Association between candidate gene markers and harness racing traits in Italian trotter horses

Association between candidate gene markers and harness racing traits in Italian Trotter horses

Stefania Dall'Olio, Samuele Bovo, Silvia Tinarelli, Giuseppina Schiavo, Barbara Padalino, Luca Fontanesi

Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, Viale Giuseppe Fanin 46, 40127 Bologna, Italy

Correspondence:

S. Dall'Olio; e-mail: stefania.dallolio@unibo.it

Running head: SNPs associated with Italian Trotter horse performances

Abstract

Knowledge of genes involved in the variability of performance of racehorses is, in general, limited. In this study, six single nucleotide polymorphisms (SNPs) of five genes were genotyped in 412 Italian Trotter horses (ITHs) and possible associations between five segregating SNPs and harness racing performance (best racing time in career, earnings, placings, starts, and wins) were investigated. These markers were in genes related to exercise physiology (creatine kinase muscle, *CKM*; and cytochrome c oxidase, subunit 4, isoform 2, *COX412*), gait patterns (*DMRT3*), carbohydrates metabolism (*GCK*) and morphological phenotypes (*LCORL*). Two SNPs of the *GCK* gene were used to predict haplotypes. The mutate A allele of the *DMRT3* was almost fixed in the ITHs. Single marker association analyses showed that the *COX412* SNP was significantly associated with variation of number of starts (P = 0.0003), and wins (P = 0.0045). This *COX4I2* polymorphic site seems therefore to be a promising marker for racing longevity and durability. A haplotype of the *GCK* was instead positively associated with best racing time (P= 0.0118), earnings (P = 0.0370) and wins (P = 0.0061). This is the first study highlighting the effect of *GCK* on harness racing performance in horses. Confirmation of these results in independent populations would be important to evaluate the inclusion of the SNPs in marker-assisted selection programmes in ITHs.

Keywords: Association study; racing performance; genes; racehorse; *Equus caballus*.

1. Introduction

Many horse breeds are used for harness racing in different parts of the world (reviewed in Thiruvenkadan et al., 2009). The Italian Trotter horse breed (ITHs or Italian Standardbred) was developed over the last century by crossing English Thoroughbred stallions with native mares with an

aptitude for trotting. This Italian breed also experienced genetic influence from imported trotters such as French Trotter (characterized by resistance to fatigue also termed as stamina), Russian Trotter (speed in long-distance) and American Standardbred (speed) (Bigi and Zanon, 2020).

The current studbook of the ITHs is closed and allows the new registration of horses born from already registered parents; however, trotters registered elsewhere can be admitted if their racing records have achieved some specific requirements. In 2019 the population size of the ITHs was 2586 and the number of foals born was 2004 (National Trotter Horse Breeder Association; ANACT; http://www.anact.it). The breeding program of the ITHs includes phenotypic racing performances such as best racing time in career, total earnings, and rank traits as main selection criteria (Pieramati et al., 2007). Despite selection success for harness racing of the ITHs (e.g. Varenne's success who raced in seven countries, won 61 of his 73 races, and winnings of over 6 million € is thought to be a record), relatively little is known about the genetic factors that underpin trot racing performances of this breed.

The identification of polymorphic genetic markers associated with variability of horse racing performances is of great interest for the racing industry both economically and for animal wellbeing (Raudsepp et al., 2019). Association studies between genetic markers and horse racing performances have been performed in a number of horse breed (Pereira et al., 2016; Raudsepp et al., 2019). In Thoroughbreds, a breed specialized for gallop racing, a few markers associated with racing performances were in the creatine kinase muscle-specific (*CKM*) gene, involved in energy metabolism and resistance to fatigue, and in the cytochrome C oxidase, subunit 4, isoform 2 (*COX412*) gene, related to cellular respiration and exercise physiology (Gu et al., 2010). The doublesex and mab-3 related transcription factor 3 (*DMRT3*) gene is the main gene involved in neuro-sensorial coordination and gait patterns of horses, showing effects on harness racing performances (Andersson et al., 2012). Variability in the ligand-dependent nuclear receptor corepressor-like (*LCORL*) gene was associated with withers height and others morphological phenotypes (Makvandi-Nejad et al.,

2012; Metzger et al., 2013; Signer-Hasler et al., 2012; Tetens et al., 2013). The BIEC2-808543 single nucleotide polymorphism (SNP), located upstream of the *LCORL* gene, has been identified as a diagnostic marker associated with variation of morphological traits in several horse breeds (Makvandi-Nejad et al., 2012; Signer-Hasler et al., 2012; Tozaki et al., 2016).

A few polymorphisms have been found to occur at different frequencies across different horse breeds, such as the glucokinase (*GCK*) gene, related to glucose homeostasis and carbohydrate metabolism, and additional investigations could be carried out to determine possible associations with racing performance traits (Dall'Olio and Manieri, 2006, 2007). The main goal of this study was to investigate possible associations between markers in five candidate genes (*CKM*, *COX4I2*, *DMRT3, GCK* and *LCORL*) and the racing ability of the Italian Trotters.

2. Materials and methods

All procedures involving animals followed Italian and European Union regulations for animal care. Animal Care and Use Committee approval was not needed for this study because data were extracted from an already existing database. The authors had no direct control over the care of the animals included in this study. Hair samples were collected with owners' written consent and on a voluntary basis.

2.1. Sampling and phenotypes of the Italian Trotter horses

Hairs were collected from 412 ITHs, 256 females and 156 males (122 stallions and 34 geldings), descended from 136 stallions and 372 mares. Individual harness racing traits of the ITHs were derived from a public online database of the National Union for the Increasing of Horse Breeds (UNIRE; http://www.unire.gov.it). To have complete performance records for all horses, retrospective data of performances of horses born before

2010 were used. ITHs are allowed to race from 2 up to to 7 years of age if females and up to 10 years of age if males (https://www.politicheagricole.it/). Racing parameters investigated were: record or best racing time in career (the lowest recorded time performed on 1600-2000 meters, expresses as seconds/km), total earnings (Euro), starts or number of races (including the zero starts), number of placings (total number of times a horse finished a race in top-five places) and number of wins (times when a horse won the race).

Pedigree, date of birth, racing distance (m), kind of start and date of the best racing record were taken in the online database of UNIRE. All best racing times were obtained from auto-start races, as more than 99% of racing time were from this kind of start in the Italian Trotters (Pieramati et al., 2011). A sample of Thoroughbred (TH; 20 horses) was used for allele frequency determination. Those samples were collected in different horse stables in the North of Italy out a population of about 500 horses making sure that the sampled horses were offspring of different sires.

2.2. SNPs genotyping

Genomic DNA was extracted from hair roots using a standard protocol. Primer pairs were designed using Primer 3 (http://frodo.wi.mit.edu/primer3/input.htm) and dCaps Finder 2.0 (http://helix.wustl.edu/dcaps/; Neff et al., 2002) to amplify genomic regions encompassing *CKM*, *COX412, DMRT3, GCK* and BIEC2-808543 (thereafter referred as *LCORL*) SNPs. Details of polymorphisms (EquCab3.0 horse genome assembly. Ensembl database, release 100, April 2020) and PCR conditions were reported in Table S1. Amplicons were subjected to restriction fragment length polymorphism (RFLP) analyses (Table S1).

2.3 Statistical analyses

79 Hardy-Weinberg equilibrium of the genotyped SNPs was evaluated using the HWE software program (Linkage Utility Programs, Rockefeller
80 University, New York, NY, USA). Spearman correlations between racing traits of the ITHs were obtained with the procedure (PROC) CORR of SAS,
81 version 9.4 (SAS Institute Inc. Cary, NC).
82 Data were filtered according to a minimum number of five completed races (Pieramati et al., 2011) and gender, geldings (n= 34) were eliminated,
83 consequently, out of the initial 412 horses only 313 horses (208 females and 105 stallions) were retained for further analyses. For best racing time,
84 only the records obtained on a distance of 1600-1660 m were retained. Each trait was tested for normality using PROC UNIVARIATE of SAS. To
85 improve the approximation to the normal distribution, dependent variables were logarithmically transformed. Best racing time and earnings were
86 transformed in ln(best racing time-68.2) and ln(earnings), respectively, whereas other traits were log10 transformed (Árnason, 2001; Jäderkvist-
87 Fegraeus et al., 2017).
88 Single marker association analyses with transformed racing traits were carried out by the PROC GLM of SAS. All dependent variables were
89 adjusted for the fixed effects of marker genotypes (three levels), sex (females and males) and year of birth (six levels: <1999, 2000-2001, 2002-2003,
90 2004-2005, 2006-2007, 2008-2010). Year of birth was added in the model to take into account both genetic progress and improved environmental
91 conditions for training methods and the quality of the tracks (Thiruvenkadan et al., 2009). Number of starts was included as a covariate in the model
92 because an increase of starts was associated with an improvement of different harness traits. Initially, earning was used as a covariate for number of
93 starts, but it was then removed because it was not significant. Best racing time was adjusted also based on the age when the record was performed
94 (three levels: 2-3 years, 4 years, more than 5 years) and on the racing distance (two levels: 1600-1620 m, 1640-1660 m) as it is well known that
95 trotters' performance varies depending on the race distance and their age (Padalino et al., 2005). Additive effects were reported when the contrasts

between the two homozygous genotypes were significantly different from zero (P< 0.05). Dominant effects were reported when the contrasts between the average of the two homozygous genotypes and the heterozygous genotype were significantly different from zero (P< 0.05). Bonferroni correction was used to adjust P-nominal values for five segregating SNPs: threshold for significance was 0.05/5= 0.01. Results with P-nominal values ranging from 0.01 to 0.05 were considered as suggestive evidence of association.

The genotypes of two SNPs of the *GCK* gene, a SNP in the 5'-untranslated region (5'-UTR) of exon 1 B corresponding to upstream promoter, located at 460 bp upstream of the ATG start codon (g.14963036G>A, Dall'Olio and Minieri, 2007), and a missense mutation in exon 2 (g.14920889T>C giving the substitution of lysine with glutamic acid; Shubitowski et al., 2001) were used for haplotype reconstruction using PHASE (version 2.1) software (Stephens and Donnelly; 2003). Haplotypes with frequency > 0.05 were considered in the association analysis. Evaluation of the haplotype substitution effects on the performances was obtained using the PROC REG of SAS. The model in the association analysis included the number of each haplotype (0, 1 and 2 copies) and the same variation factors of the single marker association analyses.

3. Results

3.1. Description of harness racing traits of the Italian Trotter horses

Descriptive statistics of untransformed racing traits of 412 ITHs (not edited data) are shown in Table S2. The number of starts ranged from 0 to 204. The best racing time ranged from 71 s/km to 86 s/km (i.e. 1'.11''.00 and 1'.26''.00). Best racing time was obtained on a variety of distances (mean ± standard deviation equal to 1624 ± 68 m, range from 1600 m to 2100 m distance), of which 71.3 % at 1600-1620 m and 26.9 % at 1640-1660 m. The range of total earning was from 0.00 Euro to 1,144,344.80 Euro. Spearman's correlations between racing parameters based on time, earnings

and ranks are shown in Table S3. As expected, all traits were found to be significantly correlated ($P < 0.0001$) and correlations ranged from r= -0.40 (best racing time with both placings and starts) to r= +0.96 (placings with starts). Edited untransformed data (at least 5 starts) by sex are reported in Table S4. Due to the different carreer lenghts, males (M) were better than females (F) for number of starts (M= 65.30 ± 41.22, F= 51.04 ± 35.80), earnings (M = 101,687.68 ± 183,366.20, F = 39,087.57 ± 44,795.17), placings (M = 26.45 ± 19.44, F= 22.78 ± 18.58) and wins (M = 9.45 ± 7.92, F = 6.24 ± 5.76). However, as expected, it is worth noting that males had fastest best race time/recors than females (M = 73.83 ± 1.71 s/km, F = 74.98 ±1.64 s/km).

3.2. SNP and haplotype frequencies in Italian Trotters and Thoroughbreds

Allele frequencies of the SNPs analysed in the two racehorses are shown in Fig.1 and Table S5. Investigated SNPs did not deviate from Hardy-Weinberg equilibrium ($P > 0.05$). The major allele of the *CKM* gene was G in the ITHs (0.59) whereas the two alleles had equal frequency in the TH horses. Allele T of the *COX4I2* gene in ITHs and THs had frequency of 0.66 and 0.50, respectively. The majority of the ITHs was found to be homozygous for the mutant A allele (gait-keeper allele) of the *DMRT3* (genotypic frequency of 0.94). All THs were homozygous for the wild type C allele (non gait-keeper allele) at this SNP position. Allele T of the SNP located upstream the *LCORL* gene was the major allele in both ITH and TH breeds (0.72 and 0.80, respectively). Using the two polymorphisms of the *GCK* gene (g.14963036G>A and g.14920889T>C), four haplotypes were inferred in the ITHs. The most frequent haplotype was G:T (0.40), followed by G:C (0.33), A:C (0.24) and A:T (0.02).

3.3. SNPs and haplotypes associated with a few harness racing traits

Probability values (nominal P values) of factors included in the model of the single marker association analyses and coefficient of determination (R-square in percentage) are reported in Table S6. Sex had significant effects on all traits (as also expected from the description of the racing performances reported above), whereas year of birth was an important source of variation for all traits except for placings and start numbers (Table S6). The number of starts had significant effects on all traits except for best racing time. Year at record and race distance were significant on best racing time. Coefficients of determination was higher for number of placings (75.0-75.9 %) and number of wins (45.6-49.3%).

The T>C SNP of *COX4I2* was associated with number of starts ($P = 0.0003$), and suggestively associated with wins ($P = 0.0045$) (Tables 1 and S7). Least square means ± standard error of harness racing parameters obtained using genotypes of *COX4I2* gene are reported in Table 1. Number of starts was higher in genotype C/C compared to genotypes T/T ($P < 0.001$) and C/T ($P < 0.05$). For the number of wins, the comparison between T/T and C/C genotypes was significant ($P < 0.01$). The *COX4I2* polymorphism exhibited a significant additive effect (-13.33±3.49; P= 0.0002) for starts, whist a significant dominance effect (1.34±0.64; P= 0.0383) for wins was found.

GCK haplotype G:C, including G allele of the SNP in 5'-UTR of exon 1B and C (p.Glu51) allele, was associated with three traits (Tables 2 and S8): best racing time, earnings (both $P < 0.05$) and wins ($P < 0.0$). In the contrast analysis, horses with two copies of G:C haplotype showed better performances such as lower best racing time and both higher earnings and number of wins, compared to one copy and other haplotypes. The G:T and A:C haplotypes were not associated with any traits (Table 2).

4. Discussion

146 The initiale objective of this study was to evaluate possible associations between five candidate gene markers (SNPs in *CKM*, *COX4I2*, *DMRT3*,
147 *GCK* and *LCORL*) and harness racing performance parameters in ITHs. *COX4I2* SNP resulted associated with number of starts, and wins. While
148 number of starts may be regarded as a possible indicator of the soundness of a horse's basic conformation and a measure of sporting
149 longevity/durability (Árnason, 1994; Solé et al., 2017; Thiruvenkadan et al., 2009), wins seems to reflect the horse temperament and the will to win
150 (Thiruvenkadan et al., 2009).

151 *GCK* was associated not only with wins but also with best racing time, and earnings. Best racing time (record) is the direct measure of the speed
152 and the ability of trotting, and this trait is consequently very important to become popular in the racing industry (Árnason, 2001, Thiruvenkadan et
153 al., 2009). Earnings, the measure of racing success and win money, is instead the owners' and breeders' goal. Estimated heritability for best racing
154 time and earnings in the ITHs was found to be 0.43 and 0.21, respectively (Pieramati et al., 2007). In the literature racing time has always be found to
155 have the highest heritability among racing performance traits (Gómez et al., 2011; Thiruvenkadan et al., 2009).

156 In our study, allele frequencies of markers were estimated not only in ITHs but also in a sample of THs. Many studies have indeed investigated
157 whether the studied markers were associated with performance in Thoroughbreds. As expected, opposite alleles of *C>A* mutation of *DMRT3* were
158 relatively fixed in ITHs and THs. In the literature, the *DMRT3* gene has been linked to the growth of central circuit spinal interneurons during mammal
159 development and it was identified to be responsible for coordinating the movement of the limbs (Andersson et al., 2012). The mutant A allele of the
160 *DMRT3* gene causes indeed a premature stop codon resulting in a transcribed not functional protein that is permissive for the ability to perform
161 alternate gaits, showing a favourable effect on harness racing performances (Andersson et al., 2012).

The high frequency of mutant A allele in the ITHs (0.96) confirmed high frequency or fixation of this allele as it is reported in the literature for other breeds used for harness racing (Andersson et al., 2012; Negro Roma et al., 2016; Petersen et al., 2013; Promerova et al., 2014; Ricard, 2015). The wild type C allele was fixed in the Thoroughbred analysed as well as in other non-gaited horses bred for high-speed gallop, dressage, and show jumping (Anderssen et al., 2012; Promerová et al., 2014). The other polymorphisms investigated were found to be polymorphic in both racehorse breeds.

No associations of *CKM* and *LCORL* SNPs for different performances could be detected in the ITHs. The *CKM* gene is important for muscle energy metabolism during exercise as encoded enzyme playing a crucial role for the transport of high-energy phosphate to the cells where energy is demanded and several data supported that CKM may be responsible for fatigue resistance (Schröder et al., 2011). The SNP analysed of the *CKM* gene did not segregate in the Spanish Trotter horses (Negro Rama et al., 2016). In THs, A allele has been shown to be favourable for racing performance, with A/A and A/G genotypes being superior to homozygous G/G (Gu et al., 2010).

The *LCORL* SNP resulted associated with variability of wither height and other morphological traits in different horse breeds (Makvandi-Nejad et al., 2012; Signer-Hasler et al., 2012; Metzger et al., 2013; He et al., 2015; Takasuga, 2016; Tozaki et al., 2016). Many morphological traits are useful to categorize domestic horse breeds into diverse types (e.g. dolichomorphic, mesomorphic and brachimorphic types) (Dall'Olio et al., 2010). For instance, in ITHs and THs which are both dolichomorphic type breeds, T allele was major as reported in other light breeds (Metzger et al., 2013). The opposite C allele was major and near to fixation in brachymorphic type (or heavy**)** horses (Metzger et al., 2013).

The association analyses showed that the *COX4I2* gene marker was associated with two traits (starts, and wins) in the ITHs. The association with the number of starts is novel and it is worth highlithing that horses with allele C completed about 13 racing more than horses with allele T. This

179 could be due to the improved muscle metabolism favoured by this allele. *COX4I2* encodes indeed an enzyme that is one of the ten nuclear-coded
180 polypeptide chains of cytochrome c oxidase (COX), the terminal enzyme in the mitochondrial respiratory chain (Napiwotzki and Kadenbach, 1998).
181 The subunit 4 (COX4) exerts a physiological regulative function to adjust cellular energy production to demand and two isoforms, the ubiquitous
182 isoform 1 and the tissue-specific isoform 2, that are encoded by two different genes (*COX4I1* and *COX4I2*, respectively), exist.

183 The investigated polymorphism in the *COX4I2* gene is located within putative binding sites (glucocorticoid response element and p53 tumour
184 suppressor) that are erased with allele C (Gu et al., 2010). In a case-control study designed in Thoroughbreds, the frequency of the *COX4I2* T allele
185 was higher in the group with the best racecourse performance (elite in Flat race) compared with a group that had never won a race. In the same study,
186 a weak but significant association was found between the *COX4I2* genotypes and the distance of the highest grade Group race won by each individual
187 (Gu et al., 2010). No significant associations between the T>C SNP and best racing time of Spanish Trotters horses and speed index of Quarter Horses
188 have been identified (Negro Rama et al., 2016; Pereira et al., 2016).

189 In association analysis between genetic markers and equestrian athletic performance, major problems concern the variety of sports and
190 performances, and the same traits can be measured and analysed in many different ways (Rivero, 2007; Schröder et al., 2011; Thiruvenkadan et al.,
191 2009; Wylie and Newton, 2017; Raudsepp et al., 2019). Therefore, it is difficult to compare multiple studies analysing the same gene variants, as the
192 performance traits cannot be comparable. For example, gallop and trot are different gaits that require specific coordination patterns (as cited before)
193 and different selection purposes could have given functional and structural adaptations, including changes in physical conditions such as muscle
194 metabolism (i.e. density and volume of mitochondria, ability to increase the blood oxygen-carrying capacity, intramuscular storage of energy substrate,

195 stamina), muscle strength (power generation) and psychological factors such as temperament, and emotional reactions (Schröder et al., 2011; Rivero
196 and Hill., 2016).

197 The SNP of the 5'-UTR of exon 1B and missense mutation of the *GCK* gene have been analysed in several horse breeds reared in Italy (Dall'Olio
198 and Manieri, 2006; 2007), while, to our knowledge, this is the first association study with harness racing performance. Haplotype analyses of the *GCK*
199 gene allowed to discover association with three traits (best racing time, earnings, and wins). These findings may be related to the muscle fibers
200 presented in the horses and their large glycogen storage (Rivero and Hill., 2016). The use of two alternative promoters (upstream or 1B and downstream
201 or 1L) and alternative splicing of the *GCK* gene results in distinct isoforms that exhibit tissue-specific expression. In the pancreas and other
202 neuroendocrine cells (i.e. enteric and nervous system), the enzyme acts as the primary glucose sensor and is involved in glucose-stimulated insulin
203 secretion and glucose homeostasis. In the liver, the enzyme is important for glucose uptake and its conversion to glycogen (Matschinsky and Wilson;
204 2019).

205 In humans, mutations in the *GCK* gene have been associated with multiple types of diabetes and other metabolic disorders (Matschinsky and
206 Wilson; 2019). In cattle, the *GCK* has been detected as a putative candidate gene of signature selection (Sorbolini et al., 2016). Surprisingly, we found
207 an association with the evaluated characteristics when considering haplotypes, and the same did not occur when analyzing the SNPs individually. This
208 could be due to the additional linkage information that is catured by the haplotype analysis, as already demonstrated in other association analyses (e.g.
209 Bovo et al., 2020). However, further studies in independent populations of ITHs are needed to confirm the association of the *GCK* gene with harness
210 racing performance of horses.

211 Our results should be interpreted with cautions because this was a retrospective study with a few potential limits. Possible health disorders (i.e.
212 lameness, respiratory problems, injuries) which could have affected performance were not available and not considered. Other variation factors, such
213 as trainers and drivers were not included in the model. The study should be replicated and data related to performance should be collected prospectively
214 including information related to health, training and nutrition of the horses, which are well-known factors affecting performance (Thiruvenkadan et
215 al., 2009). The before mentioned limits may have contributed to the fact that the obtained R-square was below 76% indicating that the our model fails
216 to capture part of the variance in the dependent variables tested. However, the tested traits are polygenic and other genetic effects (i.e. polygenic effect,
217 individual effect) should be addressed. Future studies should consequently aim to collect data to performance models allowing to achieve higher R-
218 squares. Notwithstanding those limitations, this is the first study investigating possible associations between genetic markers and performance traits
219 in Italian Trotters.
220
221 **5. Conclusions**
222 The study investigated possible associations between markers in five candidate genes (*CKM*, *COX4I2*, *DMRT3*, *GCK* and *LCORL*) and the
223 racing ability of the Italian Trotters. Our results showed that a polymorphim in the *COX4I2* gene may be a promising marker for several harness racing
224 indicators including longevity and durability in Italian Trotter horses. Moreover, to the best of our knowledge, this study reported for the first time
225 significant associations between *GCK* and harness racing phenotypes in horses. However, additional genetic markers-phenotypes (harness racing
226 performance recorded prospectively) association studies across multiple replication cohorts with different performance backgrounds should be carried
227 out before incorporating these markers in breeding plans.

CRediT authorship contribution statement

Stefania Dall'Olio: Conceptualization, Supervision, Writing - Original Draft, Writing - Review & Editing, Funding acquisition. **Samuele Bovo:** SNP genotyping, Methodology, Data curation, Writing - Review & Editing **Silvia Tinarelli:** SNPs genotyping**,** Methodology, Data curation, Review & Editing. **Giuseppina Schiavo:** Formal analysis, Investigation, Methodology, Writing - Review & Editing. **Barbara Padalino:** Data curation, Review & Editing. **Luca Fontanesi:**, Supervision, Writing - Review & Editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The Authors thank all horse breeders donating hair samples from their horses. We thank the Italian Trotter Studbook for providing data. This study was funded by the University of Bologna RFO program [2017-2018].

References

Andersson, L.S., Larhammar, M., Memic, F., Wootz, H., Schwochow, D., Rubin, C.J., Patra, K., Árnason, 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., Lindgren, G., Andersson, L.,

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.

Árnason, T., 1994. The importance of different traits in genetic improvement of trotters. Proceedings of World Congress on Genetics Applied to Livestock Production. University of Guelph, Guelph, Canada. August 7–12, 1994;17: 462-469.

Árnason, T., 2001. Trends and asymptotic limits for racing speed in standardbred trotters. Livest. Prod. Sci. 72, 135–145. https://doi.org/10.1016/S0301-6226(01)00274-3.

Bigi, D., Zanon, A., 2020. Atlante delle razze autoctone. Bovini, equini, ovicaprini, suini allevati in Italia. Milan: Edagricole-New Business Media. ISBN-10**:** 8850655614; ISBN-13**:** 978-8850655618.

Bovo, S., Ballan, M., Schiavo, G., Gallo, M., Dall'Olio, S., Fontanesi, L. 2020. Haplotype-based genome-wide association studies reveal new loci for haematological and clinical-biochemical parameters in Large White pigs. Anim. Genet. 51, 601-606. https://doi.org/10.1111/age.12959.

Dall'Olio, S., Minieri, L., 2007. Analysis of upstream promoter region and corresponding 5' UTR of glucokinase (*GCK*) gene in horse breeds. Ital. J. Anim. Sci. 6 (SUPPL. 1), 106-108. https://doi.org/10.4081/ijas.2007.1s.106.

Dall'Olio, S., Fontanesi, L., Nanni Costa, L., Tassinari, M., Minieri, L., Falaschini, A., 2010. Analysis of horse myostatin gene and identification of single nucleotide polymorphisms in breeds of different morphological types. J. Biomed. Biotechnol. 2010. pii: 542945. https://doi.org/10.1155/2010/542945.

Dall'Olio, S., Minieri, L., 2006. Partial sequencing and polymorphisms of glucokinase gene (GCK) in horses reared in Italy. Proc. Italian Society for Veterinary Science 60, 499-599.

262 Gómez, M.D., Varona, L., Molina, A., Valera, M., 2011. Genetic evaluation of racing performance in trotter horses by competitive models. Livest.
263 Sci. 140, 155-160. https://doi.org/10.1016/j.livsci.2011.03.024.
264 Gu, J., MacHugh, DE., McGivney, BA., Park, SD., Katz, LM., Hill, EW., 2010. Association of sequence variants in CKM (creatine kinase, muscle)
265 and COX4I2 (cytochrome c oxidase, subunit 4, isoform 2) genes with racing performance in Thoroughbred horses. Equine Vet. J. Suppl. 569-
266 575. https://doi.org/10.1111/j.2042-3306.2010.00181.x.
267 He, S., Zhang, L., Li, W., Liu, M., 2015. BIEC2-808543 SNP in the LCORL Gene is Associated with Body Conformation in the Yili Horse. Anim.
268 Biotechnol. 26, 289-291. https://doi.org/10.1080/10495398.2014.995303.
269 Jäderkvist-Fegraeus, K., Lawrence, C., Petäjistö, K., Johansson, MK., Wiklund, M., Olsson, C., Andersson, L., Andersson, LS., Røed, KH., Ihler,
270 CF., Strand, E., Lindgren, G.,Velie, BD., 2017. Lack of significant associations with early career performance suggest no link between the
271 DMRT3 "Gait Keeper" mutation and precocity in Coldblooded trotters. PLoS One. 12, e0177351. https://doi.org/
272 10.1371/journal.pone.0177351.
273 Makvandi-Nejad, S., Hoffman, G.E., Allen, J.J., Chu, E., Gu, E., Chandler, A.M., Loredo, A.I., Bellone, R.R., Mezey, J.G., Brooks, S.A., Sutter, N.B.,
274 2012. Four loci explain 83% of size variation in the horse. PLoS One 7, e39929. https://doi.org/10.1371/journal.pone.0039929.
275 Matschinsky, FM., Wilson, DF., 2019. The Central Role of Glucokinase in Glucose Homeostasis: A Perspective 50 Years After Demonstrating the
276 Presence of the Enzyme in Islets of Langerhans. Front. Physiol. 10, 148. https://doi.org/10.3389/fphys.2019.00148.
277 Metzger, J., Schrimpf, R., Philipp, U., Distl, O., 2013. Expression levels of LCORL are associated with body size in horses. PLoS One 8, e56497.
278 https://doi.org/10.1371/journal.pone.0056497.

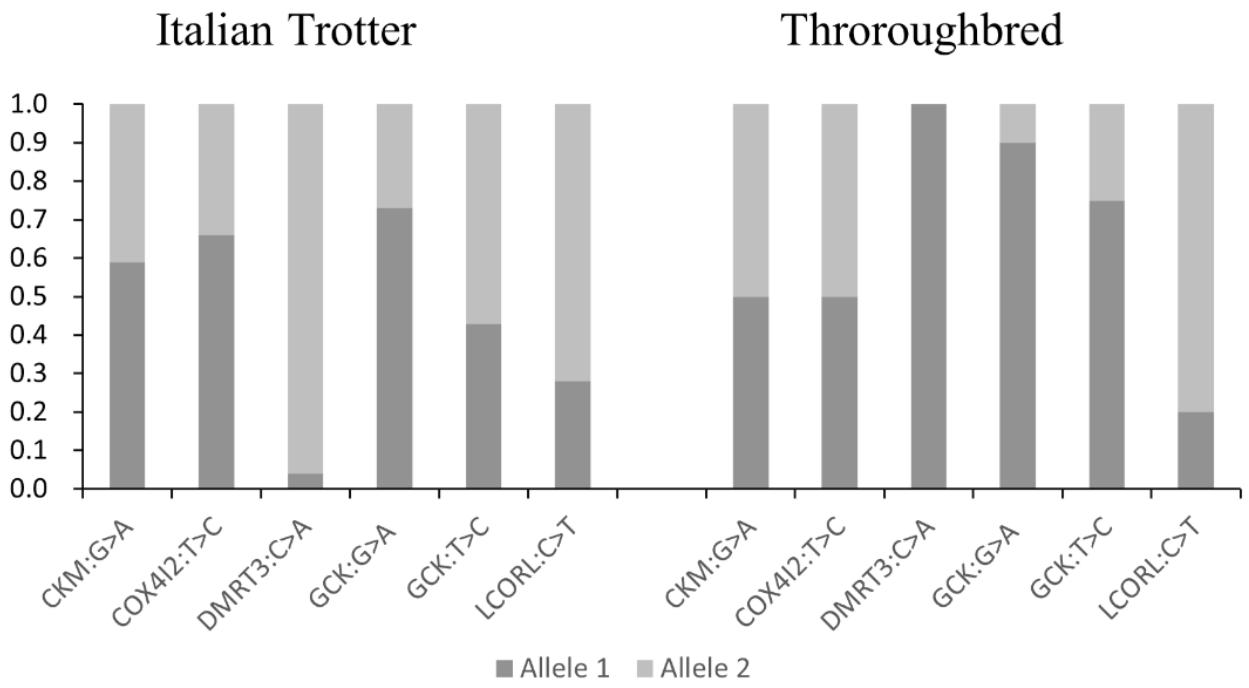
279 Napiwotzki, J., Kadenbach, B., 1998. Extramitochondrial ATP/ADP-ratios regulate cytochrome *c* oxidase activity via binding to the cytosolic domain
280 of subunit IV. Biol. Chem. 379, 335–339. https://doi.org/10.1515/bchm.1998.379.3.335.
281 Neff, MM., Turk, E., Kalishman, M., 2002. Web-based Primer Design for Single Nucleotide Polymorphism Analysis. Trends Genet. 18, 613-615.
282 https://doi.org/10.1016/s0168-9525(02)02820-2.
283 Negro Rama, S., Valera, M., Membrillo, A., Gómez, MD., Solé, M., Menendez-Buxadera, A., Anaya, G., Molina, A., 2016. Quantitative analysis of
284 short- and long-distance racing performance in young and adult horses and association analysis with functional candidate genes in Spanish
285 Trotter horses. J. Anim. Breed. Genet. 133, 347-356. https://doi.org/10.1111/jbg.12208.
286 Padalino, B., De Palo, P., Tateo, A., Centoducati, P., 2005. Management of Standardbred trotter: an investigation in Apulia region. Ippologia 16, 4,
287 15-18.
288 Pereira, G.L., De Matteis, R., Meira, C.T., Regitano, L.C.A., Silva, J.A.IIV., Chardulo, L.A.L., Curi, R.A., 2016. Comparison of Sequence Variants
289 in the PDK4 and COX4I2 Genes Between Racing and Cutting Lines of Quarter Horses and Associations With the Speed Index. J. Equine Vet.
290 Sci. 39, 1-6. https://doi.org/10.1016/j.jevs.2015.07.001.
291 Petersen, JL., Mickelson, JR., Rendahl, AK., Valberg, SJ., Andersson, LS., Axelsson, J., Bailey, E., Bannasch, D., Binns, MM., Borges, AS., Brama,
292 P., da Câmara Machado, A., Capomaccio, S., Cappelli, K., Cothran, EG., Distl, O., Fox-Clipsham, L., Graves, KT., Guérin, G., Haase, B.,
293 Hasegawa,T., Hemmann, K., Hill, EW., Leeb, T., Lindgren, G., Lohi, H., Lopes, MS., McGivney, BA., Mikko, S., Orr, N., Penedo, MC., Piercy,
294 RJ., Raekallio, M., Rieder, S., Røed, KH., Swinburne, J., Tozaki, T., Vaudin, M., Wade, CM., McCue, ME., 2013. Genome-wide analysis
295 reveals selection for important traits in domestic horse breeds. PLoS Genet. 9, e1003211. https://doi.org/10.1371/journal.pgen.1003211.

296 Pieramati, C., Fusaioli, L., Scacco, L., Buttazzoni L., Silvestrelli M., 2007. On the use of elo rating on harness racing results in the genetic evaluation
297 of trotter. Ital. J. Anim. Sci. 6 (Suppl. 1), 189-191. https://doi.org/10.4081/ijas.2007.1s.189.
298 Pieramati, C., Giontella, A., Sarti, F.M., Silvestrelli, M., 2011. Assessment of limits for racing speed in the Italian trotter population. Ital. J. Anim.
299 Sci. 10, e53. https://doi.org/10.4081/ijas.2011.e53.
300 Promerová, M., Andersson, LS., Juras, R., Penedo, MC., Reissmann, M., Tozaki, T., Bellone, R., Dunner, S., Hořín, P., Imsland, F., Imsland, P.,
301 Mikko, S., Modrý, D., Roed, KH., Schwochow, D., Vega-Pla, JL., Mehrabani-Yeganeh, H., Yousefi-Mashouf, NG., Cothran, E., Lindgren, G.,
302 Andersson, L., 2014. Worldwide frequency distribution of the 'Gait keeper' mutation in the DMRT3 gene. Anim. Genet. 45, 274-282.
303 https://doi.org/10.1111/age.12120.
304 Raudsepp, T., Finno, CJ., Bellone, RR., Petersen, JL., 2019. Ten years of the horse reference genome: insights into equine biology, domestication and
305 population dynamics in the post-genome era. Anim. Genet. 50, 569-597. https://doi.org/10.1111/age.12857.
306 Ricard, A., 2015. Does heterozygosity at the DMRT3 gene make French trotters better racers? Genet. Sel. Evol. 47, 10. https://doi.org/10.1186/s12711-
307 015-0095-7.
308 Rivero, JL., 2007. A scientific background for skeletal muscle conditioning in equine practice. J. Vet. Med. A Physiol. Pathol. Clin. Med. 54, 321-
309 332. https://doi.org/10.1111/j.1439-0442.2007.00947.x.
310 Rivero, JL., Hill, EW., 2016. Skeletal muscle adaptations and muscle genomics of performance horses. Vet. J. 209, 5-13.
311 https://doi.org/10.1016/j.tvjl.2015.11.019.

312 Schröder, W., Klostermann, A., Distl, O., 2011. Candidate genes for physical performance in the horse. Vet. J. 190, 39-48.
313 https://doi.org/10.1016/j.tvjl.2010.09.029.
314 Shubitowski, D.M., Venta, PJ., Douglass, CL., Zhou, RX., Ewart, SL., 2001. Polymorphism identification within 50 equine gene-specific sequence
315 tagged sites. Anim. Genet. 32, 78-88. https://doi.org/10.1046/j.1365-2052.2001.00738.x.
316 Signer-Hasler, H., Flury, C., Haase, B., Burger, D., Simianer, H., Leeb, T., Rieder, S., 2012. A genome-wide association study reveals loci influencing
317 height and other conformation traits in horses. PLoS One. 7, e37282. https://doi.org/10.1371/journal.pone.0037282.
318 Solé, M., Valera, M., Gómez, MD., Sölkner, J., Molina, A., Mészáros, G., 2017. Heritability and factors associated with number of harness race starts
319 in the Spanish Trotter horse population. Equine Vet. J. 49, 288-293. https://doi.org/10.1111/evj.12632.
320 Sorbolini, S., Gaspa, G., Steri, R., Dimauro, C., Cellesi, M., Stella, A., Marras, G., Marsan, PA., Valentini, A., Macciotta, NP., 2016. Use of canonical
321 discriminant analysis to study signatures of selection in cattle. Genet. Sel. Evol. 48, 58. https://doi.org/10.1186/s12711-016-0236-7.
322 Stephens, M., Donnelly, P., 2003. A comparison of bayesian methods for haplotype reconstruction. Am. J. Hum. Genet. 73, 1162-1169.
323 https://doi.org/10.1086/379378.
324 Takasuga, A., 2016. PLAG1 and NCAPG-LCORL in livestock. Anim. Sci. J. 87, 159-167. https://doi.org/10.1111/asj.12417.
325 Tetens, J., Widmann, P., Kühn, C., Thaller, G., 2013. A genome-wide association study indicates LCORL/NCAPG as a candidate locus for withers
326 height in German Warmblood horses. Anim. Genet. 44, 467-471. https://doi.org/10.1111/age.12031.
327 Thiruvenkadan, N., Kandasamy, S., Panneerselvam, 2009. Inheritance of racing performance of trotter horses: An overview. Liv. Sci. 124, 163-181.
328 https://doi.org/10.1016/j.livsci.2009.01.010.

329 Tozaki, T., Sato, F., Ishimaru, M., Kikuchi, M., Kakoi, H., Hirota, KI., Nagata, SI., 2016. Sequence variants of BIEC2-808543 near LCORL are
330 associated with body composition in Thoroughbreds under training. J. Equine Sci. 27, 107-114. https://doi.org/10.1294/jes.27.107.
331 Wylie, CE., Newton, JR., 2017. A systematic literature search to identify performance measure outcomes used in clinical studies of racehorses. Equine
332 Vet. J. 50, 304-311. https://doi.org/10.1111/evj.12757.

Fig. 1. Allele frequencies of six analysed single nucleotide polymorphisms in Italian Trotter and Throroughbred racehorses. Allele 1 (the first listed allele) and allele 2 (the second listed allele) of the genotyped gene markers are indicated: *CKM*:g.16079732G>A, *COX4I2:*g.23314524T>C, *DMRT3:*g.22391254C>A, *GCK:*g.14963036G>A, *GCK*:14920889T>C and *LCORL:*g.107374136C>T.



338 **Table 1**. Effects of the *COX4I2* T>C genotypes on harness racing performance in the Italian Trotter horses (estimated least squares means ± standard

339 error per different genotypes are presented).

Traits[1]	**Least squares means ± standard error/genotype**[2]			**P**[3]
	T/T (n= 143)	**C/T (n= 124)**	**C/C (n=33)**	
Best racing time, s/km	74.78±0.14	74.60±0.15	74.98±0.26	0.2882
Earnings, Euro	77370.14±9858.84	90497.59±10255.83	67209.23±19192.89	0.3356
Placings, n°	23.02±0.56	23.54±0.58	27.75±1.09	0.1306
Starts, n°	50.93±3.28 ^{b,A}	61.85±3.40 ^{a}	77.60±6.28 ^{b,B}	0.0003***
Wins, n°	6.90±0.44 ^{c}	8.92±0.46	8.25±0.87 ^{d}	0.0045**

340 [1] Transformed values were used for statistical analyses.

341 [2] Different lower case superscript letters in the same line indicate statistically significant or suggestive differences of genotype classes (a,b: $P < 0.05$;

342 c,d: $P < 0.01$; A,B: $P < 0.001$).

343 [3] Probability of genotype: *$P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Traits[1]	**Least squares means ± standard error/genotype**[2]			**P**[3]
	T/T (n= 143)	**C/T (n= 124)**	**C/C (n=33)**	

Best racing time, s/km	74.78±0.14	74.60±0.15	74.98±0.26	0.2882
Earnings, Euro	77370.14±9858.84	90497.59±10255.83	67209.23±19192.89	0.3356
Placings, n°	23.02±0.56	23.54±0.58	27.75±1.09	0.1306
Starts, n°	50.93±3.28 [b,A]	61.85±3.40 [a]	77.60±6.28 [b,B]	0.0003***
Wins, n°	6.90±0.44 [c]	8.92±0.46	8.25±0.87 [d]	0.0045**

[1] Transformed values were used for statistical analyses.

[2] Different lower case superscript letters in the same line indicate statistically significant or suggestive differences of genotype classes (a,b: $P < 0.05$; c,d: $P < 0.01$; A,B: $P < 0.001$).

[3] Probability of genotype: *$P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Table 2. *GCK* haplotypes substitution effects on harness racing performance in the Italian Trotter horses (estimated least square means ± standard error per haplotype copy number are presented).

Haplotypes [1]	**Traits** [2]	**Least square means ± standard error**[3]			**P**[4]
		Haplotype copy			
		0	**1**	**2**	
[G:T]	N°/haplotype copy	86	123	37	
	Best racing time, s/km	74.75±0.18	74.77±0.16	74.78±0.27	0.9131
	Earnings, Euro	81206.56±9131 7.11	65226.83±780 7.27	57812.31±13 774.87	0.5894
	Placings, n°	22.75±0.72	24.02±0.61	24.51±1.07	0.7572
	Starts, n°	60.74±4.29	58.57±3.60	58.40±6.35	0.9102
	Wins, n°	8.09±0.56	7.86±0.47	7.99±0.82	0.9455
[G:C]	N°/haplotype copy	105	114	27	
	Best racing time, s/km	74.76±0.17 [a]	74.89±0.16 [c]	74.13±0.30 [b,d]	0.0118*

	Earnings, Euro	68192.52±8114 .02 [a]	59451.16±792 8.22 [a]	120734.87±1 5872.48 [b]	0.0370*
	Placings, n°	23.68±0.65	23.78±0.63	23.02±1.26	0.6826
	Starts, n°	60.22±3.81	58.71±3.73	57.89±7.47	0.9384
	Wins, n°	8.12±0.48[a]	7.24±0.47 [c]	10.53±0.95 [b,d]	0.0061**
[A:C]	N°/haplotype copy	141	95	10	
	Best racing time, s/km	74.69±0.14	74. 84±0.17	74.98±0.48	0.5644
	Earnings, Euro	71141.33±7438 .39	64180.09±866 2. 78	103743.58±2 6113.52	0.9871
	Placings, n°	23.93±0.58	23.36±0.67	22.78±2.04	0.5374
	Starts, n°	58.36±3.41	58.92±3.97	76.10±11.90	0.3111
	Wins, n°	8.22±0.44	7.75±0.52	6.27±1.56	0.4109

[1] *GCK* haplotypes are defined according to the following marker order**:** g.14963036G>A and g.14920889T>C (EquCab3.0 horse genome assembly. Ensembl database, release 100, April 2020).

[2] Transformed values were used for statistical analyses.

[3] Different lower case superscript letters in the same line indicate statistically significant differences of genotype classes (a,b: P <0.05; c,d: P < 0.01).

[4] Probability of genotype: *P < 0.05, ** P < 0.01, *** P < 0.001.

358 **Supplementary materials**

359 **Table S1.** Genotyped markers with primers, PCR conditions, amplified regions and PCR-RFLP details.

Marker, horse chromosome (ECA), polymorphism[1]	**F: forward sequence (5'-3') of primer[2] R: reverse sequence (5'-3') of primer[2]**	**Size (bp), gene region[3]**	**PCR[4]**	**PCR-RFLP,[5] restriction enzyme**	**Allele fragments[6]**
CKM, (ECA10), g.16079732G>A	F: CGCGATGTAAACCCAAATTC R: CTGGAGTGGGCACCGGGATAGAGT	172, I 4	55/30	*Hinf*I	G= 100 bp + 47 bp + 25 bp A= 100 bp + 72 bp.
COX4I2, (ECA22), g.23314524T>C	F: TGGTGGGCCTTCCAGAGGAGGGTA R: CTCACCGCCTCTCTCTGTTC	169, I 2	68/30	*Rsa*I	T= 169 bp C= 145 bp + 24 bp
DMRT3, (ECA23), *g.22391254C>A* (p.Ser301Stop)	F: AATCTTCCCCAACCAGAAGC R: CAACACGAGGCTCTCGGGCTCTGTC	151, E 5	60/30	*Taq*I	C= 125 bp + 26 bp A= 151 bp.
GCK, (ECA4), g.14963036G>A	F: CAGTCCCAGTTTTATGCATGG R: CCTCCATCCTGGCTCGTC	670, 5'-UTR of E 1B	62/45	*Bsh1236*I	G= 486 bp + 184 bp A= 670 bp

GCK, (ECA4), g.14920889T>C (p.Lys51Glu)	F: GCCTGAGGCTAGAGACCCGC R: AGGGGTACCTCTTCCTGCTC	180, E 2	53/30	*Bsh1236*I	T= 180 bp, C= 161 bp +19 bp
LCORL, (ECA3), g.107374136C>T, BIEC-808543	F: CCAAATTTGCCTGGCTAAG R: GGCTTTGACCGGATAGCATA	152, flanking region	58/30	*Alu*1	C= 122 bp + 30 bp, T= 152 bp.

[1] Polymorphisms were mapped based on EquCab3.0 horse genome assembly (Ensembl database (release 100, April 2020).

[2] The nucleotide (mismatched) of the primer that inserted an artificial restriction site was underlined. [3] I: intron, 5'-UTR: 5'-untraslated region, E: exon.

[4] PCR conditions: annealing temperature (°C) and extension time (seconds). PCR reactions were carried out in a 20 µl reaction volume, that included 2-4 µl of DNA template (10-80 ng), 10 pmol of each primer, 250 mM of each dNTP, 1.5 mM $MgCl_2$ and 1 U of EuroTaq DNA polymerase (EuroClone, Milan, Italy). The PCR cycles included a first denaturation step at 95 °C for 5 min, 35 cycles (30 sec at 95 °C, 30 sec at 53-68 °C, and 30 or 45 sec at 72 °C) and a final step at 72 °C for 9 min.

[5] PCR-RFLP: restriction fragment length polymorphism.

[6] The PCR products and fragments obtained from digestions were electrophoresed on 2.5% (*GCK*:g.14963036G>A) or 3.5% agarose gels and visualized with 1×GelRed Nucleid Acid Gel Stain (Biotium Inc., Hayward, CA, USA).

371 **Table S2**. Descriptive statistics of untransformed harness racing performance of 412 Italian Trotter
372 horses.

Trait	Mean	Standard deviation	Min	Max	Median
Best race time, s/km	74.74	1.96	71.00	86.00	75
Earnings, Euro	49,111.36	103,729.93	0	1,144,344.80	21,916.60
Placings, n°	23.51	20.50	0	116	19
Starts, n°	49.40	42.81	0	204	42
Wins, n°	6.98	6.94	0	47	5

373

374 **Table S3**. Spearman correlation between untransformed harness racing traits of 412 Italian Trotter
375 horses. The values for the phenotypic correlations are shown in the upper rows, the lower rows report
376 the P values.

	Best racing time	**Earnings**	**Placings**	**Starts**	**Wins**
Best racing time	1	-0.72	-0.40	-0.40	-0.65
		P < 0.0001	P < 0.0001	P < 0.0001	P < 0.0001
Earnings		1	+0.81	+0.80	+0.92
			P < 0.0001	P < 0.0001	P < 0.0001
Placings			1	+0.96	+0.73
				P < 0.0001	P < 0.0001
Starts				1	+0.75
					P < 0.0001
Wins					1

377

Table S4. Descriptive statistics, after editing, of untransformed harness racing performance of 313 Italian Trotter horses by sex (208 females and 105 males).

Trait, Sex[1]	Mean	Standard deviation	Min	Max	Median
Best race time, s/km. F	74.98	1.64	71.01	82.05	75.02
Best race time, s/km. M	73.83	1.71	71.02	80.02	74.01
Earnings, Euro. F	39,087.57	44,795.17	0.00	328,099.50	25,707.45
Earnings, Euro. M	101,687.68	183,366.20	0.00	1,144,344.80	44,0361.10
Placings, n°. F	22.78	18.58	0	88	18.0
Placings, n°. M	26.45	19.44	0	81	22
Starts, n°. F	51.04	35.80	5	167	45
Starts, n°. M	65.30	41.22	6	170	55
Wins, n°. F	6.24	5.76	0	35	5
Wins, n°. M	9.45	7.92	0	47	8

[1] F= females, M= males.

382 **Table S5**. Allele and genotype frequencies of the genotyped single nucleotide polymorphisms in
383 Italian Trotter and Throroughbred racehorses.

Marker/Breeds	Allele frequency[1]		Genotype frequency		
	1	2	11	12	22
CKM g.16079732G>A					
Italian Trotter	0.59	0.41	0.33	0.51	0.15
Throroughbred	0.50	0.50	0.12	0.75	0.12
COX4I2 g.23314524T>C					
Italian Trotter	0.66	0.34	0.44	0.44	0.12
Throroughbred	0.50	0.50	0.25	0.50	0.25
DMRT3 g.22391254C>A (p.Ser301Stop)					
Italian Trotter	0.04	0.96	0.03	0.03	0.94
Throroughbred	1.00	0.00	1.00	0.00	0.00
GCK g.14963036G>A					
Italian Trotter	0.73	0.27	0.54	0.39	0.07
Throroughbred	0.90	0.10	0.80	0.20	0.00
GCK g.14920889T>C (p.Lys51Glu)					
Italian Trotter	0.43	0.57	0.18	0.50	0.31
Throroughbred	0.75	0.25	0.60	0.30	0.10
LCORL g.107374136C>T					
Italian Trotter	0.28	0.72	0.03	0.39	0.58
Throroughbred	0.20	0.80	0.00	0.400	0.600

384 [1] allele1>allele2 (the first and the second allele indicated in the SNP nomenclature).

385

386 **Table S6**. P-values for fixed effects and covariate used in the models for the association analyses

387 between markers[1] and harness racing performances in the Italian Trotter horses.

Dependent variables[2]	**Factor of variations**	***CKM*[3]**	***COX4I2*[3]**	***GCK***		***LCORL*[3]**
				c.-460T>C	**g.82A>G**	
Best racing time, s/km	Genotype	0.5976	0.2882	0.2628	0.6472	0.5228
	Sex	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	Year of birth	0.0095	0.0104	0.0904	0.0102	0.0006
	Year at record	<0.0001	<0.0001	0.0001	<0.0001	<0.0001
	Distance	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	Start numbers [4]	0.1742	0.1003	0.0627	0.1314	0.0339
	R2 of model, %	33.6	32.4	32.6	32.6	34.6
Earnings, Euro	Genotype	0.0579	0.3356	0.6078	0.2396	0.3502
	Sex	0.0189	0.0053	0.0162	0.0318	0.0334
	Year of birth	<0.0001	<0.0001	0.0002	<0.0001	0.0004
	Start numbers [4]	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	R2, %	32.0	33.7	32.6	32.4	32.6
Placings, n°	Genotype	0.5606	0.1306	0.3552	0.9578	0.5317
	Sex	0.0265	0.0325	0.0682	0.0105	0.0482
	Year of birth	0.9905	0.9870	0.9569	0.9898	0.9822
	Start numbers [4]	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	R2, %	75.3	75.2	75.9	75.4	75.0
Starts, n°	Genotype	0.0844	0.0003***	0.3076	0.6364	0.4831
	Sex	0.0005	0.0034	0.0204	0.0028	0.0009

	Year of birth	0.1120	0.1296	0.2093	0.2342	0.1346
	R2, %	7.7	11.3	6.0	5.5	7.2
Wins, n°	Genotype	0.4235	0.0045**	0.7406	0.9552	0.4081
	Sex	0.0011	0.0007	0.0014	0.0011	0.0003
	Year of birth	<0.0001	<0.0001	0.0006	<0.0001	<0.0001
	Start numbers[4]	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	R2, %	45.6	49.3	46.1	46.0	46.8

388 [1] Markers are defined in Table S1.

389 [2] Transformed values were used for statistical analyses.

390 [3] Probability of genotype: *P < 0.05, ** P < 0.01, *** P < 0.001.

391 [4] Covariate.

392

393 **Table S7**. Effects of the marker genotypes on harness racing performance in the Italian Trotter horses

394 (estimated least squares means ± standard error per different genotypes are presented).

Gene markers[1]	**Least squares means ± standard error/genotype**[3]			**P**[5]
Traits[2]	**1/1**[4]	**1/2**[4]	**2/2**[4]	
CKM:G>A				
Best racing time, s/km	79.85±0.16	74.66±0.14	74.61±0.23	0.5976
Earnings, Euro	79071.87±11653.55	88273.90±9477.94	73267.60±16779.57	0.0579
Placings, n°	23.38±0.65	23.06±0.53	24.95±0.93	0.5606
Starts, n°	66.10±3.91	55.40±3.22	58.63±5.69	0.0844
Wins, n°	7.45±0.53 [a]	8.28±0.43	7.65±0.77	0.4235
*COX4I2:*T>C				
Best racing time, s/km	74.78±0.14	74.60±0.15	74.98±0.26	0.2882
Earnings, Euro	77370.14±9858.84	90497.59±10255.83	67209.23±19192.89	0.3356
Placings, n°	23.02±0.56	23.54±0.58	27.75±1.09	0.1306
Starts, n°	50.93±3.28 [b,A]	61.85±3.40 [a]	77.60±6.28 [b,B]	0.0003***
Wins, n°	6.90±0.44 [c]	8.92±0.46	8.25±0.87 [d]	0.0045**
GCK:G>A				
Best racing time, s/km	74.67±0.15	74.74±0.17	75.20±0.35	0.2628
Earnings, Euro	85920.14±10290.50	81212.52±12134.81	85310.93±27382.82	0.7068
Placings, n°	23.93±0.56	23.24±0.66	21.74±1.50	0.3552
Starts, n°	58.46±3.39	56.87±4.00	71.70±8.98	0.3076
Wins, n°	8.27±0.45	8.07±0.53	7.32±1.19	0.7406

GCK:T>C				
Best racing time, s/km	74.21±0.21	74.78±0.14	74.69±0.17	0.6472
Earnings, Euro	66640.80±14948.31	76090.14±9444.65	98150.44±11976.79	0.2396
Placings, n°	23.64±0.85	23.81±0.54	23.04±0.68	0.9578
Starts, n°	59.48±5.28	57.75±3.40	62.65±4.22	0.6364
Wins, n°	8.15±0.70	7.93±0.44	7.91±0.56	0.9552
LCORL:C>T				
Best racing time, s/km	74.69±0.34	74.68±0.16	74.86±0.14	0.5228
Earnings, Euro	106970.35±24464.66	93179.34±11592.50	68945.85±10062.15	0.3502
Placings, n°	23.29±1.38	22.95±0.65	23.50±0.57	0.5317
Starts, n°	54.91±8.32	56.50±3.94	61.83±3.40	0.4831
Wins, n°	7.26±1.14	8.45±0.54	7.66±0.47	0.4081

[1] Gene markers are: *CKM*:g.16079732G>A, *COX4I2:*g.23314524T>C, *DMRT3:*g.22391254C>A, *GCK:*g.14963036G>A, *GCK*:14920889T>C and *LCORL:*g.107374136C>T.

[2] Traits transformed values were used for statistical analyses.

[3] Different lower case superscript letters in the same line indicate statistically significant or suggestive differences of genotype classes (a,b: $P < 0.05$; c,d: $P < 0.01$; A,B: $P < 0.001$).

[4] allele1>allele2 (the first and the second allele indicated in the SNP nomenclature).

[5] Probability of genotype: *$P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Table S8. P-values for fixed effects and covariates used in the models for the association analyses between *GCK* haplotypes and harness racing performance in the Italian Trotter horses.

Dependent variables[1]	**Factor of variations**	***GCK* haplotypes**[2, 3]		
		[G:T]	**[G:C]**	**[A:C]**
Best racing time (s/km)	Genotype	0.9131	0.0118*	0.5644
	Sex	<0.001	<0.001	<0.001
	Year of birth	0.1141	0.1759	0.1158
	Year at record	0.002	0.003	0.0001
	Distance	<0.001	<0.001	<0.001
	Start numbers [4]	0.0707	0.0468	0.0887
	R2 of model, %	31.6	34.1	31.0
Earnings, Euro	Genotype	0.5894	0.0370*	0.9871
	Sex	0.1191	0.1117	0.1032
	Year of birth	0.0016	0.0008	0.0020
	Start numbers [4]	<0.0001	<0.0001	<0.0001
	R2, %	32.7	34.3	32.4
Placings, n°	Genotype	0.7572	0.6826	0.5374
	Sex	0.0917	0.0805	0.0995
	Year of birth	0.9569	0.9657	0.9620
	Start numbers [4]	<0.0001	<0.0001	<0.0001
	R2, %	75.4	75.4	75.5
Starts, n°	Genotype	0.9102	0.9384	0.3111
	Sex	0.0239	0.0221	0.0202
	Year of birth	0.1833	0.1980	0.2215
	R2, %	5.1	5.1	6.0

Wins, n°	Genotype	0.9455	0.0061**	0.4109
	Sex	0.0152	0.0161	0.0138
	Year of birth	0.0034	0.0012	0.0028
	Start numbers[4]	<0.0001	<0.0001	<0.0001
	R2, %	46.2	48.5	46.6

404 [1] Transformed values were used for statistical analyses.

405 **[2]** *GCK* haplotypes are defined according to the following marker order**:** g.14963036G>A and

406 g.14920889T>C.

407 [3] Probability of genotype: *P < 0.05, ** P < 0.01.

408 [4]Covariates.

409