



Original Research

Genotype Imputation and Accuracy Evaluation in Racing Quarter Horses Genotyped Using Different Commercial SNP Panels



Guilherme L. Pereira^{a,b,*}, Tatiane C.S. Chud^a, Priscila A. Bernardes^a, Guilherme C. Venturini^b, Luís A.L. Chardulo^b, Rogério A. Curi^b

^a Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista (Unesp), Jaboticabal, SP, Brazil

^b Faculdade de Medicina Veterinária e Zootecnia, Universidade Estadual Paulista (Unesp), Botucatu, SP, Brazil

ARTICLE INFO

Article history:

Received 28 March 2017

Received in revised form 25 June 2017

Accepted 31 July 2017

Available online 3 August 2017

Keywords:

Chromosome

Population structure

IBD

MAF

Concordance rate

ABSTRACT

The availability of large-scale SNP genotyping panels for horses has opened up new possibilities for genetic studies of the species. The first panel version contained around 54,000 SNPs (54k). The current second generation panel possesses approximately 65,000 SNPs (65k), which 19,000 of them are new markers and 45,000 were present in the first panel. The objectives of this study were to perform two-step genotype imputation between individuals in a relatively small population sample of racing Quarter Horses genotyped with the 54k or 65k panel and to evaluate the accuracy of imputation in a practical situation. Genotypes were imputed using 116 horses genotyped with the 54k SNP array and 233 animals genotyped with the 65k array. For the simulations, random samples were chosen to compose the imputed (about 50% samples) and reference populations in two scenarios. Scenario A simulated the genotype imputation in the first step (from 65k to 54k) and the scenario B in the second step (from 54k to 65k). For each scenario, the individuals genotyped with the same panel were considered as the reference population but that were included in the imputed group had their exclusive genotypes masked for subsequent evaluation of imputation accuracy. Thus, after quality control, the 54k and 65k panels contained 7,048 and 16,940 exclusive markers, respectively. The mean concordance rate was 0.9815 and 0.9751 for scenarios A and B, and the mean allelic r^2 was 0.9791 and 0.974, respectively. No important influences of the genomic relationship coefficients on imputation accuracy were observed between the imputed and reference samples. Thus, two-step imputation of exclusive genotypes between the 54k and 65k equine panels resulted in a considerable increase of SNPs for all samples (12%–26%) without losses in the quality of information.

© 2017 Elsevier Inc. All rights reserved.

Animal welfare/ethical statement: All procedures involving the animals were conducted according to Brazilian guidelines on animal welfare (protocol 157/2014) and were approved by the Ethics Committee on Animal Use (CEUA, FMVZ, Unesp, Botucatu, SP). The authors declare to be in accordance with the national statutes of animal welfare and ethics.

Conflict of interest statement: The authors declare no conflicts of interest.

* Corresponding author at: Guilherme L. Pereira, Faculdade de Medicina Veterinária e Zootecnia, Universidade Estadual Paulista, Rua Prof. Dr. Walter Mauricio Correra, s/n - Rubião Junior, Botucatu, SP 18618-970, Brazil.

E-mail address: guipicoia@hotmail.com.br (G.L. Pereira).

1. Introduction

The advent of next generation sequencing technology has permitted the sequencing and mapping of thousands of polymorphisms spread across the equine genome [1,2], which resulted of the creation of large-scale SNP genotyping panels, opening up new possibility for genetic studies of the species.

In its first version, the genotyping platform for the species, the Equine SNP50 BeadChip produced by Illumina, Inc (USA),

contained about 54,000 SNPs (54k). The current second generation of the Illumina chip (Equine SNP70 BeadChip; Illumina, Inc) possesses approximately 65,000 SNPs (65k); of these, 19,000 are new markers and 45,000 were present in the Equine SNP50 BeadChip. The genotyping of SNPs with the two different equine panels in the same experiment or in different experiments can result in an loss of data (about 9k and 20k for the panels of 54k and 65k in no QC data, respectively), especially when it is necessary to unify data for genetic studies and to use only SNPs in common between the two panels. However, the imputation of SNP genotypes based on population linkage disequilibrium (LD) data can offer a potential solution to this problem, that is, it allows the compatibility and continuity of existing studies.

In general, genotype imputation permits to combine data derived from population samples genotyped with panels of different densities [3,4], to generate high-density (HD) panels for animals genotyped with low-density panels [5,6], and to predict missing genotypes due to the application of quality control (QC) of the data or reading errors. In this respect, to increase the success of genotype imputations, several studies have investigated the factors that directly influence the accuracy of imputation such as the size of the reference population [7,8], the degree of relationship between the reference and imputed population [8,9], panel density [7,10], and minor allele frequency (MAF) [5,8,10]. However, imputation in real situations of sets genotyped with different panels may be a challenge in terms of the inference of accuracy. On this point, some analyses and simulations can be performed using a single data set to choose the most efficient strategy for imputation, as well as to obtain parameters of reliability of the imputations performed under these circumstances.

The Quarter Horse is one of the most popular horse breeds in the world because of its versatility, robustness, and docile temperament and is widely used for working cattle, functional equestrian tests, and speed tests [11]. As the fastest breed in the world, the racing line of Quarter Horses shows better performance in short-distance races than any other horse breed [12].

Due the lack of research involving genotypes imputation in equine breeds and the importance of the racing line of Quarter Horses for world horse, the objectives of this study were to perform two-step genotype imputation between individuals in a relatively small population sample genotyped with the 54k or 65k panel and to evaluate the accuracy of imputation in a practical situation.

2. Material and Methods

2.1. Animals

A total of 360 male and female Quarter Horses of the racing line, registered with the Brazilian Association of Quarter Horse Breeders (ABQM in the Portuguese acronym), were used. Blood was collected from 120 of these horses in 2011 and from the remaining 240 animals during the first half of 2015. These animals, 78 males and 282 females, were born to 83 stallions and 249 mares, resulting in an average of 4.3 offspring/stallion and of 1.4 offspring/mare. The animals belonged to 159 breeders and were

housed at the Sorocaba Jockey Club (Sorocaba, São Paulo, Brazil) and on 25 farms located in towns of the state of São Paulo, Brazil. The collection and, consequently, the presence of full-sibs in the sample were avoided.

All procedures involving the animals were conducted according to Brazilian guidelines on animal welfare (protocol 157/2014) and were approved by the Ethics Committee on Animal Use (CEUA, FMVZ, Unesp, Botucatu, SP).

2.2. SNP Genotyping and QC

Among the 360 horses used in this study, 120 were genotyped in 2011 using the Equine SNP50 BeadChip (Illumina, Inc). The remaining DNA samples ($n = 240$) were genotyped in 2015 using the Equine SNP70 BeadChip (Illumina, Inc). The two arrays were read with the HiScan system (Illumina, Inc). Quality control analysis of the genotype data obtained with the 54k or 65k arrays ($n = 120$ and $n = 240$) was performed using the *snpStats* package (Clayton, 2015) of the R program (R Core Team, 2016). Animals with a call rate < 0.9 were excluded from the data set (Table 1). SNPs located on the X chromosome, SNPs with a call rate < 0.9 , and SNPs with a P -value $< 1 \times 10^{-5}$ for Hardy–Weinberg equilibrium were eliminated (Table 1). The MAF was not used initially as an exclusion criterion of SNPs since different ranges of MAF were applied to verify the efficiency of two-step genotype imputation between the SNP50 and SNP70 equine chips.

2.3. Population Analysis

For the analysis of population stratification, all SNPs shared between the two panels that passed QC were submitted to pruning for the selection of markers that were not in LD, considering $r^2 < 0.2$. The probability of being identical by state (IBS) among all pairs of samples was calculated using the autosomal SNPs that remained after pruning. Finally, for visual analysis, a multidimensional scaling plot was obtained with the MDS plot option of the PLINK 1.07 program [13].

The relationship coefficients, that is, the probability of individuals sharing markers that are identical by descent (IBD), were calculated for all pairs of samples considering only SNPs shared between the two panels. Markers that are IBD can be used to identify individuals that are closer than

Table 1
Number of markers of the Equine SNP50 BeadChip and Equine SNP70 BeadChip and the number of SNPs excluded per quality control criterion used.

Criteria used	SNP50 (n)	SNP70 (n)
Criterion Used for the Exclusion of Samples		
Call rate < 0.9	4	7
Remaining samples	116	233
Criterion used for the exclusion of SNPs		
Genotyped SNPs	54,602	65,157
Located on the X chromosome	3,223	3,411
Call rate < 0.9	148	504
Hardy–Weinberg equilibrium ($p < 1 \times 10^{-5}$)	4,519	4,638
Remaining SNPs	46,712	56,604

it would be expected in a homogenous sample, that is, are related [13]. The IBD estimates were given by Z0, Z1, and Z2, which represent the probability that two markers share no IBD marker, one IBD marker, and two IBD markers, respectively. The proportion of IBD between samples was given by $PI_HAT = p(Z2) + p(Z1)$.

The pruning procedures and estimation of IBS and IBD were performed with the PLINK 1.07 program [13]. The mean IBD proportion of each sample and the mean relationship between samples of the imputed set (which had masked markers) and the reference set were obtained with the R program [14].

2.4. Genotype Imputation and Evaluation of Imputation Accuracy

Genotype imputation was performed using the Flmpite v.2.2 program [5]. As described above, the individuals used in this study were genotyped exclusively with one or the other panel. Thus, final imputation was carried out in two steps (i.e., from the 65k to the 54k chip and vice-versa).

Since the data set did not contain animals genotyped with both panels, the imputation accuracy of the markers between panels was estimated through simulations involving different scenarios. Two imputation scenarios were proposed to evaluate the capacity of correctly imputing exclusive SNPs of each panel. First, the 65k panel was used as the reference and the 54k panel as the imputed set (scenario A). This scenario considered all markers present in the 65k panel and only those shared with the 54k panel. To assess the imputation accuracy of SNPs exclusively found in SNP70, approximately 50% of the samples genotyped with this chip were chosen randomly. The remaining samples genotyped with the SNP70 chip had

Table 2

Minimum, mean, and maximum relationship coefficients obtained by means of the IBD matrix for all samples and for samples imputed in each step (54k–65k and 65k–54k).

Samples	Minimum	Mean	Maximum
Total	0.00360	0.09620	0.21100
65k–54k ^a (scenario A)	0.02072	0.09909	0.18300
54k–65k ^a (scenario B)	0.01449	0.10030	0.21890

Abbreviation: IBD, identical by descent.

^a Mean IBD values obtained for imputed samples in relation to the reference samples.

their exclusive markers masked (removed) and were included in the samples to be imputed, that is, samples genotyped with the 54k panel. The same procedure was adopted in the second step (scenario B), which consisted of inverting the panels used as the reference and imputed set.

The accuracy of imputation was evaluated using the concordance rate (CR), which corresponds to the proportion of correctly imputed genotypes. In addition, the allelic r^2 (allelic correlation) was estimated, which is determined by the square of the correlation between the dosage of the imputed alleles (allele of smaller effect) and the allele dosage of the original genotype. The accuracy values were obtained per SNP (means per MAF range and chromosomes) and per sample.

3. Results and Discussion

3.1. Population Structure and Analysis of Relationship

Analysis of population stratification was performed to verify the existence of clusters exclusively consisting of individuals genotyped with the same chip in the total sample of animals. Different substructures were observed in the MDS plot (Fig. 1), but none of them involved only samples genotyped with one or the other array in the densest regions of the graph.

This result indicates the absence of exclusive clusters of animals genotyped with the 54k or 65k panel, that is, the samples genotyped with the panels of lower and higher density do not represent two distinct populations. The occurrence of distinct populations in the reference and imputed sets may result in low accuracy values, that is, the imputed genotypes are not reliable.

Table 3

Number of markers of the Equine SNP50 BeadChip and Equine SNP70 BeadChip and those shared between the two panels after quality control and number of remaining markers after two-step imputation.

Panels	SNPs After QC (n)
SNP50	46,712
Exclusive SNP50 ^a	7,048
SNP70	56,604
Exclusive SNP70 ^a	16,940
Shared panel (SNP50 and SNP70)	39,664
Panel after imputation	63,652

Abbreviation: QC, quality control.

^a SNPs that were imputed to the other panel.

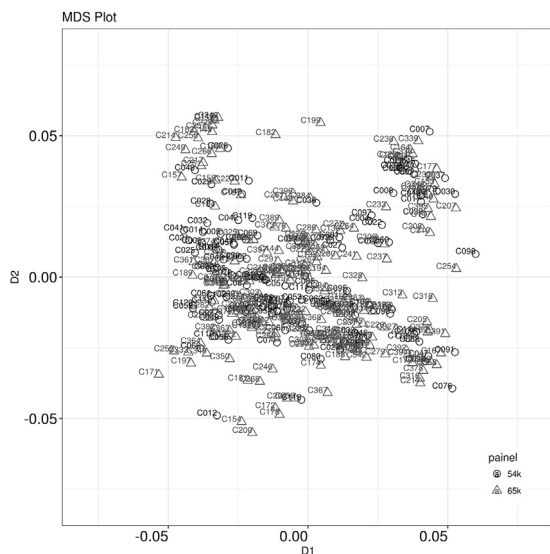


Fig. 1. MDS plot of the total data set considering 39,664 SNPs shared between the two equine arrays. The circles and lozenges correspond to animals genotyped with the 54k and 65k arrays, respectively. MDS, multidimensional scaling; SD, single nucleotide polymorphism.

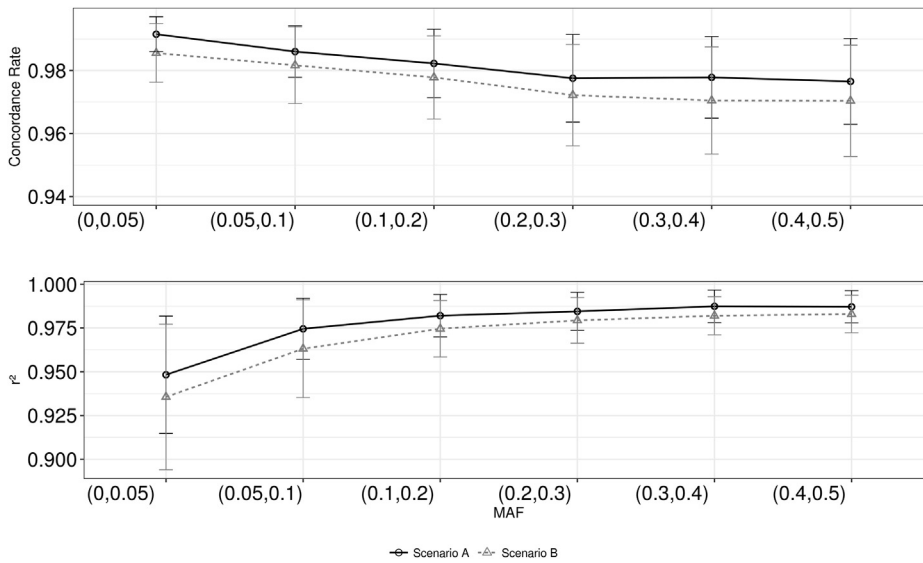


Fig. 2. Imputation efficiency for two scenarios and different ranges of MAF (x-axis) measured by mean concordance rate (CR) and allelic r^2 (y-axis). Each line and point indicate a different imputation scenario. MAF, minor allele frequency.

The IBD estimates for the determination of the average relationship coefficient based on SNP markers were given by the average probability to contain one or two IBD markers (IBD proportion). In general, the genomic relationship average values for each pair of individuals agreed with the pedigree information. Animals with an IBD proportion above the average (0.0962) exhibited a higher level of relationship (first and second degree) compared to all individuals of the population, generally grandparents and parents. The mean and maximum relationship of the imputed population compared to the reference population was higher in scenario B than in scenario A, while the minimum relationship was higher in scenario A (Table 2). The mean and maximum relationship coefficients obtained for the imputed populations in each scenario were similar to those found for the total data set (Table 2).

3.2. Imputation Accuracy per SNP

Two-step imputation of SNPs resulted in a considerable increase of information for each panel. Table 3 shows the number of SNPs shared and exclusive of each panel, as well as the total number of SNPs after two-step imputation. The gain in the number of markers in relation to the final imputed panel (65k) was 26% for SNP50, exceeding the original number of SNPs and 12% for SNP70.

In scenario A (65k–54k), 113 individuals were considered for the reference population and 236 for the imputed population, including 116 and 120 individuals genotyped with the 54k and 65k arrays, respectively. In scenario B (54k–65k), 50 individuals were considered for the reference population and 299 for the imputed population, including 66 and 233 individuals genotyped with the 54k and 65k arrays, respectively. The use of approximately 50%

Table 4

Imputation accuracies measured by the concordance rate (CR) and allelic r^2 considering all imputed samples and samples with relationship above and below the average IBD for scenarios A and B in racing Quarter Horses.

Scenario	Step	Imputed Population	Mean CR	SD	Minimum	Maximum	Median
A	65k–54k	Total (n = 120)	0.9814	0.01173	0.9152	0.9989	0.9851
		High relationship (n = 56)	0.9841	0.00769	0.9671	0.9952	0.9864
		Low relationship (n = 58)	0.9794	0.01013	0.9534	0.9961	0.9813
B	54k–65k	Total (n = 66)	0.9751	0.01362	0.9408	0.9949	0.9774
		High relationship (n = 30)	0.9801	0.01108	0.9536	0.9949	0.9825
		Low relationship (n = 34)	0.9704	0.01412	0.9408	0.9911	0.9738
Scenario	Step	Imputed population	Mean r^2	SD	Minimum	Maximum	Median
A	65k–54k	Total (n = 120)	0.9868	0.00779	0.9538	0.9997	0.9894
		High relationship (n = 56)	0.9884	0.00631	0.9733	0.9967	0.9904
		Low relationship (n = 58)	0.9848	0.00789	0.9643	0.9974	0.9872
B	54k–65k	Total (n = 66)	0.9811	0.01098	0.9527	0.9968	0.9828
		High relationship (n = 30)	0.9849	0.00930	0.9630	0.9968	0.9870
		Low relationship (n = 34)	0.9775	0.01127	0.9527	0.9943	0.9792

Abbreviations: IBD, identical by descent; SD, standard deviation.

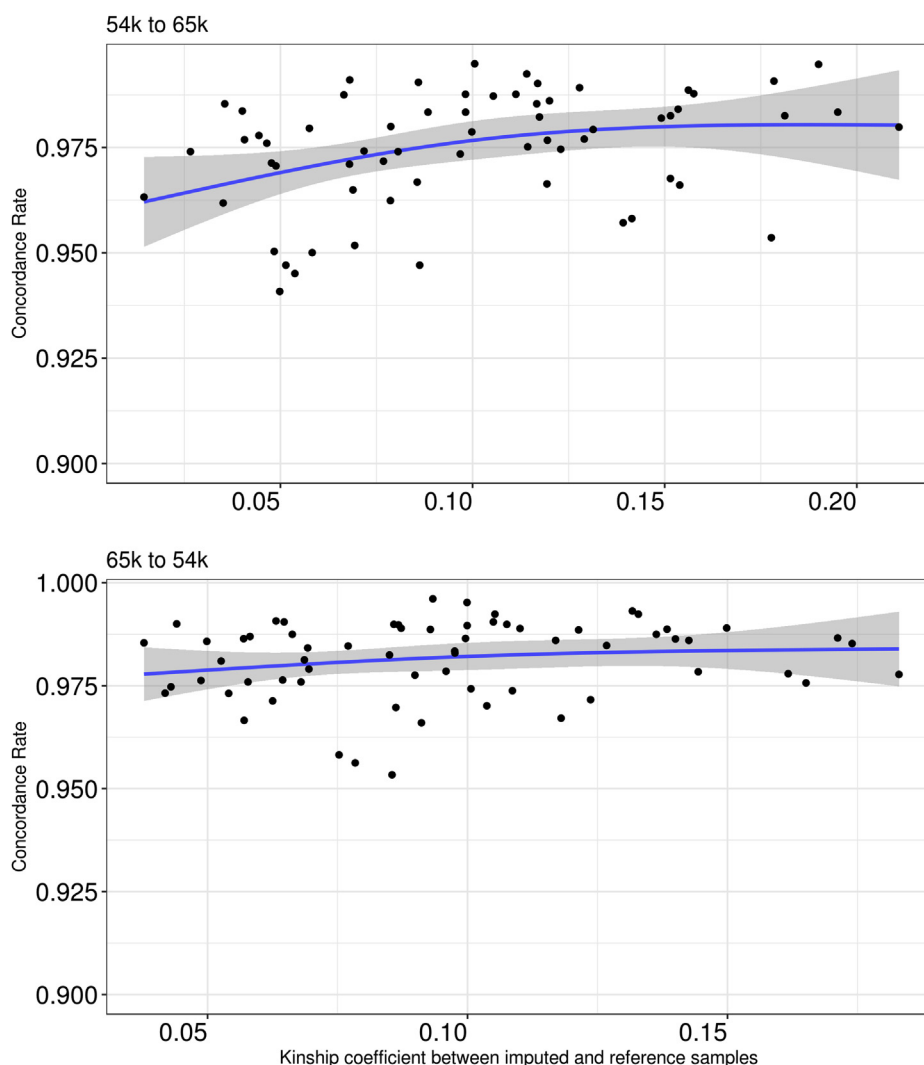


Fig. 3. Regression of the concordance rate (CR) (y-axis) and relationship coefficient given by the average IBD between samples of the imputed and reference populations (x-axis) for two scenarios in racing Quarter Horses. IBD, identical by descent.

of the samples as reference set and 50% as imputed set permitted to maintain a satisfactory number of individuals for both the imputation process and the evaluation of accuracy. For the imputed populations, only genotypes shared between the two panels were considered (39,664 SNPs). In each scenario, individuals genotyped with the same array were considered as the reference population but that were included in the imputed group had their exclusive genotypes masked for the subsequent evaluation of imputation accuracy. Thus, after QC, the 54k and 65k panels contained 7,048 and 16,940 exclusive markers, respectively.

The mean CR was 0.9815 ± 0.02882 and 0.9751 ± 0.03731 for scenarios A and B, respectively, and the mean allelic r^2 was 0.9791 ± 0.04209 and 0.9740 ± 0.04911 . The imputation accuracies from 65k to 54k were higher than vice-versa. The imputation accuracy was also analyzed for each scenario based on the mean CR and allelic r^2 for each range of MAF.

As expected for CR, higher accuracies were obtained for lower MAF (Fig. 2). The allelic correlation (r^2) has been commonly used to estimate the imputation accuracy for alleles with low MAF to minimize the dependence on allele frequency [5]. However, the calculation of r^2 per SNP (and not per sample) requires a large number of animals to compose the imputed set, in order to obtain nonbiased estimates [8]. Moreover, the size of the sample needs to be sufficiently large to contain the minimum number of alleles with low MAF to be imputed; otherwise, the r^2 values might be estimated with large standard errors [15]. Consequently, the allelic correlation estimates may be poorly reliable for imputation scenarios with few animals.

On the other hand, the CR is not affected by the size of the population. However, CR tends to be overestimated, especially in the case of rare alleles or those with low MAF [16]. Other factors that can affect CR are the genetic composition of the reference and imputed samples and the

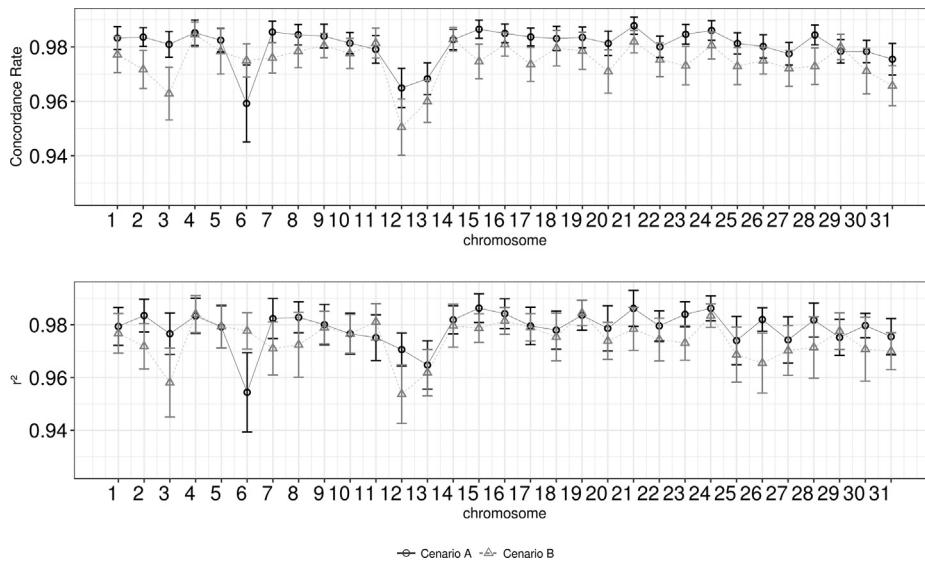


Fig. 4. Imputation accuracy per equine autosomal chromosome (31) (x-axis) for two scenarios measured by mean concordance rate (CR) and allelic r^2 (y-axis) of true and imputed genotypes. Each point corresponds to a different imputation scenario.

marker density of the arrays to be imputed. Chud et al [9] in a composite beef cattle breed found CR ranged from 0.75 to 0.97 for imputations from the 3k to HD panel (777k) and from 80k to HD, respectively. On the other hand, very small numbers of markers (1–2k) can slightly reduce imputation accuracy as observed by Ogawa et al [17] in cattle. Very low CR values for ranges of MAF less than 0.1 can lead to low reliability of imputed genotypes considering that these values are generally overestimated.

3.3. Imputation Accuracy per Sample

To imputed samples, accuracies were obtained based on mean CR and allelic r^2 (Table 4). Next, to investigate the impact of relationship on imputation accuracy, accuracy was estimated for subsets divided according to average IBD for imputed samples obtained in each scenario, in which individuals with a relationship coefficient above the average were classified as high relationship and those with a coefficient below the average as low relationship (Table 4). Due to the small number of samples in the present study, building additional scenarios based on the number of individuals that compose each population (reference or imputed) would not be relevant.

No significant differences in mean CR and allelic r^2 were observed between imputed samples of either scenario or between individuals classified as high and low relationship. The mean CR and allelic r^2 for high-relationship samples had lower standard deviations than those for low-relationship samples in scenarios A and B. However, higher standard deviations were observed in scenario B, which are possibly related to the relatedness structure between imputed and reference samples and the smaller number of samples present.

No influence of relationship on imputation accuracy was observed in scenario A (65k–54k) ($P > .1$). However,

in scenario B, a higher imputation accuracy was found for animals with higher average relationship coefficients compared to the reference population ($P < .01$) (Fig. 3).

The coefficients of determination (R^2) were low in both situations (0.029 and 0.113 for scenarios A and B, respectively), indicating that in this study, the accuracy of imputation was not strongly influenced by the degree of relatedness between individuals of the imputed and reference populations. In contrast to these findings, Chud et al [9] demonstrated a strong association of the degree of relationship between individuals of the population with imputation accuracy. Although the Flmpuete software combines imputation methodologies (family and population), the population-based method differs from most of the other softwares because it assumes that all animals are related to some extent. According to Kong et al [18], pedigree information is an important factor for the identification of haplotypes and for imputation. However, significant gains in accuracy have only been obtained in situations in which the individuals genotyped with different panels are genetically related and the imputations occur from low-density panels (384, 3k, or 6k) to HD panels [19].

Studies evaluating the impact of relatedness information on genotype imputations in cattle considered thousands of animals and several complete generations in their pedigrees [5,8,19]. In the present study, the inclusion of pedigree information would be irrelevant considering the small number of SNPs to be imputed in relation to the number shared between genotyped samples and the relatively small size of the total data set. In addition, the failure of breeder associations to make pedigree files systematically available, as well as the occurrence of registration errors, may result in losses of accuracy during the imputation process.

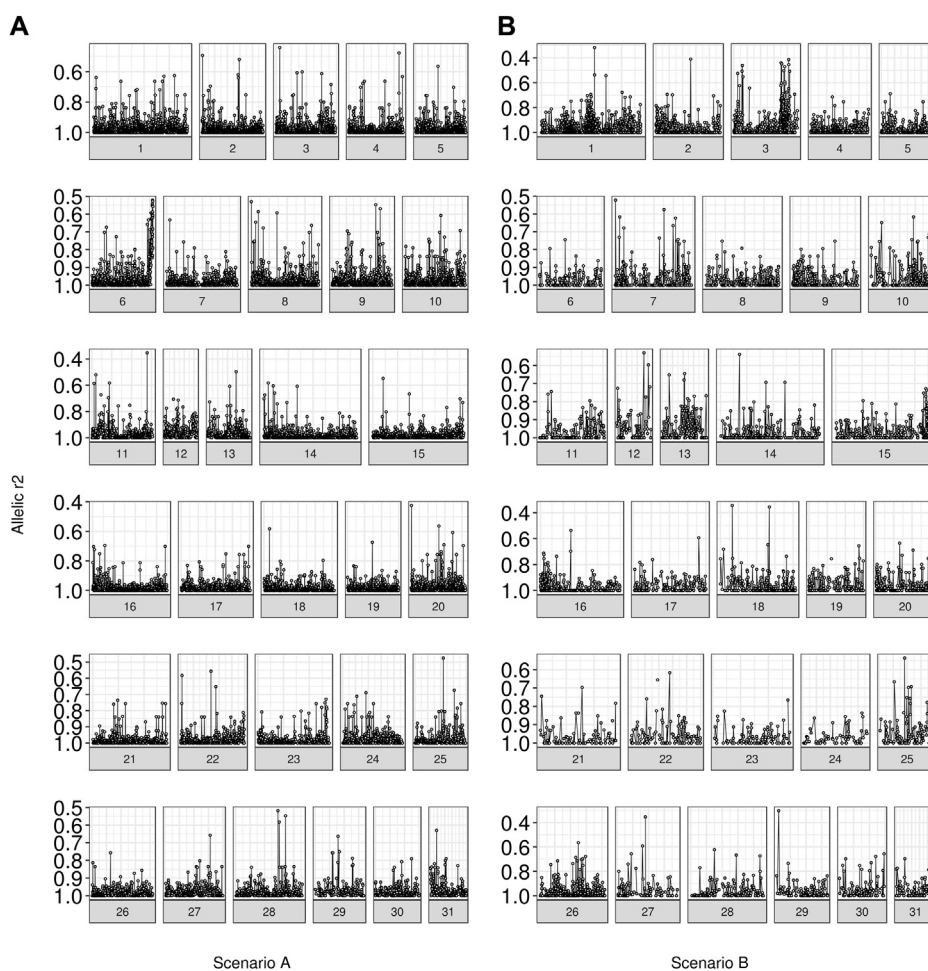


Fig. 5. Imputation accuracy measured by concordance rate (CR) (y-axis) per SNP and per chromosome (x-axis) for two scenarios (A) and (B) in racing Quarter Horses.

The fact that the animals studied here are relatively close genetically, especially when the minimum relationship between the imputed and reference sets of 0.02072 for scenario A and of 0.01449 for scenario B was considered (Table 3)—indicating an important degree of relationship even between less closely related individuals contributes to genotype imputation with high accuracy.

3.4. Imputation Accuracy per Chromosome

The mean CR per chromosome ranged from 0.9505 to 0.9845 and from 0.9592 to 0.9877 for scenarios B and A, respectively, and the mean allelic r^2 from 0.9536 to 0.9840 and from 0.9544 to 0.9862 (Fig. 4). Low values of CR and r^2 were observed for ECA 3, 12, 13, and 25 in scenario B and for ECA 6 and 12 in scenario A. The mean CR and r^2 values found for each chromosome are similar to those reported by Piccoli et al [19] and Ventura et al [8]. However, the average accuracies per chromosome do not follow the patterns described in a study on beef cattle [19]. In that study, chromosome length was associated with imputation accuracy, in which the highest CR was obtained for chromosome 1 (the largest

bovine chromosome) and the lowest for chromosome 28 (the smallest bovine chromosome).

Fig. 5 shows the plot of CR per SNP for each chromosome. In scenario A, a low imputation accuracy was observed for a large part of the terminal region of ECA 6 (lowest mean CR for this scenario) and for the central region of ECA 12 and 13. In scenario B, greater dispersion of the accuracy values was generally observed for practically all chromosomes, especially after ECA 17. A low imputation accuracy was found for the initial and terminal regions of ECA 3 and the central region of ECA 12 (chromosomes for which the lowest mean CR values were obtained in this scenario).

Ventura et al [8] analyzed the proximal and telomeric regions of each of the 26 autosomal chromosomes of sheep (using 100 SNPs per end). The authors found that most chromosomes had problems at least at one of the ends, except for chromosome 26 which had an accuracy higher than 60% at each end. This fact was not observed in this study, in which SNPs with low imputation accuracy were distributed across the chromosomes.

The imputation of exclusive genotypes between chips performed in this work differs from others because it has a

very practical approach. The absence of genotyped animals with both arrays (common in genotyped equine research in different years), either lack of samples of biological material (DNA, blood, tissues...) from genotyped horses with old chips or lack of arrays that were discontinued, did not allow the accuracy of imputation that was estimated directly, which would be ideal. However, the imputation accuracies obtained for imputation between exclusive markers follow common simulations used in other works, in which part of the genotypes are removed and later imputed. In this sense, the approach used in this work provides reliable results of imputation accuracy of genotypes in the racing line of Quarter Horses.

4. Conclusions

The gain in quality genotype information for all individuals of racing Quarter Horses used in this study obtained by two-step imputation (from 65k to 54k and from 54k to 65k) was considerable, especially in view of the small size of the sample, a common situation in studies on horses compared to those involving other livestock species such as cattle.

The strategies for genotype imputation and accuracy evaluation adopted in the present study were effective and may be useful in similar situations, that is, a relatively small sample size, animals genotyped exclusively with one or the other array, and inconsistencies on pedigree information, but with genetically related individuals.

Acknowledgments

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp), São Paulo, Brazil—process number: 2014/20207-1.

References

- [1] Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, et al. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 2009;326:865–7.
- [2] Doan R, Cohen ND, Sawyer J, Ghaffari N, Johnson CD, Dindot SV. Whole-genome sequencing and genetic Variant analysis of a Quarter horse mare. *BMC Genomics* 2012;13:78.
- [3] Hayes BJ, Bowman PJ, Daetwyler HD, Kijas JW, Van Der Werf JHJ. Accuracy of genotype imputation in sheep breeds. *Anim Genet* 2011;43:72–80.
- [4] Larmer L, Sargolzaei M, Ventura R, Schenkel F. Imputation accuracy from low to high density using within and across breed reference populations in Holstein, Guernsey and Ayrshire cattle. In: Technical report to the Dairy cattle breeding and genetics Committee on February 28, 2012. Guelph, ON, Canada: University of Guelph; 2012.
- [5] Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 2014;15:478.
- [6] Boichard D, Chung H, Dassonneville R, David X, Eggen A, Fritz S, et al. Design of a bovine low-density SNP array optimized for imputation. *PLoS One* 2012;7:e34130.
- [7] Sargolzaei M, Chesnais JP, Schenkel FS. Flmpu - an efficient imputation algorithm for dairy cattle populations. *J Dairy Sci* 2011; 94(E-Suppl. 1):421.
- [8] Ventura RV, Miller SP, Dodds KG, Auvray B, Lee M, Bixley M, et al. Assessing accuracy of imputation using different SNP panel densities in a multi-breed sheep population. *Genet Sel Evol* 2016;48:71.
- [9] Chud TCS, Ventura RV, Schenkel FS, Carvalheiro R, Buzanskas ME, Rosa JO, et al. Strategies for genotype imputation in composite beef cattle. *BMC Genet* 2015;16:99.
- [10] Zhang Z, Druet T. Marker imputation with low-density marker panels in Dutch Holstein cattle. *J Dairy Sci* 2010;93:5487–94.
- [11] ABQM. Associação Brasileira dos Criadores de Cavalos Quarto de Milha, 2015. Available at: <http://www.abqm.com.br/item1.htm>. Accessed 15 March 2016.
- [12] America's Horse Daily. All About the Racing American Quarter Horse, 2012. Available at: http://americashorsedaily.com/all-about-the-racing-american-quarter-horse/#.Urx8N_RDvik. Accessed 12 November 2013.
- [13] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81: 559–75.
- [14] R core Team. R: a language and environment for statistical computing. R foundation for Statistical Computing. <http://www.R-project.org/>.
- [15] Browning SR, Browning BL. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 2009;84:210–23.
- [16] Ma P, Brødum RF, Zhang Q, Lund MS, Su G. Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish Red Cattle. *J Dairy Sci* 2013;96:4666–77.
- [17] Ogawa S, Matsuda H, Taniguchi Y, Watanabe T, Sugimoto Y, Iwaisaki H. Estimation of variance and genomic prediction using genotypes imputed from low-density marker subsets for carcass traits in Japanese black cattle. *Anim Sci J* 2016;87:3–12.
- [18] Kong A, Masson G, Frigge ML, Gylfason A, Zusmanovich P, Thorleifsson G, et al. Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat Genet* 2008;40:1068–75.
- [19] Piccoli M, Braccini J, Cardoso F, Sargolzaei M, Schenkel F. Impact of imputation in Braford and Hereford beef cattle. *BMC Genet* 2014; 15:157.