1 element are in high linkage disequilibrium (59). The *MSTN* allele associated with sprinting (ERE-1) was also associated with increased number of starts for two- and three-year-old Thoroughbred horses (53). In summary, genotypes at this locus will not determine whether or not a horse is a champion; however, they appear to indicate the best racing distance for the horse (sprint versus endurance). Heterozygotes for the *MSTN* variants were found among Thoroughbred champions at all distances, whereas endurance horses were not homozygous for the insert, and sprinters had at least one copy of the insert (**Figure 3**).

CANDIDATE GENE STUDIES: EVIDENCE FOR *COX4I2*

Gu et al. (63) chose 23 candidate genes based upon their described function in exercise physiology and identified 41 variants in 14 of these genes for further study. The distribution of variants was compared for Elite versus other Thoroughbreds that competed but had never won. Significance was found in the initial survey of 148–150 horses for two variants, one in *COX4I2* (EquCab 2.0: ECA2 g.22684390C>T) and another in *CKM* (EquCab 2.0: ECA10 g.15884567A>G). The associations became stronger when comparing the Elite winners to the set of nonwinners. However, after adding a second cohort of 130 horses, only *COX4I2* remained statistically significant (**Table 3**). The authors hypothesize that variation in each of the genes could contribute to performance by altering energy metabolism. *COX4I2* (*cytochrome oxidase subunit 4I2*) encodes a protein that is one part of the multi-subunit enzyme, cytochrome c oxidase, which catalyzes electron transfer from reduced cytochrome c to oxygen during mitochondrial respiration.

CANDIDATE GENE STUDIES: EVIDENCE FOR *PDK4*

Hill et al. (64) investigated another 17 candidate genes selected from regions implicated by a previous study (48). They identified 57 variants across these genes and compared Elite versus non-Elite horses, as well as differences in best distance among Elite performers. Statistically significant

Downloaded from www.annualreviews.org
Guest (guest)
IP: 18.220.137.164
On: Fri, 03 May 2024 20:09:24

Table 3 Most significant candidate genes/mitochondria for impact on performance

Gene	Author	Trait	Variant	Distribution	Impact
MSTN	Hill et al. (49)	Best distance	EquCab 2.0: ECA18 g.66493737C>T and promotor insert; insert thought causal	See **Figure 2**	"C" and the promotor insert associated with sprinting ability
COX4I2	Gu et al. (63)	Elite versus non-Elite	EquCab 2.0: ECA2 g.22684390C>T presumed linked	Genotypes Ratios: CC:CT:TT TBE: 4:44:32 TBO: 10:30:15	"TT" and "CT" favorable for Elite
PDK4	Hill et al. (64)	Elite versus non-Elite	EquCab 2.0: ECA4 g.38973231 A>G presumed linked	Genotypes Ratio: AA:AG:GG TBE: 35:92:55 TBO: 10:35:50	"AA and "AG" favorable for Elite
Mitochondria	Harrison & Turion Gomez (65)	Elite wins at distances from 1,400 m–2,800 m	13 mitotypes based on selected mtDNA sequences	Frequency for favorable 5 of 13 mitotypes 0.516; based on 1,000 Thoroughbreds	$P < 0.05$–0.001 for favorable effect of 5 mitotypes and race distance
Mitochondria	Lin et al. (66)	Performance: Elite races versus non-Elite	Whole-genome sequence of mitochondria for 195 horses	Frequency of L3b and L3b1a = 0.208 for 1,123 horses	Negative effect for mitotype L3b ($P = 0.0003$) and L3b1a ($P = 0.0007$)

Abbreviations: mtDNA, mitochondrial DNA; TBE, Thoroughbred Elite; TBO, Thoroughbred "other" or non-Elite.

results ($P < 0.05$) were found for 13 SNPs in 9 genes, *ACSS1*, *TCN*, *ACTN2*, *PTGS1*, *PDK4*, *ACN9*, *PON1*, *ADHFE1*, and *COX4I1*. The most significant result implicated *PDK4* (EquCab2.0: ECA4 g.38973231 A>G) when comparing Elite versus non-Elite horses. Variants from four genes (*COX4I1*, *PDK4*, *ACSS1*, and *AN9*) were selected for validation studies, but only the tests with *PDK4* produced significant results following validation (**Table 3**). *PDK4* (*pyruvate dehydrogenase kinase 4*) encodes a mitochondrial protein that inhibits the pyruvate dehydrogenase network and contributes to regulation of glucose metabolism.

Mitochondria

Mitochondria are organelles in the cell responsible for aerobic respiration, specifically, the generation of ATP energy via glucose oxidation. Aerobic respiration is very important for races longer than 1,600 m, whereas anaerobic respiration is most effective in short-distance races. Harrison & Turrion-Gomez (65) compared 13 mitochondrial genotypes (mitotypes) for Elite horses at distances ranging from 1,400 m to 2,800 m and identified statistically significant differences for five of the mitotypes (four mitotypes at $P < 0.05$ and one mitotype at $P < 0.001$) with respect to performance (**Table 3**). However, Lin et al. (66) investigated whole-genome sequences of mitochondria for Thoroughbreds and did not find mitotypes associated with superior performance but identified one associated with poorer performance ($P = 0.0003$). The authors speculated the effect could be due to a deleterious variant (T1458C), which may affect 16s ribosomal RNA structure (**Table 3**).

SUMMARY FOR CANDIDATE GENES

The variants and genes listed in **Table 3** are those that were published and rose to the level of statistical significance in association tests with performance traits. Although the argument is persuasive for the *MSTN* variation exerting an effect on racing performance, we cannot be certain that the other genes are actually responsible for the associations. In the case of the two genes, *COX4I2* and *PDK4*, and the mitochondrial variants, the functional cause of the association could

Downloaded from www.annualreviews.org
Guest (guest)
IP: 18.222.137.xxx
On: Fri, 03 May 2024 20:09:24

lie in a nearby region of DNA. In other words, the variants examined may be in linkage disequilibrium with the actual variant that changes the phenotype. The actual variant(s) responsible for the altered function cannot be determined without additional data.

Other candidate genes were implicated with racing phenotypes but did not achieve statistical significance. That does not necessarily mean those genes are excluded from being important to racing performance. Achieving statistical significance depends upon the magnitude of the impact each variant has on the phenotype, the strength of the association, and the size of the study. Regardless, the effects of these and other, as-yet-unproven variants on performance appear to be modest. Furthermore, although some genotypes were more prevalent among successful horses, winning horses were found with a wide variety of genotypes. This is likely due to the complexity of the trait (racing success) as well as the complexity of the genetics of performance. Just as the Thoroughbred's racing prowess is due to evolutionary developments in diverse systems (e.g., cardiovascular, musculoskeletal), it appears many genes have detectable (statistically significant) effects, but individually, none has an overwhelming impact on racing success.

INVESTIGATIONS OF GENETIC REGULATORY ELEMENTS AND PERFORMANCE

Another approach to identifying candidate genes is to find those genes whose transcription (RNA production, i.e., expression) is relatively increased or decreased during exercise. Denham et al. (67) reviewed studies of gene expression associated with training and exercise in horses. They documented that, as a result of training and/or exercise, some genes are upregulated and some are downregulated in all horses. However, if there are differences, and those differences are heritable, then measures of gene expression may serve as a more specific phenotypic measure than race time, money earned, or handicapping. Farries et al. (68) used gene expression levels in muscle as a phenotype and then applied GWAS using SNP chips. They took muscle biopsies from untrained horses before and following exercise and performed RNA sequencing to assess changes in gene expression for 111 horses. The responses were measured, and associations were identified using a modification of GWAS that identified associated transcripts termed expression quantitative trait loci (eQTL). They reported thousands of eQTL, either near the SNP variants (*cis*) or distant to or on a different chromosome than the associated SNP variants (*trans*). The next challenge is to identify which eQTL impact performance phenotypes. Han et al. (57) noted that one of the significant associations in a GWAS was near an eQTL for the gene *ARHGAP26*. A productive future approach may be to couple genetic marker studies with gene expression data (54, 69) to provide discrete phenotypic measures to match the discrete genetic measures.

OTHER ASPECTS: ELIMINATING THE NEGATIVE

Gaffney & Cunningham (10) and Eckhardt et al. (11) suggested that performance barriers may be related to constraints on physiology and metabolism or may be attributed to unsoundness resulting from the negative interaction of fragility with speed. Identifying those elements that constrain performance and focusing selection in those areas might provide effective approaches for breeders. Estimates of the heritability of undesirable conformational traits range from those that are lowly heritable (e.g., base narrow and turned-out feet = 0.16, turned-in feet = 0.17) to several that are not only high in heritability (offset knees = 0.42, back at the knee = 0.66) but also prevalent (turned-out feet in 30% of the nearly 4,000 Thoroughbred yearlings studied) (70). Although there was a tendency for horses with conformational faults to be less likely to race than those without, one study found no significant associations between performance and conformation (70). Similarly, no association was found between forelimb conformation and racing performance

Downloaded from www.AnnualReviews.org
Guest (guest)
IP: 18.220.137.164
On: Fri, 03 May 2024 20:09:24

in two-year-olds (71). At the same time, some of the most common reasons that horses do not compete are related to lameness and musculoskeletal injuries. Jeffcott et al. (72) reported that 52% of horses in training at Newmarket during 1980 experienced lameness requiring veterinary attention. Of those, 11% sustained injuries that ended their careers. Wilsher et al. (73) studied approximately 1,000 two- and three-year-olds racing in Australia; 4% of two-year-olds and 9% of three-year-olds had career-ending injuries. Several genetic studies have shown that the occurrence of fractures has a low to moderate heritability. Oki et al. (74) reported a heritability of 0.17–0.19 for injury to the superficial digital flexor tendon. Welsh et al. (75) reported heritability values of 0.31–0.34 for injuries to the superficial digital flexor tendon and 0.21–0.37 for distal limb fractures. GWAS implicated genetic variants on ECA18 as associated with fracture; Blott et al. (76) suggested the strongest association was with the gene *ZNF804A*, whereas Tozaki et al. (77) suggested that a pleiotropic effect including *MSTN* was responsible for the strong association.

Two other health conditions that have been analyzed for heritable influences are recurrent laryngeal neuropathy and exercise-induced pulmonary hemorrhage (EIPH). The former is a common cause of laryngeal hemiplasia, often referred to as roaring. Two GWAS have been conducted on this condition. One, in Warmblood horses, found weak evidence for a genetic effect of chromosome 21 (78). The second study implicated genes on chromosome 1 that had been associated previously with size (*LCORL/NCAPG*) (79); large horses have been suspected to be predisposed to recurrent laryngeal neuropathy.

EIPH results from rupture of blood vessels in lung alveoli and, if excessive, can affect athletic performance (80). EIPH appears to be common based on endoscopic examination of Thoroughbred lungs after exercise, but sometimes blood can be seen in the nostrils, a condition called epistaxis. Epistaxis is considered an extreme manifestation of EIPH. A heritability value of 0.27 was estimated for lifetime risk of epistaxis (81), although no gene(s) have been identified that may contribute to the condition.

CONCLUSIONS

Genetic variation exists for racing performance among Thoroughbreds, and breeders stand to benefit from discovery of better ways to practice selection. An overarching question is whether or not the maximum limit for racehorse performance has been reached by some equine athletes. Obviously, there will be some limit on what a horse may achieve. It seems inconceivable that a horse will ever run a mile in a minute. But is it possible that existing racing records will be broken? In 1973, Secretariat won the Belmont Stakes by 31 lengths in 2 min 24 s, eclipsing the previous record by 3.2 s. Since then, the best time in the Belmont Stakes was recorded by Easy Goer in 1989 at 2 min 26 s, a full 2 s slower than Secretariat's time in 1973. What will it take to break Secretariat's record? Certainly, such a horse must have an abundance of genetic advantages, no structural weaknesses, superior management and training, a skilled jockey, and good fortune. We anticipate the tools developed in these and future genetic studies will contribute to the development and discovery of other record-breaking horses.

Clearly, we are at the early stages of identifying important genetic markers for racing performance at the DNA level. Lots of variation exists, and molecular genetic studies identify effects from many areas of the genome. This field will benefit from joining performance measures with gene expression studies and genetics. Not surprisingly, several studies have noted large influences of management on racehorse success; genetic influences could not be measured without accounting for the effect of the trainer and jockey. The role of genetics may not be to identify winners but, rather, to identify those horses that will not benefit from the skills of the trainer and jockey owing to insufficient genetic merit or heritable defects.

Downloaded from www.AnnualReviews.org
Guest (guest)
IP: 18.222.x.x
On: Fri, 03 May 2024 20:09:24

SUMMARY POINTS

1. Record times are established less frequently in prominent races than they were 100 years ago. This poses the question of whether horses are approaching a physiological barrier for racing performance.
2. Horses race varying distances at different ages and on different types of tracks. There is no single phenotypic measure for racing success.
3. Race time, money earned, and handicapping have all been used as measures of racing success. Many studies of Thoroughbred populations across the world have demonstrated low to moderate heritability of racing success using all of these measures.
4. Molecular genetic studies demonstrate association of racing performance traits with many regions of the genome and even provide evidence for effects of some specific genes. However, none of the variants individually determines success.
5. *MSTN* is well established as influencing best racing distance (sprinting versus endurance) but, alone, does not determine racing success.
6. Some horses may have superior genetics for racing but fail owing to other genetic deficits. Some variants have been identified with unsoundness, which has measurable heritability.
7. Genetic factors contributing to racing success are complex; those influencing health, behavior, and soundness may play a role in addition to genes impacting musculoskeletal, cardiovascular, and other systems directly involved in athletic performance.

FUTURE ISSUES

1. Race time, money earned, and handicapping are phenotypes with a large subjective component for Thoroughbreds. As research moves forward, we must identify objective measures of athletic performance for our genetic studies. Improved understanding of the equine adaptations by physiological systems will provide better targets for molecular genetic research.
2. Whole-genome sequencing is becoming much less expensive and can be used in place of single-nucleotide polymorphism genotyping for genome-wide association studies. Indeed, whole-genome sequencing allows for identification of the actual variants causing the effect rather than simply identifying associated chromosomal regions.
3. Current genetic methodologies that entail investigations for the effects of one gene at a time are, by themselves, inadequate to understand the entirety of these complex traits. Indeed, success may be determined by specific combinations of genetic determinants. Development of methods to discover complex interactions among genes with respect to performance will benefit this research.
4. Improved annotation of the horse genome to include regulatory elements active in various tissues and to refine the annotation of protein-coding genes will provide new and potentially better genetic targets for study.

Downloaded from www.AnnualReviews.org
Guest (guest)
IP: 18.220.137.164
On: Fri, 03 May 2024 20:09:24

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Willett P. 1970. *The Thoroughbred*. New York: G.P. Putnams's Sons
2. Binns M, Morris T. 2010. *Thoroughbred Breeding: Pedigree Theories and the Science of Genetics*. Pomfret, VT: J.A. Allen
3. Cunningham EP, Dooley JJ, Splan RK, Bradley DG. 2001. Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses. *Anim. Genet.* 32:360–64
4. Wallner B, Palmieri N, Vogl C, Rigler D, Bozlak E, et al. 2017. Y chromosome uncovers the recent oriental origin of modern stallions. *Curr. Biol.* 27:2029–35.e5
5. Felkel S, Vogl C, Rigler D, Dobretsberger V, Chowdhary BP, et al. 2019. The horse Y chromosome as an informative marker for tracing sire lines. *Sci. Rep.* 9:6095
6. Bower MA, Campana MG, Whitten M, Edwards CJ, Jones H, et al. 2011. The cosmopolitan maternal heritage of the Thoroughbred racehorse breed shows a significant contribution from British and Irish native mares. *Biol. Lett.* 7:316–20
7. Cosgrove EJ, Sadeghi R, Schlamp F, Holl HM, Moradi-Shahrbabak M, et al. 2020. Genome diversity and the origin of the Arabian horse. *Sci. Rep.* 10:9702
8. Petersen JL, Mickelson JR, Rendahl AK, Valberg SJ, Andersson LS, et al. 2013. Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLOS Genet.* 9:e1003211
9. Estes JA. 1958. Pedigrees. In *Stud Managers' Handbook*, pp. 82–101. Lexington: Univ. Ky. (article reprinted from 1952 column from *The Blood Horse*)
10. Gaffney B, Cunningham EP. 1988. Estimation of genetic trend in racing performance of thoroughbred horses. *Nature* 332:722–24
11. Eckhardt RB, Eckhardt DA, Eckhardt JT. 1988. Are racehorses becoming faster? *Nature* 335:773
12. Gardner DS. 2006. Historical progression of racing performance in the Thoroughbred horse and man. *Equine Vet. J.* 38:581–83
13. Denny MW. 2008. Limits to running speed in dogs, horses and humans. *J. Exp. Biol.* 211:3836–49
14. Hámori D, Halász G. 1959. Der Einfluß der Selektion auf die Entwicklung der Schnelligkeit des Pferdes. *Z. Tierzüchtung Züchtungsbiol.* 73:47–59
15. Cunningham EP. 1990. *The Genetics of Performance in Thoroughbreds*. Hobart, Aust.: World Congr. Blood-horse Breed.
16. Langlois B. 1980. Heritability of racing ability in thoroughbreds—a review. *Livest. Prod. Sci.* 7:591–605
17. Ricard A. 1998. *Developments in the genetic evaluation of performance traits in horses. Presented at 6th World Congress on Genetics Applied to Livestock Production*, Jan. 11–16, Univ. New Engl., NSW, Aust.
18. Langlois B, Blouin C. 2007. Annual, career or single race records for breeding value estimation in race horses. *Livest. Sci.* 107:132–41
19. Oki H, Sasaki Y, Willham RL. 1994. Genetics of racing performance in the Japanese Thoroughbred horse: II. Environmental variation of racing time on turf and dirt tracks and the influence of sex, age, and weight carried on racing time. *J. Anim. Breed. Genet.* 111:128–37
20. Oki H, Sasaki Y, Lin CY, Willham RL. 1995. Influence of jockeys on racing time in Thoroughbred horses. *J. Anim. Breed. Genet.* 112:171–75
21. Langlois B. 1996. A consideration of the genetic aspects of some current practices in Thoroughbred horse breeding. *Ann. Zootech.* 45:41–51
22. Tolley E. 1985. A review of the inheritance of racing performance in horses. *Anim. Breed. Abstr.* 53:163–85
23. Ricard A, Bruns E, Cunningham EP. 2000. Genetics of performance traits. In *The Genetics of the Horse*, ed. AT Bowling, A Ruvinsky, pp. 411–38. Wallingford, UK: CAB Int.
24. Thiruvenkadan AK, Kandasamy N, Panneerselvam S. 2009. Inheritance of racing performance of Thoroughbred horses. *Livest. Sci.* 121:308–26

Downloaded from www.AnnualReviews.org
Guest (guest)
IP: 18.222.[illegible]
On: Fri, 03 May 2024 20:09:24

25. Velie BD, Hamilton NA, Wade CM. 2015. Heritability of racing performance in the Australian Thoroughbred racing population. *Anim. Genet.* 46:23–29
26. Oki H, Sasaki Y, Willham RL. 1997. Estimation of genetic correlations between racing times recorded at different racing distances by restricted maximum likelihood in Thoroughbred racehorses. *J. Anim. Breed. Genet.* 114:185–89
27. Moritsu Y, Funakoshi H, Ichikawa S. 1994. Genetic evaluation of sires and environmental factors influencing best racing times of Thoroughbred horses in Japan. *J. Equine Sci.* 5:53–58
28. Oki H, Sasaki Y, Willham RL. 1995. Genetic parameter estimates for racing time by restricted maximum likelihood in the Thoroughbred horse of Japan. *J. Anim. Breed. Genet.* 112:146–50
29. Mota MD, Abrahão AR, Oliveira HN. 2005. Genetic and environmental parameters for racing time at different distances in Brazilian Thoroughbreds. *J. Anim. Breed. Genet.* 122:393–99
30. Ekız B, Koçak Ö. 2007. Estimates of genetic parameters for racing times of Thoroughbred horses. *Turk. J. Vet. Anim. Sci.* 31:1–5
31. Estes JA. 1948. More notes on ranking of sires in proportion to opportunity. *Blood Horse* 52:650–51
32. Estes JA. 1948. Statistics on prominent sires, adjusted for changing dollar. *Blood Horse* 52:470–71
33. Estes JA, Baumohl A. 1960. Racing class and sire success. *Blood Horse* 60:48–53
34. Estes JA, Baumohl A. 1960. Dams of stakes winners. *Blood Horse* 60:99–105
35. Burns EM, Enns RM, Garrick DJ. 2006. The effect of simulated censored data on estimates of heritability of longevity in the Thoroughbred racing industry. *Genet. Mol. Res.* 5:7–15
36. Moritsu Y, Terai A, Tashiro T. 1998. Relationship between sire breeding values for the rating score on turf and dirt racing tracks in Thoroughbred racehorses. *J. Equine Sci.* 9:89–92
37. Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, et al. 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 326:865–67
38. Kalbfleisch TS, Rice ES, DePriest MS Jr., Walenz BP, Hestand MS, et al. 2018. Improved reference genome for the domestic horse increases assembly contiguity and composition. *Commun. Biol.* 1:197
39. Burns EN, Bordbari MH, Mienaltowski MJ, Affolter VK, Barro MV, et al. 2018. Generation of an equine biobank to be used for Functional Annotation of Animal Genomes project. *Anim. Genet.* 49:564–70
40. Donnelly CG, Bellone RR, Hales EN, Nguyen A, Katzman SA, et al. 2021. Generation of a biobank from two adult Thoroughbred stallions for the Functional Annotation of Animal Genomes initiative. *Front. Genet.* 12:650305
41. Kingsley NB, Kern C, Creppe C, Hales EN, Zhou H, et al. 2019. Functionally annotating regulatory elements in the equine genome using histone Mark ChIP-Seq. *Genes* 11:3
42. Schaefer RJ, Schubert MK, Bailey EK, Bannasch DL, Barrey EP, et al. 2017. Developing a 670k genotyping array to tag similar to 2M SNPs across 24 horse breeds. *BMC Genom.* 18:565
43. McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, et al. 2012. 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
44. Bray MS, Hagberg JM, Pérusse L, Rankinen T, Roth SM, et al. 2009. The human gene map for performance and health-related fitness phenotypes: the 2006–2007 update. *Med. Sci. Sports Exerc.* 41:35–73
45. Mosher DS, Quignon P, Bustamante CD, Sutter NB, Mellersh CS, et al. 2007. A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. *PLOS Genet.* 3:e79
46. McPherron AC, Lee SJ. 1997. Double muscling in cattle due to mutations in the myostatin gene. *PNAS* 94:12457–61
47. Schröder W, Klostermann A, Distl O. 2011. Candidate genes for physical performance in the horse. *Vet. J.* 190:39–48
48. Gu J, Orr N, Park SD, Katz LM, Sulimova G, et al. 2009. A genome scan for positive selection in Thoroughbred horses. *PLOS ONE* 4:e5767
49. Hill EW, Gu J, Eivers SS, Fonseca RG, McGivney BA, et al. 2010. A sequence polymorphism in *MSTN* predicts sprinting ability and racing stamina in Thoroughbred horses. *PLOS ONE* 5:e8645
50. Binns MM, Boehler DA, Lambert DH. 2010. Identification of the myostatin locus (MSTN) as having a major effect on optimum racing distance in the Thoroughbred horse in the USA. *Anim. Genet.* 41(Suppl. 2):154–58

Downloaded from www.AnnualReviews.org
Guest (guest)
IP: 18.220.137.164
On: Fri, 03 May 2024 20:09:24

51. Tozaki T, Miyake T, Kakoi H, Gawahara H, Sugita S, et al. 2010. A genome-wide association study for racing performances in Thoroughbreds clarifies a candidate region near the *MSTN* gene. *Anim. Genet.* 41(Suppl. 2):28–35
52. Shin DH, Lee JW, Park JE, Choi IY, Oh HS, et al. 2015. Multiple genes related to muscle identified through a joint analysis of a two-stage genome-wide association study for racing performance of 1,156 Thoroughbreds. *Asian-Aust. J. Anim. Sci.* 28:771–81
53. Farries G, McGettigan PA, Gough KF, McGivney BA, MacHugh DE, et al. 2018. Genetic contributions to precocity traits in racing Thoroughbreds. *Anim. Genet.* 49:193–204
54. Farries G, Gough KF, Parnell AC, McGivney BA, McGivney CL, et al. 2019. Analysis of genetic variation contributing to measured speed in Thoroughbreds identifies genomic regions involved in the transcriptional response to exercise. *Anim. Genet.* 50:670–85
55. McGivney BA, Hernandez B, Katz LM, MacHugh DE, McGovern SP, et al. 2019. A genomic prediction model for racecourse starts in the Thoroughbred horse. *Anim. Genet.* 50:347–57
56. Velie BD, Hamilton NA, Wade CM. 2016. Heritability of racing durability traits in the Australian and Hong Kong Thoroughbred racing populations. *Equine Vet. J.* 48:275–79
57. Han H, McGivney BA, Farries G, Katz LM, MacHugh DE, et al. 2020. Selection in Australian Thoroughbred horses acts on a locus associated with early two-year old speed. *PLOS ONE* 15:e0227212
58. Hill EW, McGivney BA, Gu J, Whiston R, MacHugh DE. 2010. A genome-wide SNP-association study confirms a sequence variant (g.66493737C>T) in the equine myostatin (*MSTN*) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses. *BMC Genom.* 11:552
59. Petersen JL, Valberg SJ, Mickelson JR, McCue ME. 2014. Haplotype diversity in the equine *myostatin* gene with focus on variants associated with race distance propensity and muscle fiber type proportions. *Anim. Genet.* 45:827–35
60. Santagostino M, Khoriauli L, Gamba R, Bonuglia M, Klipstein O, et al. 2015. Genome-wide evolutionary and functional analysis of the Equine Repetitive Element 1: An insertion in the myostatin promoter affects gene expression. *BMC Genet.* 16:126
61. Rooney MF, Hill EW, Kelly VP, Porter RK. 2018. The "speed gene" effect of myostatin arises in Thoroughbred horses due to a promoter proximal SINE insertion. *PLOS ONE* 13:e0205664
62. Bower MA, McGivney BA, Campana MG, Gu J, Andersson LS, et al. 2012. The genetic origin and history of speed in the Thoroughbred racehorse. *Nat. Commun.* 3:643
63. Gu J, MacHugh DE, McGivney BA, Park SD, Katz LM, Hill EW. 2010. Association of sequence variants in CKM (creatine kinase, muscle) and COX4I2 (cytochrome c oxidase, subunit 4, isoform 2) genes with racing performance in Thoroughbred horses. *Equine Vet. J. Suppl.* 569–75
64. Hill EW, Gu J, McGivney BA, MacHugh DE. 2010. Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. *Anim. Genet.* 41(Suppl. 2):56–63
65. Harrison SP, Turrion-Gomez JL. 2006. Mitochondrial DNA: an important female contribution to thoroughbred racehorse performance. *Mitochondrion* 6:53–63
66. Lin X, Zheng HX, Davie A, Zhou S, Wen L, et al. 2018. Association of low race performance with mtDNA haplogroup L3b of Australian thoroughbred horses. *Mitochondrial DNA A* 29:323–30
67. Denham J, McCluskey M, Denham MM, Sellami M, Davie AJ. 2020. Epigenetic control of exercise adaptations in the equine athlete: Current evidence and future directions. *Equine Vet. J.* 53:431–50
68. Farries G, Bryan K, McGivney CL, McGettigan PA, Gough KF, et al. 2019. Expression quantitative trait loci in equine skeletal muscle reveals heritable variation in metabolism and the training responsive transcriptome. *Front. Genet.* 10:1215
69. Bryan K, McGivney BA, Farries G, McGettigan PA, McGivney CL, et al. 2017. Equine skeletal muscle adaptations to exercise and training: evidence of differential regulation of autophagosomal and mitochondrial components. *BMC Genom.* 18:595
70. Love S, Wyse CA, Stirk AJ, Stear MJ, Calver P, Voute LC, Mellor DJ. 2006. Prevalence, heritability and significance of musculoskeletal conformational traits in Thoroughbred yearlings. *Equine Vet. J.* 38:597–603

Downloaded from www.AnnualReviews.org
Guest (guest)
IP: 18.22[illegible]
On: Fri, 03 May 2024 20:09:24

71. Santschi EM, White BJ, Peterson ES, Gotchey MH, Morgan JM, Leibsle SR. 2017. Forelimb conformation, sales results, and lifetime racing performance of 2-year-old Thoroughbred racing prospects sold at auction. *J. Equine Vet. Sci.* 53:74–80
72. Jeffcott LB, Rossdale PD, Freestone J, Frank CJ, Towers-Clark PF. 1982. An assessment of wastage in Thoroughbred racing from conception to 4 years of age. *Equine Vet. J.* 14:185–98
73. Wilsher S, Allen WR, Wood JL. 2006. Factors associated with failure of Thoroughbred horses to train and race. *Equine Vet. J.* 38:113–18
74. Oki H, Miyake T, Kasashima Y, Sasaki Y. 2008. Estimation of heritability for superficial digital flexor tendon injury by Gibbs sampling in the Thoroughbred racehorse. *J. Anim. Breed. Genet.* 125:413–16
75. Welsh CE, Lewis TW, Blott SC, Mellor DJ, Stirk AJ, Parkin TD. 2014. Estimates of genetic parameters of distal limb fracture and superficial digital flexor tendon injury in UK Thoroughbred racehorses. *Vet. J.* 200:253–56
76. Blott SC, Swinburne JE, Sibbons C, Fox-Clipsham LY, Helwegen M, et al. 2014. A genome-wide association study demonstrates significant genetic variation for fracture risk in Thoroughbred racehorses. *BMC Genom.* 15:147
77. Tozaki T, Kusano K, Ishikawa Y, Kushiro A, Nomura M, et al. 2020. A candidate-SNP retrospective cohort study for fracture risk in Japanese Thoroughbred racehorses. *Anim. Genet.* 51:43–50
78. Dupuis MC, Zhang Z, Druet T, Denoix JM, Charlier C, et al. 2011. Results of a haplotype-based GWAS for recurrent laryngeal neuropathy in the horse. *Mamm. Genome* 22:613–20
79. Boyko AR, Brooks SA, Behan-Braman A, Castelhano M, Corey E, et al. 2014. Genomic analysis establishes correlation between growth and laryngeal neuropathy in Thoroughbreds. *BMC Genom.* 15:259
80. Morley PS, Bromberek JL, Saulez MN, Hinchcliff KW, Guthrie AJ. 2015. Exercise-induced pulmonary haemorrhage impairs racing performance in Thoroughbred racehorses. *Equine Vet. J.* 47:358–65
81. Velie BD, Raadsma HW, Wade CM, Knight PK, Hamilton NA. 2014. Heritability of epistaxis in the Australian Thoroughbred racehorse population. *Vet. J.* 202:274–78

Downloaded from www.AnnualReviews.org
Guest (guest)
IP: 18.220.137.164
On: Fri, 03 May 2024 20:09:24