

Genomic analysis of gaits and racing performance of the French trotter

Anne Ricard^{1,2}  | Arnaud Duluard³

¹Université Paris-Saclay, INRAE, AgroParisTech, GABI, Jouy-en-Josas, France

²Pole Développement Innovation Recherche, IFCE, Gouffern en Auge, France

³Département Elevage & Santé Animale, LeTrot, Paris, France

Correspondence

Anne Ricard, Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France.
Email: anne.ricard@inra.fr

Funding information

Fonds Eperon; Institut Français du Cheval et de l'Équitation

Abstract

The aim was to disentangle gait characteristics from other qualities needed for racing performances with a genomic analysis of French trotters (FT). A sample of 1,390 horses were recruited, from which 46% were genotyped with Illumina chip of 54,602 SNPs, 49% with Affymetrix chip of 670,806 SNPs and 586 had a completed questionnaire on trotting technique. Racing performances cover the period 1996 to 2018. There were 252,368 FT-born; 96,617 qualified and 83,962 which participated in a race. After quality control, 377,611 SNPs were retained and imputed. Questionnaire described trotting technique over 13 questions which were summarized, after principal component analysis in 3 traits: pacer, heavy trot/gallop and other defects. GWAS and genomic evaluation were performed using single-step approach. We found 25 QTL for racing performances and 9 for trotting technique. Only DMRT3 mutation was significant for both traits. To tend to pace avoid the defect at gallop and lead to a better early career for earnings, less percentage of disqualified races at all ages and more harness than under saddle career. This is the portrait of AA genotype at DMRT3. We found 5 other QTL, not linked to gait traits, which might improve selection of genetically independent performance traits of earnings per races and percentage of finished races. For only earnings at different ages and in under saddle or harness races, genomic evaluation remains the best way to predict performances.

KEYWORDS

genomic, GWAS, horse, racing performance, trotter

1 | INTRODUCTION

The breeding of trotting racehorses has undergone profound changes since the publication of the existence of a gene with a major effect on gaits (Andersson et al., 2012). In American Standardbred, the specialized breed to race as pacers or as trotters (Cothran et al., 1987), the DMRT3 mutation (23:g.22999655C>A) is fixed and nearly fixed in Swedish and

Norwegian Standardbred (Promerova et al., 2014). However, in other trotter breeds, there is still polymorphism. The frequency of the A allele was determined to be 35% in the Swedish Norwegian Coldblooded Trotter (Fegraeus, Lawrence, et al., 2017), 63% in Finnhorses (Fegraeus et al., 2015) and 87% in the Spanish Trotter (Rama et al., 2016). In the French Trotter (FT), the frequency of allele A was estimated at 76% (Ricard, 2015). The mutation causes a premature stop

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. *Journal of Animal Breeding and Genetics* published by John Wiley & Sons Ltd

codon (*DMRT3_Ser201STOP*), which results in a truncated DMRT3 protein. The mutation, a change from cytosine (C) to adenine (A), is permissive for performing pace, a lateral two-beat gait. Evidence for a critical role of DMRT3 in coordination of limb movements was obtained in the knockout mouse model by characterization of the locomotion and spinal cord function (Andersson et al., 2012). In this first publication, the significant positive effect of genotype AA on harness racing performance was highlighted in the Swedish Standardbred. However, in all other breeds, the positive effect of the A allele on racing performance has not always been demonstrated (Fegraeus, Lawrence, et al., 2017; Jaderkvist, Andersson, et al., 2014; Ricard, 2015). The effect depended on the character of the measured racing performance (e.g. early or not) and the association of A with C, thus with a heterozygous genotype. For FT, the selection strategy for increasing the frequency of allele A is not straightforward as the CA genotype was found to have higher earnings per finished race after the age of 4 years (Ricard, 2015). The aim of this study was to provide practical solutions for breeding of FT in the recent context of merchandizing of genomic tests.

DMRT3 is a gene whose mutation has a strong effect on gait lateralization and consequently on performance in trotting races, because trotting is a symmetrical gait as opposed to gallop. The objective of our study was to determine the importance of the characteristics of the trot of FT for performances in races and possibly to find other quantitative trait loci (QTL) responsible for these characteristics. Finally, we wanted to determine other genomic regions associated with the various traits of racing performances. Combining all these results, the ultimate goal is identifying the optimal evaluation strategy for selection of reproducers.

2 | MATERIALS AND METHODS

2.1 | Data

A cohort of 704 FT has been recruited from their trainers during 2016, 2017 and the beginning of 2018. The trainers responded to a questionnaire for each horse that included 13 trot characteristics. The questionnaire was completed for 83% of these horses (586). From this cohort, 90% (637) horses were genotyped using the Affymetrix Axiom Equine genotyping array with 670,806 single nucleotide polymorphisms (SNPs). The horses of this sample were born between 2001 and 2016, with the majority born between 2009 and 2013 (87%). Among the horses, 43% were females, 41% were geldings and 16% were males.

From earlier studies (Brard & Ricard, 2015; Ricard, 2015), a sample of 686 horses had been previously genotyped with Illumina Equine SNP50 chip (54,602 SNPs). This cohort of horses was assessed with a protocol which aimed to identify

TABLE 1 Sample of trotters genotyped or gait surveyed

| | Surveyed | Without questionnaire | Total |
|----------------------|----------|-----------------------|-------|
| Genotyped Illumina | 0 | 686 | 686 |
| Genotyped Affymetrix | 519 | 118 | 637 |
| Without genotype | 67 | 0 | 67 |
| Total | 586 | 804 | 1,390 |

osteocondrosis markers. It consisted of 98 sires born before 2003 and 588 progeny tested for osteochondrosis and born mostly from 2002 to 2008. They were 39% females, 31% geldings and 30% males.

The final sample, genotyped or gait surveyed, contained 1,390 FT, from which 1,323 were genotyped (95%) and 586 had a completed questionnaire (42%). Ultimately, 519 horses (37%) were both genotyped and assessed with a questionnaire. Table 1 summarized the data.

LeTrot, the French association in charge of the development of trotting races in France, and of the protection of the FT breed in its specificity, provided performance data in races for FT. Performances included all results in races from 1996 to 2018. There were 252,368 FT-born from 1994 to 2015. Among them, 38% were qualified and, therefore, authorized to participate in races.

Pedigrees were provided by IFCE (Institut Français du cheval et de l'Equitation), the French technical institute in charge of the equine industry, including breeding. Over 6 generations of pedigrees of all trotters born between 1994 and 2015 represented a total of 314,396 horses.

2.2 | Imputation

To examine all 1,323 genotyped horses, imputation was performed using Fimpute 2.2 software (Sargolzaei et al., 2014). First, quality control removed SNPs, which did not reach quality test requirements. The tests used included the minimum allele frequency test (MAF) $\geq 5\%$, the Hardy–Weinberg p -value test $\geq 10^{-6}$, call rate $\geq 90\%$, a valid position on autosomal chromosome and p -value of test of different MAFs between the two chips $\geq 10^{-5}$. Reference map used was that of Equine 3.0 reference sequence if the SNP position is available or Equine 2.0 if not. After quality control, 377,611 SNPs were retained: 0.4% solely on the Illumina Equine SNP50 chip, 89.6% on the Affymetrix Axiom Equine chip and 10.0% on both. Imputation was performed on all these SNPs, adding the information of pedigree data for over 4 generations (6,905 horses). Quality of imputation has been already checked in Chassier et al. (2018). On a slightly lower sample of the present data, the mean concordance rate per individual was 0.9903 and the mean r^2 was .9905. The SNP of the DMRT3 mutation was only present in the Affymetrix Axiom Equine

chip, and therefore, the allele was imputed for the 687 horses genotyped with the Illumina Equine SNP chip. The results of patented genotype at DMRT3 for the 637 horses genotyped with the Affymetrix Axiom Equine chip was given to the owners thanks to the permission and collaboration with EQUIBIOGENES and CAPILET GENETICS (<https://www.equibiogenes.com/> and <http://www.capiletgenetics.com/sv>).

2.3 | Model for racing performances

Racing performances were analysed using all the data for horses born from 1994 to 2015, including the 1,390 horses of the studied sample. Racing performance traits of horses used in the study were the same as previously described (Ricard, 2015):

- Qualification status (Q), which is obtained once during the horse's life, is a binary variable (0 = unqualified, 1 = qualified in the qualification test before entering races). Qualification test is a special short race of 2000 m which must be accomplished in less than a given time according to age (from 2 to 5) and year, changing with the improvement of the population. For example, qualification time was 1'20" for horses born from January to May 2017 at the age of 3 in 2020.
- The logarithm of annual earnings, divided by the annual number of finished races (i.e. races in which the horse was not disqualified) at 2, 3 and 4 years of age, and the logarithm of the sum of earnings between the age of 5 and 10 years, divided by the number of finished races over the same period, were calculated separately for harness racing (LnEH2, LnEH3, LnEH4, LnEH5-10) and racing under saddle (LnES3, LnES4, LnES5-10). Disqualification occurs when the horse does not perform regular trot, for example gallop and pace.
- Proportion of finished races (F) without disqualification was treated as a repetitive binary variable for each race started (0 = disqualified, 1 = finished race) at the ages of 3 and 4, and between the ages of 5 and 10 years for harness racing only (FH3, FH4, FH5-10).
- Number of races started (S) at 3 and 4 years of age and between the ages of 5 and 10 years for harness racing only. A "probit" transformation was applied to the number of races started because its distribution was far from normal (SH3, SH4, SH5-10).

A mixed linear animal model was used to analyse racing performances. Fixed effects included a combination of year of birth and gender effect (44 levels) and a cohort effect (3 levels: horses recruited for osteochondrosis analysis, horses recruited for questionnaire data and other horses). Random effect included the animal genetic value. The relationship

matrix between genetic values was constructed using both genealogy data and genomic data in a single-step GBLUP method (Aguilar et al., 2010; Christensen & Lund, 2010; Fernando et al., 2014; Legarra et al., 2014) and using the rules outlined by VanRaden (2008) for genomic relationship matrix. For FH3, FH4 and FH5, the performance was repeated several times. Therefore, a random permanent environmental effect was added to the model. The software used was BLUPF90 (Aguilar et al., 2018).

To detect significant SNPs, single-trait analysis was performed using previously estimated variance components (Ricard, 2015). GWAS was performed with back solutions of SNP effect, and *p*-value was calculated as described in Aguilar et al. (2019). Significant SNPs were identified using *p*-values. Genomic control (Devlin & Roeder, 1999) was applied to the *p*-values to prevent incorrect distribution of the test statistics. The effective number of independent tests in our data was calculated as the inverse of the mean of linkage disequilibrium (r^2) between all available pairs of SNPs by chromosome (Li et al., 2012). Linkage disequilibrium between chromosomes was assumed to be negligible. The genome-wide significance was then set at 1% and divided by the number of independent tests. Significant SNPs were grouped according to the distance between them to define a lower number of significant regions. As long as distance between successive significant SNPs is <1,000 kb, they were assumed to be part of the same region. In these regions, the highest significant SNP was retained defining the quantitative trait nucleotides (QTNs).

The back solution of the SNP effect only provides the effect of allele. For the QTNs, a supplementary analysis was performed to estimate the effects of the 3 possible genotypes. The objective was to detect a possible dominance effect for the heterozygote type. For this analysis, genotypes were considered as fixed effects and only phenotypes for genotyped animals were included in the analysis. To consider all performances, phenotypes were computed as deregressed proofs calculated from the estimated breeding values and reliabilities obtained with the first full model without genomic data. Proofs were calculated according to Ricard et al. (2013). Then, genomic best linear unbiased prediction (gBLUP) (VanRaden, 2008) was used with ASReml software (Gilmour et al., 2014), which provided the Wald *F* statistics for the fixed effects, that is the genotypes. To perform gBLUP with ASReml, the inverse of the genomic matrix calculated with the BLUPF90 program was used. Results are the solutions of fixed effects and Wald *F* statistics for all QTNs and traits.

A cross-validation was used to compare different strategies of genetic evaluation to predict racing performances. The training set included all horses born until 2010 ($n = 192,167$) and their performance results until 2012 (first age for qualification of horses born in 2010). The validation set included all genotyped horses born from 2011 ($n = 446$) and their

performance results until 2018. Three genetic evaluations were calculated as follows:

- Genetic evaluation with only genealogy data to compute the relationship matrix. Fixed effect included the combination of year and sex (34 levels) effect and a cohort effect.
- Genomic evaluation using ssGBLUP with genealogy and genomic data to compute the relationship matrix. Fixed effect included the combination of year and sex (34 levels) effect and a cohort effect.
- Genomic evaluation using ssGBLUP with genealogy and genomic data to compute relationship matrix. Fixed effect was the combination of year and sex (34 levels) effect, a cohort effect and genotypes of significant QTNs.

For the cross-validation, multitrait analyses were used. To reduce the size of the system of equations, several multitraits were used, including the following: (a) Q, LnEH3, LnEH4, LnEH5-10; (b) LnEH3, FH3, SH3; (c) LnEH4, FH4, SH4; (d) LnEH5-10, FH5-10, SH5-10; and (e) LnEH3, LnEH4, LnEH5-10, LnES3, LnES4, LnES5-10.

2.4 | Model for trotting technique questionnaires

The questionnaire used included 13 questions about each of the horses. Questionnaire responses were converted into binary variables (0/1), with 0 representing the absence of default or characteristic and 1 representing its presence. They were as follows:

- Pacer
- Defective gait: pace
- Heavy trot
- Defective gait: gallop
- Bad mouth
- Necessity of equipment (e.g. overcheck)
- Trot crookedly
- Mowing hindlimbs
- Paddling
- Knee hitting
- Defective gait: traquenard
- Better on right-handed course or left-handed course
- Unequal contact in both reins

A mixed linear animal model was used to estimate heritability and the environmental and genetic effects on questionnaire responses. Fixed environmental effects included gender, birth year and trainer effect (23 levels). Random effects included the animal's genetic value. The relationship matrix between genetic values was constructed using genealogy data

(9,092 ancestors over 6 generations). Probit transformation was used to transform binary variables. The software used was BLUPF90 (Aguilar et al., 2018). From this analysis, we constructed residual trotting technique variables equal to probit variable response to questionnaire, minus the estimated effect of gender, year of birth and trainer.

Principal component analysis was performed on the residual trotting technique variables using the FactoMineR package (Lê et al., 2008). The results led to the creation of three synthetic trotting technique variables to summarize the characteristics of each horse's trotting technique as measured by the questionnaire. The synthetic variable was constructed from the grouping of elementary variables. The synthetic variable was the sum of elementary binary variables capped at 1, 2 and 3 respectively. Elementary variables were re-grouped as follows:

- Pacer, defective gait: pace (2 variables).
- Defective gait: gallop, heavy trot, bad mouth (3 variables)
- Other defects (8 variables).

Synthetic variables were used for GWAS analysis with the same methodology described for racing performance variables. A mixed linear animal model was used. Fixed environmental effects included gender, birth year and trainer effect (23 levels). Random effects included the animal's genetic value. The relationship matrix between genetic values was constructed using genealogy data, and genomic data in a single-step GBLUP method and back solutions of SNP effect and *p*-value were calculated as described in Aguilar et al. (2019). Significant SNPs were identified using *p*-values after genomic control.

We tested the capacity of the trotting technique questionnaire to predict the genotype of the horse for the DMRT3 mutation. We used the random forest classification analysis using all questionnaire binary variable and the genotype at DMRT3 on the 519 horses, with both questionnaire and genotype using the “randomForest” function of the “randomForest” package in R (Breiman, 2001; Liaw & Wiener, 2018). The default parameters were used, that is the number of trees was 500 and the number of variables tried at each split was ($\sqrt{n} = 3$). The training set included two-thirds of the sample chosen at random, and the remainder was used as the validation set. We performed 100 resamplings.

We tested the capacity of trotting technique questionnaire to predict racing performances. We used the SAS software SAS/GLM (SAS/STAT, 2019) to perform general linear model analysis on racing performances. The dependent variables were residuals of performances, that is racing performance variables, minus the estimated effect of the combination of the year and gender and of the cohort (see above for the full model used to analyse racing performance). The independent variables were the three synthetic trotting

technique variables defined after the principal component analysis.

3 | RESULTS

3.1 | General statistics on the data

Genotyped horses were significantly more frequently qualified compared with the whole FT population: 88% versus 38%. Their racing performances were slightly higher than those of all racing horses: z -score around 0.6 for earnings (LnEH2, LnEH3, LnEH4, LnEH5-10, LnES3, LnES4, LnES5-10) but close to zero for all other traits (FH3, FH4, FH5, SH3, SH4, SH5). Horses with questionnaire were almost all qualified (98%). The superiority for racing performances of this surveyed sample was in the same order as that of the genotyped sample (z -score for earnings around 0.7 and close to 0 for other traits).

For DMRT3 mutation, the proportion of AA genotype in the sample of the 1,323 genotyped horses was 59%, and there were 37% CA and 4% CC, with A allele frequency of 77%. The A allele frequency increased with time (Figure 1).

3.2 | GWAS on racing performances

The mean of linkage disequilibrium over the genome was determined to be 0.0005920. The number of independent tests was then estimated to be $1/0.0005920 = 1689$. Then, the p -value threshold was set to $1\%/1689 = 5.9 \times 10^{-6}$ or $-\log(p\text{-value}) = 5.2$. There were 195 SNPs which exceeded the threshold. These SNPs belonged to 25 regions on the genome (Table 2). There were 19 regions with only one

significant SNP and 6 with more than one. Five regions with more than one SNP covered from 11,181 to 1,126,145 bp. The 6th region was a large region near DMRT3 involving 146 SNPs ranging from -2.2 to $+3.6$ Mb around the position 23:22,999,655. The DMRT3 mutation was not the highest p -value ($-\log(p\text{-value}) = 14.68$) in the region but close.

3.3 | Genotype effects of DMRT3 and QTNs on racing performance

The effects of DMRT3 genotypes on all racing performances are presented in Figure 2 expressing the difference with respect to the AA genotype. For qualification status (Q), the AA was significantly the most favourable genotype, followed by CA and then CC. There were no horses with CC genotype in races at 2 years old, and AA was the best genotype at this age compared with CA for earnings. Then, for earnings, from 3 to 5–10 years old, the advantage of AA compared with CA was not significant and the CA performed better than AA after 4 years, but with no significant effect, except for racing under saddle. The CC genotype was associated with the worst earning results, except for the LnES5-10, but this result was not significant. For the proportion of finished races, the genotype AA demonstrated significantly higher results at all ages compared with CC. Further, the probability of participating in a race was highest for the AA genotype.

Effects all the 25 QTNs on all racing performances are summarized in Figure 3. Each QTN was chosen according to p -value threshold 5.9×10^{-6} established in 3.2 for one trait, but in this figure the p -values reported for all traits were an uncorrected raw p -value. All QTNs had an effect on one of earnings traits (LnEH2, LnEH3, LnEH4, LnEH5-10, LnES3, LnES4 or LnES5-10). Among them, 13 had an effect on both

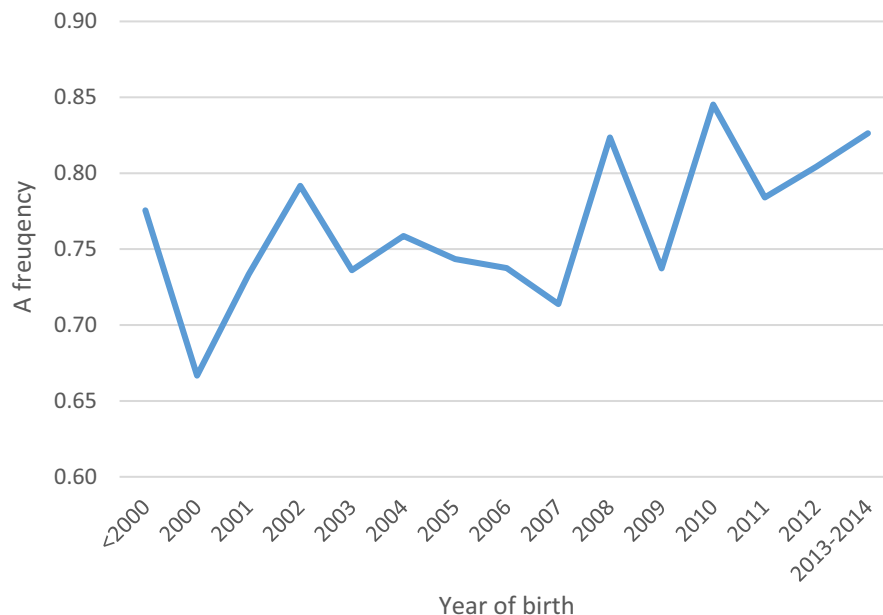


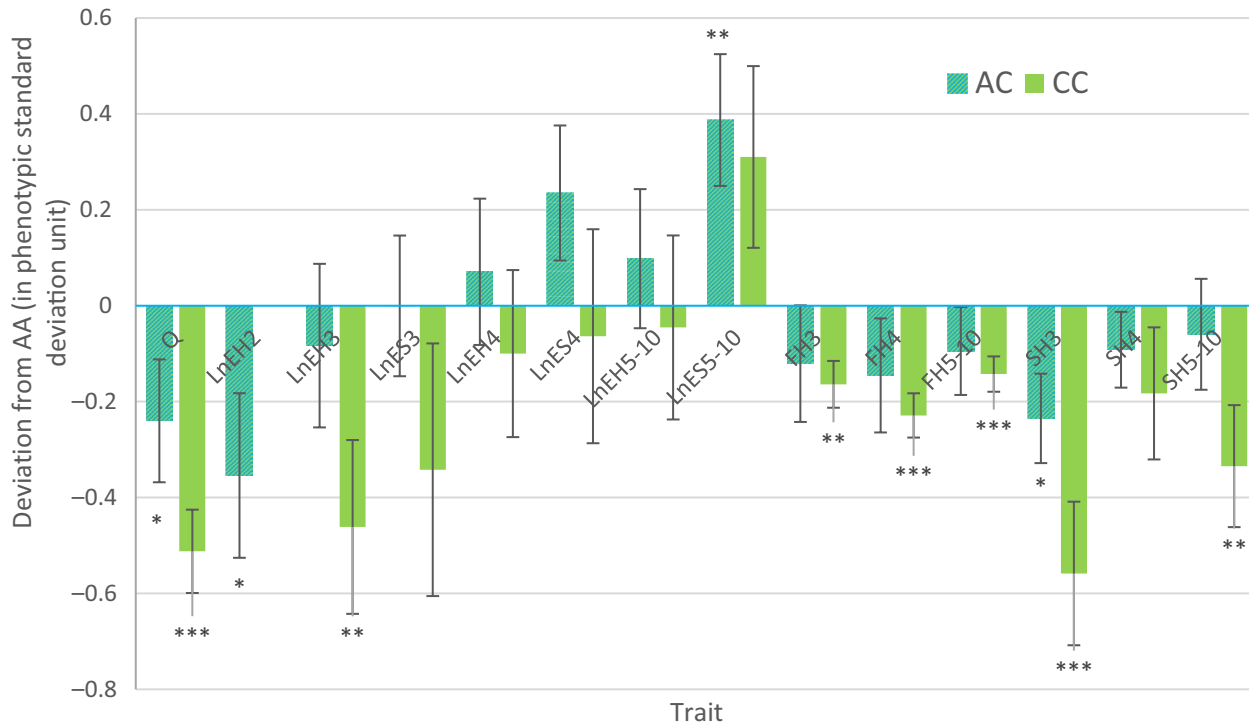
FIGURE 1 Frequency of A allele mutation of DMRT3 gene in French Trotters sample according to birth year [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 The 25 regions detected by GWAS for racing performances

| SNP with highest <i>p</i> -value in the region | | | | | | | Region | | | |
|--|---------------------|-----|------------------------|-----|--------------|-------------------------|----------|-------------|-------------|--|
| SNP name (Affymetrix) | SNP name (Illumina) | Chr | Physical position (bp) | MAF | Minor allele | -log (<i>p</i> -value) | <i>N</i> | Start (bp) | End (bp) | Traits involved ^a |
| AX-103109120 | | 1 | 97,926,701 | 44% | G | 5.32 | 1 | 97,926,701 | 97,926,701 | LnES5-10 |
| AX-103109976 | | 1 | 107,688,415 | 28% | T | 5.31 | 1 | 107,688,415 | 107,688,415 | LnES4 |
| AX-104483448 | | 1 | 157,413,078 | 10% | G | 5.32 | 1 | 157,413,078 | 157,413,078 | LnEH4 |
| AX-103206587 | | 2 | 70,783,851 | 48% | G | 5.68 | 1 | 70,783,851 | 70,783,851 | LnEH5-10 |
| AX-103387382 | | 2 | 108,263,302 | 38% | T | 6.47 | 1 | 108,263,302 | 108,263,302 | LnEH4 |
| AX-103524072 | | 2 | 113,687,283 | 32% | G | 5.76 | 1 | 113,687,283 | 113,687,283 | LnEH5-10 |
| AX-103055777 | | 3 | 9,044,075 | 34% | C | 6.31 | 1 | 9,044,075 | 9,044,075 | LnEH5-10 |
| AX-103402679 | | 4 | 40,784,504 | 9% | G | 5.58 | 1 | 40,784,504 | 40,784,504 | LnEH4 |
| AX-104659319 | BIEC2-867640 | 4 | 59,519,415 | 9% | C | 5.33 | 1 | 59,519,415 | 59,519,415 | LnES5-10 |
| AX-103248261 | | 6 | 83,845,383 | 36% | C | 6.62 | 1 | 83,845,383 | 83,845,383 | Q |
| AX-104480266 | | 7 | 48,364,149 | 43% | C | 7.07 | 1 | 48,364,149 | 48,364,149 | LnEH3, LnEH4 |
| AX-103944680 | | 7 | 74,283,640 | 22% | C | 5.26 | 1 | 74,283,640 | 74,283,640 | LnEH2 |
| AX-104315706 | | 8 | 58,818,493 | 43% | G | 6.59 | 10 | 58,646,376 | 59,772,521 | LnES3 |
| AX-104634643 | BIEC2-1073869 | 9 | 10,208,149 | 7% | C | 6.50 | 4 | 10,162,835 | 10,208,149 | SH4 |
| AX-104121015 | | 10 | 74,041,814 | 15% | G | 5.24 | 1 | 74,041,814 | 74,041,814 | LnEH5-10 |
| AX-103963491 | | 11 | 18,013,970 | 42% | A | 5.51 | 1 | 18,013,970 | 18,013,970 | LnEH5-10 |
| AX-103609883 | | 11 | 21,979,718 | 37% | T | 7.12 | 12 | 21,877,594 | 21,982,576 | LnES4, LnES5-10 |
| AX-104368747 | | 11 | 45,228,405 | 43% | G | 5.45 | 1 | 45,228,405 | 45,228,405 | LnEH3 |
| AX-103214187 | | 14 | 14,798,378 | 6% | T | 5.38 | 2 | 14,798,378 | 14,809,559 | LnEH2 |
| AX-104713486 | | 14 | 20,139,466 | 24% | T | 5.71 | 1 | 20,139,466 | 20,139,466 | LnEH4 |
| AX-103860575 | | 17 | 73,263,145 | 11% | T | 5.32 | 1 | 73,263,145 | 73,263,145 | LnEH3 |
| AX-103137314 | | 18 | 49,432,839 | 19% | C | 6.27 | 1 | 49,432,839 | 49,432,839 | LnEH3, LnEH4, LnEH5-10 |
| AX-103424215 | | 20 | 31,895,674 | 49% | T | 6.23 | 2 | 31,821,280 | 31,895,674 | LnEH4 |
| AX-104802280 | | 23 | 22,478,301 | 22% | C | 15.09 | 146 | 20,821,392 | 26,578,170 | Q, LnEH2, LnES5-10, FH3, FH4, FH5-10, SH3, SH4 |
| AX-104608702 | | 24 | 492,148 | 15% | A | 5.24 | 1 | 492,148 | 492,148 | LnEH4 |

Abbreviations: bp, base pairs; Chr, chromosome; MAF, minor allele frequency; *N*, number of SNPs in the region.

^aName of traits: Q = qualification, LnEHx = logarithm of earnings per finished harness races at the age of *x* with *x* = 2, 3, 4 and from 5 to 10 years, LnESx = logarithm of earnings per finished races under saddle at the age of *x* with *x* = 3, 4 and from 5 to 10 years, FHx = proportion of finished Harness races at the age of *x* with *x* = 3, 4 and 5–10, SHx = number of started harness races at the age of *x* with *x* = 3, 4 and 5–10.



Q = qualification, LnEH x = logarithm of earnings per finished harness races at the age of x with $x = 2, 3, 4$ and from 5 to 10 years old, LnES x = logarithm of earnings per finished races under saddle at the age of x with $x = 3, 4$ and from 5 to 10 years old, FH x = Proportion of finished Harness races at the age of x with $x = 3, 4, 5-10$, SH x = Number of started harness races at the age of x with $x = 3, 4, 5-10$.

FIGURE 2 Effect of AC and CC genotypes as difference respect to AA genotype at DMRT3 genotype locus on racing performances. [Colour figure can be viewed at wileyonlinelibrary.com]

earnings in harness races and races under saddle, 9 on harness races only and 3 on races under saddle only. The same genotype was always in the same direction for all earnings traits: always positive or negative with the notable exception of DMRT3. The rare allele had positive effect on earnings for 14 of the QTNs. Most (17) had also a similar effect on qualification. Few (8) had also an effect on the percentage of finished races but not at all ages, except for DMRT3. When the QTN had an effect on the percentage of finished races, it was in the same direction than earnings for 5 and in the other direction for 2 (not counting DMRT3). Significant effect on number of starts occurred principally for the trait at the age of 5 to 10 years. Other than DMRT3, 9 QTNs are concerned, 7 in the same direction than earnings and 2 in the opposite.

3.4 | DMRT3 region

The DMRT3 region can be divided into two parts: SNPs located in position prior to DMRT3 mutation with a relatively high linkage disequilibrium with DMRT3 and one located after the DMRT3 mutation with less linkage (Appendix S1). Manhattan plot over the region for all racing performances

is reported in Appendix S2. Many SNPs in the region were highly significant for Q. For earnings, many SNPs in the region were highly significant for LnEH2; however, from 3 years of age to 10 years, none reached the significant threshold, except for LnES5-10. The effect of many SNPs in the DMRT3 region on probability to finish a race was significant at all ages. Figure 4 plots the estimated effects of alleles of the 146 SNPs in DMRT3 region for pairs of traits related to earnings. Favourable alleles on early performance (Q/LnE2) became negative in late performance (LnH5-10 and LnS5-10). Figure 5 plots absolute effect of alleles of the 146 SNPs in DMRT3 region according to linkage disequilibrium with DMRT3 mutation. There was no clear evidence of SNPs whose effect would not be related to its linkage disequilibrium with DMRT3, except perhaps some points for LnEH3.

3.5 | Questionnaire: analysis of elementary variables

The plot illustrated in Figure 6 demonstrated the variable circle of the principal component analysis for the first two dimensions. Only 32% of the variability was explained by these two



for DMRT3 region, the QTN retained was the known mutation not the highest p-value

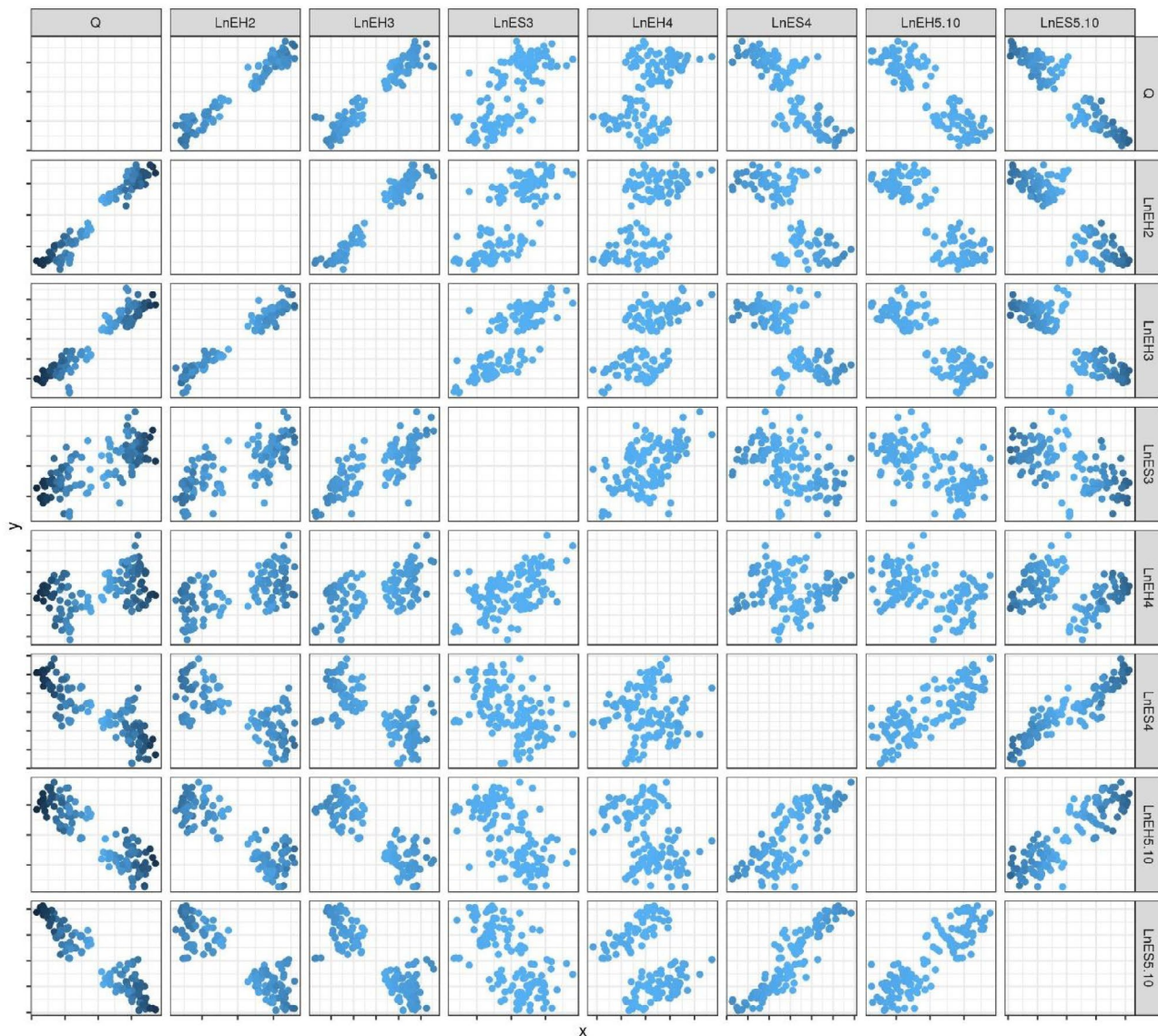
Q = qualification, LnEHx = logarithm of earnings per finished harness races at the age of x with x = 2, 3, 4 and from 5 to 10 years old, LnESx = logarithm of earnings per finished races under saddle at the age of x with x = 3, 4 and from 5 to 10 years old, FHx = Proportion of finished Harness races at the age of x with x = 3, 4, 5-10, SHx = Number of started harness races at the age of x with x = 3, 4, 5-10.

FIGURE 3 Effects of the genotypes of the QTNs on all racing performances. [Colour figure can be viewed at wileyonlinelibrary.com]

first components. We needed 6 components to explain 65% of the variance meaning that each defect was rather a different trait. From the plot, we used 3 main groups of variables to qualify the trotting technique of the horse. The first main group was the ability to pace, including pacer and horses defective at pace. The second group of variables grouped horses defective at gallop, with heavy trot and a bad mouth. The third group included all other locomotion defects. In later analysis, questionnaires were then summarized with these 3 groups. The 3 synthetic

variables were constructed from the grouping of elementary variables: they were the sum of elementary binary variables, capped to 1, 2 and 3, respectively. They will be named “pacer,” “heavy trot gallop,” and “other defects.”

There was a negative association between horses qualified as “pacer” and “heavy trot gallop.” There were no horses qualified as “pacer” which was also qualified as “defective gait: gallop.” There was a positive association between “other defects” and both “pacer” and “heavy trot gallop”: 77% of horses



Q = qualification, LnEHx = logarithm of earnings per finished harness races at the age of x with $x = 2, 3, 4$ and from 5 to 10 years old, LnESx = logarithm of earnings per finished races under saddle at the age of x with $x = 3, 4$ and from 5 to 10 years old.

FIGURE 4 Plot of allele effect of the 146 significant SNPs in DMRT3 region for pairs of racing performance traits. [Colour figure can be viewed at wileyonlinelibrary.com]

scored higher than 0 for “pacer” and 78% of scored higher than 0 for “heavy trot gallop” were scored higher than 0 for “other defects” compared with 64% and 52% for horses scored 0.

Some of the trotting technique questionnaire variables were heritable, as well as synthetic variables (Table 3.)

3.6 | Questionnaire: GWAS analysis

The frequency of the DMRT3 genotypes for trained horses was very different according to responses to the trotting technique questionnaire (Table 4). However, the random forest

classification based on questionnaire responses was unable to correctly predict the horses' genotypes at DMRT3. The mean over samplings of correctly predicted genotypes was 83% for the training samples and 62% for the validation samples, but never exceeded 77%. Mostly, the model failed to identify the CA genotype, with 50% of true CA genotype assigned to AA genotype. It failed also to identify CC genotypes, which were never predicted, and in 49% of the cases, were assigned to AA. Therefore, our results demonstrate that genotyping is necessary, regardless of the knowledge of the trotting technique of the horse.

Among the studied population, therefore trained and almost all qualified (98%), we found that 100% of horses

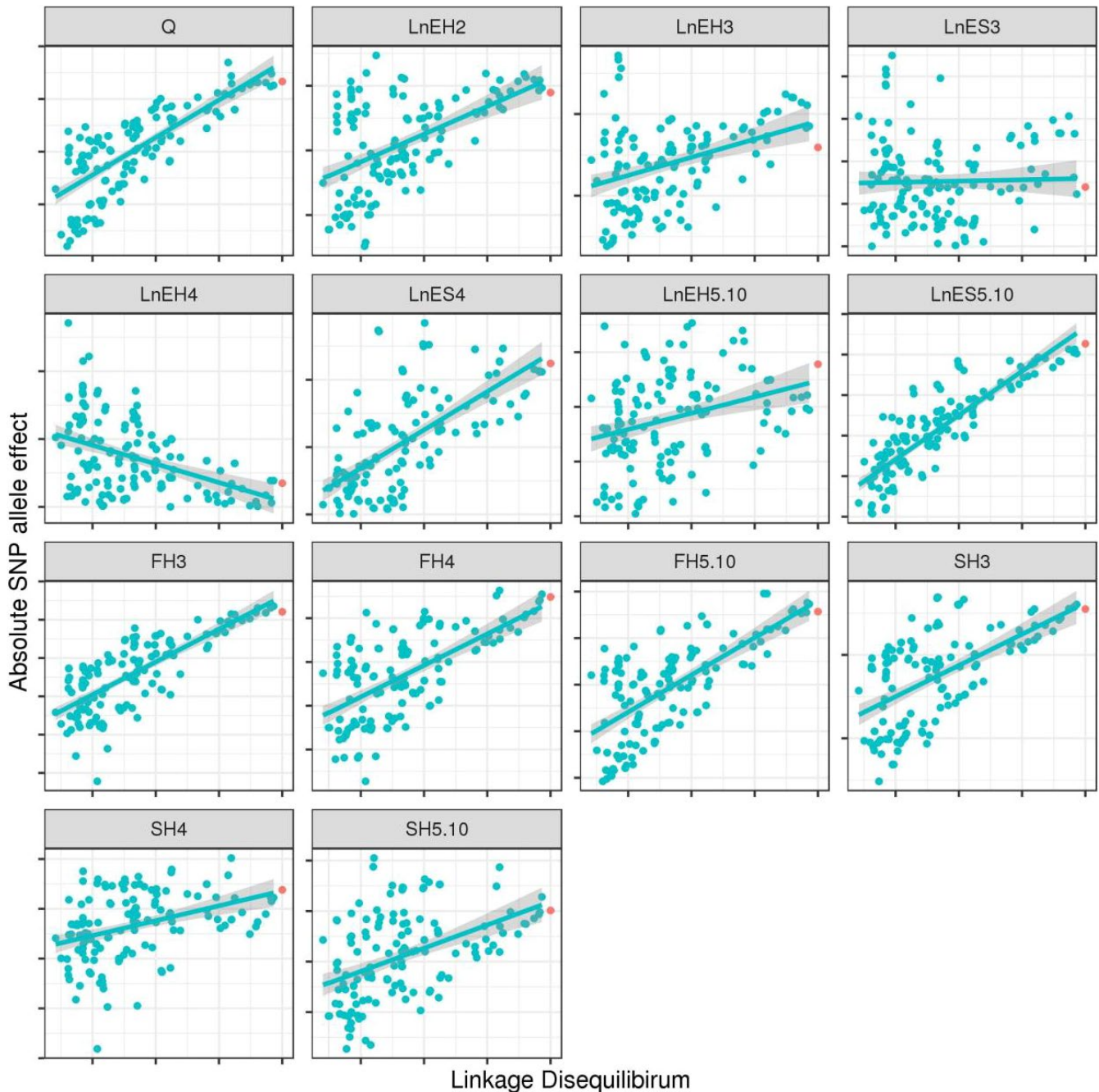


FIGURE 5 Absolute value of allele effects of the 146 significant SNPs in DMRT3 region according to linkage disequilibrium with DMRT3 genotype (red point DMRT3 mutation). Name of the traits: Q = qualification, LnEHx = logarithm of earnings per finished harness races at the age of x with x = 2, 3, 4 and from 5 to 10 years, LnESx = logarithm of earnings per finished races under saddle at the age of x with x = 3, 4 and from 5 to 10 years, FHx = proportion of finished Harness races at the age of x with x = 3, 4 and 5–10, SHx = number of started harness races at the age of x with x = 3, 4 and 5–10 [Colour figure can be viewed at wileyonlinelibrary.com]

defective at pace and 94% of horses qualified as pacer were AA, even if all AA horses were not qualified as pacer or defective at pace. None of the horses with genotype CC and only 6% of CA were qualified as pacer and none CC or CA as defective at pace. AA versus CA or CC is clearly a mark of the ability to pace. In the same way, but less systematic, CA and CC horses were more frequent among horses defective at gallop: 90% of horses with defective gait at gallop

were of these genotypes. Therefore, the presence of at least one C copy is a marker of inability to pace and tendency to gallop (or to have heavy trot and bad mouth). Other defect in trotting technique is distributed more equally in the 3 genotypes. Genotype AA at DMRT3 increased the probability to pace, avoid “heavy trot gallop” and had no significant effect on other defects. Heterozygote CA was closer to CC than AA.

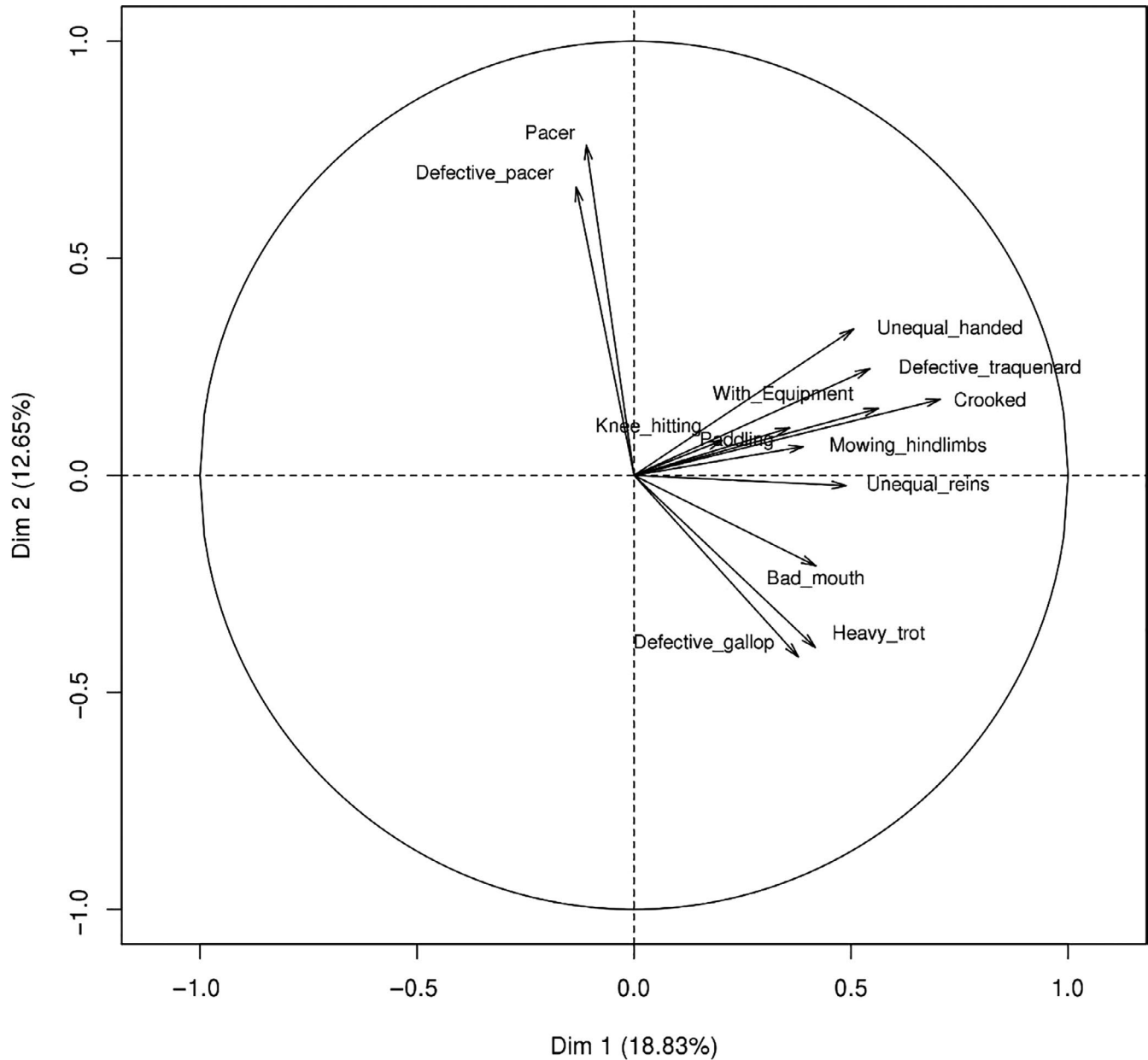


FIGURE 6 Principal component analysis of trotting technique responses to questionnaire ($n = 586$)

Table 5 summarized GWAS analysis with 9 QTNs detected. For horses qualified as “pacer,” in addition to the region of DMRT3, there was a QTN on chromosome 10 and 3 QTNs on chromosome 28 representing 7 SNPs above the significant threshold. For all, the rare allele had a positive effect on the ability to pace, contrary to the effect of DMRT3 were the frequent homozygote horse (AA) paced. The 3 QTNs on chromosome 28 were distant but in linkage disequilibrium ($r^2 = .66, .32, .46$). The QTNs had no effect of the other trotting technique traits, contrary to DMRT3 which had also an effect on “heavy trot gallop.” For the trait “heavy trot gallop,” in addition to DMRT3, there were 3 QTNs: two on chromosome 6 and one on chromosome 19. For two of these three QTNs, the rare allele increased the frequency of “heavy trot gallop” horses. So, there was only one frequent allele (C, 64%) which leads to gallop:

the one on chromosome 6:28,033,876. For “other defects,” there were 2 QTNs. One was the 6:28,033,876 where the frequent allele leads to more defects, as for the trait “heavy trot gallop.” And the second one on chromosome 3 was in the same direction (frequent allele, increase in defects).

DMRT3 was the only QTN detected for both racing performances and trotting technique.

3.7 | Questionnaire: effect on racing performance

The frequency of horses scored 1 for “pacer” was higher at 2 years old (22%) and much lower in races under saddle

TABLE 3 Heritability of questionnaire variables on trotting technique

| | Heritability | SE |
|-------------------------------|--------------|------|
| Elementary variables (0/1) | | |
| Pacer | 0.19 | 0.12 |
| Defective gait: pace | 0.26 | 0.14 |
| Heavy trot | 0.00 | 0.00 |
| Defective gait: gallop | 0.10 | 0.11 |
| Bad mouth | 0.25 | 0.14 |
| Necessity of equipment | 0.08 | 0.10 |
| Trot crookedly | 0.05 | 0.09 |
| Mowing hindlimbs | 0.02 | 0.08 |
| Padding | 0.00 | 0.00 |
| Knee hitting | 0.10 | 0.10 |
| Defective gait: traquenard | 0.11 | 0.11 |
| Better on one hand | 0.05 | 0.09 |
| Unequal contact in both reins | 0.38 | 0.17 |
| Synthetic variables | | |
| Pacer (0/1) | 0.18 | 0.12 |
| Heavy trot gallop (0/1/2) | 0.18 | 0.12 |
| Other defects (0/1/2/3) | 0.20 | 0.13 |

(<5%, two low number of pacer horses to estimate an effect on earnings in races under saddle). The effect to be scored 1 for “pacer” was always favourable whatever the performance trait (Figure 7). Significant effect was found on earnings at 2 and 3 years old and on number of starts at 3 and 4 years old. The frequency of horses scored higher or equal to 1 for “heavy trot gallop” was higher in races under saddle (65%) and lower at 2 years old (35%). The effect to be scored higher or equal to 1 for “heavy trot gallop” was unfavourable for late performances (LnEH5, FH4, FH5-10, SH5-10), especially for the proportion of finished races. For the trait “other defects,” there were no differences in frequency of horses scored higher or equal to 1 at the different ages of performance and between harness and under saddle races. To be scored to 1 or more was unfavourable for performances at 3 years old (LnEH3, FH3, SH3), but then, there was no effect on later performances.

3.8 | Genetic evaluation

Results on genetic evaluation for racing performance traits are outlined in Figure 8. Comparisons involved 4 evaluations: “QTNs,” the sum of QTN estimates obtained from ss-GBLUP with QTNs as fixed effects; “genealogy,” the BLUP evaluation with a relationship matrix based on genealogy; “genomic,” the ssGBLUP evaluation with a relationship

matrix based on genealogy and genomic without QTNs; and “genomic + QTNs,” the sum of QTN estimates and ss-GBLUP genomic value. Results for qualification (Q) were not relevant because the sample was highly selected on this trait (98% of horses qualified in the validation set). The use of only QTN estimates yielded the lowest correlations, and sometimes, even negative correlations, except for FH4 and FH5-10. Genomic and genealogic evaluation often yielded close correlations, generally slightly higher for the genomic evaluation. The mean of absolute differences in correlation between these two types of evaluations was 0.01, with the maximum for FH4 with the correlation of 0.11 with genomic evaluation and 0.07 with genealogic evaluation. The addition of QTNs to genomic evaluation yielded lower or higher correlation in equal cases. The gain was often less important (average of +0.02) than the loss (average of -0.05).

4 | DISCUSSION

The aim of the questionnaire submitted to the trainers was to qualify the trotting technique of the horse. Then, the knowledge of trotting technique of the sample of horses surveyed was to allow to disentangle the importance of gaits from other qualities for success in racing.

Trotting technique was divided into three major components: (a) ability to pace, (b) heavy trot and tendency to gallop, (c) other diverse incorrect trotting technique leading eventually to the necessity of equipment. To be a pacer gives an advantage for qualification and early career at 2 and 3 years old. The pacer horse is probably easy to manage for harness trotting, which explains the easy qualification, stronger presence at 2 years old, higher earnings per finished harness race in the beginning of the career until 3 years old and higher number of starts in harness races at 3 and 4 years old. Pacing could have been considered a handicap for trotting but, in fact, do not lead to an increase of percentage of races disqualified. On the other hand, horses with heavy trot and tendency to gallop keep a handicap throughout their career with a higher percentage of disqualified races, especially from 4 years old and a lower earnings per finished races from 5 years old. These horses are preferentially orientated to races under saddle. Horses with other trotting defects undoubtedly require an apprenticeship, with bad season at the age of 3 (for all racing performance traits), but the handicap disappears afterwards.

The only QTNs which link trotting technique and racing performances are DMRT3 mutation. We did not find other QTNs for both trotting technique and racing performances. We did not find any effect of SNPs found by GWAS on gait traits in gaited horses (Amano et al., 2018; Bussiman et al., 2020; Fegraeus, Hirschberg, et al., 2017; Staiger et al., 2016) or those found in Standardbred to differentiate trotters from pacers (McCoy et al., 2019).

| | All | DMRT3 genotype | | |
|----------------------------------|------------------|----------------|--------------|-------------|
| | (n = 586) | AA (n = 304) | CA (n = 191) | CC (n = 24) |
| Elementary variables (0/1) | | | | |
| Pacer | 14% | 22% | 2% | 0% |
| Defective gait: pace | 4% | 7% | 0% | 0% |
| Heavy trot | 36% | 25% | 50% | 50% |
| Defective gait: gallop | 12% | 2% | 25% | 29% |
| Bad mouth | 31% | 22% | 39% | 42% |
| Necessity of equipment | 25% | 20% | 32% | 29% |
| Trot crookedly | 20% | 19% | 21% | 4% |
| Mowing hindlimbs | 8% | 6% | 12% | 17% |
| Paddling | 15% | 14% | 16% | 8% |
| Knee hitting | 13% | 12% | 15% | 21% |
| Defective gait: traquenard | 15% | 16% | 12% | 8% |
| Better on one hand | 40% | 41% | 36% | 38% |
| Unequal contact in both reins | 10% | 8% | 10% | 13% |
| Synthetic variables ^a | | | | |
| Pacer (0-1/1) | 86-14%/14% | 23% | 2% | 0% |
| Heavy trot gallop (0-1-2/≥1) | 47-33-20%/53% | 40% | 71% | 71% |
| Other defects (0-1-2-3/≥1) | 34-26-17-22%/66% | 60% | 71% | 75% |

^aDetails of synthetic distribution variables were given for the whole sample, and only ≥ 1 by DMRT3 genotypes.

The mutant allele of DMRT3 was detected in Icelandic horses exhibiting pace (Andersson et al., 2012). Since then, the effect of the genotype at DMRT3 was investigated in multiple gaited breeds, exhibiting not only pace, but also various alternative gaits other than walk, trot or gallop. In some breeds, the ability to pace is generalized and the A allele of DMRT3 mutation is fixed (Giantsis et al., 2018; Staiger et al., 2016). Breeds with the fixed C allele do not exhibit pacing or other particular gaits (Pereira et al., 2016; Regatieri et al., 2017). In all the other breeds, the polymorphism for the mutation of DMRT3 remains. In these cases, the relationship with pace was sometimes exclusive with all pacer AA and all non-pacer CC (Novoa-Bravo et al., 2018). However, in most breeds, there were several AA horses that were unable to pace, even if none of the CC horse were able to pace (Amano et al., 2018; Fegraeus, Hirschberg, et al., 2017; Jaderkvist et al., 2015; Kristjansson et al., 2014) or all AA horses were able to pace, and some CC horses too (Jaderkvist, Kangas, et al., 2014). However, there has been an established relationship between genotype at DMRT3 and the ability to pace (Han & Peñagaricano, 2016). For alternative

TABLE 4 Frequency of responses equal to 1 to questionnaire on trotting technique for horses of the sample and according to DMRT3 genotype

gaits such as marcha batida, marcha picada in Brazilian Mangalarga Marchador horses (Bussiman et al., 2019, 2020; Fonseca et al., 2017; Patterson et al., 2015), rack and slow gait in American Saddlebred horses (Regatieri et al., 2016), Tölt in Icelandic horse (Fegraeus, Hirschberg, et al., 2017; Jaderkvist et al., 2015; Kristjansson et al., 2014), and Trocha and Colombian trot in Colombian horses (Novoa-Bravo et al., 2018), the relationship with DMRT3 genotype was not so obvious and many other markers were found. We had attempted to predict the DMRT3 genotype using only the knowledge of trotting technique and failed, even if a strong difference in genotype frequency was found between main trotting technique characteristics. Almost all pacers were AA genotype, but all AA genotype were not pacer. This can be partly due to the unbalanced frequency of the alleles. Almost all gallop horses were CA or CC but not all CA or CC were susceptible to gallop and other trotting techniques defects were not related to DMRT3. That is why we could not succeed to predict DMRT3 genotype from questionnaires. This is in accordance with the complex relationship between locomotion traits and DMRT3 genotypes of all

TABLE 5 The 9 regions detected by GWAS for trotting technique synthetic traits

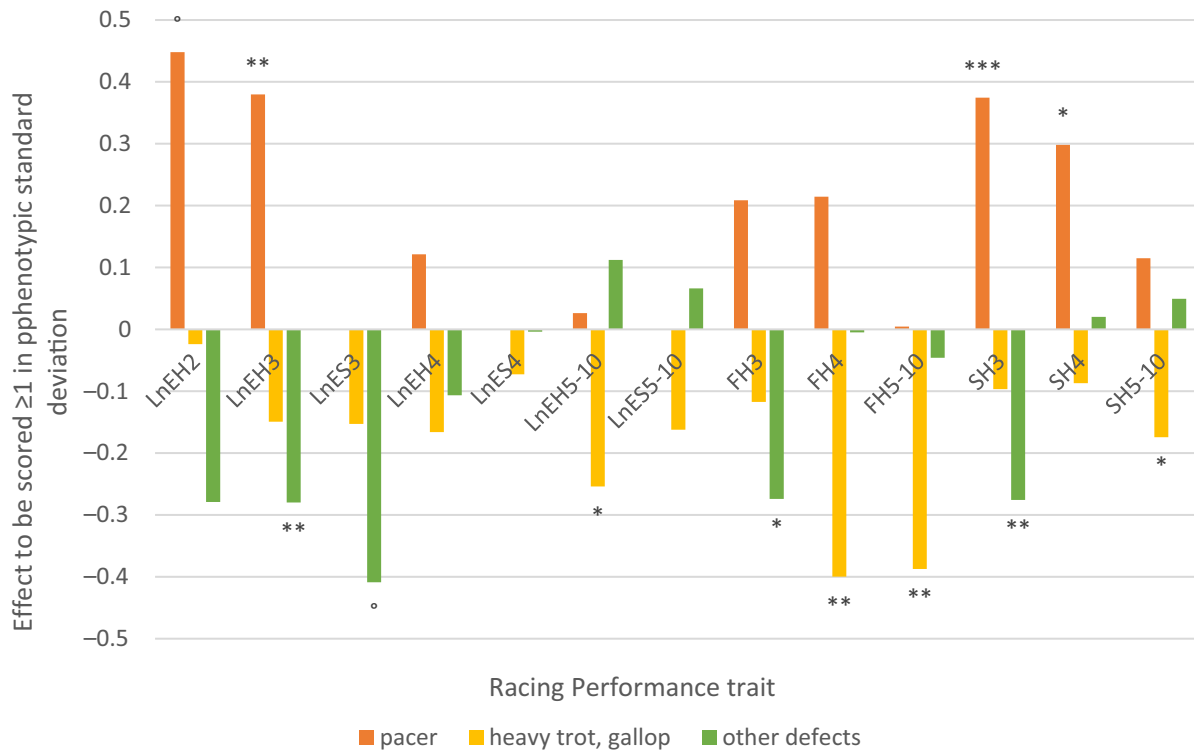
| SNP with highest <i>p</i> -value in the region | | Region | | | | | | | | | |
|--|---------------------|--------|------------------------|-----|--------------|------------------------|-------------------|---------------|----------|-------------|-------------|
| SNP name (Affymetrix) | SNP name (Illumina) | Chr. | Physical position (bp) | MAF | Minor allele | -log(<i>p</i> -value) | | | | | |
| | | | | | | Pacer | Heavy trot gallop | Other defects | <i>N</i> | Start (bp) | End (bp) |
| AX-103316285 | | 3 | 117,737,775 | 39% | G | | | 24.4 | 1 | 117,737,775 | 117,737,775 |
| AX-103765309 | | 6 | 23,004,247 | 8% | A | | 5.5 | | 1 | 23,004,247 | 23,004,247 |
| AX-104291172 | | 6 | 28,033,876 | 36% | T | | 14.2 | 9.0 | 1 | 28,033,876 | 28,033,876 |
| AX-103124547 | | 10 | 80,332,761 | 5% | G | 5.3 | | | 1 | 80,332,761 | 80,332,761 |
| AX-103701100 | BIEC2-445347 | 19 | 52,921,610 | 11% | C | | 5.3 | | 1 | 52,921,610 | 52,921,610 |
| AX-103675442 | BIEC2-724869 | 23 | 22,578,702 | 22% | G | 8.6 | 16.1 | | 119 | 19,618,796 | 2,3,666,303 |
| AX-104504719 | | 28 | 4,460,564 | 11% | T | 5.7 | | | 1 | 4,460,564 | 4,460,564 |
| AX-103598216 | | 28 | 6,666,464 | 8% | G | 6.0 | | | 1 | 6,666,464 | 6,666,464 |
| | | 28 | 8,749,086 | 15% | A | 6.6 | | | 5 | 8,140,111 | 8,749,086 |

Abbreviations: bp, base pairs; Chr, chromosome; MAF, minor allele frequency; *N*, number of SNPs in the region.

these studies. In the Standardbred population, the A allele is fixed. McCoy et al. (2019) succeeded to distinguish strictly trotters and pacers with random forest performed with 303 variants obtained by whole-genome sequencing in regions selected from GWAS analysis. Although pacers and trotters belong to the same Standardbred breed, they are genetically distinct populations. This explains their easy separation in two groups, which is not the case for trotting technique in FT, where no different lines such as “pacers” existed. In FT, the gait questionnaire did not prevent the owner from learning about their horse's DMRT3 genotype if they wished to know it.

Mechanism of DMRT3 genotype CA or CC on racing performances was a combination of effects of trotting technique related to pace and defect at gallop: less earnings at 2, and 3 years old (“pace” effect) higher percentage of disqualified races at all ages (“gallop” effect). But it cannot explain the advantage of CA horses in races under saddle and late harness races for earnings.

Results found for DMRT3 effect on racing performance confirmed those demonstrated by Ricard (2015) on FT with partially using the same study sample. DMRT3 cannot be studied in the Standardbred population from the United States, either trotter or pacer, because they are all of the same AA genotype (Promerova et al., 2014). However, the first favourable effect on trotting racing performance was found in the Swedish Standardbred, where the CA genotype occurred due to mating with FT (Andersson et al., 2012; Jaderkvist, Andersson, et al., 2014). In local trotter breeds, polymorphisms still exist. In Finnhorses, Fegraeus et al. (2015) found a favourable effect on all racing performance criteria of the AA genotype compared with CA and CC genotypes, regardless of the horses' age (3–6 or 10 years). Moreover, the effect of the AA genotype in Swedish Norwegian Coldblooded trotter (CBT) was controversial. Firstly, Jaderkvist, Andersson, et al. (2014, 2014) found a superiority of the AA genotype compared with CA and CC. Nevertheless, the significant difference at the age of 3 years disappeared mostly at 3–6 years of age and a non-significant tendency of the CA genotype being superior to AA for earnings over 3 to 6 years of age appeared. The following study on CBT (Fegraeus, Lawrence, et al., 2017) found no effect of the AA genotype on precocity of the racing performances and only two traits with significant differences between genotypes: AA having fastest times and CC having the highest number of disqualifications. At the age of 3–6 years, the CA genotype was found to be superior to the CC genotype (except for frequency of disqualification), but AA showed no superiority over CA. More this study found a low proportion of racehorses with the AA genotype, suggesting an association between AA genotype and poor durability. In the Spanish trotter, Rama et al. (2016) found AA to be best for speed in short and long distance races at young and adult age. In these breeds, the frequency of each genotype



Q = qualification, LnEHx = logarithm of earnings per finished harness races at the age of x with x = 2, 3, 4 and from 5 to 10 years old, LnESx = logarithm of earnings per finished races under saddle at the age of x with x = 3, 4 and from 5 to 10 years old, FHx = Proportion of finished Harness races at the age of x with x = 3, 4, 5-10, SHx = Number of started harness races at the age of x with x = 3, 4, 5-10.

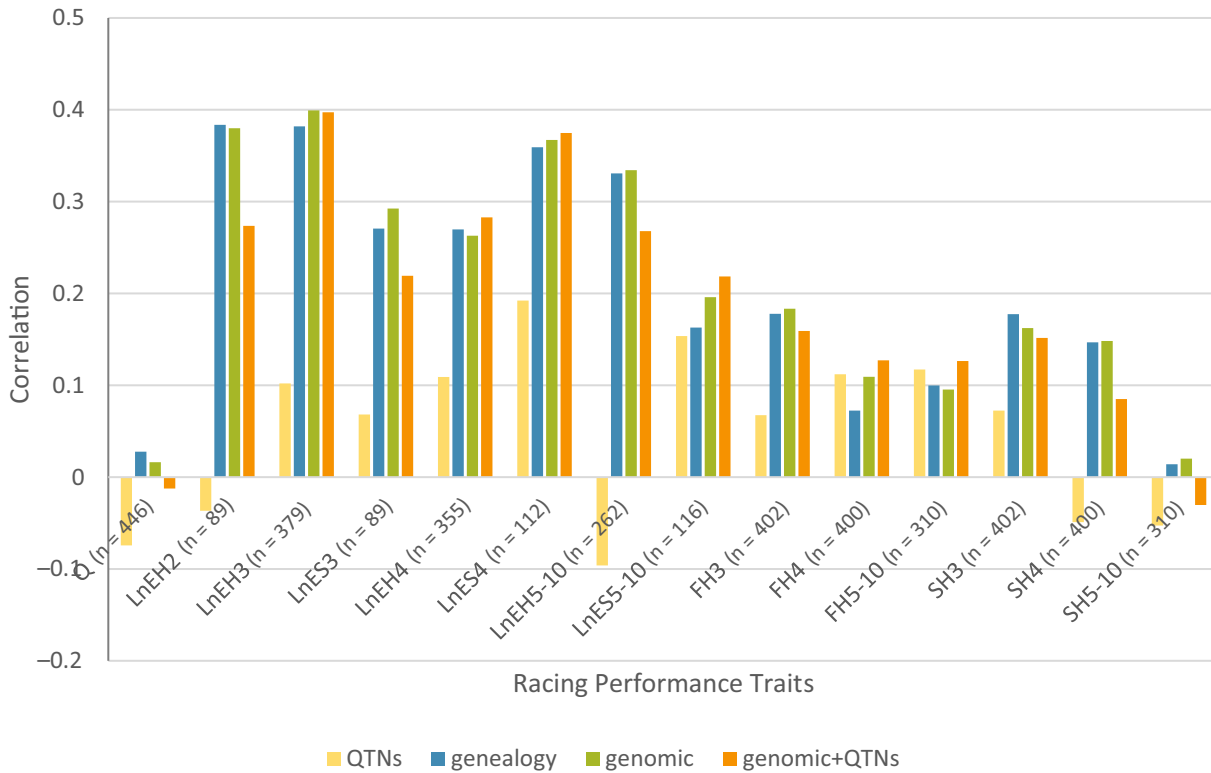
FIGURE 7 Effect to be scored ≥ 1 (with defect) as difference from 0 (without defect) for trotting technique synthetic variables on racing performances. [Colour figure can be viewed at wileyonlinelibrary.com]

was rather different; in CBT, the genotype frequency of AA, CA and CC was 9%, 51% and 40%, respectively, whereas it was determined to be 37%, 52% and 11% in Finnhorses and 76%, 21% and 2% in Spanish Trotter. The FTs in our study represent an intermediate case, with the genotype frequency of AA, CA and CC of 59%, 37% and 4%.

The results of our analysis of the region around DMRT3 showed no other clear effects in addition to the effect of the linkage disequilibrium with DMRT3 genotype. These results are in accordance with those previously demonstrated by Staiger et al. (2017), which also showed that the mutation had initially appeared around 1,000 years ago. Their results provided conclusive evidence that the DMRT3:Ser301STOP mutation is causal, as no other sequence polymorphisms showed an equally strong association with locomotion traits. In this study, FT was found within the 2 most common haplotypes, with either the C or A allele of the DMRT3 mutation. Therefore, the SNPs we found around DMRT3 were, perhaps, merely signals due to linkage disequilibrium. However, Velie et al. (2018) found a significant association between specific SNPs and career earnings, best km time and number of gallops, on the chromosome 23 between

positions 21,064,571 and 23,333,501 to other genes in the region. However, only two of these SNPs were significant in our analysis. One was in strong linkage disequilibrium with DMRT3 (0.77), but the other was not (0.31). The SNP AX-102982528 with the nearest gene of ENSECAG00000023609 may be a promising lead to follow. Other significant SNPs in Velie et al. (2018) linked to DOCK8 (dedicator of cytokinesis 8), PTAR1 or PIP5K1B (phosphatidylinositol-4-phosphate 5-kinase type-1 beta) were not found to be significant or had a MAF that was too low to be retained.

Therefore, selection for A Allele of DMRT3 mutation will lead to horses more able to pace, less to gallop, but with equal other problems in trotting technique than other genotypes, with higher probability to be qualified and to do a 2-year successful season of racing, and lower percentage of disqualified races. But it also leads to horses less suitable for races under saddle and horses which won in the second part of the career after 5 years old. This balance in the effect of DMRT3 explained why the gene is still segregating in FT population while affecting the main selection objective since decades. Therefore, depending of the wanted racing programme, the genomic information may help differently the selection of FT.



Footnote: QTNs = genetic evaluation as sum of genotypes estimates of significant QTNs, genealogy = multiple trait genetic evaluation using pedigree relationship matrix, genomic = multiple trait genetic evaluation using pedigree and genomic relationship matrix in a Single Step, genomic+QTNs = as genomic plus sum of QTNs genotypes effects.

Q = qualification, LnEHx = logarithm of earnings per finished harness races at the age of x with x = 2, 3, 4 and from 5 to 10 years old, LnESx = logarithm of earnings per finished races under saddle at the age of x with x = 3, 4 and from 5 to 10 years old, FHx = Proportion of finished Harness races at the age of x with x = 3, 4, 5-10, SHx = Number of started harness races at the age of x with x = 3, 4, 5-10.

FIGURE 8 Correlation between genetic evaluation and racing performances on validation sample (n = 446) with different evaluation models. [Colour figure can be viewed at wileyonlinelibrary.com]

The first deal for breeding of French trotters may be to improve trotting technique, but, as we did not find QTNs other than DMRT3 for both trotting technique variables and racing performance, improving trotting technique with such QTNs does not assure improving racing performance but perhaps an easy training. A strong QTNs on chromosome 28 increase the ability to pace. The frequency of the allele which increases the ability to pace is low (<15%), so the room for improvement is large. But the allele effect does not prevent significantly from galloping. Perhaps the favourable effect of DMRT3 allele A on performance is its prevention against gallop rather than its positive influence on pace. In that case, the benefic influence of A 28:8,749,086 may not be as suitable as it seems. To prevent from gallop, QTNs found are already in frequent proportion (92%, 89%) except one on chromosome 6:28,033,876. For this QTN, the favourable allele reduces

tendency to gallop and to have other problems and have a frequency of 36% which could be increased.

The second deal is to focus directly on racing performances. Success in races is a complex trait. A carrier in trotting races required first to be qualified and then to succeed at 3 and 4 years old to be able to continue at 5 years and older (strong selection at 5 years old due to total earning conditions to enter a race). The success in one season in trotting races is the product of earning per finished races, number of finished races per start and number of starts. Each component of the performance is important to obtain finally high total earnings. Genetic correlations between all these traits have already been calculated in FT (Ricard, 2015). Moderate positive genetic correlations were found between qualification (0.42–0.62) and all earnings traits. Between earnings traits, except a moderate

correlation with LnE2, all other correlations between earnings at different ages were very high (0.83–0.93). High genetic correlations were also found between harness and under saddle traits (0.66–0.87). Genetic correlations between earnings traits and proportion of finished races were close to zero or even slightly negative (–0.10 to –0.22). Genetic correlations with number of starts were close to zero, except for the range of 5 to 10 years old where the trait represented more the longevity of the horse over all these years and where the genetic correlation was high with earnings (0.69). If the purpose of the selection is to improve each of these traits, without objective to break the existing genetic correlation between them, “genealogy” or “genomic” evaluation was proved to be more efficient than all other QTN strategies (“QTNs” or “genomic + QTNs”). There were several QTN regions identified for various racing performance traits. However, from a practical point of view, “genealogy” evaluation outperformed the knowledge of several QTNs to predict performances in races. “Genealogy” evaluation provided a satisfactory level of prediction, and “genomic” evaluation was slightly better in most cases. This is the strength of multiple trait approach for “genealogy”/“genomic” evaluation. The tiny exception was the trait percentage of finished races. Slightly negative genetic correlations made classical selection of breeding values difficult. In that case, the help of the 5 QTNs with a favourable effect on both earnings and percentage of finished races may be interesting (in 3 cases, the rare homozygote is favourable so the room of improvement is important). We did not find any help from QTL found in other studies. Velie et al. (2018) performed GWAS for race performances in Norwegian Swedish Coldblooded. Further, Fegraeus et al. (2018) selected SNPs from delta fixation index analysis comparing Coldblooded trotters, North Swedish draught horses and Standardbreds. Then, they analysed differences in race performances for Coldblooded trotters for the genotypes of the most promising one near EDN3 gene. Rama et al. (2016) tested the effect of genotypes at 4 major genes other than DMRT3 on race performances of the Spanish trotter. None of the significant SNPs obtained by these authors were significant in our study. More than 50% of the significant SNPs were not in the SNP list retained in our study, principally due to low MAF or low other quality controls. We also checked SNPs found in GWAS analysis in gaited horses (Amano et al., 2018; Bussiman et al., 2020; Fegraeus, Hirschberg, et al., 2017; Staiger et al., 2016) and those found in Standardbred to differentiate trotters from pacers (McCoy et al., 2019) with no success. Particular adaptations of each breed or population to the same racing performance depended on different genes according to its overall genetic background. The DMRT3 mutation influences a number of breeds for the large effect on laterality of gaits. However, this is a rare

example of genes which influences different breeds in the same way. For the details of complex traits such as race success, many different qualities and local adaptation are needed for specific breeds.

5 | CONCLUSION

Trotting technique, as measured by our questionnaire, influences racing performances. The main characteristic of trotting technique is the link between the ability to pace which avoid the risk to gallop in races. Horses which are able to pace and with no defect at gallop had easier qualification, earlier good performances and a lower percentage of disqualified races, even late in the career. The only QTL found with an effect on trotting technique and racing performances was the major gene mutation in DMRT3 shown to be responsible for pacing. Rather than favouring pace, the favourable effect of genotype AA of DMRT3 in trotting races was to prevent from gallop. We could not explain the remaining slightly favourable effect of heterozygote CA on late performances. CA horse was similar to CC horse for the characterization of trotting technique at first, but perhaps it is possible to teach CA horses more easily to prevent from gallop than CC with time.

Selection for the A allele of DMRT3-orientated race profiles included earlier performances and lower races under saddle. Genomic evaluation with multiple traits models improved more efficiently earnings per finished races for both harness races and races under saddle than this selection for one mutation. However, with this quantitative genetic evaluation, it remains difficult to improve both earnings per finished races and percentage of finished races because they are genetically independent traits. We found other QTL, presumably not responsible of gait traits but of other qualities needed to perform in races, which were able to improve both earnings per finished races and percentage of finished races. These QTL may be new tools for improvement of French trotters.

ACKNOWLEDGEMENTS

This work was supported by IFCE (Institut Français du Cheval et de l'équitation) and Fonds Eperon. We thank all owners of the horses and Le Trot which provided performance data.

CONFLICT OF INTEREST

The authors declare that they do not have any conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not

publicly available due to privacy restrictions: genotyping and questionnaires were authorized by owners only for the research involved. Results of races are publicly available on <https://www.letrot.com/fr/>.

ORCID

Anne Ricard  <https://orcid.org/0000-0002-2113-5153>

REFERENCES

- Aguilar, I., Legarra, A., Cardoso, F., Masuda, Y., Lourenco, D., & Misztal, I. (2019). Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. *Genetics Selection Evolution*, *51*, 28. <https://doi.org/10.1186/s12711-019-0469-3>
- Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., & Lawlor, T. J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science*, *93*, 743–752. <https://doi.org/10.3168/jds.2009-2730>
- Aguilar, I., Tsuruta, S., Masuda, Y., Lourenco, D., Legarra, A., & Misztal, I. (2018). *BLUPF90 suite of programs for animal breeding with focus on genomics*. Paper presented at the Proceedings of the world congress on genetics applied to livestock production: 11–16 February 2018; Auckland 2018.
- Amano, T., Onogi, A., Yamada, F., Kawai, M., Shirai, K., & Ueda, J. (2018). Genome-wide association mapping and examination of possible maternal effect for the pace trait of horses. *Animal Genetics*, *49*, 461–463. <https://doi.org/10.1111/age.12711>
- Andersson, L. S., Larhammar, M., Memic, F., Wootz, H., Schwochow, D., Rubin, C. J., Patra, K., Arnason, T., Wellbring, L., Hjälm, G., Imsland, F., Petersen, J. L., McCue, M. E., Mickelson, J. R., Cothran, G., Ahituv, N., Roepstorff, L., Mikko, S., Vallstedt, A., ... Kullander, K. (2012). Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. *Nature*, *488*, 642–646. <https://doi.org/10.1038/nature11399>
- Brard, S., & Ricard, A. (2015). Should we use the single nucleotide polymorphism linked to DMRT3 in genomic evaluation of French trotter? *Journal of Animal Science*, *93*, 4651–4659. <https://doi.org/10.2527/jas.2015-9224>
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. <https://doi.org/10.1023/a:1010933404324>
- Bussiman, F. D. O., Aparecida dos Santos, B., Abreu Silva, B. D. C., Mamani, G. C. M., Grigoletto, L., Pereira, G. L., Ferraz, G. D. C., Ferraz, J. B. S., Mattos, E. C., Eler, J. P., Ventura, R. V., Curi, R. A., & Balieiro, J. C. D. C. (2020). Genome-wide association study: Understanding the genetic basis of the gait type in Brazilian Mangalarga Marchador horses, a preliminary study. *Livestock Science*, *231*, 103867. <https://doi.org/10.1016/j.livsci.2019.103867>
- Bussiman, F. O., dos Santos, B. A., Silva, B. C. A., Perez, B. C., Pereira, G. L., Chardulo, L. A. L., Eler, J. P., Ferraz, J. B., Mattos, E. C., Curi, R. A., & Balieiro, J. C. C. (2019). Allelic and genotypic frequencies of the DMRT3 gene in the Brazilian horse breed Mangalarga Marchador and their association with types of gait. *Genetics and Molecular Research*, *18*, grm18217. <https://doi.org/10.4238/gmr18217>
- Chassier, M., Barrey, E., Robert, C., Duluard, A., Danvy, S., & Ricard, A. (2018). Genotype imputation accuracy in multiple equine breeds from medium- to high-density genotypes. *Journal of Animal Breeding and Genetics*, *135*, 420–431. <https://doi.org/10.1111/jbg.12358>
- Christensen, O. F., & Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution*, *42*, 2. <https://doi.org/10.1186/1297-9686-42-2>
- Cothran, E. G., MacCluer, J. W., Weitkamp, L. R., & Bailey, E. (1987). Genetic differentiation associated with gait within American Standardbred horses. *Animal Genetics*, *18*, 285–296. <https://doi.org/10.1111/j.1365-2052.1987.tb00772.x>
- Devlin, B., & Roeder, K. (1999). Genomic control for association studies. *Biometrics*, *55*, 997–1004. <https://doi.org/10.1111/j.0006-341X.1999.00997.x>
- Fegraeus, K. J., Hirschberg, I., Arnason, T., Andersson, L., Velie, B. D., Andersson, L. S., & Lindgren, G. (2017). To pace or not to pace: A pilot study of four- and five-gaited Icelandic horses homozygous for the DMRT3 'Gait Keeper' mutation. *Animal Genetics*, *48*, 694–697. <https://doi.org/10.1111/age.12610>
- Fegraeus, K. J., Johansson, L., Mäenpää, M., Mykkänen, A., Andersson, L. S., Velie, B. D., Andersson, L., Árnason, T., & Lindgren, G. (2015). Different DMRT3 genotypes are best adapted for harness racing and riding in Finnhorses. *Journal of Heredity*, *106*, 734–740. <https://doi.org/10.1093/jhered/esv062>
- Fegraeus, K. J., Lawrence, C., Petäjistö, K., Johansson, M. K., Wiklund, M., Olsson, C., Andersson, L., Andersson, L. S., Røed, K. H., Ihler, C. F., Strand, E., Lindgren, G., & Velie, B. D. (2017). Lack of significant associations with early career performance suggest no link between the DMRT3 "Gait Keeper" mutation and precocity in Coldblooded trotters. *PLoS One*, *12*, e0177351. <https://doi.org/10.1371/journal.pone.0177351>
- Fernando, R. L., Dekkers, J. C. M., & Garrick, D. J. (2014). A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. *Genetics Selection Evolution*, *46*(1), 46–50. <https://doi.org/10.1186/1297-9686-46-50>
- Fonseca, M. G., Ferraz, G. D., Lage, J., Pereira, G. L., & Curi, R. A. (2017). A genome-wide association study reveals differences in the genetic mechanism of control of the two gait patterns of the Brazilian Mangalarga Marchador breed. *Journal of Equine Veterinary Science*, *53*, 64–67. <https://doi.org/10.1016/j.jevs.2016.01.015>
- Giantsis, I. A., Diakakis, N. E., & Avdi, M. (2018). genetic composition and evaluation of the status of a non-descript indigenous horse population from Greece, the Macedonian Pacer. *Journal of Equine Veterinary Science*, *71*, 64–70. <https://doi.org/10.1016/j.jevs.2018.10.003>
- Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., & Thompson, R. (2014). *ASReml user guide release 4.1 structural specification*. VSN International Ltd.
- Han, Y., & Peñagaricano, F. (2016). Unravelling the genomic architecture of bull fertility in Holstein cattle. *BMC Genetics*, *17*, 143. <https://doi.org/10.1186/s12863-016-0454-6>
- Jäderkvist Fegraeus, K., Velie, B. D., Axelsson, J., Ang, R., Hamilton, N. A., Andersson, L., Meadows, J. R. S., & Lindgren, G. (2018). A potential regulatory region near the EDN3 gene may control both harness racing performance and coat color variation in horses. *Physiological Reports*, *6*(10), e13700. <https://doi.org/10.14814/phy2.13700>
- Jäderkvist, K., Andersson, L. S., Johansson, A. M., Árnason, T., Mikko, S., Eriksson, S., Andersson, L., & Lindgren, G. (2014). The DMRT3 'Gait keeper' mutation affects performance of Nordic and Standardbred trotters. *Journal of Animal Science*, *92*, 4279–4286. <https://doi.org/10.2527/jas2014-7803>

- Jaderkvist, K., Holm, N., Imsland, F., Arnason, T., Andersson, L., Andersson, L. S., & Lindgren, G. (2015). The importance of the DMRT3 'Gait keeper' mutation on riding traits and gaits in Standardbred and Icelandic horses (vol 176, pg 33, 2015). *Livestock Science*, *180*, 268. <https://doi.org/10.1016/j.livsci.2015.08.005>
- Jaderkvist, K., Kangas, N., Andersson, L. S., & Lindgren, G. (2014). Gaitedness is associated with the DMRT3 'Gait keeper' mutation in Morgan and American Curly horses. *Animal Genetics*, *45*, 908–909. <https://doi.org/10.1111/age.12228>
- Kristjansson, T., Bjornsdottir, S., Sigurdsson, A., Andersson, L. S., Lindgren, G., Helyar, S. J., Klonowski, A. M., & Arnason, T. (2014). The effect of the 'Gait keeper' mutation in the DMRT3 gene on gaiting ability in Icelandic horses. *Journal of Animal Breeding and Genetics*, *131*, 415–425. <https://doi.org/10.1111/jbg.12112>
- Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software*, *25*, 18. <https://doi.org/10.18637/jss.v025.i01>
- Legarra, A., Christensen, O. F., Aguilar, I., & Misztal, I. (2014). Single Step, a general approach for genomic selection. *Livestock Science*, *166*, 54–65. <https://doi.org/10.1016/j.livsci.2014.04.029>
- Li, M.-X., Yeung, J. M. Y., Cherny, S. S., & Sham, P. C. (2012). Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Human Genetics*, *131*, 747–756. <https://doi.org/10.1007/s00439-011-1118-2>
- Liaw, A., & Wiener, M. (2018). *Breiman and Cutler's random forests for classification and regression, package random forest*.
- McCoy, A. M., Beeson, S. K., Rubin, C.-J., Andersson, L., Caputo, P., Lykkjen, S., Moore, A., Piercy, R. J., Mickelson, J. R., & McCue, M. E. (2019). Identification and validation of genetic variants predictive of gait in standardbred horses. *Plos Genetics*, *15*(5), e1008146. <https://doi.org/10.1371/journal.pgen.1008146>
- Negro Rama, S., Valera, M., Membrillo, A., Gómez, M. D., Solé, M., Menendez-Buxadera, A., Anaya, G., & Molina, A. (2016). Quantitative analysis of short- and long-distance racing performance in young and adult horses and association analysis with functional candidate genes in Spanish Trotter horses. *Journal of Animal Breeding and Genetics*, *133*, 347–356. <https://doi.org/10.1111/jbg.12208>
- Novoa-Bravo, M., Fegraeus, K. J., Rhodin, M., Strand, E., Garcia, L. F., & Lindgren, G. (2018). Selection on the Colombian paso horse's gaits has produced kinematic differences partly explained by the DMRT3 gene. *PLoS One*, *13*(8), e0202584. <https://doi.org/10.1371/journal.pone.0202584>
- Patterson, L., Staiger, E. A., & Brooks, S. A. (2015). DMRT3 is associated with gait type in Mangalarga Marchador horses, but does not control gait ability. *Animal Genetics*, *46*, 213–215. <https://doi.org/10.1111/age.12273>
- Pereira, G. L., de Matteis, R., Regitano, L. C. A., Chardulo, L. A. L., & Curi, R. A. (2016). MSTN, CKM, and DMRT3 gene variants in different lines of Quarter Horses. *Journal of Equine Veterinary Science*, *39*, 33–37. <https://doi.org/10.1016/j.jevs.2015.09.001>
- Promerová, M., Andersson, L. S., Juras, R., Penedo, M. C., Reissmann, M., Tozaki, T., Bellone, R., Dunner, S., Hořín, P., Imsland, F., Imsland, P., Mikko, S., Modrý, D., Roed, K. H., Schwochow, D., Vega-Pla, J. L., Mehrabani-Yeganeh, H., Yousefi-Mashouf, N., Cothran, E. G., ... Andersson, L. (2014). Worldwide frequency distribution of the 'Gait keeper' mutation in the DMRT3 gene. *Animal Genetics*, *45*, 274–282. <https://doi.org/10.1111/age.12120>
- Regatieri, I. C., Eberth, J. E., Sarver, F., Lear, T. L., & Bailey, E. (2016). Comparison of DMRT3 genotypes among American Saddlebred horses with reference to gait. *Animal Genetics*, *47*, 603–605. <https://doi.org/10.1111/age.12458>
- Regatieri, I. C., Pereira, G. L., Neto, A. R. T., Ferraz, G. C., Curi, R. A., & Queiroz-Neto, A. (2017). Polymorphisms in MCT1, CD147, PDK4, and DMRT3 genes in Arabian and Quarter Horses. *Journal of Equine Veterinary Science*, *48*, 161–165. <https://doi.org/10.1016/j.jevs.2016.06.077>
- Ricard, A. (2015). Does heterozygosity at the DMRT3 gene make French trotters better racers? *Genetics Selection Evolution*, *47*, 10. <https://doi.org/10.1186/s12711-015-0095-7>
- Ricard, A., Danvy, S., & Legarra, A. (2013). Computation of deregressed proofs for genomic selection when own phenotypes exist with an application in French show-jumping horses. *Journal of Animal Science*, *91*, 1076–1085. <https://doi.org/10.2527/jas.2012-5256>
- Sargolzaei, M., Chesnais, J. P., & Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, *15*, 478. <https://doi.org/10.1186/1471-2164-15-478>
- SAS/STAT (2019). *SAS/STAT 14.1, Copyright © [2002-2012] SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc.*
- Staiger, E. A., Abri, M. A., Silva, C. A. S., & Brooks, S. A. (2016). Loci impacting polymorphic gait in the Tennessee Walking Horse. *Journal of Animal Science*, *94*, 1377–1386. <https://doi.org/10.2527/jas.2015-9936>
- Staiger, E. A., Almén, M. S., Promerová, M., Brooks, S., Cothran, E. G., Imsland, F., Jäderkvist Fegraeus, K., Lindgren, G., Mehrabani Yeganeh, H., Mikko, S., Vega-Pla, J. L., Tozaki, T., Rubin, C. J., & Andersson, L. (2017). The evolutionary history of the DMRT3 'Gait keeper' haplotype. *Animal Genetics*, *48*, 551–559. <https://doi.org/10.1111/age.12580>
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Velie, B. D., Fegraeus, K. J., Solé, M., Rosengren, M. K., Røed, K. H., Ihler, C.-F., Strand, E., & Lindgren, G. (2018). A genome-wide association study for harness racing success in the Norwegian-Swedish coldblooded trotter reveals genes for learning and energy metabolism. *BMC Genetics*, *19*, 80. <https://doi.org/10.1186/s12863-018-0670-3>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Ricard A, Duluard A.

Genomic analysis of gaits and racing performance of the French trotter. *J Anim Breed Genet*.

2021;138:204–222. <https://doi.org/10.1111/jbg.12526>