

GENETIC RELATIONSHIPS AND RACING SUCCESS: A STUDY ON SIBLING PERFORMANCE IN THOROUGHBRED HORSES

¹ *Gene-Use.co AS, Oslo, Norway, geneuse.ai*

ABSTRACT

Thoroughbred racing performance is influenced by a complex interplay of polygenic genetic effects and environmental factors. Contemporary pedigree- and genome-based studies indicate only moderate additive genetic control, with heritability (h^2) for time-based traits typically between 0.07 and 0.25 and rising to about 0.60 only for best-distance preference. Against this background, we quantified the additional contribution of close pedigree relationships by analysing race records from more than 60 000 Thoroughbreds. Performance similarity of full- and three-quarter siblings was compared with that of randomly paired contemporaries using two non-parametric tests: the Mann–Whitney U test for ranking differences at two years of age and the Kolmogorov–Smirnov test for entire performance distributions across ages and ability thresholds. Sibling pairs showed significantly smaller rank distances and more similar distributional profiles than random pairs, with the effect strongest among elite performers. These results demonstrate that, even on top of the modest overall heritability, close genetic relatedness measurably increases similarity in racing outcomes. The findings refine expectations for selective breeding by highlighting that genetic gains are real but incremental, and they underline the continuing importance of management and training in maximising racehorse performance.

1. INTRODUCTION

Thoroughbred horse racing success reflects both genetics and environment, but modern quantitative studies agree that the genetic component is moderate rather than overwhelming. In the British population, whole-genome relationship models place h^2 for sprint, middle-distance and long-distance speed at 0.124, 0.122 and 0.074, respectively. Similarly, in the extensive Australian dataset analysed by Velie et al. (2015), log cumulative earnings and earnings per start showed heritabilities of 0.19–0.23, while the more specific trait “best race distance” reached 0.61 ± 0.03 . Studies from Japan corroborate the picture: Oki et al. (1995) reported h^2 values that decline from ≈ 0.25 for short-course times to ≈ 0.08 for staying races, and Tozaki et al. (2012) found $h^2 \approx 0.11$ –0.25 for earnings-based traits depending on model specification. A recent review synthesising genomic work emphasises a highly polygenic architecture, with thousands of loci of small effect and limited scope for rapid genetic improvement (Bailey et al., 2022). Whole-genome sequencing of 185 North-American thoroughbreds spanning five generations identified more than 14 million segregating autosomal SNPs and extensive runs of homozygosity, further confirming that each variant contributes only a minute fraction of the phenotypic variance (Bailey et al., 2024).

Environmental factors (track surface, jockey skill, training regime, season and horse age) contribute at least as much to race outcome variance as genetics and often interact with genetic predispositions. Nevertheless, the consistency of moderate heritability estimates across continents confirms that genetic variation does influence performance. While many studies estimate population-wide heritability, the incremental effect of close pedigree relationships (e.g. full or $\frac{3}{4}$ siblings) on performance similarity has received less quantitative attention. Because full siblings share, on average, 50 % of their segregating alleles and $\frac{3}{4}$ siblings share 37.5 %, comparing their race records with those of unrelated contemporaries provides a natural experiment to probe the impact of shared genetics under real racing conditions.

Here, we analyse more than 60 000 racehorses to test whether full- and three-quarter siblings perform more similarly than randomly paired horses from the same cohorts. We apply Mann–Whitney U test (Mann & Whitney, 1947) to rank data up to two years of age, and Kolmogorov–Smirnov test (Kolmogorov, 1933; Smirnov, 1948) to whole-distribution data across all ages and four ability thresholds.

By integrating these non-parametric approaches with a large, modern dataset, we provide a refined estimate of how much additional performance similarity can be attributed to close genetic relatedness beyond the modest baseline heritability reported for the breed.

2. METHODS

This study examines the performance of full siblings and 3/4 siblings in comparison to random pairs of Thoroughbred horses using a dataset of over 60,000 records from Australia and New Zealand. The focus is on whether genetic relationships lead to more similar racing outcomes. A control group of randomly selected horses was used as a benchmark for comparison, ensuring consistent race conditions and age distributions across the study groups.

2.1 Sample Definitions:

The terms "Full siblings" refer to horses sharing both parents, while "3/4 siblings" indicate horses sharing one parent and one grandparent. Figure 1 illustrates the sibling definitions applied in the study. The inclusion of 3/4 siblings allows for a more nuanced understanding of how partial genetic relationships influence racing performance.

Performance was assessed under four rating thresholds:

- No threshold: All subjects included, regardless performance
- Good: only top 50% subjects included
- Excellent: only top 20% subjects included
- Exceptional: only top 5% subjects included

These rating categories were selected to distinguish varying levels of performance across the sibling groups and random pairs.

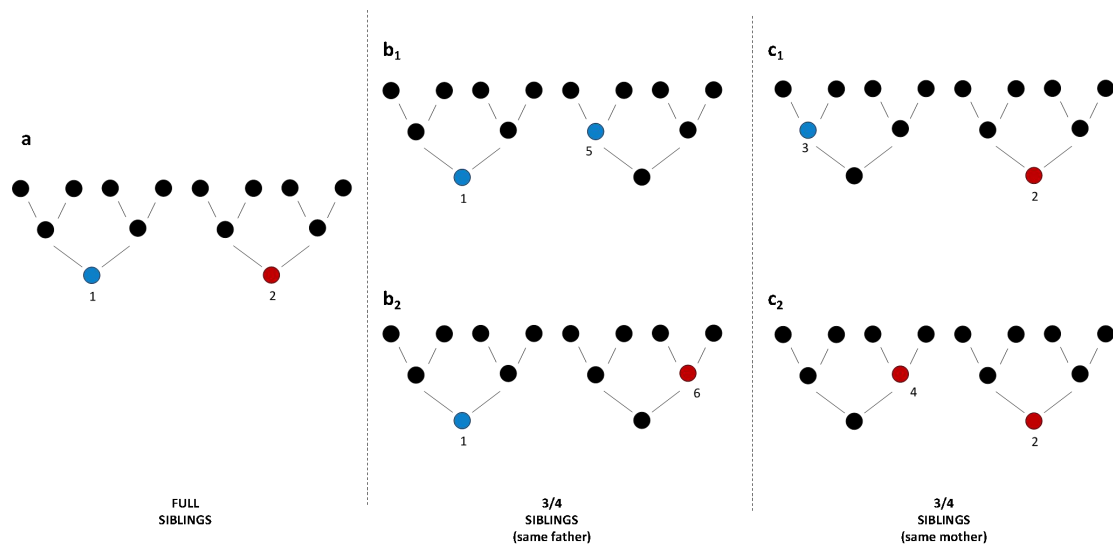


Figure 1

2.2 Mann-Whitney U Test

The Mann-Whitney U test, also known as the Wilcoxon rank-sum test, was employed to compare the performance distributions between two independent groups: full siblings and random horses, as well as 3/4 siblings and random horses. The Mann-Whitney U test ranks all observations together and calculates the U statistic based on the sum of ranks for each group. In this study, n_1 represents the number of sibling pairs analyzed and n_2 represents the number of random pairs. The U statistic is the smaller of the two possible U values, and a significant difference between the expected and obtained U values indicates that genetic factors (e.g., shared parentage or grandparentage) may be influencing performance. The test is particularly suitable for non-parametric data, allowing for comparisons of medians and performance rankings without assuming normal distribution.

2.2 Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov test was also utilized in this study to assess performance differences between full siblings and random horses across multiple age groups. This test compares the empirical distribution functions (CDFs) of two samples and calculates the largest absolute difference between the CDFs, D value. Larger D values indicate greater differences between the two distributions.

The K-S test was applied to explore how performance similarities between siblings change as horses age, providing insight into the longevity and persistence of genetic effects. In this context, it was particularly useful for detecting both shifts in the central tendency (e.g., mean performance) and differences in the overall distribution of performance across age groups. Statistical significance was determined by comparing the D statistic to critical values for the respective sample sizes, with corresponding p-values calculated to confirm whether the observed differences were due to random variation.

2.3 Complementary Insights from Both Tests

The combination of these two tests offers complementary perspectives on genetic influence. The Mann-Whitney U test highlights differences in median rankings, while the Kolmogorov-Smirnov test reveals broader distributional differences. Together, they provide a comprehensive view of the role that genetic relationships play in Thoroughbred racing performance. Figure 2 illustrates the sampling, test and comparison process.

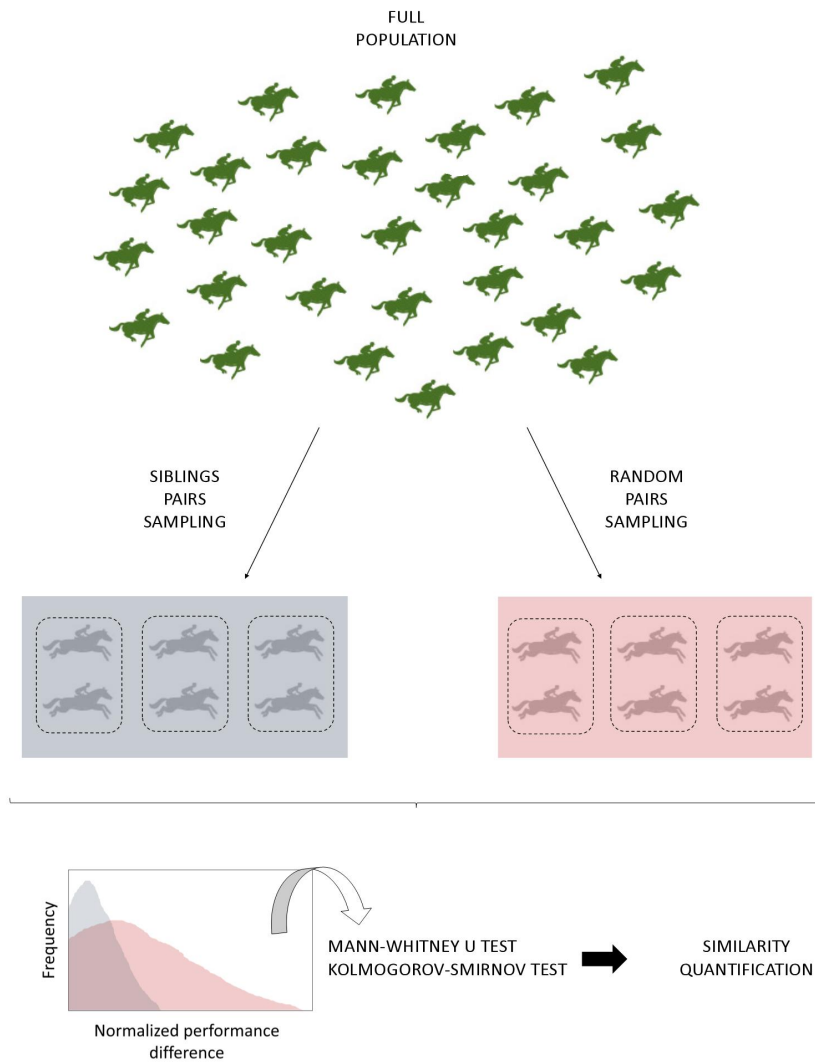


Figure 2

3. RESULTS & DISCUSSION

3.1 Mann-Whitney U test up to age 2

Full Siblings. The Mann-Whitney U test results showed that full siblings consistently displayed more similar performance outcomes compared to random horse pairs when comparing across all performance groups (see column a in Figure 3 left).

3/4 Siblings. Similar trends were observed in 3/4 siblings when analyzing all performances, with full siblings showing stronger performance similarity (see Figure 3 left, where all relationships show positive percentage differences with the random distribution). It is noteworthy that 3/4 siblings show close or even stronger similarities than full siblings (see

Figure 3 right). At higher performances, these observations could be mixed due to stronger inbreeding.

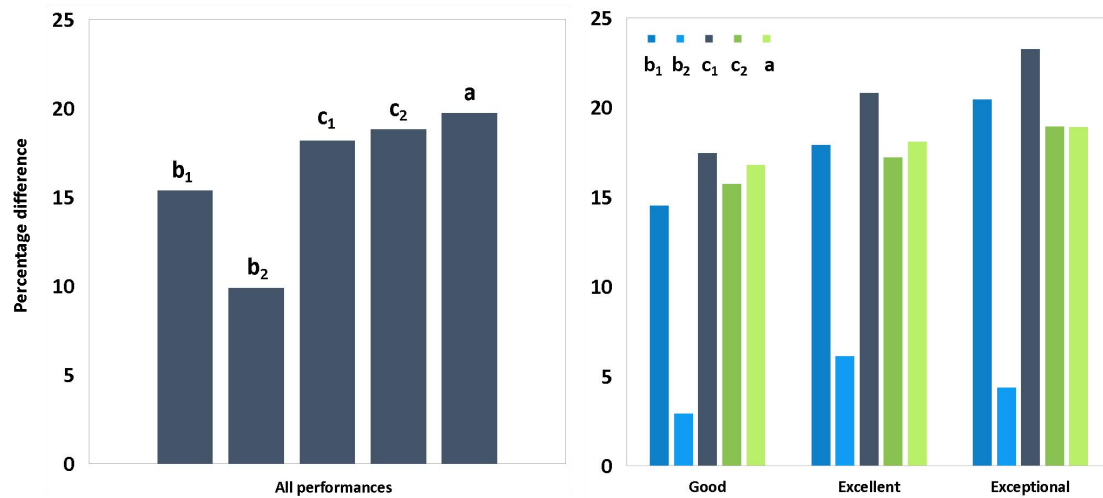


Figure 3

These results emphasize the importance of even partial genetic relationships, particularly among elite racehorses

3.2 Kolmogorov-Smirnov Tests Across (all ages)

We employed Kolmogorov-Smirnov test to further explore performance similarities between siblings across multiple age groups. The K-S statistic, which measures the largest difference between cumulative distribution functions, was consistently higher for full siblings across all age groups compared to random horses. Figure 4 provides a visual representation of the K-S test results across different age groups. The top graph shows the K-S statistic for each group, highlighting the magnitude of the genetic effect, while the bottom graph shows the corresponding p-values for group in relation to the significance threshold (0.05). High K-S values indicate larger similarities in performance between siblings, demonstrating a stronger genetic effect. Low p-values (below 0.05) indicate statistical significance, confirming that the observed performance differences between siblings and random horses are not due to chance. As shown in the top panel, the K-S values generally increase for all groups, suggesting that the genetic influence on performance similarities is more pronounced in horses sharing ancestors. In particular, the populations a, b₁, c₁ and c₂ are clearly more similar in performance than any random population. In the case of b₂, the sample is not statistically significant (as reflected by its large p-value). The results confirm that siblings exhibit greater similarity in performance distributions compared to random pairs, therefore indicating that purely genetic factors have a sustained impact on performance, even for mature horses.

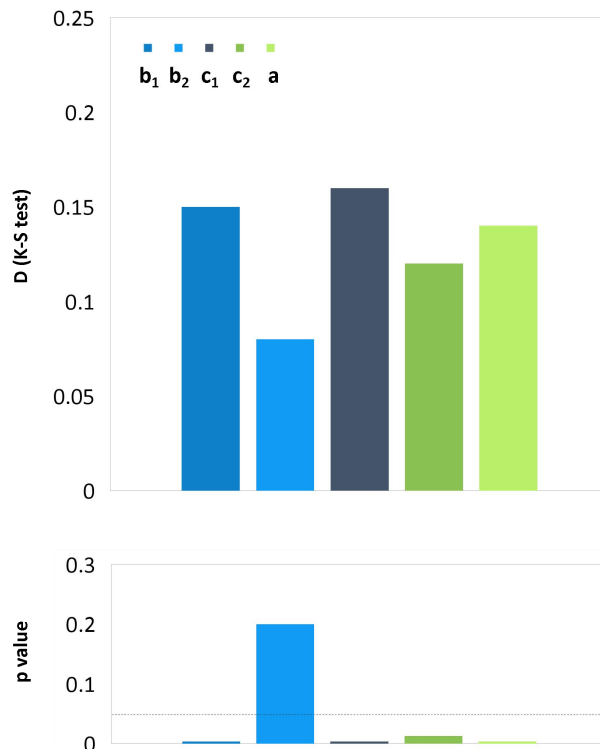


Figure 4

3.3 Analysis and comparison

Full siblings share approximately 50% of their genetic material from each parent, and this shared material can include genes that influence critical performance traits like speed, stamina, muscle composition, and overall health. Many of these traits are highly heritable; for instance, performance parameters such as speed, stamina, and best racing distance have been found to be highly heritable, with the best racing distance exhibiting heritability close to 100%. Specific genetic factors, such as myostatin variants, are strongly associated with racing distance, while other genomic regions show weak to modest associations with racing performance (Bailey et al., 2021). This indicates that multiple genetic factors contribute to Thoroughbred racing success, and genetic variation related to athleticism remains important. This strong genetic foundation reinforces the notion that genetic relationships play a significant role in Thoroughbred racing outcomes. If breeders had accurate estimates of these highly heritable performance parameters, they could more effectively breed horses for specific racing conditions and events (Williamson & Beilharz, 1998).

While genetics clearly exert a strong influence, it is important to recognize the contribution of environmental factors such as training, diet, race management, and other external influences.

These factors also play a critical role in race performance, meaning that a horse's overall racing ability is a product of both inherited genetic traits and environmental conditions.

The results from our analyses consistently showed that full siblings had significantly smaller performance differences (both absolute and normalized) compared to random pairs of horses. This pattern held true even for high-performing horses (from "good" to "exceptional"), further highlighting the role of shared genetics in performance outcomes. The differences between the expected and obtained U values were consistent across increasing performance thresholds, indicating that the observed genetic effects are not due to random chance, but rather represent a fundamental influence on performance traits.

The additional Kolmogorov-Smirnov tests, conducted across different age groups, provided further insights into the role of genetic relationships in performance. The K-S tests showed that genetic relationships have a sustained influence on performance outcomes, even as horses mature. The consistently low p-values across groups confirmed the statistical significance of these findings.

For 3/4 siblings, while the genetic influence was less pronounced compared to full siblings, they still demonstrated a stronger similarity in performance than random pairs. This was particularly evident in elite horses, suggesting that even partial genetic relationships can affect performance, especially among the top performers.

3.4 Broader implications for breeding programs

These results provide important insights for breeding strategies in the Thoroughbred industry. The findings demonstrate that selective breeding, which emphasizes the transmission of advantageous genetic traits, can significantly improve racing outcomes. Full sibling analysis suggests that breeders should prioritize pairing horses with known pedigrees that have a track record of producing high performers. The impact of partial genetic relationships also suggests that breeders should consider 3/4 siblinghood when designing mating strategies, particularly for elite-level races. Ultimately, this study highlights the importance of incorporating genetic data into breeding and training programs, ensuring that thoroughbreds are not only bred for performance but also for long-term genetic sustainability.

4. CONCLUSIONS

This study provides evidence that genetic relationships, particularly shared parentage, significantly impact thoroughbred racing performance. The Mann-Whitney U and Kolmogorov-Smirnov tests consistently demonstrated that full sibling pairs exhibit more similar race outcomes compared to randomly selected pairs, with this similarity becoming increasingly evident as performance thresholds rise. The results also suggest that partial

genetic relationships, as seen in 3/4 siblings, contribute to racing performance, particularly among elite horses. While full siblings displayed the highest degree of similarity, 3/4 siblings (both maternal and paternal) also showed significant performance similarities, particularly at higher performance levels. This finding indicates that genetic advantages can extend beyond full parentage and still provide performance benefits, especially in top-tier racing environments.

From a practical perspective, these insights offer valuable guidance for the design of selective breeding programs. Breeders are encouraged to prioritize not only full sibling relationships but also partial genetic ties when aiming to produce high-performing racehorses. By employing both full and partial sibling data, the thoroughbred industry can enhance the likelihood of breeding champions while maintaining genetic diversity for long-term success.

5. REFERENCES

1. Bailey, E., Reid, R., & Corbin, L. J. (2022). Genetics of Thoroughbred Racehorse Performance. *Annual Review of Animal Biosciences*, 9, 253–277. <https://doi.org/10.1146/annurev-animal-061220-023100>.
2. Bailey, E., Finno, C.J., Cullen, J.N., Kalbfleisch, T. & Petersen, J.L. (2024). Analyses of whole-genome sequences from 185 North American Thoroughbred horses spanning five generations. *Scientific Reports*, 14, 22930. <https://doi.org/10.1038/s41598-024-73645-9>
3. Langlois, B. (1980). Heritability of racing ability in thoroughbreds. *Livestock Production Science*, 7(6), 591–605. [https://doi.org/10.1016/0301-6226\(80\)90030-8](https://doi.org/10.1016/0301-6226(80)90030-8)
4. Mota, M., Oliveira, H. N., & da Silva, R. G. (1998). Genetic and environmental factors that affect the best time of Thoroughbred horses in Brazil. *Journal of Animal Breeding and Genetics*, 115(2), 123–129. <https://doi.org/10.1111/j.1439-0388.1998.tb00333.x>
5. Oki, H., Sasaki, Y., & Willham, R. L. (1994). Genetic and environmental factors that affect the best time of Thoroughbred horses in Japan. *Journal of Animal Breeding and Genetics*, 111(2–3), 128–137. <https://doi.org/10.1111/j.1439-0388.1994.tb00464.x>
6. Oki, H., Sasaki, Y., & Willham, R. L. (1995). Genetic parameter estimates for racing time by restricted maximum likelihood in the Thoroughbred horse of Japan. *Journal of Animal Breeding and Genetics*, 112(2), 146–150. <https://doi.org/10.1111/j.1439-0388.1995.tb00548.x>
7. Tozaki, T., Miyake, T., Kakoi, H., Gawahara, H., Sugita, S., Hasegawa, T., & Mukoyama, H. (2012). Heritability estimates for racing performance in Japanese Thoroughbred racehorses using linear and non-linear model analyses. *Journal of Animal*

- Breeding and Genetics, 129(5), 347–355. <https://doi.org/10.1111/j.1439-0388.2011.00982.x>
8. Velie, B. D., Hamilton, N. A., & Wade, C. M. (2015). Heritability of racing performance in the Australian Thoroughbred racing population. *Animal Genetics*, 46(2), 206–209. <https://doi.org/10.1111/age.12234>
 9. Williamson, S. A., & Beilharz, R. G. (1996). Heritabilities of racing performance in Thoroughbreds: a study of Australian data. *Journal of Animal Breeding and Genetics*, 113(6), 531–541. <https://doi.org/10.1111/j.1439-0388.1996.tb00667.x>
 10. Williamson, S. A., & Beilharz, R. G. (1998). The inheritance of speed, stamina and other racing performance characters in the Australian Thoroughbred. *Journal of Animal Breeding and Genetics*, 115(1–6), 1–16. <https://doi.org/10.1111/j.1439-0388.1998.tb00301.x>.
 11. Mann, H.B., & Whitney, D.R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 18(1), 50-60. <https://doi.org/10.1214/aoms/1177730491>
 12. Kolmogorov, A.N. (1933). Sulla Determinazione Empirica di una Legge di Distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, 4, 83-91.
 13. Smirnov, N.V. (1948). Table for Estimating the Goodness of Fit of Empirical Distributions. *The Annals of Mathematical Statistics*, 19(2), 279-281. <https://doi.org/10.1214/aoms/1177730256>.