



ELSEVIER

Contents lists available at ScienceDirect

## Journal of Equine Veterinary Science

journal homepage: [www.j-evs.com](http://www.j-evs.com)

Original Research

## Genetic Parameters for Racing Performance of Thoroughbred Horses Using Bayesian Linear and Thurstonian Models



Manuela Pires Monteiro da Gama<sup>a,\*</sup>, Rusbel Raúl Aspilcueta Borquis<sup>b</sup>,  
Francisco Ribeiro de Araújo Neto<sup>c</sup>, Henrique Nunes de Oliveira<sup>b</sup>,  
Gisele Machado Fernandes<sup>a</sup>, Marcílio Dias Silveira da Mota<sup>b,1</sup>

<sup>a</sup> Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brazil

<sup>b</sup> Departamento de Zootecnia, Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista "Julio de Mesquita Filho", Campus de Jaboticabal, Jaboticabal, São Paulo, Brazil

<sup>c</sup> Departamento de Ciência Animal, Instituto Federal Goiano, Campus de Rio Verde, Rio Verde, Goiás, Brazil

## ARTICLE INFO

## Article history:

Received 24 November 2015

Received in revised form 17 March 2016

Accepted 22 March 2016

Available online 4 April 2016

## Keywords:

Equine

Heritability

Gibbs sampling

TMCLA

## ABSTRACT

The objective of this study was to estimate genetic parameters for rank in Thoroughbred horses using a Bayesian linear model (BLM) and a Thurstonian model (TM) to provide data that contribute to the selection and consequent genetic improvement of the breed in Brazil. Data were provided by the company Turf Total Ltda and consisted of 250,809 records for rank obtained from 40,300 horses and from 34,316 Thoroughbred races (distances of 1,000, 1,300, 1,600, and 2,000 m) that occurred between 1992 and 2011 on six tracks. The rank records at each distance were considered different traits and were submitted to single-trait analysis using BLM and TM. Fixed effects included sex, age, post-position, race, and level of difficulty. The heritability estimates for rank ranged from 0.228 to 0.032 when BLM was used and from 0.293 to 0.047 when TM was used. These estimates tended to decrease with increasing race distance in the two analyses. The TM estimated higher heritability for rank than BLM, indicating the possible use of this model in selection programs of Thoroughbred racehorses in Brazil.

Published by Elsevier Inc.

### 1. Introduction

In Brazil, the selection of Thoroughbred horses is based on the performance of the animal and of its parents. The pedigree plays an important role in the choice of parents of the next generation. Racing time, that is, the performance of the animal, is the main trait taken into consideration to evaluate and choose an animal as a reproducer.

Furthermore, the animal's performance is directly related to its earnings.

In addition to earnings and racing time, the final rank of the animal in a race is one of the most common performance measures of Thoroughbred racehorses. The use of rank traits may better adjust the information collected during Brazilian races because the ranks of all animals of the competition are recorded, whereas racing time and earnings are not always available for all horses competing in a race. The use of ranks permits to compare animals that participate in the same race with those of another race. However, in this case, the level of difficulty of a race should be determined and included in the analysis model [1].

Rank corresponds to differences in the level of the animals of that competition but does not evaluate velocity

\* Corresponding author at: Manuela Pires Monteiro da Gama, Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Avenida Bandeirantes, 3900, Monte Alegre, 14049-900, Ribeirão Preto, São Paulo, Brazil.

E-mail address: [manugama@gmail.com](mailto:manugama@gmail.com) (M.P.M.da. Gama).

<sup>1</sup> Marcílio Dias Silveira da Mota—In memoriam.

itself. This criterion eliminates some of the problems associated with the assessment of performance for different distances and track conditions, atypical inferior performance, and problems with abnormal performance criteria [2–8]. Rank is a good selection criterion for genetic improvement programs because it is heritable, according to genetic parameter estimates [9]. In two different studies, genetic parameter estimates for rank ranged 0.07–0.10 and 0.07–0.12, respectively [5,9]. These heritabilities were 0.18 in French Thoroughbreds and 0.06 in racehorses from the United Kingdom and Ireland [8], 0.14 in Thoroughbreds [10], 0.13 in Quarter horses [11], and 0.12 [5] and 0.25 in Arab horses [12]. In view of the above considerations, the aim of the present study was to the estimate genetic parameters for rank using a Bayesian linear model (BLM) and a Thurstonian model (TM) in Thoroughbred breeding programs in Brazil.

## 2. Material and Methods

### 2.1. Origin and Description of the Data

The database used was provided by Turf Total Ltda and consisted of 272,277 records for rank obtained from 40,300 different Thoroughbreds horses participating in races that occurred from 1992 to 2011 on tracks in Gávea (Rio de Janeiro State), Campos (Rio de Janeiro State), Cidade Jardim (São Paulo State), Tarumã (Paraná State), Cristal (Rio Grande do Sul State), and São Vicente (São Paulo State). The contemporary groups were defined using animals that participated in the same race (combination of day, hour, and city where the race occurred).

The database initially contained 30 different distances, which were restricted to four in the present study because they are the most traditional and most important distances in the country and therefore accounted for most observations in the data file. Table 1 shows the number of races, number of animals and observations per sex, and mean number of animals per race according to distance. There was a larger number of observations for 1,300-m races because many breeders decide whether their horses will participate in longer distance races based on the results of the animals obtained at this distance. The choice of distance and level of difficulty is made by the breeder based on the animal's performance at shorter distances and in previous races.

The pedigree file contained 58,694 animals, which were used in all analyses. As can be seen in Table 1, the mean number of animals per race was always about seven and the participation of females tended to decrease with increasing race distance and was only higher at 1,000 m.

For the genetic analysis of rank, races performed at different distances were considered different traits. This strategy was adopted because studies investigating other performance traits [13] demonstrated that races at different distances show genetic correlations less than unity, indicating that the addition of this effect as fixed effect alone would not be sufficient to describe all differences between animals. Thus, single-trait analysis was performed for each trait using BLM and TM.

### 2.2. Analysis of Rank Using a Bayesian Linear Model

The BLM model can be written in matrix form as:

$$y = X\beta + Za + Wc + \varepsilon \quad (1)$$

where  $y$ ,  $\beta$ ,  $a$ ,  $c$ , and  $\varepsilon$  are vectors of observations and fixed (sex, age, postposition, race, and level of difficulty), direct additive genetic, permanent environmental, and residual effects, respectively, and  $X$ ,  $Z$ , and  $W$  are incidence matrices that associate  $\beta$ ,  $a$ , and  $c$  with the observations. Uniform and Gaussian a priori distributions were assumed for the fixed and random effects, respectively:

$$\beta \propto \text{constant} \quad (2)$$

$$a | \sigma_a^2 \sim N[0, (A\sigma_a^2)] \quad (3)$$

$$c | \sigma_p^2 \sim N[0, (I_n\sigma_p^2)] \quad \varepsilon | \sigma_r^2 \sim N[0, (I_n\sigma_r^2)] \quad (4)$$

where  $A$  and  $I_n$  are relationship and identity matrices, respectively, and  $\sigma_a^2$ ,  $\sigma_p^2$ , and  $\sigma_r^2$  are the additive genetic, permanent environmental, and residual variance, respectively. A priori distributions derived from inverted chi-square distributions were used for the variance components. Thus:

$$\sigma_a^2 | S_a, v_a \sim \chi^{-2}[S_a v_a, v_a] \quad (5)$$

$$\sigma_p^2 | S_p, v_p \sim \chi^{-2}[S_p v_p, v_p] \quad (6)$$

$$\sigma_r^2 | S_r, v_r \sim \chi^{-2}[S_r v_r, v_r] \quad (7)$$

where  $S_a$  and  $v_a$ ,  $S_p$  and  $v_p$ , and  $S_r$  and  $v_r$  are a priori values and degrees of freedom for direct additive,

**Table 1**

Number of races, observations per sex, number of animals per sex, and mean number of animals per race according to distance.

Distance [m]	No. of Races	No. of Male Observations	No. of Males	No. of Female Observations	No. of Females	Mean Number of Animals/Race
1,000	6,793	23,915	8,007	28,847	9,375	7.76
1,300	15,282	63,270	14,610	47,606	12,692	7.25
1,600	9,133	46,523	10,339	20,002	6,627	7.28
2,000	3,108	14,051	4,640	7,540	2,677	6.95

permanent environmental, and residual variances, respectively.

### 2.3. Analysis of Rank Using a Thurstonian Model

The variance components were estimated based on the rank of the animals in each race using the TM first presented by Tavernier, although it is not explicitly called as Thurstonian [1,14]. The model included an underlying variable responsible for rank and can be written as follows:

$$l_{ik} = x'_{ik}\beta + z'_{ik}a + z'_{ik}p + w'_{ik}h + e_{ik}, \quad (8)$$

where  $i$  is the animal;  $\beta$  are fixed effects (sex, age, post-position, race, and level of difficulty);  $a$  is the vector of random additive genetic effects;  $p$  is the vector of random permanent environmental effects (shared by the same horse in different races);  $h$  is the vector of the random effect of race;  $e$  is the vector of residual effects, and  $x_{ik}$ ,  $z_{ik}$ , and  $w_{ik}$  are incidence matrices.

The distributions of the fixed effects of age and post-position according to distance are shown in Tables 2 and 3, respectively. Age was divided into 18 classes, with the first class including animals aged 3–4 years, the second class including animals aged 4–5 years, and so forth until the 18th class with animals aged 20 years or older.

The level of difficulty of a race depends on the profile of the competing animals and the amount of prize money paid. Faster horses, with previous performance already guaranteed, tend to attend more prestigious and high-prize money races. The profile of the animal, which interferes with its performance and hence the choice of race to be held, depends on the age of the competing animal, sex, number of previous victories, track where the race occurred, and type of race.

To include the level of difficulty of each race in the model used for the evaluation of rank, different weights were attributed to the races depending on this consideration and divided into 13 classes as shown in Table 4.

**Table 2**  
Distribution of age classes according to distance.

Age Class	1,000 m	1,300 m	1,600 m	2,000 m
1	34,402	53,169	32,130	10,687
2	11,599	28,102	17,073	6,595
3	1,663	17,423	9,827	2,783
4	1,301	8,097	4,594	966
5	429	2,619	1,616	365
6	114	764	498	93
7	37	199	170	42
8	3	33	30	2
9	0	11	2	2
10	0	12	7	0
11	11	18	11	2
12	6	18	15	5
13	3	19	23	5
14	2	13	20	6
15	3	25	6	9
16	5	16	6	6
17	12	5	2	0
18	22	22	21	14
Total	52,612	110,565	66,051	21,581

**Table 3**  
Distribution of postposition of the animals according to distance.

Postposition	1,000 m	1,300 m	1,600 m	2,000 m
1	5,906	12,990	7,361	2,618
2	5,889	13,079	7,590	2,623
3	5,922	13,042	7,518	2,548
4	5,888	13,016	7,595	2,552
5	5,848	12,949	7,537	2,527
6	5,725	12,416	7,118	2,302
7	5,246	10,831	6,127	1,905
8	4,161	8,479	4,940	1,416
9	3,097	5,725	3,554	967
10	2,048	3,656	2,405	658
11	1,330	2,254	1,601	473
12	790	1,214	1,000	346
13	439	595	598	222
14	205	242	362	137
15	71	59	220	93
16	25	16	137	73
17	14	1	82	56
18	5	0	59	42
19	2	1	25	11
20	1	0	22	12
Total	52,612	110,565	66,051	21,581

### 2.4. Gibbs Sampling and Post-Gibbs Analysis

All analyses were carried out using the TMCLA program [15,16], in which a Gibbs sampler was implemented for BLM and TM. A chain of 700,000 iterations was used for each analysis, with a burn-in period of 1,000 iterations and a thinning interval of 100 iterations. Convergence and the total number of effective samples were determined with the Gibanal program [17]. To establish an association between distances, Pearson's correlations were calculated between estimated breeding values (EBVs) obtained in the single-trait analyses for each distance (using either BLM or TM).

## 3. Results

Tables 5 and 6 show the estimates of variance components and heritability for rank obtained with the BLM and TM, respectively. The heritability estimates obtained with the two models showed a similar trend, with heritability

**Table 4**  
Distribution of the level of race difficulty according to distance (the lower the class, the lower the difficulty).

Difficulty Class	1,000 m	1,300 m	1,600 m	2,000 m
1	24,601	43,291	17,341	3,906
2	2,001	4,4044	1,303	119
3	8,881	17,423	13,107	3,546
4	1,084	2,815	1,477	478
5	4,269	13,973	7,989	2,834
6	428	1,173	415	107
7	3,031	9,340	5,430	2,015
8	257	878	497	32
9	5,303	6,376	9,149	3,349
10	1,468	2,157	4,072	1,371
11	553	167	2,414	1,085
12	192	0	851	1,141
13	544	0	2,006	1,598
Total	52,612	110,565	66,051	21,581

**Table 5**

Mean and standard deviations (between brackets) of posterior distributions of additive genetic ( $\sigma_a^2$ ), permanent environmental ( $\sigma_p^2$ ), and residual ( $\sigma_r^2$ ) variances and heritability ( $h^2$ ) for rank at the distances studied obtained with the Bayesian linear model.

Distance [m]	$\sigma_a^2$	$\sigma_p^2$	$\sigma_r^2$	$h^2$
1,000	1.835 [0.117]	1.568 [0.098]	4.651 [0.040]	0.228 [0.013]
1,300	0.765 [0.057]	0.676 [0.048]	5.191 [0.029]	0.115 [0.008]
1,600	0.394 [0.055]	0.984 [0.057]	5.694 [0.041]	0.056 [0.008]
2,000	0.221 [0.064]	1.032 [0.089]	5.688 [0.075]	0.032 [0.009]

decreasing with increasing racing distance. The additive genetic, permanent environmental, and residual and phenotypic variances were higher for BLM at all distances studied, although higher heritabilities were estimated with the TM. The heritability estimates obtained by the latter method were 28.5%, 27.8%, 46.4%, and 46.9% higher than those obtained by BLM at distances of 1,000, 1,300, 1,600, and 2,000 m, respectively.

Using BLM, there was an eightfold decrease in additive genetic variance from the distance of 1,000 m (1.835) to 2,000 m (0.221). This decrease was more pronounced when the additive genetic variance was estimated with the TM (11-fold, from 0.661 to 0.060). Similarly, comparing the distance of 1,000 and 2,000 m, there was a sevenfold decrease in heritability when the BLM was used (0.228 and 0.032, respectively), whereas heritability was six times lower when the TM was used (0.293 and 0.047). For both models, the permanent environmental variance was higher for the distance of 1,000 m and lower for the distance of 1,300 m.

The correlations between EBVs obtained in each analysis were of moderate to low magnitude, ranging from 0.50 to 0.08 for BLM and from 0.52 to 0.10 for TM (Table 7). The highest correlations were observed between races of 1,300 and 1,600 m and the lowest correlations between races of 1,000 and 2,000 m. The models used resulted in a decrease in the Pearson's correlation estimates with increasing difference between racing distances.

#### 4. Discussion

The heritabilities for rank estimated with the two models are higher than those reported in the literature. In a study involving Iranian Thoroughbred horses [18], the estimates ranged from 0.17 (1,000 m) to 0.11 (1,600 m). The heritability was 0.15 for the general database. In Thoroughbreds from Spain, heritability for rank ranged from

**Table 6**

Mean and standard deviations (between parentheses) of posterior distributions of additive genetic ( $\sigma_a^2$ ), permanent environmental ( $\sigma_p^2$ ), and residual ( $\sigma_r^2$ ) variances and heritability ( $h^2$ ) for rank at the distances studied obtained with the Thurstonian model.

Distance [m]	$\sigma_a^2$	$\sigma_p^2$	$\sigma_r^2$	$h^2$
1,000	0.661 [0.042]	0.592 [0.038]	1 [0]	0.293 [0.016]
1,300	0.205 [0.015]	0.188 [0.013]	1 [0]	0.147 [0.010]
1,600	0.110 [0.014]	0.221 [0.014]	1 [0]	0.082 [0.010]
2,000	0.060 [0.018]	0.224 [0.228]	1 [0]	0.047 [0.014]

**Table 7**

Pearson's correlations of estimated breeding values for rank obtained with the Bayesian linear model (above the diagonal) and the Thurstonian model (below the diagonal).

Trait	Distance [m]			
	1,000	1,300	1,600	2,000
1,000	—	0.48	0.39	0.08
1,300	0.47	—	0.50	0.15
1,600	0.39	0.52	—	0.30
2,000	0.10	0.18	0.30	—

0.10 to 0.07 according to the model used [9]. In another study [19], heritability ranged from 0.16 (1,000 m) to 0.06 (>1,800 m) for Polish Thoroughbreds and from 0.08 (1,800 m) to 0.02 (1,400 m) for Arab horses. For Quarter horses in Brazil, the estimated heritability for rank was 0.13 [11], a value similar to that reported for Arab horses (0.12) [5]. In another study on Arab horses, the estimated heritability was higher (0.25) [12]. Taken together, these results demonstrate a clear tendency of heritability for rank to decrease with increasing racing distance. The heritability estimates (obtained with both models) indicate that the shorter the distance, the greater are the differences that can be attributed to the average effect of genes and greater genetic gains can therefore be obtained by selection.

The Pearson's correlations between EBVs for different distances showed a trend similar to that observed for other performance traits when distances are considered in single-trait analyses [13,18,19]. The low magnitude of the estimated correlations demonstrates that the effect of genes acting together at different distances is small and that selection at any distance will not result in indirect genetic gains at other distances. Hill et al [20], studying a polymorphism in the MSTN gene, provided evidence that genetic differences in muscle development between animals influence animal performance at different distances. The use of molecular markers such as that described by the authors might be interesting in genetic evaluations of rank to increase the prediction accuracy of EBVs, especially at longer distances (at which lower heritabilities are observed).

#### 5. Conclusions

The two models used indicate that, at shorter racing distances, the selection for rank is more effective in Brazilian Thoroughbred horses because of higher additive genetic variance and heritability. The higher heritability estimates obtained with the TM suggest the possible use of this approach in programs designed to select for rank of the animals, especially at longer distances. The correlations obtained at different distances are of moderate to low magnitude, indicating that the set of genes acting at a given distance does not necessarily act at other distances.

#### Acknowledgment

The authors thank Dr Anne Ricard for providing the TMCLA program for this study.

## References

- [1] Tavenier A. Genetic evaluation of horses based on ranks in competitions. *Genet Select Evol* 1991;23:159–73.
- [2] Langlois B. Heritability of racing ability in Thoroughbreds—a review. *Livest Prod Sci* 1980;7:387–408.
- [3] Williamson SA, Beilharz RG. Heritabilities of racing performance in Thoroughbreds: a study of Australian data. *J Anim Breed Genet* 1996;113:505–24.
- [4] Williamson SA, Beilharz RG. The inheritance of speed, stamina and other racing performance characteristics in the Australian Thoroughbred. *J Anim Breed* 1998;115:1–16.
- [5] Belhajyahia T, Blouin C, Langlois B, Harzalla H. Breeding evaluation of Arab horses from their racing results in Tunisia by a BLUP with an animal model. *Anim Res* 2003;52:481–8.
- [6] Sobczynska M, Lukaszewicz M. Genetic parameters of racing merit of Thoroughbred horses in Poland. *J Anim Breed Genet* 2004;121:302–6.
- [7] Svobodova S, Blouin C, Langlois B. Estimation of genetic parameter of Thoroughbred racing performance in Czech Republic. *Anim Res* 2005;54:499–509.
- [8] Bokor A, Blouin C, Langlois B. Possibility of selecting racehorses on jumping ability based on their steeplechase race results in France, the United Kingdom and Ireland. *J Anim Breed Genet* 2007;124:124–32.
- [9] Chico MD. Genetic analysis of Thoroughbred racing performance in Spain. *Ann Zoot* 1994;43:393–7.
- [10] Lee KJ, Park KD, Kang MG, Kim DR, Moon YY. Estimation of genetic parameters for racing performance of Thoroughbred horses. *Korean J Anim Sci* 1995;37:11–8.
- [11] Villela LCV, Mota MDS, Oliveira HN. Genetic parameters of racing performance traits of Quarter horses in Brazil. *J Anim Breed Genet* 2002;119:229–34.
- [12] Sobczynska M, Kownacki M. Genetic aspects of racing performance in Polish purebred Arab horses. I. Genetic parameters. *J Appl Genet* 1997;38:179–86.
- [13] Mota MDS. Genetic correlation between performance at different racing distances in Thoroughbreds. *Livest Sci* 2006;104:227–32.
- [14] Gianola D, Simianer HA. Thurstonian model for quantitative genetic analysis of ranks: a Bayesian approach. *Genetics* 2006;174:1613–24.
- [15] Ricard A, Legarra A. Validation of models for analysis of ranks in horse breeding evaluation. *Genet Select Evol* 2010;42:3.
- [16] Ricard A, Legarra A. TMCLA software. 2012 Available at: <https://exchange-fichiers.inra.fr/get?k=Iq31nKFOBkRHoBnRxx2>.
- [17] Van Kaam JBCHM. Gibanal 2.9: analyzing program for Markov Chain Monte Carlo Sequences. Wageningen, The Netherlands: Department of Animal Science, Wageningen Agricultural University; 1998.
- [18] Bakhtiari J, Kashan NEJ. Estimation of genetic parameters of racing performance in Iranian Thoroughbred horses. *Livest Sci* 2009;120:151–7.
- [19] Sobczynska M. Genetic correlations between racing performance at different racing distances in Thoroughbreds and Arab horses. *Czech J Anim Sci* 2006;51:523–8.
- [20] Hill EW, Gu J, Eivers SS, Fonseca RG, McGivney BA, et al. A sequence polymorphism in MSTN predicts sprinting ability and racing stamina in Thoroughbred horses. *PLoS One* 2010;5:e8645.