https://doi.org/10.17221/120/2019-CJAS

Genotype imputation strategies for Portuguese Holstein cattle using different SNP panels

*Alessandra Alves Silva[1], Fabyano Fonseca Silva[1], Delvan Alves Silva[1], Hugo Teixeira Silva[1], Cláudio Napolis Costa[2], Paulo Sávio Lopes[1], Renata Veroneze[1], Gertrude Thompson[3,4], Julio Carvalheira[3,4]**

[1]*Department of Animal Science, Federal University of Viçosa, Viçosa, Brazil*

[2]*Embrapa Dairy Cattle, Juiz de Fora, Brazil*

[3]*Research Center in Biodiversity and Genetic Resources (CIBIO-InBio), University of Porto, Porto, Portugal*

[4]*Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, Porto, Portugal*

**Corresponding author: jgc3@cibio.up.pt*

Citation: Silva A.A., Silva F.F., Silva D.A., Silva H.T., Costa C.N., Lopes P.S., Veroneze R.V., Thompson G., Carvalheira J. (2019): Genotype imputation strategies for Portuguese Holstein cattle using different SNP panels. Czech J. Anim. Sci., 64, 377–386.

Abstract: Although several studies have investigated the factors affecting imputation accuracy, most of these studies involved a large number of genotyped animals. Thus, results from these studies cannot be directly applied to small populations, since the population structure affects imputation accuracy. In addition, factors affecting imputation accuracy may also be intensified in small populations. Therefore, we aimed to compare different imputation strategies for the Portuguese Holstein cattle population considering several commercially available single nucleotide polymorphism (SNP) panels in a relatively small number of genotyped animals. Data from 1359 genotyped animals were used to evaluate imputation in 7 different scenarios. In the S1 to S6 scenarios, imputations were performed from LDv1, 50Kv1, 57K, 77K, HDv3 and Ax58K panels to 50Kv2 panel. In these scenarios, the bulls in 50Kv2 were divided into reference (352) and validation (101) populations based on the year of birth. In the S7 scenario, the validation population consisted of 566 cows genotyped with the Ax58K panel with their genotypes masked to LDv1. In general, all sample imputation accuracies were high with correlations ranging from 0.94 to 0.99 and concordance rate ranging from 92.59 to 98.18%. SNP-specific accuracy was consistent with that of sample imputation. S4 (40.32% of SNPs imputed) had higher accuracy than S2 and S3, both with less than 7.59% of SNPs imputed. Most probably, this was due to the high number of imputed SNPs with minor allele frequency (MAF) < 0.05 in S2 and S3 (by 18.43% and 16.06% higher than in S4, respectively). Therefore, for these two scenarios, MAF was more relevant than the panel density. These results suggest that genotype imputation using several commercially available SNP panels is feasible for the Portuguese national genomic evaluation.

Keywords: dairy cattle; genomic evaluation; imputation accuracy

In practice, several studies in animal breeding with genomic data such as genomic prediction or genome-wide association studies (GWAS) use imputed data (Jattawa et al. 2016; Wang et al. 2016). Thus, imputation strategies have become an important approach to make efficient use of all available information. Moreover, there are several commercial single nucleotide polymorphism (SNP) panels with different densities in dairy cattle (Nicolazzi et al. 2015) providing many genotypic

Partially supported by CAPES/FCT (99999.008462/2014-03) and CNPq/INCT-CA.

https://doi.org/10.17221/120/2019-CJAS

datasets that are routinely shared between countries to minimize the costs of genomic information, reinforcing the need and importance for reliable imputation strategies.

Briefly, imputation refers to statistical and computational tools applied to infer SNP genotypes which are not obtained from a low-density panel using information from a reference population genotyped with a higher density panel (Ventura et al. 2014). Several studies have investigated the factors affecting imputation accuracy in dairy cattle population, such as the number of reference individuals, the relationship between reference population and target population, minor allele frequency (MAF), linkage disequilibrium (LD), and the difference between marker densities of the reference and imputed sets (Larmer et al. 2014; Boison et al. 2015; Garcia-Ruiz et al. 2015). However, most of these studies involved a large number of genotyped animals using low-density panels derived from only an HD panel. Thus, results from these studies cannot be directly applied to small populations, since the population structure affects imputation accuracy. In addition, these factors affecting imputation accuracy may also be intensified in small populations. Therefore, we aimed to test different imputation strategies for Portuguese Holstein population considering several commercially available SNP panels with different densities in a small number of genotyped animals.

MATERIAL AND METHODS

Data from 1359 genotyped animals (793 bulls and 566 cows) were used in this study. Of these, 50.85% were foreign bulls, mainly American and Canadian ones (82.63%), while 49.15% of the animals were Portuguese (83.83% cows and 16.17% bulls). In summary, the bulls were genotyped using different SNP panels: LDv1 (GeneSeek Genomic Profiler, Neogen Corp., Lincoln, USA), 50Kv1 and 50Kv2 (Bovine SNP50v.1 and Bovine SNP50v.2 BeadChips, Illumina, San Diego, USA), 57K (USDA Illumina, San Diego, USA), 77K and HDv3 (GeneSeek Genomic Profiler, Neogen Corp., Lincoln, USA). The cows were genotyped using the Ax58K panel (Affymetrix, Santa Clara, USA). The numbers of genotyped animals by panel and year of birth are shown in Table 1 and the number of SNPs for each panel is shown in Table 2.

Marker positions and chromosomes in the map for each panel were standardized according to the UMD v3.1 assembly (Zimin et al. 2009). Quality control (QC) analyses for SNP and samples were done separately for each panel using PLINK v1.07 (Purcell et al. 2007). In QC for samples, errors of sex disagreement were checked by heterozygosity on the X chromosome; animals with the call rate < 0.90 were discarded; parent–offspring pairs were tested for Mendelian inconsistencies; and deviations from heterozygosity were controlled by removing animals with ± 3 standard deviations. A total of

Table 1. Number of genotyped animals by panel according to birth year

Birth year	50Kv2	LDv1	50Kv1	57K	Ax58K[1]	77K	HDv3
1966–2003	352	–	23	76	–	31	18
2004	9	–	6	–	–	–	6
2005	3	–	22	–	–	–	7
2006	7	–	25	–	–	–	8
2007	2	–	20	–	–	–	7
2008	5	–	25	1	–	–	13
2009	6	1	15	2	–	–	6
2010	23	3	8	1	71	–	11
2011	35	–	–	–	96	–	1
2012	10	–	–	–	166	3	1
2013	–	–	–	–	193	–	–
2014	1	–	–	–	39	–	–
2015	–	–	–	–	1	–	–
Total	453	4	144	80	566	34	78

[1]females

https://doi.org/10.17221/120/2019-CJAS

Table 2. Description of evaluated imputation scenarios

Scenario	Description[1]	Animals *n* (validation/target)	SNPs *n* before QC	SNPs *n* after QC	SNPs *n* common to 50Kv2
S1	LDv1	101/4	8 610	7 306	6 304
S2	50Kv1	101/144	54 001	39 910	35 866
S3	57K	101/80	56 947	40 602	35 692
S4	77K	101/34	76 883	69 403	23 051
S5	HDv3	101/78	139 376	118 298	33 383
S6	Ax58K	101/566	57 497	42 123	32 476
S7	Ax58K-a[2]	566/566	57 497	42 123	5 437

SNPs = single nucleotide polymorphisms, QC = quality control analysis

[1]reference panel for all scenarios studied was the 50Kv2 (54 609 SNPs before QC) including 38 624 SNPs after QC analysis

[2]validation population consisted of 566 cows which had their genotypes (32 476 SNPs in common to 50Kv2) masked to the LDv1 panel

1530 animals were considered in QC, in which 11.2% were excluded, thus 1359 animals were effective for imputation analyses. In QC for SNPs, markers with the call rate < 0.95, minor allele frequency (MAF) < 0.02, and Hardy-Weinberg equilibrium with χ^2 lower than 10^{-6} were excluded. SNPs with positions unknown or located on sex chromosome were not considered in the analysis. The number of SNPs after QC for each panel is described in Table 2.

When using a diversity of genotype panels, mainly from different technologies (e.g., Illumina and Affymetrix), it is important to assess their consistency by performing a population structure analysis. A genotype data set contained 3239 SNPs that were shared between all panels that were selected and used to evaluate the population structure by principal components analysis. For this analysis, principal components were calculated from the genomic relationship matrix (**G**) obtained according to VanRaden (2008) as follows:

$$\mathbf{G} = (\mathbf{M} - 2\mathbf{P})\,(\mathbf{M} - 2\mathbf{P})'/2\Sigma p_i(1 - p_i)$$

where:

M = matrix of minor allele (with dimensions equal to the number of animals by the number of SNP markers)

p_i = frequency of allele A of the $i^{\text{-th}}$ SNP

P = matrix (with dimensions equal to the number of animals by the number of SNPs) with each row containing the p_i values

The **G** matrix and principal components were obtained using the PreGSF90 software (Misztal et al. 2014). The first and the second principal component calculated based on the **G** matrix are shown in Figure 1. The animals from different panels showed high connectivity, indicating high consistency between the studied panels.

The genotypes from different panels were imputed to 50Kv2 because this panel included the highest number of males. In addition, several studies have shown that increasing the SNP density above 50 000 markers (50K) added small gains in the reliability of genomic prediction for Holstein cattle (VanRaden et al. 2011, 2013). Imputations were performed using the FImpute 2.2 software (Sargolzaei et al. 2014). According to Boison et al. (2015) and Jattawa et al. (2016), this software presents high imputation accuracies and high computational performance in cattle populations. The FImpute option combining family and population-based algorithms was considered here.

Seven imputation scenarios based on panel densities were investigated. In the S1 to S6 scenarios, imputations were performed from LDv1, 50Kv1, 57K, 77K, HDv3 and Ax58K panels to 50Kv2 panel. In these scenarios, the bulls in 50Kv2 panel were divided into reference (n = 352) and validation (n = 101) populations based on their year of birth. The validation bulls had their 50K genotypes masked to each imputed panel (S1 to S6). The last scenario (S7) was defined to assess the quality of the imputation from Ax58K to 50Kv2 (Affymetrix and Illumina technologies, respectively), where the validation population consisted of 566 cows from Ax58K panel. Their 58K genotypes were masked to LDv1 panel and the reference population was given by 352 bulls from 50Kv2. The seven scenarios are shown in Table 2.

Imputation accuracies (per sample and SNP-specific) were assessed using Spearman's correlation coefficient (r) between the imputed

https://doi.org/10.17221/120/2019-CJAS

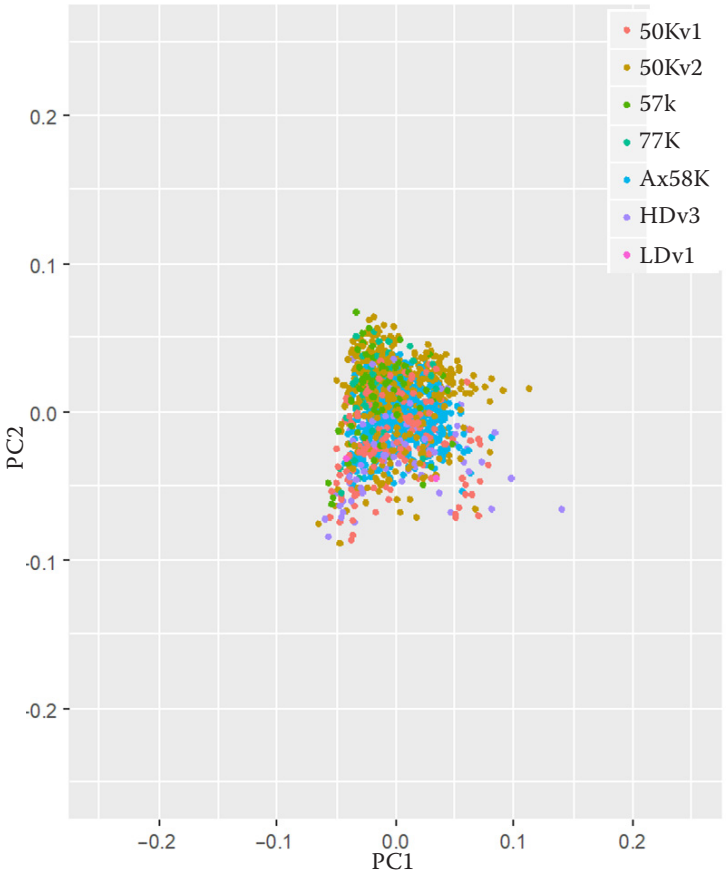

Figure 1. Plot of the first 2 principal components (PC) of the genomic relationship matrix between animals of each panel

and true SNP markers and the concordance rate (CR) as a proportion of correctly imputed SNPs in relation to all imputed SNPs.

The effect of having relatives in the reference population for imputation was also investigated. Based on the **G** matrix (3239 SNPs shared between all panels), we calculated the average of the top 10 relationships and the average of all relationships higher than zero between each imputed animal and those in the reference population. The relationship between relatedness and imputation accuracy was assessed by regressing r on the average of the top 10 relationships.

To evaluate the effect of MAF on imputation accuracy, the SNPs to be imputed were classified according to two levels of MAF obtained from the reference population: MAF < 0.1 and MAF < 0.05. Linkage disequilibrium (LD) between markers was measured using r^2 (Hill and Robertson 1968).

RESULTS AND DISCUSSION

Our study evaluated the accuracy of imputation for Portuguese Holstein cattle using several commercially available SNP panels with different densities and a relatively low number of genotyped animals. Imputation was performed using the FImpute software (Sargolzaei et al. 2014), and we evaluated sample imputation accuracy, effect of relatedness on sample imputation accuracy, SNP-specific imputation accuracy, effect of MAF on SNP-specific imputation, and linkage disequilibrium.

Sample imputation accuracy and effect of relatedness on accuracy. Concordance rate and squared Pearson correlation coefficient have been reported in several studies (Boison et al. 2015; Jattawa et al. 2016; Ventura et al. 2016) as a measure of imputation accuracy. Nevertheless, the Pearson correlation coefficient assumes that the two samples are normally distributed. If the assumption of normality is violated, the Pearson correlation coefficient may produce unreliable results. Differently, Spearman's rank correlation coefficient is a non-parametric measure of correlation, calculated from ranks and it depicts a monotonic relationship (Goktas and Isci 2011). Therefore, Spearman's correlation seems to be more appropriate to measure accuracy for data obtained from genotypes.

https://doi.org/10.17221/120/2019-CJAS

Table 3. Sample imputation accuracy for all scenarios evaluated. Spearman's correlation coefficient (*r*) between imputed and true SNPs markers and CR, as a proportion of correctly imputed SNPs versus true SNPs

Scenario	SNPs imputed *n*	*r*				CR			
		mean	min.	max.	SD	mean	min.	max.	SD
S1	32 320 (83.68%)	0.96	0.93	0.99	0.012	95.39	91.46	98.27	1.45
S2	2 758 (7.14%)	0.98	0.80	0.99	0.032	97.25	73.46	99.27	4.26
S3	2 932 (7.59%)	0.98	0.80	0.99	0.032	97.24	75.03	99.28	4.02
S4	15 573 (40.32%)	0.99	0.95	0.99	0.008	98.11	93.37	99.22	1.07
S5	5 241 (13.57%)	0.99	0.93	0.99	0.011	98.18	91.18	99.43	1.48
S6	6 148 (15.92%)	0.98	0.93	0.99	0.010	97.87	91.64	99.15	1.31
S7	27 039 (83.26%)	0.94	0.79	0.98	0.020	92.59	76.54	97.40	2.41

SNPs = single nucleotide polymorphisms, CR = concordance rate, min = minimum, max = maximum, SD = standard deviation

Table 3 shows the sample imputation accuracy (*r* and CR) for each evaluated scenario. In general, sample imputation accuracies were within the range of those reported for dairy cattle (Boison et al. 2015; Jattawa et al. 2016). The accuracy means were high for all scenarios, ranging from 0.94 to 0.99. The CR was also high ranging from 92.59 to 98.18%. The lowest values of accuracy were observed in S1 and S7 scenarios, which had a higher number of imputed SNPs (83.68 and 83.26%, respectively). Although the mean of imputation accuracy was similar over the scenarios, the standard deviation was higher for S2 and S3, which had the lower number of imputed SNPs (7.14 and 7.59%, respectively). On the other hand, the scenarios that had a higher number of imputed SNPs (> 13.57%) presented the lowest standard deviations. Although studies have shown that imputation accuracy increases according to a reduction in the SNP number to be imputed (Khatkar et al. 2012; Chud et al. 2015), this pattern was not clearly observed here, since the best results were observed for S4, in which a higher number of SNPs was imputed (40.32%). Several factors may have influenced these results, such as the size and structure of the reference population, level of the relationship between the animals to be imputed and the reference population size, the position of SNP on the chromosome and its frequency (MAF) in the population (Ventura et al. 2014).

The S7 scenario was performed to evaluate imputation accuracy between panels from different companies (Illumina and Affymetrix), in which 27 039 SNPs were imputed. Accuracies were high (mean of 0.94 for *r* and of 92.59% for CR) and similar to those found in the other scenarios that consisted of panels from Illumina only (Table 3). Similar results were observed by Zhou et al. (2014) and Berry et al. (2016). In addition, lower-density chips have become widely used for genotyping of females as an appropriate strategy for genomic selection. Therefore it is essential to evaluate the imputation accuracy for this group. The validation in females (S7 scenario) also indicates that females may be imputed using males in the reference population. Similar results were reported by Chud et al. (2015), when these authors showed that the imputation of female genotypes could be carried out using only males in the reference population. In addition, they referred to the highest mean genomic relatedness between reference and target populations observed in this scenario, which may have contributed to imputation accuracy.

Studies have shown that imputation accuracy is strongly associated with the level of the relationship between the animals from imputed and reference data set (Carvalheiro et al. 2014; Boison et al. 2015; Ventura et al. 2016). In general, imputation accuracy is measured only as a function of the density of SNPs to be imputed using a validation population composed of animals from the reference panel, in which their SNPs were masked to the target SNP panel. Nevertheless, the validation population may not represent the real population to be imputed. Therefore, to assist the comparison of imputation accuracies in the different scenarios, a summary of the genomic relationship between the reference and validation population has been exploited (Daetwyler et al. 2013; Boison et al. 2015).

In our study the average relatedness between validation and reference individuals as well as between target and reference individuals is shown in Table 4. The higher genomic relatedness with reference individuals was observed in S3 and S4 scenarios, in

https://doi.org/10.17221/120/2019-CJAS

Table 4. Genomic relationships between animals in the imputed and reference data set (values are means + standard deviations)

Panel (Imput/Ref)	Genomic relationships[1]	
	top 10	relationships > 0
Val/Ref	0.11 (0.05)	0.031 (0.008)
S1/Ref	0.09 (0.01)	0.028 (0.003)
S2/Ref	0.13 (0.04)	0.034 (0.006)
S3/Ref	0.18 (0.06)	0.046 (0.010)
S4/Ref	0.19 (0.05)	0.049 (0.012)
S5/Ref	0.14 (0.06)	0.035 (0.009)
S6/Ref	0.10 (0.03)	0.031 (0.005)

Imput = scenarios imputed, Val = validation population, Ref = reference population

[1]top 10 = average genomic relationships of top 10 and average relationships > 0 between imputed individuals and reference individuals

which 58.75 and 73.53% of individuals had at least one genotyped half-sib in the reference population, respectively. The remaining scenarios including validation had less than 15.38% of individuals with at least one genotyped half-sib in the reference population. The mean of the top 10 genomic relationships between animals of imputed scenarios and reference population was similar (ranging from 0.09 to 0.19) to the mean of genomic relationships between animals of validation and reference populations (Table 4). This result indicated that the animals selected to compose the validation population are representative of all evaluated scenarios, thus the imputation accuracies can be compared.

In addition, the impact of relatedness between validation and reference animals on imputation accuracy is shown in Figure 2. The greatest influence of relatedness with the reference population on the r was observed for S1 ($P < 0.01$), in which a higher number of SNPs was imputed. On the other hand,

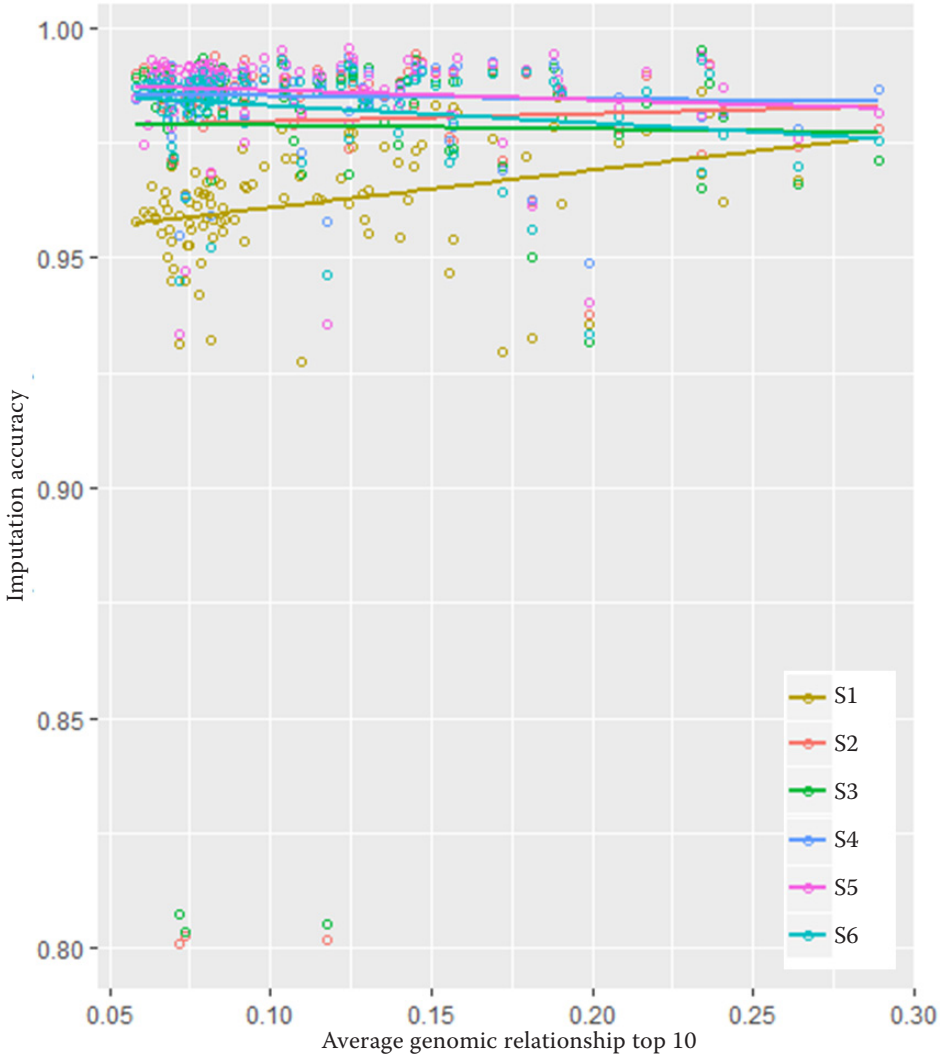

Figure 2. Imputation accuracy (r) as a function of genomic relatedness (top 10)

https://doi.org/10.17221/120/2019-CJAS

no such a behaviour was observed ($P > 0.05$) in the remaining scenarios (S2, S3, S4, S5, and S6), in which a lower number of SNPs was imputed. These results indicate that relatedness has a greater influence on imputation accuracy when higher SNP densities are imputed. Similar results were also found by Carvalheiro et al. (2014) and Chud et al. (2015).

SNP-specific imputation accuracy and effect of MAF. In order to graphically display the results of imputation accuracy by chromosomes, we made a circular plot using the *circlize* package (Gu et al. 2014) of R software (R Core Team, 2018). The average accuracy (*r* and CR) for SNPs by chromosomes is shown in Figure 3, Supplementary Table S1 and Table S2 in Supplementary Online Material (SOM). In general, the *r* and CR accuracies among chromosomes ranged from 0.88 to 0.93 (93.31 to 96.41%) for S1; 0.88 to 0.95 (95.78 to 98.15%) for S2; 0.91 to 0.96 (95.61 to 98.35%) for S3; 0.94 to 0.97 (97.57 to 98.66%) for S4; 0.92 to 0.97 (97.47 to 98.85%) for S5; 0.93 to 0.97 (97.26 to 98.72%) for S6; and from 0.84 to 0.89 (90.79 to 94.26%) for S7. As expected, the S1 and S7 scenarios presented the lowest accuracy due to the higher number of imputed SNPs (83.68% and 83.26%, respectively). On the other hand, although with minimal differences, S4 containing 40.32% of imputed SNPs presented higher accuracy when compared to S2 or S3 scenarios, in which only 7.14% and 7.59% of SNPs were imputed, respectively.

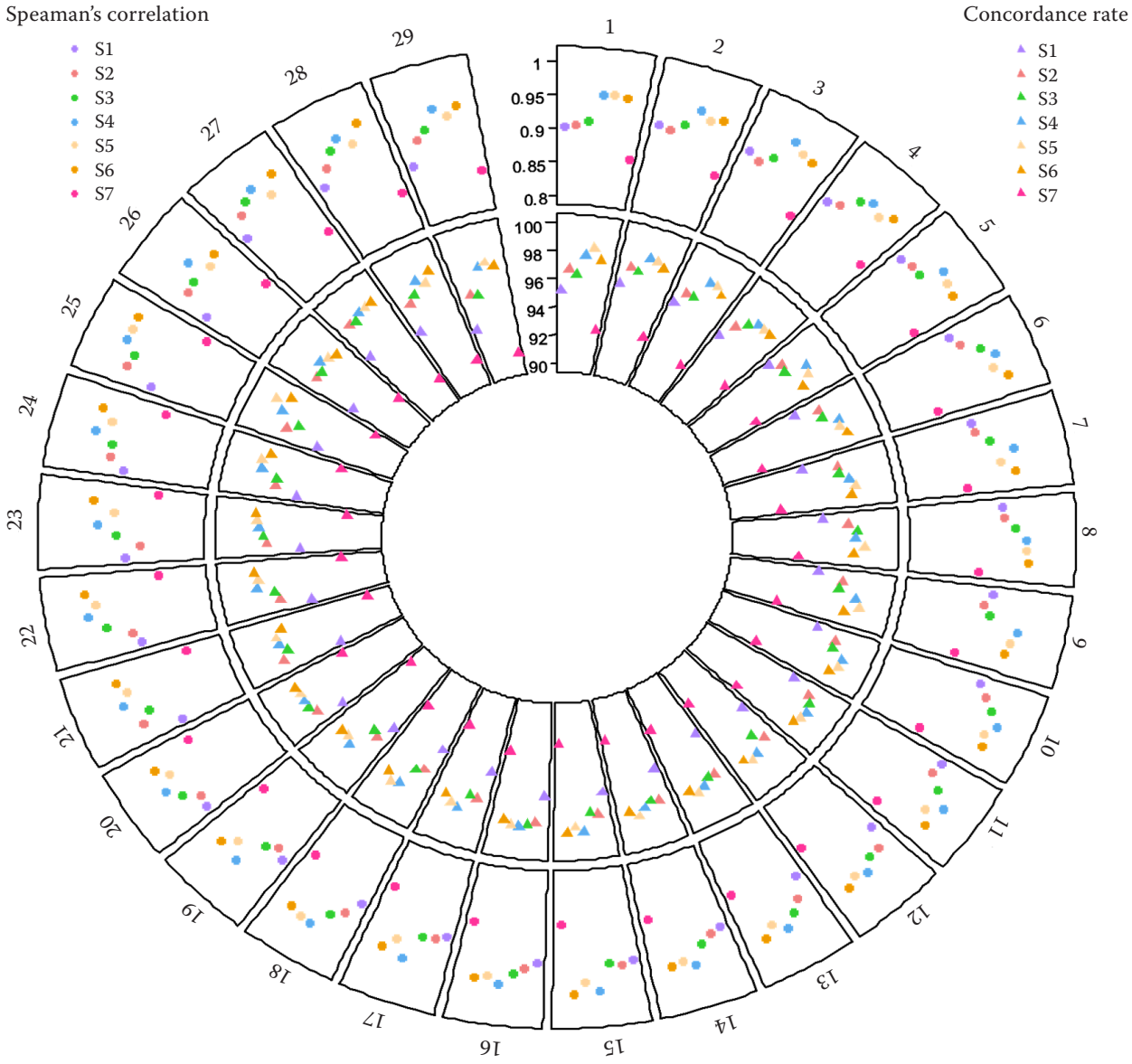

Figure 3. SNP-specific imputation accuracy for all scenarios evaluated

https://doi.org/10.17221/120/2019-CJAS

Table 5. Number of imputed single nucleotid polymorphisms (SNPs) and number of SNPs at different levels of minor allele frequency (MAF) evaluated

Scenario	Imputed SNPs *n*	MAF	
		< 0.1	< 0.05
S1	32 320	5 743 (17.77%)	2 094 (6.48%)
S2	2 758	1 176 (42.64%)	769 (27.88%)
S3	2 932	1 065 (36.32%)	748 (25.51%)
S4	15 573	3 453 (22.17%)	1 471 (9.45%)
S5	5 241	1 790 (34.15%)	982 (18.74%)
S6	6 148	1 158 (18.84%)	638 (10.38%)

Most probably this occurred due to the imputation of SNPs with low frequency in the population. To clarify this point, we calculated the average MAF (< 0.1 and < 0.05) in the reference population for each set of SNPs imputed in each scenario and we observed that the number of imputed SNPs with MAF lower than 0.05 was 18.43% and 16.06% higher in S2 and S3 compared to S4, respectively (Table 5). Therefore, for S2 and S3, MAF was more relevant to imputation accuracy than panel density. This is consistent with a previous study of Ventura et al. (2016), in which the authors investigated imputation accuracy for rare alleles according to the MAF level (ranging from 0 to 0.05) and observed low accuracy for rare alleles (up to 57.8%). In addition, Boison et al. (2015) showed that the Illumina 50Kv2 panel presented a higher proportion of markers with low MAF compared to the panels from GeneSeek. Probably, a strategy to reduce the effect of MAF on imputation accuracy would be to increase the size of the reference population (Heidaritabar et al. 2015).

In agreement with Carvalheiro et al. (2014), the *r* values were higher than the corresponding CR values for sample imputation accuracy because the penalty for one incorrectly imputed allele is relatively higher for CR than for *r* (Table 3). On the other hand, the opposite behaviour was observed in SNP-specific imputation accuracy, in which the *r* values were lower than the corresponding values of CR (Figure 3, Supplementary Table S1 and Table S2 in SOM). There are no reports comparing both criteria in SNP-specific imputation accuracy, but most probably this occurred due to the MAF effect. As reported by Hickey et al. (2012), a marker with very low MAF increases the probability of being homozygous for the common allele, which in general produces an increase in CR. The opposite behaviour is expected for *r*, where the imputation accuracy is lower for markers with low MAF.

Linkage disequilibrium. The mean of LD estimated for 29 chromosomes considering the animals of the reference population is shown in Figure 4. It was observed that 52.82% of the SNPs had low levels of LD (< 0.10). The mean LD for the population was 0.107. The highest mean LD was observed on chromosome 14 (r^2 = 0.135), while chromosome 27 had the lowest mean LD (r^2 = 0.083). These results are in agreement with those reported by Salem et al. (2018), when the authors studied the level of LD in Portuguese Holstein cattle.

In summary, we evaluated the imputation accuracy for Portuguese Holstein cattle using several commercially available SNP panels with different densities in a relatively small number of genotyped animals. Sample imputation accuracy was higher than 0.93 for Spearman's correlation and higher than 92% for concordance rate. In addition, the relatedness has a large influence on accuracy when higher SNPs densities are imputed. SNP-specific imputation accuracy was higher than 0.86 for Spearman's correlation and higher than 92% for

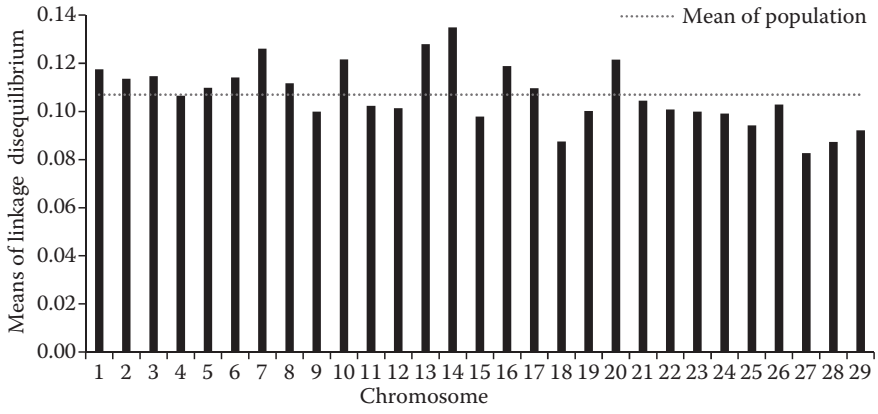

Figure 4. Means of linkage disequilibrium (r^2) between adjacent SNPs markers separated by at most 1 Mb within each chromosome

https://doi.org/10.17221/120/2019-CJAS

concordance rate. Moreover, MAF was more relevant to accuracy than the panel density, probably due to the small number of animals in the reference population used in this study.

CONCLUSION

Our results suggest that genotype imputation for Portuguese Holstein cattle using several commercially available SNP panels with different densities in a relatively small number of genotyped animals is feasible and may be advantageous to the national genomic evaluations of dairy cattle.

Acknowledgement. The authors acknowledge Portuguese Dairy Cattle Breeders Association (ANABLE) and Embrapa Dairy Cattle for providing the used data.

REFERENCES

Berry D.P., O'Brien A., Wall E., McDermott K., Randles S., Flynn P., Park S., Grose J., Weld R., McHugh N. (2016): Inter- and intra-reproducibility of genotypes from sheep technical replicates on Illumina and Affymetrix platforms. Genetics Selection Evolution, 48, 1–5.

Boison S.A., Santos D.J.A., Utsunomiya A.H.T., Carvalheiro R., Neves H.H.R., O'Brien A.M.P., Garcia J.F., Solkner J., da Silva M.V.G.B. (2015): Strategies for single nucleotide polymorphism (SNP) genotyping to enhance genotype imputation in Gyr (Bos indicus) dairy cattle: Comparison of commercially available SNP chips. Journal of Dairy Science, 98, 4969–4989.

Carvalheiro R., Boison S.A., Neves H.H.R., Sargolzaei M., Schenkel F.S., Utsunomiya Y.T., O'Brien A.M.P., Solkner J., McEwan J.C., Van Tassell C.P., Sonstegard T.S., Garcia J.F. (2014): Accuracy of genotype imputation in Nelore cattle. Genetics Selection Evolution, 46, 1–11.

Chud T.C.S., Ventura R.V., Schenkel F.S., Carvalheiro R., Buzanskas M.E., Rosa J.O., de Alvarenga Mudadu M., da Silva M.V.G.B., Mokry F.B., Marcondes C.R., Regitano L.C.A., Munari D.P. (2015): Strategies for genotype imputation in composite beef cattle. BMC Genetics, 16, 1–10.

Daetwyler H.D., Calus M.P.L., Pong-Wong R., de los Campos G., Hickey J.M. (2013): Genomic prediction in animals and plants: Simulation of data, validation, reporting, and benchmarking. Genetics, 193, 347–365.

Garcia-Ruiz A., Ruiz-Lopez F.J., Wiggans G.R., Van Tassell C.P., Montaldo H.H. (2015): Effect of reference population size and available ancestor genotypes on imputation of Mexican Holstein genotypes. Journal of Dairy Science, 98, 3478–3484.

Goktas A., Isci O. (2011): A comparison and normality test of some measures of association via simulation for rectangular doubly ordered cross tables. Metodoloski Zvezki, 8, 17–37.

Gu Z., Gu L., Eils R., Schlesner M., Brors B. (2014): circlize implements and enhances circular visualization in R. Bioinformatics, 30, 2811–2812.

Heidaritabar M., Calus M.P.L., Vereijken A., Groenen M.A.M., Bastiaansen J.W.M. (2015): Accuracy of imputation using the most common sires as reference population in layer chickens. BMC Genetics, 16, 1–14.

Hickey J.M., Crossa J., Babu R., de los Campos G. (2012): Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. Crop Science, 52, 654–663.

Hill W.G., Robertson A. (1968): Linkage disequilibrium in finite populations. TAG Theoretical and Applied Genetics, 38, 226–231.

Jattawa D., Elzo M.A., Koonawootrittriron S., Suwanasopee T. (2016): Imputation accuracy from low to moderate density single nucleotide polymorphism chips in a Thai multibreed dairy cattle population. Asian-Australasian Journal of Animal Sciences, 29, 464–470.

Khatkar M.S., Moser G., Hayes B.J., Raadsma H.W. (2012): Strategies and utility of imputed SNP genotypes for genomic analysis in dairy cattle. BMC Genomics, 13, Article No. 538.

Larmer S.G., Sargolzaei M., Schenkel F.S. (2014): Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across Canadian dairy breeds. Journal of Dairy Science, 97, 3128–3141.

Misztal I., Tsuruta S., Lourenco D.A.L., Aguilar I., Legarra A., Vitezica Z.G. (2014): Manual for BLUPF90 Family of Programs. University of Georgia, Athens, USA. Available at http://nce.ads.uga.edu/wiki/doku.php?id=application_programs (accessed Feb 15, 2018).

Nicolazzi E.L., Biffani S., Biscarini F., Orozco P., Wengel T., Caprera A., Nazzicari N., Stella A. (2015): Software solutions for the livestock genomics SNP array revolution. Animal Genetics, 46, 343–353.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J., Sham P.C. (2007): PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics, 81, 559–575.

R Core Team (2018): R: A Language and Environment for Statistical Computing. R Foundation for Statistical

https://doi.org/10.17221/120/2019-CJAS

Computing, Vienna, Austria. Available at https://www.R-project.org/ (accessed May 10, 2018).

Salem M.M.I., Thompson G., Chen S., Beja-Pereira A., Carvalheira J. (2018): Linkage disequilibrium and haplotype block structure in Portuguese Holstein cattle. Czech Journal of Animal Science, 63, 61–69.

Sargolzaei M., Chesnais J.P., Schenkel F.S. (2014): A new approach for efficient genotype imputation using information from relatives. BMC Genomics, 15, Article No. 478.

VanRaden P.M. (2008): Efficient methods to compute genomic predictions. Journal of Dairy Science, 91, 4414–4423.

VanRaden P.M., O'Connell J.R., Wiggans G.R., Weigel K.A. (2011): Genomic evaluations with many more genotypes. Genetics Selection Evolution, 43, 1–11.

VanRaden P.M., Null D.J., Sargolzaei M., Wiggans G.R., Tooker M.E., Cole J.B., Sonstegard T.S., Connor E.E., Winters M., van Kaam J.B.C.H.M., Valentini A., Van Doormaal B.J., Faust M.A., Doak G.A. (2013): Genomic imputation and evaluation using high-density Holstein genotypes. Journal of Dairy Science, 96, 668–678.

Ventura R.V., Lu D., Schenkel F.S., Wang Z., Li C., Miller S.P. (2014): Impact of reference population on accuracy of imputation from 6K to 50K single nucleotide polymorphism chips in purebred and crossbreed beef cattle. Journal of Animal Science, 92, 1433–1444.

Ventura R.V., Miller S.P., Dodds K.G., Auvray B., Lee M., Bixley M., Clarke S.M., McEwan J.C. (2016): Assessing accuracy of imputation using different SNP panel densities in a multi-breed sheep population. Genetics Selection Evolution, 48, 1–20.

Wang Y., Lin G., Li C., Stothard P. (2016): Genotype imputation methods and their effects on genomic predictions in cattle. Springer Science Reviews, 4, 79–98.

Zhou L., Heringstad B., Su G., Guldbrandtsen B., Meuwissen T.H.E., Svendsen M., Grove H., Nielsen U.S., Lund M.S. (2014): Genomic predictions based on a joint reference population for the Nordic Red cattle breeds. Journal of Dairy Science, 97, 4485–4496.

Zimin A.V., Delcher A.L., Florea L., Kelley D.R., Schatz M.C., Puiu D., Hanrahan F., Pertea G., Van Tassell C.P., Sonstegard T.S., Marcais G., Roberts M., Subramanian P., Yorke J.A., Salzberg S.L. (2009): A whole-genome assembly of the domestic cow, Bos taurus. Genome Biology, 10, Article R42.

Received: 2019–05–22
Accepted: 2019–08–16