A cohort study of racing performance in Japanese Thoroughbred racehorses using genome information on ECA18

T. Tozaki*, E. W. Hill[†], K. Hirota*, H. Kakoi[‡], H. Gawahara[‡], T. Miyake[§], S. Sugita[¶], T. Hasegawa[¶], N. Ishida[¶], Y. Nakano** and M. Kurosawa*

*Department of Molecular Genetics, Laboratory of Racing Chemistry, Utsunomiya, Tochigi 320-0851, Japan. [†]Animal Genomics Laboratory, School of Agriculture, Food Science and Veterinary Medicine, University College Dublin, Dublin 4, Ireland. [‡]Department of Genetic Analysis, Laboratory of Racing Chemistry, Utsunomiya, Tochigi 320-0851, Japan. [§]Comparative Agricultural Sciences, Graduate school of Agriculture, Kyoto University, Sakyo, Kyoto 606-8502, Japan. [¶]Equine Research Institute, Japan Racing Association, Utsunomiya, Tochigi 320-0856, Japan. **Department of Pharmacogenomics, School of Pharmaceutical Sciences, Showa University, Shinagawa, Tokyo 142-8555, Japan

Summary

Using 1710 Thoroughbred racehorses in Japan, a cohort study was performed to evaluate the influence of genotypes at four single nucleotide polymorphisms (SNPs) on equine chromosome 18 (ECA18), which were associated in a previous genome-wide association study for racing performance with lifetime earnings and performance rank. In males, both g.65809482T>C and g.65868604G>T were related to performance rank ($P = 0.005$). In females, g.65809482T>C ($P = 1.76\text{E-6}$), g.65868604G>T ($P = 6.81\text{E-6}$) and g.66493737C>T ($P = 4.42\text{E-5}$) were strongly related to performance rank and also to lifetime earnings ($P < 0.05$). When win-race distance (WRD) among all winning racehorses and best race distance (BRD) among elite racehorses were considered as the phenotypes, significant associations ($P < 0.001$) were observed for all four SNPs. The favourable race distance of both elite (BRD) and novice racehorses (WRD) was also associated with genotypes in the ECA18 region, indicating the presence of a gene in this region influencing optimum race distance in Thoroughbred racehorses. Therefore, the association with performance rank is likely due to the bias in the race distances. The location of the SNPs within and proximal to the gene encoding myostatin (*MSTN*) strongly suggests that regulation of the *MSTN* gene affects racing performance. In particular, the g.65809482T>C, g.65868604G>T and g.66493737C>T SNPs, or their combinations, may be genetic diagnostic markers for racing performance indicators such as WRD and BRD.

Keywords horse, myostatin, racing performance, retrospective cohort, Thoroughbred.

Introduction

Thoroughbred horses originated from a small number of Arab, Barb and Turk stallions and native British mares ~300 years ago (Cunningham *et al.* 2001; Hill *et al.* 2002; Bower *et al.* 2010). During the past 300 years, Thoroughbred horses have been selectively bred to improve speed and stamina, making them superior competitive racehorses. Thoroughbred horses have a very high skeletal muscle mass comprising over 55% of their total body mass (Gunn 1987). The aerobic capacity of Thoroughbred horses ($VO_{2max} > 200$ ml O_2/kg/min) is superior to that of other athletic species of similar size (Jones *et al.* 1989; Jones & Lindstedt 1993; Young *et al.* 2002). Such traits have been enhanced by intense artificial selection for sequence variants contributing to exceptional racing performance (Gu *et al.* 2009). Various measures for the evaluation of racing performance in Thoroughbred horses, such as earnings, race times and handicap ratings, have been used to estimate heritabilities (Tolley *et al.* 1985; Gaffney & Cunningham 1988; Williamson & Beilharz 1998; Mota *et al.* 2005; Sobczynska 2006), and breeding values have also been calculated to evaluate horses' potential ability to transmit genetic factors related to racing performance. While most of these studies have estimated the genetic contribution to variation in racing ability to be between 0.35 and 0.55, the heritability of best

race distance (BRD) has been estimated to be as high as 0.94 (Williamson & Beilharz 1998).

Many significant advances have been achieved in the horse genome project (http://www.uky.edu/Ag/Horsemap/welcome.html), such as the construction of half- and full-sibling linkage maps (Penedo *et al.* 2005; Swinburne *et al.* 2006), horse-human comparative maps (Tozaki *et al.* 2007; Raudsepp *et al.* 2008), and the completion of a high-quality draft horse genome sequence with over 1.1 million identified SNPs (Wade *et al.* 2009). This genetic infrastructure for the horse has enabled the identification of a genomic region on ECA18 as being associated with racing performance phenotypes. Four independent studies, including a candidate gene study (Hill *et al.* 2010a), a microsatellite-based genome-wide association study (Tozaki *et al.* 2010), and two genome-wide SNP association studies (Binns *et al.* 2010; Hill *et al.* 2010b), have identified the same genomic region on ECA18 associated with racing performance. In this region, four SNPs (g.65809482T>C, g.65868604G>T, g.66493737C>T and g.66539967A>G) were identified as candidates for genetic prediction of racing performance in Japanese racehorses based on lifetime earnings and performance rank (Tozaki *et al.* 2010). In particular, the SNP g.66493737C>T, which is located in the first intron of the *myostatin* (*MSTN*) gene, has been associated with BRD among a cohort of elite race winning Thoroughbreds (Hill *et al.* 2010a). The *MSTN* gene is known to contribute to muscle hypertrophy phenotypes in a variety of mammalian species (Grobet *et al.* 1997; McPherron & Lee 1997; Schuelke *et al.* 2004; Mosher *et al.* 2007). It was observed that Thoroughbreds with the C/C genotype at g.66493737C>T were better suited to short-distance races, C/T horses competed favourably in middle-distance races, and T/T horses had greater stamina and represented the majority of winning horses at the longer distances (Hill *et al.* 2010a).

In this study, we designed a retrospective cohort study to evaluate the influences of the four SNPs associated with racing performance in Japanese Thoroughbred racehorses. The Japan Racing Association (JRA) is the largest racing authority in Japan and hosts horse races for ∼50% (∼4000 horses) of the Thoroughbred horse population born in Japan every year. In this study, 1710 Thoroughbred racehorses born in the same year (2000) were followed over their athletic career in JRA horse races to evaluate racing performance based on lifetime earnings, JRA performance rank, BRD, and win-race distance (WRD), defined as the average distance of races won. The findings of the current study corroborate those reported by Hill *et al.* (2010a) and Tozaki *et al.* (2010).

Materials and methods

Thoroughbred horse populations

Performance information from 3927 Thoroughbred racehorses kept at JRA training centres in October 2002 was available for this study. These 3927 Thoroughbred racehorses are the same sample set used in a previous study (Tozaki *et al.* 2010). All the animals had retired by December 2006. Of the 3927 Thoroughbred horses, 3706 blood samples from 2397 males and 1309 females were available for SNP genotyping. The 3706 Thoroughbred horses were born between 1993 and 2000 and were registered as JRA racehorses by October 2002. From the 3706 horses, a sample set comprising 1710 individuals born in 2000, including 1023 males and 687 females, were used for a retrospective cohort study (Fig. 1). For BRD, 120 males and 31 females were also picked from the 3706 horses. The cohorts in the current study have some degree of overlap with the previous study; 17.8% (182 males) and 29.3% (201 females) were also genotyped in the previous study (Tozaki *et al.* 2010). The performance data for horses born between 1993 and 1999 are biased, as only horses that were performing well were still in training centres by October 2002. Therefore, such samples were omitted from this retrospective cohort study.

SNP genotyping

Genomic DNA was extracted from 3706 blood samples stored at −40 °C using the MFX-2000 MAGEXTRACTOR System according to the manufacture's protocols. Four SNPs (g.65809482T>C, g.65868604G>T, g.66493737C>T and g.66539967A>G on ECA18), which have been associated with lifetime earnings, performance rank (Tozaki *et al.* 2010) and BRD (Hill *et al.* 2010a), were genotyped by high-resolution melting and unlabelled probe methods using a LightScanner system (Idaho Technology, Inc.) according to the manufacturers' protocols. The primers, probes and annealing temperatures used for SNP genotyping are provided in Table S1. Three of the SNPs, g.65809482T>C, g.65868604G>T and g.66539967A>G, are documented as the loci BIEC2-417210, BIEC2-417274 and BIEC2-417372 in the EquCab2.0 SNP database (http://www.broadinstitute.org/), and g.66493737C>T was identified following re-sequencing and is located in the first intron of the *MSTN* gene (Hill *et al.* 2010a).

Characteristics of racing performance

To evaluate the influence of genotype differences on racing performance in Japanese Thoroughbred racehorses, we used the following phenotypes: lifetime earnings on the racetrack, final performance rank, and WRD, defined as the average of the distances of races that were won. The JRA tabulates and reports earnings using two methods. First, total prize money won during a racing career is recorded and is designated as lifetime earnings. The top five racehorses in each race are usually awarded prize money based on Japanese yen (JPY). Second, a record of earnings, which is kept of money won only for first place finishes, is used to calculate final performance
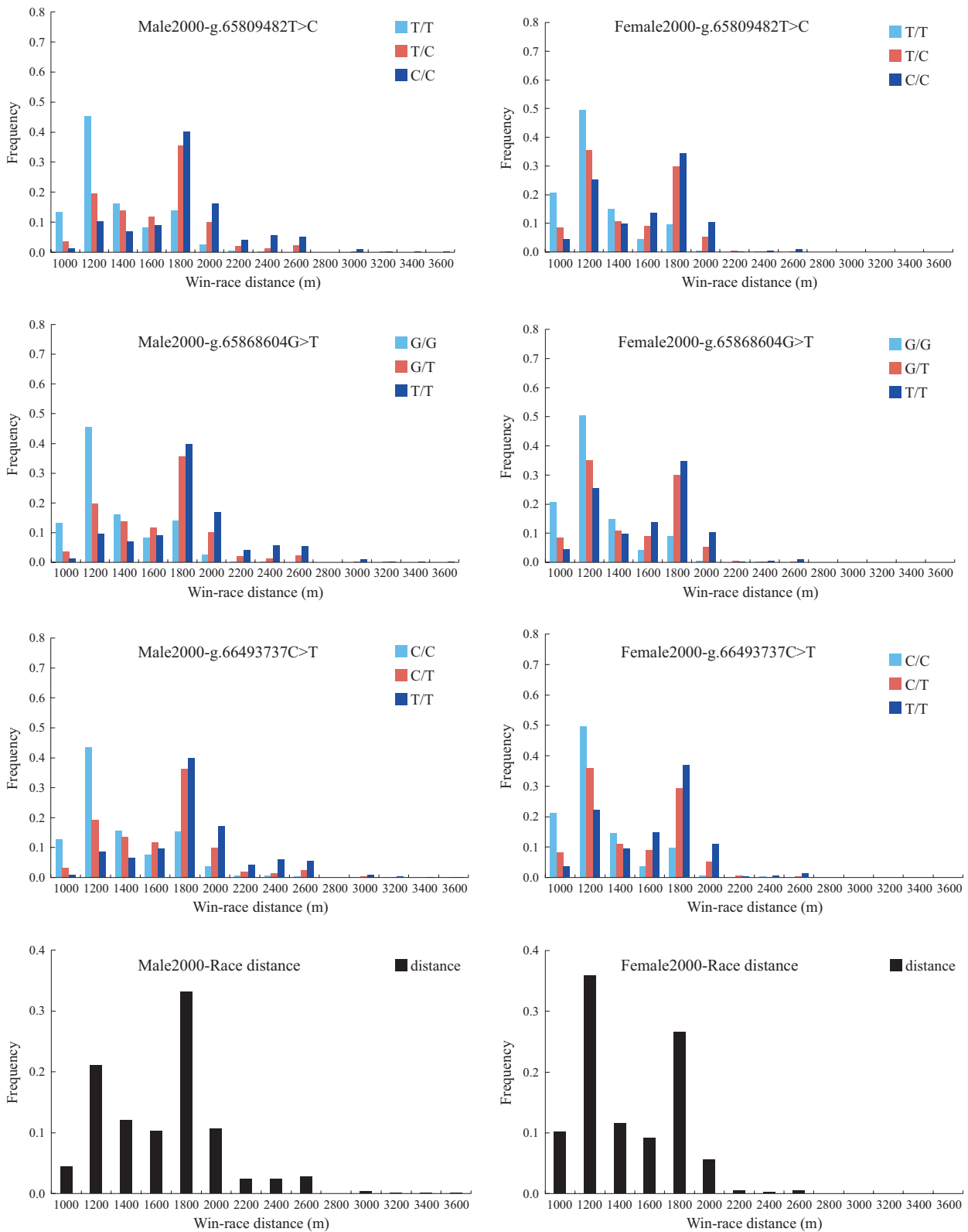
Figure 1 Genotype distributions for three SNPs (g.65809482T>C, g.65868604G>T and g.66493737C>T) for win-race distance among Japanese Thoroughbred racehorses born in 2000. The left-hand graphs show the distributions for males using data from 1070 races. The right-hand graphs show the distributions for females using data from 474 races. The bottom graphs show the distribution of the proportion of race distances competed by males and females, respectively.

rank (lifetime rank). There are five JRA ranks of total earnings. Rank 1 (0 million JPY class) is the rank for non-winning horses, followed by rank 2 (≤5 million JPY class), rank 3 (≤10 million JPY class), rank 4 (≤16 million JPY class) and rank 5 (>16 million JPY class). This second metric is referred to as performance rank. In this study, we roughly divided the five ranks into either three performance ranks [non-winning horse class (0 million JPY class), novice horse class (>0–16 million JPY class), and superior horse class (>16 million JPY class)] or two ranks (non-winning and winning, or superior and other horse groups).

The JRA also tabulates and reports racing histories for each racehorse in the JRA. For all 3706 Thoroughbred racehorses, we collected the race distances for winning racehorses and their racing grades. To calculate WRD for cohorts born in 2000, the distances of all the races that each horse won were averaged. For example, if a racehorse won 10 races, then the 10 race distances were used. To calculate BRD, the distances of each Grade race won were collected from all the racehorses used in this study, and the distance of the highest Grade race was used as the BRD for each horse. If the horse had multiple wins in races of the same grade, the most valuable race, in terms of prize money, was chosen for this study. The criterion for BRD was the same as described by Hill *et al.* (2010a). Grade races make up just 3.6% of all JRA races per annum. Therefore, in this study, of the 3706 individuals, 120 males and 31 females were elite racehorses and were used for evaluating BRD. In JRA, there are six categories of Grade race, including G1, Jpn1, G2, Jpn2, G3 and Jpn3. For example, G1 and Jpn1 races are the most valuable races in JRA and are the same grade, but Jpn1 is limited to racehorses born in Japan. Almost all G1 and Jpn1 races are competed over distances ranging from 1600 to 3200 m.

Tests of genetic association were performed using chi-square tests and the Fisher exact tests for performance rank as a categorical trait, and ANOVA and Kruskal–Wallis tests for lifetime earnings, WRD and BRD, as quantitative traits. In this study, two statistical tests based on average and median were evaluated for the quantitative traits to provide confidence for the associations. All statistical analyses were performed using IBM SPSS Statistics 19 (SPSS Japan Inc.). In the previous study (Tozaki *et al.* 2010), we observed differences in the distributions of performance indicators, such as lifetime earnings, according to gender. Therefore, in this study, the associations of the four SNPs were determined independently for each gender.

Results

Characteristics of populations used in this study

Of the 3927 Thoroughbred racehorses, 3706 individuals (2397 males and 1309 females) were available for genotyping the SNPs g.65809482T>C, g.65868604G>T,

g.66493737C>T and g.66539967A>G. These four SNPs were successfully genotyped in all 3706 individuals. The genotype frequencies and performance information for the cohort populations are shown in Table S2. In males, no significant difference was found in genotype frequency among the different birth-year groups. In females, however, clear differences in genotype frequency were observed depending on birth-year group. Furthermore, like genotype frequency, average lifetime earnings and athletic period also tended to differ across birth-year groups. The performance data for horses born between 1993 and 1999 are biased, as only well-performing horses were still in training centres by October 2002, suggesting that those populations born between 1993 and 1999 were not suitable for the current cohort study.

A retrospective cohort study using individuals born in 2000

The cohort consisted of 1023 males and 687 females (Table S2). Table 1 shows how differences in the genotype of each SNP influenced performance rank, lifetime earnings and WRD. As an overall trend, higher statistically significant differences were found in females than in males. In females, all four SNPs demonstrated a statistically significant influence on performance rank. Three SNPs, g.65809482T>C ($P = 0.003$), g.65868604G>T ($P = 0.009$) and g.66493737C>T ($P = 0.021$), also influenced lifetime earnings. In males, g.65809482T>C ($P = 0.005$) and g.65868604G>T ($P = 0.005$) influenced performance rank, but no statistically significant influence was observed for g.66493737C>T ($P = 0.074$) or g.66539967A>G ($P = 0.119$) for this trait. Unlike in females, genotype differences in males were not associated with a statistically significant difference in lifetime earnings, although non-significant differences in mean lifetime earnings were observed.

Statistically significant differences in WRD ($P < 0.001$) were observed in both males and females. For example, males with a T/T genotype at locus g.65809482T>C had an average WRD of 1308 m, while C/C individuals had a WRD of 1859 m (see Table 1). Therefore, it was expected that the T-allele at g.65809482T>C, G-allele at g.65868604G>T, C-allele at g.66493737C>T and A-allele at g.66539967A>G were suited to short-distance races, while the C-allele at g.65809482T>C, T-allele at g.65868604G>T, T-allele at g.66493737C>T and G-allele at g.66539967A>G were suited to long-distance races. In particular, g.65809482T>C, g.65868604G>T and g.66493737C>T resulted in clear differences in WRD. When the distance differences between two homozygous genotypes, such as T/T and C/C genotypes at g.65809482T>C inferring short- and long-distance alleles, respectively, were compared between males and females, the males showed a larger difference than the females. Figure 1 shows the distance-specific win frequency

Table 1 Statistical analyses for performance rank, lifetime earnings and WRD for male and female cohorts that were born in 2000.

Male (2000) (N = 1023 males)		Performance rank			P-value	Average	P-value	Average	P-value
		0	0< and ≤16 (Million JPY)	≥16	Chi-square test (Fisher's exact test)	Lifetime earnings (100 JPY)	ANOVA (Kruskal–Wallis test)	WRD (m) (N = 1070 races)	ANOVA (Kruskal–Wallis test) Differences of distance between homozygous genotype (m)
		Non-win (N = 531) 51.9%	Novice (N = 379) 37.0%	Superior (N = 113) 11.1%					
g.65809482T>C	T/T (N = 153)	64 41.8%	67 43.8%	22 14.4%		276 426		1308	
	T/C (N = 511)	255 49.9%	196 38.4%	60 11.7%	0.005 (0.005)	234 151	0.348 (<0.001)	1619	<0.001 (<0.001)
	C/C (N = 359)	212 59.1%	116 32.3%	31 8.6%		210 220		1859	551 m
g.65868604G>T	G/G (N = 155)	65 41.9%	68 43.9%	22 14.2%		273 325		1308	
	G/T (N = 511)	255 49.9%	196 38.4%	60 11.7%	0.005 (0.005)	234 162	0.391 (<0.001)	1620	<0.001 (<0.001)
	T/T (N = 357)	211 59.1%	115 32.2%	31 8.7%		211 180		1859	551 m
g.66493737C>T	C/C (N = 175)	75 42.9%	75 42.9%	25 14.3%		274 593		1369	
	C/T (N = 525)	275 52.4%	192 36.6%	58 11.0%	0.074 (0.072)	224 151	0.427 (0.002)	1627	<0.001 (<0.001)
	T/T (N = 323)	181 56.0%	112 34.7%	30 9.3%		221 921		1825	457 m
g.66539967A>G	A/A (N = 290)	134 46.2%	119 41.0%	37 12.8%		266 896		1488	
	A/G (N = 521)	276 53.0%	186 35.7%	59 11.3%	0.119 (0.117)	211 008	0.271 (0.011)	1653	<0.001 (<0.001)
	G/G (N = 212)	121 57.1%	74 34.9%	17 8.0%		236 220		1821	333 m

Table 1 (Continued)

Female (2000) (N = 687 females)		Performance rank			P-value	Average	P-value	Average	P-value
		0	0< and ≤16 (Million JPY)	≥16					ANOVA (Kruskal–Wallis test)
		Non-win (N = 434) 63.2%	Novice (N = 214) 31.1%	Superior (N = 39) 5.7%	Chi-square test (Fisher's exact test)	Lifetime earnings (100 JPY)	ANOVA (Kruskal–Wallis test)	WRD (m) (N = 474 races)	Differences of distance between homozygous genotype (m)
g.65809482T>C	T/T (N = 86)	47 54.7%	29 33.7%	10 11.6%		253 315		1242	
	T/C (N = 323)	179 55.4%	122 37.8%	22 6.8%	1.8E-06 (<0.001)	171 213	0.003 (<0.001)	1479	<0.001 (<0.001) 367 m
	C/C (N = 278)	208 74.8%	63 22.7%	7 2.5%		89 860		1609	
g.65868604G>T	G/G (N = 85)	48 56.5%	28 32.9%	9 10.6%		217 251		1219	
	G/T (N = 327)	181 55.4%	123 37.6%	23 7.0%	6.8E-06 (<0.001)	179 364	0.009 (<0.001)	1481	<0.001 (<0.001) 390 m
	T/T (N = 275)	205 74.5%	63 22.9%	7 2.5%		90 726		1609	
g.66493737C>T	C/C (N = 90)	50 55.6%	31 34.4%	9 10.0%		210 332		1237	
	C/T (N = 351)	200 57.0%	127 36.2%	24 6.8%	4.4E-05 (<0.001)	172 918	0.021 (<0.001)	1472	<0.001 (<0.001) 415 m
	T/T (N = 246)	184 74.8%	56 22.8%	6 2.4%		91 235		1652	
g.66539967A>G	A/A (N = 167)	101 60.5%	54 32.3%	12 7.2%		169 004		1333	
	A/G (N = 360)	213 59.2%	123 34.2%	24 6.7%	0.006 (0.004)	162 257	0.196 (<0.001)	1498	<0.001 (<0.001) 274 m
	G/G (N = 160)	120 75.0%	37 23.1%	3 1.9%		96 448		1607	

WRD, win-race distance.

according to the genotypes of the three SNPs that showed a clear difference in WRD. For example, males with a T/T genotype at g.65809482T>C were mainly distributed between 1000 and 1800 m, individuals with a T/C genotype were mainly distributed between 1200 and 2000 m, and individuals with C/C genotype were mainly distributed >1800 m. Similar genetic trends were observed for the other SNPs. The distribution of win frequencies indicated that differences in genotype influence WRD.

As it was observed that among females the performance rank (non-win, novice and superior) was influenced by the genotype of each SNP (Table 1), and among males two SNPs had a significant relationship, a more detailed evaluation was performed (Table 2). Performance rank was analysed based on two classifications (see Materials and methods); (i) non-winning horse vs. winning horse groups and (ii) superior horse vs. other horse groups. The first classification is important for planning a racing strategy to result in a win within 1 year after debuting, because horses competing in JRA races must win at least one race to continue racing in the JRA system. The latter classification is important for planning to produce superior racehorses (≥16 million) competing within the JRA system. In males, the T-allele at g.65809482T>C and the G-allele at

g.65868604G>T may have a marginally dominant influence (both relative risk: 1.27, 95% CI: 1.10-1.47) in the first classification, while the C-allele at g.66493737C>T and the A-allele at g.66539967A>G may have a recessive influence. However, none of the SNPs had a statistically significant influence on whether or not a horse was likely to achieve a superior performance rank. In females, all the SNPs had statistically significant influences in both categories. We observed a trend in which the T-allele at g.65809482T>C, the G-allele at g.65868604G>T, the C-allele at g.66493737C>T and the A-allele at g.66539967A>G had a dominant influence, and relative risks were higher than those for males (see Table 2). As to whether or not a horse may be a superior horse, unlike in males, higher relative risks were observed (generally ≥3) in a dominant model of the allele suited to short-distance races on WRD. These data may therefore be useful for predicting racing performance among horses competing under JRA rules.

Best race distance

As Hill *et al.* (2010a) reported that a SNP located within the *MSTN* gene influenced BRD, for the present study we

Table 2 Relative risks (RR) for performance rank based on the cohort study using JRA racehorses that were born in 2000.

Male (2000) (N = 1023 males)	Model	Non-winning vs. Winning horse groups				Superior vs. Other horse groups			
		P-value				P-value			
		Chi-square	Fisher's exact	RR	95%CI	Chi-square	Fisher's exact	RR	95%CI
g.65809482T>C	T/T + T/C:C/C	7.7E-04	<0.001	1.27	1.10–1.47	0.070	0.076	1.43	0.97–2.12
	T/T:T/C + C/C	0.007	0.008	1.26	1.08–1.46	0.154	0.162	1.37	0.89–2.12
g.65868604G>T	G/G + G/T:T/T	7.4E-04	<0.001	1.27	1.10–1.47	0.078	0.094	1.42	0.96–2.10
	G/G:G/T + T/T	0.007	0.009	1.25	1.08–1.46	0.175	0.209	1.35	0.88–2.09
g.66493737C>T	C/C + C/T:T/T	0.072	0.080	1.14	0.99–1.31	0.223	0.239	1.28	0.86–1.90
	C/C:C/T + T/T	0.009	0.010	1.24	1.07–1.43	0.133	0.145	1.38	0.91–2.08
	A/A + A/G:G/G	0.091	0.105	1.15	0.97–1.37	0.114	0.139	1.48	0.90–2.42
g.66539967A>G	A/A:A/G + G/G	0.022	0.022	1.17	1.03–1.34	0.272	0.270	1.23	0.85–1.78

Female (2000) (N = 687 females)	Model	Non-winning vs. Winning horse groups				Superior vs. Other horse groups			
		P-value				P-value			
		Chi-square	Fisher's exact	RR	95%CI	Chi-square	Fisher's exact	RR	95%CI
g.65809482T>C	T/T + T/C:C/C	1.8E-07	<0.001	1.78	1.41–2.24	0.003	0.004	3.11	1.40–6.94
	T/T:T/C + C/C	0.080	0.094	1.27	0.99–1.65	0.011	0.021	2.41	1.22–4.77
g.65868604G>T	G/G + G/T:T/T	4.4E-07	<0.001	1.74	1.39–2.20	0.004	0.004	3.05	1.37–6.81
	G/G:G/T + T/T	0.171	0.187	1.21	0.93–1.58	0.037	0.045	2.12	1.05–4.32
g.66493737C>T	C/C + C/T:T/T	2.4E-06	<0.001	1.72	1.35–2.19	0.006	0.006	3.07	1.31–7.22
	C/C:C/T + T/T	0.108	0.127	1.25	0.97–1.61	0.057	0.082	1.99	0.98–4.05
g.66539967A>G	A/A + A/G:G/G	4.0E-04	<0.001	1.62	1.21–2.16	0.018	0.018	3.64	1.14–11.67
	A/A:A/G + G/G	0.407	0.408	1.10	0.88–1.37	0.333	0.339	1.38	0.72–2.67

JRA, Japan Racing Association.

Table 3 Average best race distance (BRD) for 120 males and 31 females that were Grade race winners in JRA races and were a subset of the 3706 individuals.

	Genotypes			*P*-value	Differences of distance between homozygous genotype (m)
	Homozygote	Heterozygote	Homozygote	ANOVA (Kruskal–Wallis)	
Male (*N* = 120)					
g.65809482T>C	T/T	T/C	C/C	<0.001 (<0.001)	859
	1422	1877	2281		
g.65868604G>T	G/G	G/T	T/T	<0.001 (<0.001)	901
	1422	1874	2324		
g.66493737C>T	C/C	C/T	T/T	<0.001 (<0.001)	783
	1524	1901	2307		
g.66539967A>G	A/A	A/G	G/G	<0.001 (0.010)	438
	1800	1917	2238		
Female (*N* = 31)					
g.65809482T>C	T/T	T/C	C/C	0.577 (0.423)	0
	1600	1739	1600		
g.65868604G>T	G/G	G/T	T/T	0.482 (0.319)	40
	1560	1742	1600		
g.66493737C>T	C/C	C/T	T/T	0.482 (0.319)	40
	1560	1742	1600		
g.66539967A>G	A/A	A/G	G/G	0.716 (0.652)	−44
	1644	1731	1600		

JRA, Japan Racing Association.

conducted a similar survey using the winners of Grade races, which included JRA G1, Jpn1, G2, Jpn2, G3 and Jpn3. As shown in Table 3, 120 males and 31 females taken from all 3706 individuals were used as the sample cohorts. Three SNPs, g.65809482T>C, g.65868604G>T, and g.66493737C>T, were associated with a clear difference in BRD among males. The average BRD tended towards longer distances compared with the average WRD, presumably because of the distance distribution of the Grade races. The frequency of short-distance races was low, while the frequency of middle/long-distance races was high in the Grade races (data not shown). Although we expected similar genetic trends for BRD among females, there was no difference in BRD observed, which is likely explained by the small sample size.

Genetic diagnostic markers for optimum race distances

These results suggest that each SNP influences WRD and BRD in the JRA horse racing system. The three SNPs g.65809482T>C, g.65868604G>T and g.66493737C>T were the most likely to have an influence on WRD and BRD in Japanese Thoroughbred racehorses. These three SNPs were shown to be in strong linkage disequilibrium, with $r^2 > 0.8$ (Tozaki *et al.* 2010). Thus, we investigated the influence of their combinations. To collect data from a larger number of individuals to more accurately infer haplo-

types, we used distance data from all the races won among the winners of all genotyped individuals (*n* = 3706). Table 4 shows WRD based on individual genotypes and diplotypes predicted from multiple SNPs. Although no clear differences were observed when single SNPs and diplotypes were evaluated, we observed a marginal trend towards a larger difference between WRD-short distance and WRD-long distance both in males and females when multiple SNPs were used. While this suggests that the use of multiple SNPs facilitates prediction of optimum race distance, multiple SNPs may not necessarily be required.

Discussion

In a previous genome-wide association study, we demonstrated that four SNPs associated with racing performance were located near the *MSTN* gene on ECA18 (Tozaki *et al.* 2010). In the present retrospective cohort study, we examined which aspects of racing performance were influenced by these SNPs in the JRA horse racing system. Optimum race distances, defined as WRD, were clearly influenced by genotype in both males and females, which is consistent with the findings reported by Hill *et al.* (2010a). Importantly, this study represents a valuable independent replication study, as the sample sets used here were entirely independent of the samples used in Hill *et al.* (2010a). The study using multiple SNPs (Table 4) suggests

Table 4 Average win-race distance (WRD) by genotype of single SNPs and diplotypes of multiple SNPs for all race winners in JRA races, taken from the 3706 individuals.

SNPs	Number of counted races	Genotypes			P-value	Differences of distance between homozygous genotype (m)
		Homozygote	Heterozygote	Homozygote	ANOVA (Kruskal–Wallis)	
Male						
g.65809482–g.65868604[1]	4810	TG/TG 1342	TG/CT 1620	CT/CT 1814	<0.001 (<0.001)	472
g.65868604–g.66493737[1]	4595	GC/GC 1341	GC/TT 1622	TT/TT 1826	<0.001 (<0.001)	485
g.65809482–g.66493737[1]	4551	TC/TC 1340	TC/CT 1623	CT/CT 1824	<0.001 (<0.001)	484
g.65809482T>C	4854	T/T 1342	T/C 1621	C/C 1808	<0.001 (<0.001)	466
g.65868604G>T	4854	G/G 1342	G/T 1619	T/T 1815	<0.001 (<0.001)	473
g.66493737C>T	4854	C/C 1369	C/T 1627	T/T 1825	<0.001 (<0.001)	457
g.66539967A>G	4854	A/A 1488	A/G 1653	G/G 1821	<0.001 (<0.001)	333
Female						
g.65809482–g.65868604[1]	1726	TG/TG 1262	TG/CT 1463	CT/CT 1565	<0.001 (<0.001)	303
g.65868604–g.66493737[1]	1650	GC/GC 1252	GC/TT 1465	TT/TT 1592	<0.001 (<0.001)	340
g.65809482–g.66493737[1]	1643	TC/TC 1252	TC/CT 1464	CT/CT 1592	<0.001 (<0.001)	340
g.65809482T>C	1733	T/T 1267	T/C 1463	C/C 1565	<0.001 (<0.001)	298
g.65868604G>T	1733	G/G 1262	G/T 1463	T/T 1565	<0.001 (<0.001)	303
g.66493737C>T	1733	C/C 1263	C/T 1461	T/T 1592	<0.001 (<0.001)	330
g.66539967A>G	1773	A/A 1362	A/G 1468	G/G 1570	<0.001 (<0.001)	208

[1]Individuals with the recombinant haplotype, such as TG/TT at g.65809482–g.65868604, were excluded from the calculations for WRD, as the haplotypes 'TGC' and 'CTT' for the three SNPs were the most common in the previous study (Tozaki *et al.* 2010).

JRA, Japan Racing Association.

that a susceptible SNP or other regulatory element that influences WRD may be located in the region between g.65868604G>T and g.66493737C>T, specifically in the regulatory region, such as upstream of the *MSTN* gene. However, re-sequencing in the immediately proximal region ~2 kb upstream did not identify a variant that performed better than the g.66493737C>T SNP for the prediction of optimum race distance (Hill *et al.* 2010b). The g.66493737C>T SNP disrupts a transcription factor binding site, and functional studies will be required to

determine whether this locus represents the functional variant.

The distribution of race distances in JRA seems to be a major contributing factor to the observed influence of genotypes on performance rank and lifetime earnings. The distribution of race distances competed within the JRA was considerably unbalanced, with 1200 (short) and 1800-m (middle) races conducted more frequently (Figs 1 & S1). This imbalance indicates that individuals with the genotype suited to these distances may be more likely to achieve a

more favourable performance rank and to obtain higher lifetime earnings. While an association between genotypes and optimum race distance was observed using Japanese Thoroughbreds, the distribution of race distances in the JRA means that the short-distance alleles (*i.e.* T-allele at g.65809482T>C, G-allele at g.65868604G>T and C-allele at g.66493737C>T), as well as short- and middle-distance genotypes by their combinations, would result in the greatest opportunity to achieve a favourable performance rank (Table 2) and to acquire higher lifetime earnings (Table 1). In particular, short-distance alleles may have a dominant effect on performance rank within the JRA racing system (Table 2). These observations explain how our previous study (Tozaki *et al.* 2010) identified the candidate genomic region on ECA18 using lifetime earnings as the performance phenotype. While the actual trait associated with this region on ECA18 seems to be optimum race distance, it may be that the previous identification of association with lifetime earnings may have been observed because of the relatively low variation in non-genetic factors (*i.e.* frequency of races at certain distances), as we have used exclusively racing and pedigree information from Japanese Thoroughbreds competing in the JRA. In contrast, Hill *et al.* (2010a) did not observe a statistically significant association with racing performance for the g.66493737C>T SNP, as the horses in that study had competed in Europe, Australasia and North America, where the race distances of the elite races were not biased as in the JRA. Our successful GWAS (Tozaki *et al.* 2010) indicates the power to identify candidate genomic regions for interesting traits by collecting samples under the same environmental conditions, such as the same racing environments, even if the trait was indirectly associated with the candidate genomic region.

Figure S1 shows the distribution of WRD for all the individuals whose genotype was determined in this study. Individuals with a short-distance-suited genotype were distributed mainly between 1000 and 1800 m, while those with a middle-distance-suited genotype raced favourably between 1200 and 2000 m. The distribution of those with a long-distance-suited genotype was similar to that of middle distance, although a slight shift towards longer distances was observed. It should be noted that our findings do not mean that a horse with a long-distance-suited genotype can never win short-distance races; some horses with this genotype won shorter distances, such as 1000–1400 m, albeit with a low frequency. However, horses are more likely to compete favourably within the distance range that they are most genetically suited to.

In summary, of the four SNPs, g.65809482T>C, g.65868604G>T and g.66493737C>T were found to have similar diagnostic effects for optimum race distance. This finding is supported by the strong linkage disequilibrium ($r^2 > 0.8$) associating these three SNPs. In addition, the use of multiple SNPs, such as g.65809482T>C–g.66493737C>T and g.65868604G>T–g.66493737C>T,

may lead to a more accurate genetic prediction of WRD and BRD. This finding suggests that a cis-regulatory element or an actual susceptible SNP for racing performance is located in the candidate genomic region on ECA18 at the *MSTN* gene locus.

Acknowledgements

Conflicts of interest

Equinome Ltd. has been granted a license for commercial use of data, which is contained within patent applications: United States Provisional Serial Number 61/136553 and Irish patent application number 2008/0735. Patent Cooperation Treaty filing: 'A method for predicting athletic performance potential,' September 7, 2009. The following author is named on the applications: EH.

References

Binns M.M., Boehler D.A. & Lambert D.H. (2010) Identification of the *myostatin* locus (MSTN) as having a major effect on optimum racing distance in the Thoroughbred horse in the USA. *Animal Genetics* **41**(Suppl. 2), 154–8.

Bower M., Campana M., Whitten C. *et al.* (2011) The cosmopolitan maternal heritage of the Thoroughbred racehorse breed shows a significant contribution from British and Irish Native mares. *Biology Letters* **7**, 316–20.

Cunningham E.P., Dooley J.J., Splan R.K. & Bradley D.G. (2001) Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses. *Animal Genetics* **32**, 360–4.

Gaffney B. & Cunningham E.P. (1988) Estimation of genetic trend in racing performance of Thoroughbred horses. *Nature* **332**, 722–3.

Grobet L., Martin L.J., Poncelet D. *et al.* (1997) A deletion in the bovine *myostatin* gene causes the double-muscled phenotype in cattle. *Nature Genetics* **17**, 71–4.

Gu J., Orr N., Park S.D., Katz L.M., Sulimova G., MacHugh D.E. & Hill E.W. (2009) A genome scan for positive selection in thoroughbred horses. *PLoS ONE* **4**, e5767.

Gunn H.M. (1987) Muscle, bone and fat productions and muscle distribution of thoroughbreds and quarter horses. In: Equine exercise physiology 2: Proceedings of the Second International Conference on Equine Exercise Physiology; August 7-11 1986; San Diego, California, United States, Available at: http://www.iceep.org/pdf/iceep2/_1129101114_001.pdf.

Hill E.W., Bradley D.G., Al-Barody M., Ertugrul O., Splan R.K., Zakharov I. & Cunningham E.P. (2002) History and integrity of

thoroughbred dam lines revealed in equine mtDNA variation. *Animal Genetics* **33**, 287–94.

Hill E.W., Gu J., Eivers S.S., Fonseca R.G., McGivney B.A., Govindarajan P., Orr N., Katz L.M. & MacHugh D.E. (2010a) A sequence polymorphism in *MSTN* predicts sprinting ability and racing stamina in thoroughbred horses. *PLoS ONE* **5**, e8645.

Hill E.W., McGivney B.A., Gu J., Whiston R. & MacHugh D.E. (2010b) A genome-wide SNP-association study confirms a sequence variant (g.66493737C>T) in the equine *myostatin* (*MSTN*) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses. *BMC Genomics* **11**, 552.

Jones J.H. & Lindstedt S.L. (1993) Limits to maximal performance. *Annual Review of Physiology* **55**, 547–69.

Jones J.H., Longworth K.E., Lindholm A., Conley K.E., Karas R.H., Kayar S.R. & Taylor C.R. (1989) Oxygen transport during exercise in large mammals. I. Adaptive variation in oxygen demand. *Journal of Applied Physiology* **67**, 862–70.

McPherron A.C. & Lee S.J. (1997) Double muscling in cattle due to mutations in the *myostatin* gene. *Proceedings of the National Academy of Sciences USA* **94**, 12457–61.

Mosher D.S., Quignon P., Bustamante C.D., Sutter N.B., Mellersh C.S., Parker H.G. & Ostrander E.A. (2007) A mutation in the *myostatin* gene increases muscle mass and enhances racing performance in heterozygote dogs. *PLoS Genetics* **3**, e79.

Mota M.D., Abrahão A.R. & Oliveira H.N. (2005) Genetic and environmental parameters for racing time at different distances in Brazilian Thoroughbreds. *Journal of Animal Breeding and Genetics* **122**, 393–9.

Penedo M.C., Millon L.V., Bernoco D. *et al.* (2005) International Equine Gene Mapping Workshop Report: a comprehensive linkage map constructed with data from new markers and by merging four mapping resources. *Cytogenetics and Genome Research* **111**, 5–15.

Raudsepp T., Gustafson-Seabury A., Durkin K. *et al.* (2008) A 4,103 marker integrated physical and comparative map of the horse genome. *Cytogenetics and Genome Research* **122**, 28–36.

Schuelke M., Wagner K.R., Stolz L.E., Hübner C., Riebel T., Kömen W., Braun T., Tobin J.F. & Lee S.J. (2004) Myostatin mutation associated with gross muscle hypertrophy in a child. *New England Journal of Medicine* **350**, 2682–8.

Sobczynska M. (2006) Genetic correlation between racing performance at different racing distances in Thoroughbreds and Arab horses. *Czech Journal of Animal Science* **51**, 523–8.

Swinburne J.E., Boursnell M., Hill G. *et al.* (2006) Single linkage group per chromosome genetic linkage map for the horse, based on two-three-generation, full-sibling, crossbred horse reference families. *Genomics* **87**, 1–29.

Tolley E.S., Notter D.R. & Marlowe T.J. (1985) A review of the inheritance of racing performance in horses. *Animal Breeding Abstracts* **53**, 163–85.

Tozaki T., Swinburne J., Hirota K., Hasegawa T., Ishida N. & Tobe T. (2007) Improved resolution of the comparative horse-human map: investigating markers with in silico and linkage mapping approaches. *Gene* **392**, 181–6.

Tozaki T., Miyake T., Kakoi H., Gawahara H., Sugita S., Hasegawa T., Ishida N., Hirota K. & Nakano Y. (2010) A genome-wide association study for racing performances in Thoroughbreds clarifies a candidate region near the *MSTN* gene. *Animal Genetics* **41**(Suppl. 2), 28–35.

Wade C.M., Giulotto E., Sigurdsson S. *et al.* (2009) Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* **326**, 865–7.

Williamson S.A. & Beilharz R.G. (1998) The inheritance of speed, stamina and other racing performance characteristics in the Australian thoroughbred. *Journal of Animal Breeding and Genetics* **115**, 1–16.

Young L.E., Marlin D.J., Deaton C., Brown-Feltner H., Roberts C.A. & Wood J.L. (2002) Heart size estimated by echocardiography correlates with maximal oxygen uptake. *Equine Veterinary Journal Supplement* **34**, 467–71.

Supporting information

Additional supporting information may be found in the online version of this article.

Figure S1 Genotype distributions for three SNPs (g.65809482T>C, g.65868604G>T and g.66493737C>T) for WRD among all the Japanese Thoroughbred racehorses used in this study.

Table S1 Primer sequences for asymmetric PCR with high-resolution melting and unlabeled probe methods using a LightScanner.

Table S2 Lifetime earnings, performance rank, athletic period, and genotype frequency of the SNPs for each cohort used in this study.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.