Harari, S., Théry, A., Fortier, G., Ricard, A. (2026) Distance and type of race: different traits for racing performance in Thoroughbred horses?, *Proceedings 13th World Congress on Genetics Applied to Livestock Production* 23400: https://doi.org/10.31274/wcgalp.23400
©2026 Harari, et al. CC BY-NC-ND

Distance and type of race: different traits for racing performance in Thoroughbred horses?

S. Harari[1,2*], A. Théry[3], G. Fortier[2,4], and A. Ricard[1,5]

[1]Université Paris Saclay, INRAE, AgroParisTech, 78352 Jouy-en-Josas, France ; [2]UNICAEN, 14280 Saint-Contest, France ; [3]France Galop, Paris, France ; [4]LABÉO, 14280 Saint-Contest, France ; [5]Institut Français du Cheval et de l'Equitation, 61310 Exmes, France ;
[*]Corresponding author: suzanne.harari@inrae.fr

Introduction

In this study a model was developed to disentangle trainer effects from genetic value of the horse. Then, career diversity was genetically characterized by distinguishing participation status from earnings and by partitioning performance across racing conditions. This allowed identification of genetic influences on career trajectories and their implications for breeding, training decisions, and career management. Finally, the impact of the inbreeding coefficient was assessed on both participation status and earnings.

Materials & Methods

Performance. Official racing records on flat races in France from 2000 to 2024 were provided by France Galop. The data included 1,081,036 performances achieved by 93,609 horses. The information for each horse in each race included the name, age, sex, country and year of birth, jockey, trainer, stall number, final ranking and earning. There were 4,200 different jockeys and 3,532 different trainers. The data included 101,882 flat races and information regarding its category, track conditions both for dirt and turf track, distance, racecourse and date, prize money and racing time of the winner. The pedigree was provided by the Institut Français du Cheval et de l'Equitation on behalf of France Galop.

Horses. Horses were aged 2 to 14 years (12% aged 2, 35% 3, 21% 4, 13% 5, and 19% 6 or older) and were 21% stallions, 44% females and 35% geldings. They were born in France (74%) and abroad (26%). They were bred from 2,902 sires and 40,305 dams. Inbreeding coefficient was calculated using this whole pedigree (n=180,207 horses, 8.2 generations on average) and divided into 8 classes (0-0.05: 3% of population, 0.05-0.01: 15%, 0.01-0.015: 20%, 0.015-0.020: 19%, 0.020-0.025: 15%, 0.025-0.035: 17%, 0.035-0.05: 8%, >0.05: 3%).

Analysis 1. First aim was to characterize trainer effect to include it correctly in final model. First model (M1) was the following single trait animal mixed model:

$$\mathbf{y_1} = \mathbf{X_1 b_1} + \mathbf{Z_1 u_1} + \mathbf{Z_1 p_1} + \mathbf{T_1 t_1} + \mathbf{e_1} \quad (1)$$

where $\mathbf{y_1}$ is vector of the logarithm of the earning (LE) in each race, real or simulated by halving the amount for each successive rank for the non-winning horses, $\mathbf{b_1}$ is the vector of fixed effects including age (2, 3, 4, 5, 6 years and more), the 144 combinations between sex (stallion, female, gelding), year of birth (from 1998 to 2021 by 1 year, 2022 and 2021 grouped) and country of birth (France, abroad), inbreeding coefficient (8 levels) and stall number (18 levels), $\mathbf{u_1}$ is the vector of random animal additive genetic value with $\mathbf{V(u_1)} = \mathbf{A}\sigma^2_{u1}$, where $\mathbf{A}$ is the relationship matrix, $\mathbf{p}$ is the vector of permanent environment random effect with $\mathbf{V(p_1)} = \mathbf{I}\sigma^2_{p1}$, $\mathbf{t_1}$ is the vector of the trainer random effect with $\mathbf{V(t_1)} = \mathbf{I}\sigma^2_{t1}$, $\mathbf{e_1}$ is the vector of residuals with $\mathbf{V(e_1)} = \mathbf{I}\sigma^2_{e1}$ and $\mathbf{X_1}$, $\mathbf{Z_1}$, and $\mathbf{T_1}$ are design matrices. The second model (M2) was similar to the first one, without the trainer effect. The third model (M3) was:

$$\hat{\mathbf{t}} = \mathbf{Va} + \mathbf{Wq} + \boldsymbol{\epsilon} \quad (3)$$

where $\hat{\mathbf{t}}$ is the vector of the predicted trainer effect from M1 repeated as often as in $\mathbf{y_1}$ ($\hat{\mathbf{t}} = \mathbf{T_1}\widehat{\mathbf{t_1}}$), $\mathbf{a}$ is the vector of random animal additive genetic value with $\mathbf{V(a)} = \mathbf{A}\sigma_a^2$, $\mathbf{q}$ is the vector of permanent environmental effect with $\mathbf{V(q)} = \mathbf{I}\sigma_q^2$, $\boldsymbol{\epsilon}$ is the vector of residuals with $\mathbf{V}(\boldsymbol{\epsilon}) = \mathbf{I}\sigma_\epsilon^2$ and **V** and **W** are design matrices. The fourth model (M4) was a bivariate animal model with the first trait as in M1 and a second trait, named the trainer performance. This trait is measured for each participation of the horse depending on its trainer in that race. It is the mean per trainer of the average LE per horse when trained by the trainer, minus the average LE of the horse. The model of this trait was:

$$\mathbf{y_2} = \mathbf{X_2 b_2} + \mathbf{Z_2 u_2} + \mathbf{Z_2 p_2} + \mathbf{e_2} \quad (4)$$

where $\mathbf{y_2}$is the vector of the trainer phenotype, $\mathbf{b_2}$ the vector of fixed effects including the first year of participation of the trainer (from 2000 to 2022, 2023 and 2024 were grouped). The variance covariances were defined as: $\mathbf{V(t)} = \mathbf{I}\sigma_t^2$ (first trait only), $\mathbf{V(p)} = \mathbf{I} \otimes \mathbf{P}$, $\mathbf{V(u)} = \mathbf{A} \otimes \mathbf{G}$ and $\mathbf{V(e)} = \mathbf{I} \otimes \mathbf{R}$, with $\otimes$ the direct product, **P** the 2 x 2 matrix of permanent environment effect variance-covariance, **G** the 2 x 2 matrix of genetic variance-covariance and **R** the 2 x 2 matrix of residual variance-covariance. For all models, variance parameters were estimated using Restricted Maximum Likelihood (REML) with ASReml software (Gilmour *et al.*, 2015).

Analysis 2. Second aim was the characterisation of diversity of careers. A multivariate mixed model was used to analyse performance broken down according to race category (stakes, conditions, claiming and handicap), distance (short<1,600 m, medium, long>2,400 m), track's type (turf and dirt) and condition for turf (very firm to good, good to soft, soft and very soft to very heavy). For each type of race, four performance traits were used:

- Trait A: the horse's start status over the entire career, measured once (binary variable). This trait was considered with the threshold model,
- Trait B: LE for each horse and each race,
- Trait C: the trainer's horses' average start status minus own start status of the horse. The trainer was the first trainer of the horse or the most common throughout the horse's career if never raced in that race type of race,
- Trait D: the trainer's horses' average LE minus own average LE of the horse.

Multivariate animal models were used with the 4*N traits defined, with N the number of types of race: 4 for categories, 3 for distances, 2 for track's type, 4 for track's condition. Fixed effects for traits A and B were similar to those of M1, except age and stall number for A. For traits C and D, fixed effects included only the first year of career of the trainer. A trainer random effect was included for traits A and B, and a permanent environmental effect was included for each trait, even for traits measured once, to estimate residual covariance between the start status traits and the earnings traits. With this additional effect, residual covariances were fixed to 0. Variance parameters and effects were estimated by Markov Chain Monte Carlo Gibbs sampling to consider the binary traits with the threshold model. We used the TM software (Legarra *et al*., 2011). The number of iterations was set to 50,000 and the number of burn-in iterations to 30,000. Samples were thinned and pooled every 20 iterations.

Results & Discussion

How to manage trainer influence. All standard errors (se) were lower than 0.01. Heritability estimates were 0.06 and 0.15 respectively with (M1) and without (M2) trainer effect. In M1, trainer accounted for 0.13 of total variance. Sharman and Wilson (2023) and Velie *et al.* (2015) found a significant random effect of trainer, respectively on repeated speed (0.12 to 0.16 of total variance) and on various non-repeated performance traits. Predicted trainer effect from M1 - $\hat{\mathbf{t}}$ - was very heritable: 0.47 (M3). These results confirm the confusion between trainer and genetic effects as trainer performance is partly explained by genetics, and show the

relevance of creating a bivariate model (M4). Heritability estimate and trainer proportion for first trait of M4 (horse performance) were respectively 0.13 and 0.07. By simultaneously estimating the genetic value of the horse's performance and that of its trainer, it is possible to disentangle the respective contributions of each to the performance. As a result, similar multivariate models were used for subsequent analyses.

Diversity of careers. The results for variance components are in Table 1 (se<0.01 for all parameters). Regardless of the type of race, start status was always more heritable than LE. For LE, heritability estimates were moderate but as it is a repeated trait, reliability from own performance can reach from 0.13 to 0.31 after 5 races (LE, stakes races). Trainer effect was approximately as important as genetic effect for LE, but this varied for start status depending on the type of race. Mota *et al*. (2005) found decreasing heritability for racing time with increasing race distance, as did Sharman and Wilson (2023) for repeated racing speed with low to moderate estimates (from 0.12 to 0.07) and Sobczyńska (2006) for rank at finish (from 0.16 to 0.04), while we found stable estimates for LE in the three distances.

Table 1: Heritability, repeatability and proportion of trainer effect for horses' traits.

	Source	Race category[1]	Distance[1]	Track[1]	Track condition[1]
A Start status	h^2	0.29,0.33,0.37,0.38	0.33,0.38,0.32	0.18,0.26	0.13,0.11,0.12,0.12
	Trainer	0.11,0.16,0.19,0.02	0.25,0.27,0.37	0.19,0.20	0.33,0.34,0.37,0.39
B LE	h^2	0.13,0.15,0.09,0.07	0.11,0.11,0.13	0.13,0.12	0.09,0.09,0.09,0.10
	r	0.29,0.34,0.24,0.16	0.20,0.20,0.21	0.22,0.22	0.21,0.20,0.21,0.21
	Trainer	0.01,0.04,0.08,0.03	0.08,0.08,0.08	0.05,0.06	0.08,0.09,0.09,0.08

[1] In order: Race category: stakes, conditions, claiming and handicap; Distance: short, medium, long; Track: turf, dirt; Track condition: very firm to good, good to soft, soft, very soft to very heavy

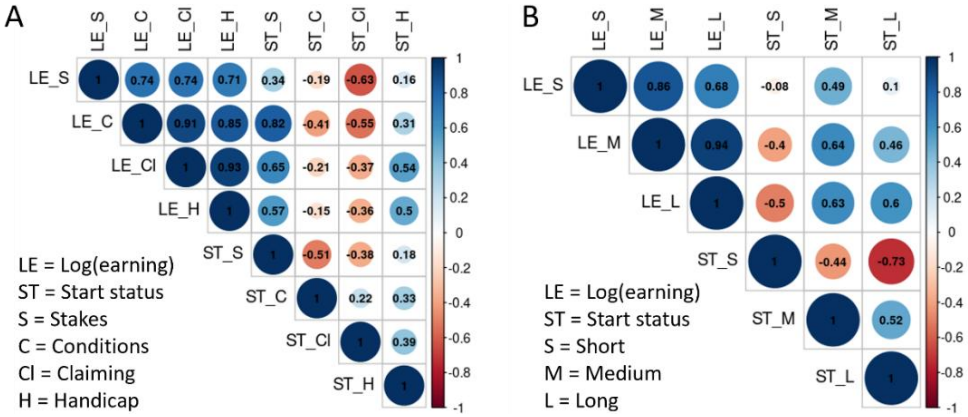

Figure 1. Genetic correlations (se<0.05) of horses' traits across category (A) and distance (B).

Analyses both across race category and distance show similar results regarding genetic correlations (Fig. 1). There were negative correlations for start status; between stakes and both conditions and claiming and between short and both medium and long-distance. Sharman and Wilson (2023) found moderate genetic correlation between short and long-distance speeds (0.47), as did Sobczyńska (2006) between 1,000 and above 1,800m races finish rank (0.54). These correlations across distance are consistent with the myostatin encoding gene (MSTN) found by Hill *et al.* (2010), involved in optimal racing distance. Positive and high (above 0.68) genetic correlations for LE can thus be surprising, but highlight genetic predisposition for start while showing that successes in different types of races have common genetic foundations

(same observations for race categories). Results showed genetics correlations of 0.19 and 0.87, respectively for start status and LE, between turf and dirt tracks while Oki *et al.* (1997) who studied racing time found 0.31 to 0.69, depending on distance. In addition, genetic correlations for different track conditions were positive for start status (0.64 to 0.80) and LE (0.84 to 0.97), which revealed that there is no genetic predisposition regarding participation or success on a given track condition. Careers trajectories are genetically segmented according to race distance and category, as there are predispositions to participate in a given type. Conversely, success is determined by the same genetic trait regardless of the race.

Impact of inbreeding. The effect of inbreeding on start status was increasing except for stakes races (Fig.2), although Hill *et al.* (2022) showed that a 10% increase in genomic inbreeding coefficient was associated with a 7% lower probability of ever racing – but we only studied horses that have raced at least once. For LE, inbreeding was not significant for the 93% of the population with coefficients between 0.5% and 5%. The effect was negative for horses with low coefficients (≤0.5%) or very high coefficients (>5%), but moderate (up to -10% of the phenotypic SD). Todd *et al.* (2018) found a negative relationship between inbreeding coefficient for both cumulative earnings and earnings per start.

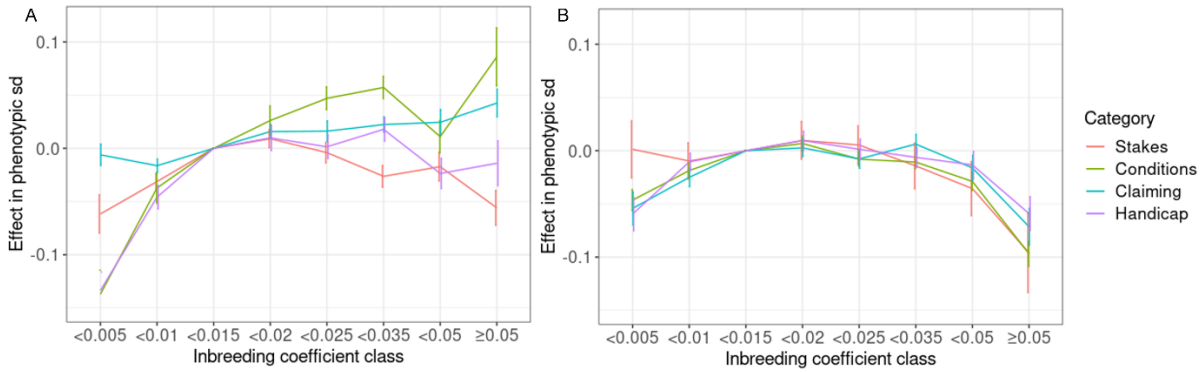

Figure 2. Effect of inbreeding coefficient on horses' traits: start status (A) and LE (B).

Conclusions

The study developed a statistical model capable of estimating genetic parameters without trainer bias by disentangling the genetic effect from the trainer effect. In addition, it showed genetic segmentation of careers based on race distance or category, but that monetary success is nearly the same genetic trait regardless of the type of race.

References

Gilmour A.R., Gogel B.J., Cullis B.R., *et al*. (2015) ASReml User Guide Release 4.1 Structural Specification, VSN International Ltd, Hemel Hempstead, HP1 1ES, UK www.vsni.co.uk

Hill E.W., Gu J., Eivers S.S., *et al*. (2010) PLoS ONE 5, e8645. https://doi.org/10.1371/journal.pone.0008645

Hill E.W., Stoffel M.A., McGivney B.A., *et al*. (2022) Proc. R. Soc. B. 289: 20220487. https://doi.org/10.1098/rspb.2022.0487

Legarra A., Varona, L., and Lopez de Maturana, E. (2011) TM Threshold model. Available at: https://alegarra.github.io/manualtm.pdf, visited on: 24-07-2025.

Mota M.D.S., Abrahão, A.R., and Oliveira, H.N. (2005) J Anim. Breed. Genet. 122:393–399. https://doi.org/10.1111/j.1439-0388.2005.00551.x

Oki H., Sasaki Y., and Willham R.L. (1997) J. Anim. Breed. Genet. 114:185–189. https://doi.org/10.1111/j.1439-0388.1997.tb00504.x

Sharman P., and Wilson A.J. (2023) Heredity 131:79–85. https://doi.org/10.1038/s41437-023-00623-8

Sobczyńska M. (2006) Czech J. Anim. Sci. 51:523–528. https://doi.org/10.17221/3973-CJAS

Todd E.T., Ho S.Y.W., Thomson P.C., *et al*. (2018) Sci Rep. 8:6167. https://doi.org/10.1038/s41598-018-24663-x

Velie B.D., Hamilton N.A., and Wade C.M. (2015) Anim. Genet. 46:23–29. https://doi.org/10.1111/age.12234