Canadian Science Publishing

ARTICLE

# Assessing the accuracy of imputation in the Gyr breed using different SNP panels

Alejandra Maria Toro Ospina, Ignacio Aguilar, Matheus Henrique Vargas de Oliveira, Luiz Eduardo Cruz dos Santos Correia, Anibal Eugênio Vercesi Filho, Lucia Galvão Albuquerque, and Josineudson Augusto II de Vasconcelos Silva



**Abstract:** The aim of this study was to evaluate the accuracy of imputation in a Gyr population using two medium-density panels (*Bos taurus* - *Bos indicus*) and to test whether the inclusion of the Nellore breed increases the imputation accuracy in the Gyr population. The database consisted of 289 Gyr females from Brazil genotyped with the GGP Bovine LDv4 chip containing 30 000 SNPs and 158 Gyr females from Colombia genotyped with the GGP indicus chip containing 35 000 SNPs. A customized chip was created that contained the information of 9109 SNPs (9K) to test the imputation accuracy in Gyr populations; 604 Nellore animals with information of LD SNPs tested in the scenarios were included in the reference population. Four scenarios were tested: LD9K_30KGIR, LD9K_35INDGIR, LD9K_30KGIR_NEL, and LD9K_35INDGIR_NEL. Principal component analysis (PCA) was computed for the genomic matrix and sample-specific imputation accuracies were calculated using Pearson's correlation (CS) and the concordance rate (CR) for imputed genotypes. The results of PCA of the Colombian and Brazilian Gyr populations demonstrated the genomic relationship between the two populations. The CS and CR ranged from 0.88 to 0.94 and from 0.93 to 0.96, respectively. Among the scenarios tested, the highest CS (0.94) was observed for the LD9K_30KGIR scenario. The present results highlight the importance of the choice of chip for imputation in the Gyr breed. However, the variation in SNPs may reduce the imputation accuracy even when the chip of the *Bos indicus* subspecies is used.

*Key words:* imputation accuracy, genomic analysis, tropical breeds.

**Résumé :** Le but de cette étude était d'évaluer la justesse de l'imputation au sein d'une population de bovins de race Gir à partir de deux puces de génotypage de densité moyenne (*Bos taurus* – *Bos indicus*) et de déterminer si l'inclusion de la race Nellore accroissait la précision de l'imputation chez la population de Girs. La base de données comprenait 289 femelles du Brésil génotypées avec la puce GGP Bovine LDv4, laquelle compte 30 000 SNP, et 158 femelles de Colombie génotypées avec la puce GGP indicus, laquelle compte 35 000 SNP. Une puce sur mesure totalisant 9109 SNP (9K) a été produite pour mesurer la justesse de l'imputation chez les populations de Girs. De plus, 604 bovins de race Nellore fournissant de l'information sur le LD des SNP testés ont été inclus dans la population de référence. Quatre scénarios ont été testés : LD9K_30KGIR, LD9K_35INDGIR, LD9K_30KGIR_NEL et LD9K_35INDGIR_NEL. Une analyse en composantes principales (PCA) a été réalisée à l'aide de la matrice génomique, la justesse de l'imputation pour chaque échantillon a été calculée à l'aide d'une corrélation de Pearson (CS) et le taux de concordance (CR) a été calculé pour les génotypes imputés. Les résultats de l'analyse PCA chez les populations colombiennes et brésiliennes ont démontré la relation génomique entre les deux populations. Les valeurs de CS et de CR variaient respectivement entre 0,88 et 0,94, ainsi qu'entre 0,93 et 0,96. Parmi les scénarios testés, la plus forte valeur de CS (0,94) a été obtenue dans le scénario LD9K_30KGIR. Les résultats obtenus montrent l'importance du choix de la puce en vue de

Received 22 May 2020. Accepted 17 April 2021.

**A.M. Toro Ospina, M.H. Vargas de Oliveira, L.E. Cruz dos Santos Correia, and L.G. Albuquerque.** Faculdade de Ciências Agrárias e Veterinárias - Unesp, CEP 14.884-900, Jaboticabal, São Paulo, Brasil.
**I. Aguilar.** Instituto Nacional de Investigación Agropecuaria, INIA, Montevideo, Uruguay.
**A.E. Vercesi Filho.** Instituto de Zootecnia, CEP 13.460-000, Nova Odessa, São Paulo, Brasil.
**J.A. de Vasconcelos Silva.** Faculdade de Medicina Veterinária e Zootecnia - Unesp, CEP 18.618-307 - Botucatu, São Paulo, Brasil.

**Corresponding author:** Alejandra Maria Toro Ospina (email: toroospina92@gmail.com).

l'imputation chez les bovins de race Gir. Cependant, la variation dans les SNP peut réduire la justesse de l'imputation même lorsque la puce de la sous-espèces *Bos indicus* est employée. [Traduit par la Rédaction]

*Mots-clés :* justesse de l'imputation, analyse génomique, races tropicales.



## Introduction

Zebu breeds (*Bos indicus*) are widely used in tropical and subtropical countries such as Brazil and Colombia for meat and milk production. Gyr cattle are a Zebu breed used for dairy production in tropical climates (Toro Ospina et al. 2020). The implementation of genomic selection after 2000 has led to the development of different single nucleotide polymorphism (SNP) panels for genotyping at lower costs. Most commercial bovine chips were developed based on *B. taurus* data, a fact that resulted in a considerable percentage of monomorphic SNPs in *B. indicus* populations (Utsunomiya et al. 2019). Although Zebu cattle play an important role in beef and dairy production in Latin America, their success depends on the availability of adequately designed SNP arrays for *B. indicus* animals (Montaldo et al. 2012). Research in genomics requires the improvement of strategies for the use of data in genetic breeding programs. One such tool is the imputation of genotypes, which is performed based on the linkage disequilibrium of a population and on family relationships (Sargolzaei et al. 2014; Boison et al. 2015). The reliability of imputation is very important so that the genotypes can be used in subsequent analyses. For this purpose, accuracy tests are performed using the correlation between true and imputed SNPs, concordance rate (CR), and allelic imputation error. Studies investigating imputation strategies using different taurine breeds (Boichard et al. 2012; Oliveira Júnior et al. 2017) and Zebu breeds such as Nellore (Chud et al. 2015) are available in the literature.

Carvalheiro et al. (2014) highlighted the importance of evaluating the accuracy of imputation to define the best imputation method and the most appropriate panel for genomic applications to the Nellore breed. However, the remaining Zebu breeds (Gyr, Brahman, and Guzerá) are also of economic importance for tropical and subtropical countries, and genomic studies on imputation strategies in these Zebu breeds are necessary to improve the genetic estimates of breeds for which less information is available (Boison et al. 2015).

By February 2017, only 4.38 million SNPs out of a total of 99.71 million had been deposited for Zebu breeds in the dbSNP database, i.e., more than 96% of published SNPs are from taurine breeds (Iqbal et al. 2019). Genotyping companies recently introduced low- and medium-density markers for Indica breeds, but these markers still suffer from verification bias (Boison et al. 2015; Utsunomiya et al. 2019). Therefore, the aim of this study was to evaluate the accuracy of imputation in a Gyr population using two medium-density panels (*Bos taurus* - *Bos indicus*) and to test whether the inclusion of the Nellore breed increases the imputation accuracy in the Gyr population.

## Materials and methods

The Brazilian dairy Gyr population used in this study comprises five farms in the states of Minas Gerais and São Paulo and is part of the population evaluated by the Program of the Brazilian Association of Dairy Gyr Breeders (ABCGIL in the Portuguese acronym). The data of the Colombian dairy Gyr population were provided by the Colombian Association of Zebu Cattle Breeders (ASOCEBU in the Spanish acronym). The data of the Nellore population were provided by the Nelore Qualitas breeding program, Brazil.

The database consisted of the genotypes of 289 Brazilian Gyr females genotyped with the GGP Bovine LDv4 chip (Illumina, San Diego, CA) (LD 30K) that contains 30 000 SNPs. The 158 Colombian Gyr females were genotyped with the GGP indicus chip (LD 35IND) (Illumina, San Diego, CA) that contains 35 000 SNPs.

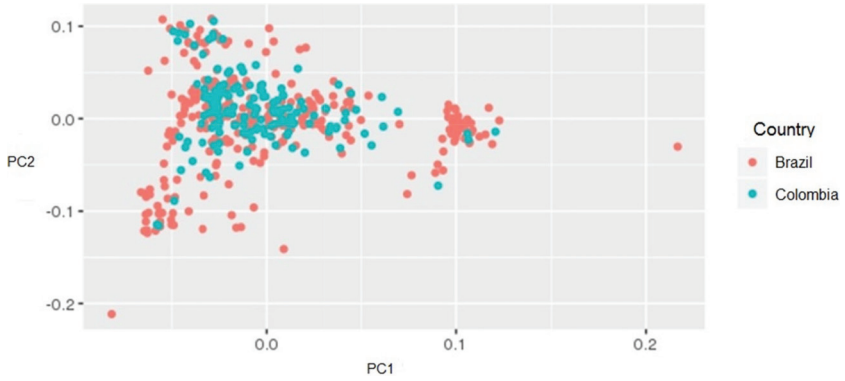
Quality control of the genotypes was performed for autosomal SNPs using the preGSf90 program (Aguilar et al. 2010). The filters applied to each scenario included the elimination of SNPs with a call rate $<0.90$, a minor allele frequency (MAF) $<0.02$, and a maximum difference between observed and expected frequency (Hardy-Weinberg equilibrium) $>0.15\%$. After quality control, 28 306 SNPs of the LD30K chip and 33 209 SNPs of the LD35IND chip remained for analysis. Different MAF intervals were tested (Purcell et al. 2007) for the customized 9K chip in the Gyr and Nellore populations.

Principal component analysis (PCA) was performed to describe the stratification of the Colombian and Brazilian Gyr populations, using the lower density 9109 SNP chip (LD 9K). PCA was computed for the genomic matrix using the *plotpca* function of the preGSf90 program (Aguilar et al. 2010).

Four imputation scenarios were investigated using different reference populations (RPop): (1) LD9K_30KGIR with RPop containing 306 Gyr animals; (2) LD9K_35INDGIR with RPop containing 176 Gyr animals; (3) LD9K_30KGIR_NEL with RPop containing 909 Gyr and Nellore animals; (4) LD9K_35INDGIR_NEL with RPop containing 780 Gyr and Nellore animals. Genotype imputation was performed with the FImpute program (Sargolzaei et al. 2014).

Four imputations per scenario were performed. Each imputation considered only ¼ of RPop, randomly choosing animals from the total RPop of the scenario and was repeated 30 times to ensure that all animals were included in the repetitions, totaling 120 imputations at the end of each scenario.

**Fig. 1.** Principal component analysis of the genomic relationship between the Colombian and Brazilian Gyr breeds, using the common SNPs (9K) of the GGP indicus and GGP Bovine LDv4 chip.






The imputation accuracy was obtained using custom programs in Linux and Python. The figures were constructed with the Biostatistical RStudio package (http://www.R-project.org). The imputation accuracies were evaluated based on Pearson's correlation (CS) between true and imputed SNPs and on CR represented by the proportion of correctly imputed genotypes. The CS was calculated between observed and imputed genotypes. The CR was calculated as the total number of markers of observed (true) genotype (0, 1, 2) for SNP of individual and the imputed genotype (0, 1, 2) for SNP of individual (Boison et al. 2015), and the allelic error rate is the percentage of incorrectly predicted alleles, i.e., mean allelic error rate (%) = number of incorrectly predicted alleles/total number of alleles imputed.

## Results and discussion

For a better understanding, we describe the assembly of the GGP Bovine LDv4 and GGP indicus chips and tested in the present study. The GGP Bovine LDv4 chip (LD30K) (Illumina, San Diego, CA) with 30 000 SNPs contains informative SNPs of the Illumina Bovine SNP50 and Bovine HD BeadChips (Illumina, San Diego, CA). The average SNP spacing of the Bovine LDv4 chip is 383 kb, with a higher concentration in the telomeric regions of the chromosome. Most SNPs were chosen specifically because of their high MAF and uniform coverage of the genome of most *B. taurus* breeds and several *B. indicus* breeds (Nellore, Brahman, Gyr, Girolando). Analysis using data of the Gyr population genotyped with LD30K showed a mean MAF of 0.3114.

The GGP indicus chip (LD35IND) (Illumina, San Diego, CA) with 35 000 SNPs mainly contains information of Brahman (0.15), Guzerá (0.15), Gyr (0.15), Nellore (0.35), Droughtmaster (0.05), and Santa Gertrudis (0.05) animals, as well as of tropical composite breeds (0.05) and other tropical indicus crosses (0.05). The MAF of the GGP indicus chip ranged from 0.1566 to 0.4318 in the breeds (Ferraz et al. 2018). Analysis using data of the Gyr population genotyped with GGP35IND showed a mean MAF of 0.2644.

In the database of the Gyr population, 18 important breeding animals of the populations were genotyped with the BovineHD BeadChip (HD) (Illumina, San Diego, CA). As part of the objective of the present study, we included a population of 604 Nellore animals genotyped with the HD chip to test the accuracy of imputation. We decided to include the animals in the imputation tests in which SNPs from the HD chip were extracted to the formats of the LD30K and LD35IND chips. Since we have two Gyr populations genotyped with different chips, our aim was to determine which chip is more accurate and to combine the information of the populations in the same density. Genotype imputation has been used to combine datasets of different breeds and (or) chips of different densities (Howie et al. 2011; Khatkar et al. 2012; Larmer et al. 2014).

By combining data from chips with different densities, we could identify how many SNPs are shared between the LD30K and LD35IND chips. There were 9109 shared SNPs, indicating the loss of a considerable amount of information of the animals. To standardize the Gyr database in a single density, we tested the imputation accuracy of the LD30K and LD35IND chips. Some studies on cattle have performed imputation tests using shared SNPs and customized chips to combine data and imputation methods (Druet et al. 2010; Ma et al. 2013; Ghoreishifar et al. 2018).

Using the 9109 shared SNPs, we created a customized chip, called 9K, for the tests. In the population genotyped with the HD chip, the SNPs were removed to obtain the format of the LD30K, LD35IND, and 9K chips for the different tests.

Figure 1 shows the first two principal components, PC1 (3.45%) and PC2 (2.53%), estimated based on the genomic relationship. The two principal components explained about 6% of the total variation. Principal component analysis of the Brazilian Gyr and Colombian Gyr populations revealed a relationship between animals and small genomic differences. The two Gyr populations

**Fig. 2.** Distribution of minor allele frequency (MAF) in the 9K chip in the Gyr and Nellore populations.

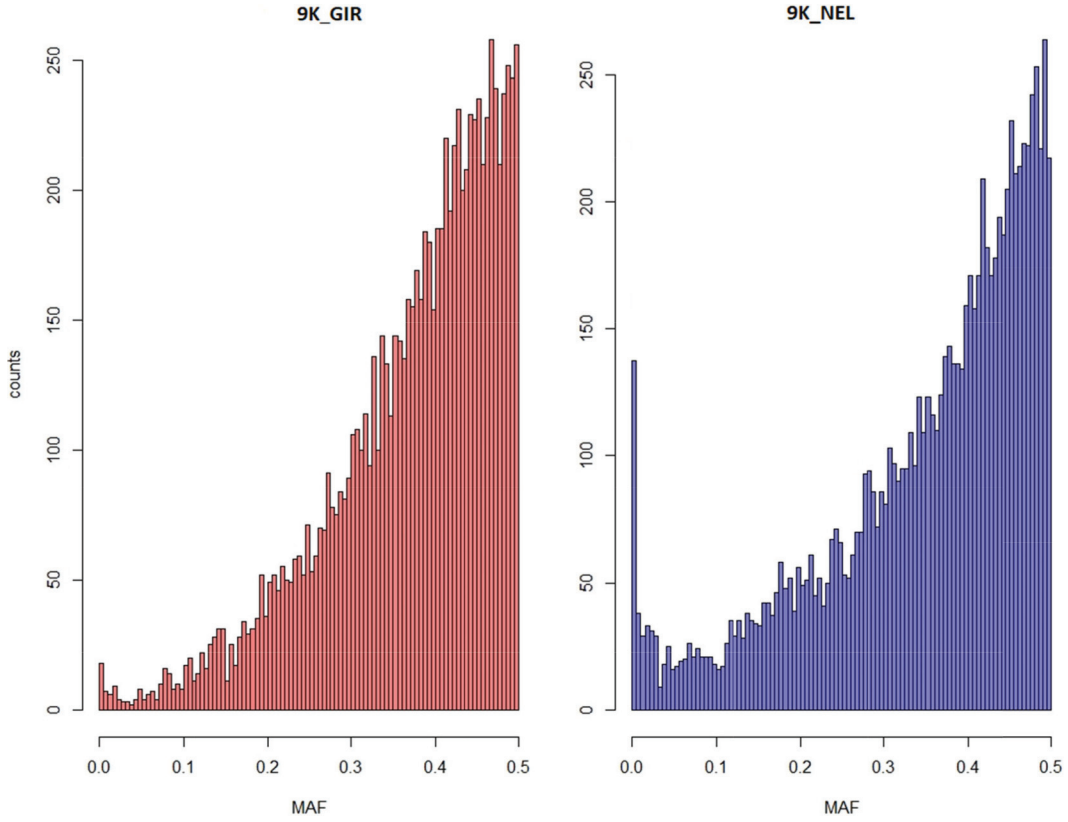



**Table 1.** Information of the genotyping chips and imputation accuracy (SD).

| | | Imputation accuracy | | |
|---|---|---|---|---|
| Scenario | Ref/imput | CS | CR | AE |
| LD9K_30KGIR* | 306/272 | 0.9488 (0.0498) | 0.9646 (0.0337) | 2.01% (1.9120) |
| LD9K_35IND_GIR* | 176/158 | 0.8771 (0.1356) | 0.9315 (0.0768) | 3.58% (2.4725) |
| LD9K_30KGIR_NEL* | 909/272 | 0.9341 (0.0754) | 0.9548 (0.0497) | 2.15% (1.8810) |
| LD9K_35IND_GIR_NEL* | 780/158 | 0.8917 (0.1184) | 0.9316 (0.0669) | 3.32% (2.6133) |

**Note:** CS, correlation; CR, concordance rate; AE, allelic error.
*Scenarios LD9K_30KGIR and LD9K_35IND_GIR show a significant difference ($p$-value $<$ 2.2e-16). When the scenarios were from the same chip, there was no significant difference ($p$-value $<$ 0.003452).

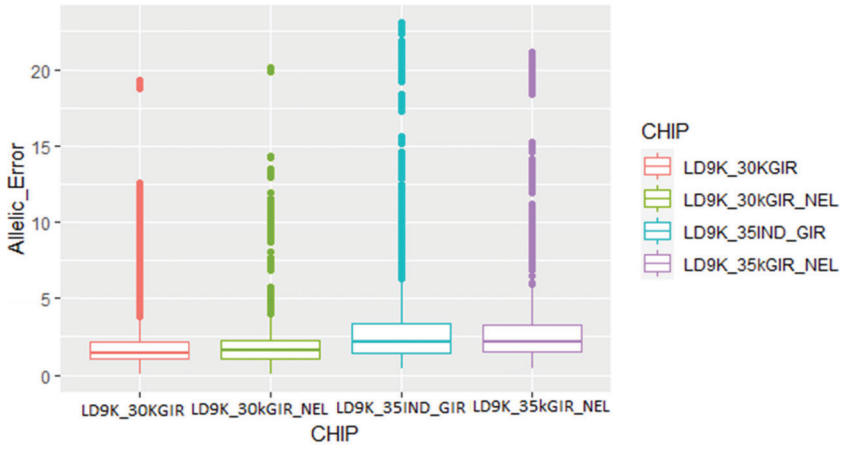
are related through the use of semen from Brazilian sires in Columbian dams (Asocebu 2020).

The 9K_NEL chip contained a higher proportion of markers with a low mean MAF (0.34) compared with 9K_GIR (0.38) (Fig. 2). The markers of the 9K chip were more adequate for the Gyr population, with the observation of a difference between the two breeds. Different results were reported by Utsunomiya et al. (2019) who observed that MAF and heterozygosity were similar, using commercial chip GGP35IND in four Zebu breeds (Brahman, Gyr, Nellore, and Sheko).

The LD9K_30KGIR scenario provided the highest CS (0.94) and CR (0.96) (Table 1), which included the Brazilian Gyr population and RPop of 306 animals. According to Bolormaa et al. (2015), the accuracy of imputation depends on the genetic distance between animals of the reference population and of the imputed population. The lowest mean allelic error rate was 2.01%, in this scenario (Fig. 3). In a study on Australian Holstein-Friesian cattle, Khatkar et al. (2012) estimated a mean allelic error rate of 2.8% when imputing genotypes from the 3K to the 50K panel. The authors attributed this finding to the larger number of SNPs with a high MAF and the relationship in the reference population.

For the LD9K_30KGIR_NEL scenario, the CS (0.93) and CR (0.93) were low with RPop of 909 animals that included the Nellore population, although this scenario comprised the largest number of reference animals. Although the Nellore and Gyr breeds belong to the subspecies *B. indicus*, a difference exists between these breeds. Villalobos-Cortés et al. (2015) studied the genetic structure and diversity of Zebu breeds and

**Fig. 3.** Mean allelic error rate (%) in the four scenarios used to test imputation accuracy.




described the independence between Gyr, Nellore, and Guzerá animals, supporting the diverse origin and different models of introgression in South America.

Some studies found no significant differences in imputation accuracy when crossbred animals were used, and there was no gain in accuracy when different breeds were included in the reference population (Piccoli et al. 2014). Berry et al. (2014) reported that the accuracy of imputation was reduced when different breeds are used in the reference population because of the low linkage disequilibrium between SNPs and short haplotypes of different breeds. These findings may explain the reduction in CS when the Nellore population was included as RPop in Gyr imputation.

The LD9K_35IND_GIR scenario provided the lowest CS (0.87) and CR (0.93) with RPop (176 animals) the least of the four scenarios evaluated (Table 1). However, Boison et al. (2015), performing imputation in different scenarios, obtained higher CS (0.91) and CR (0.92) values when genotypes were imputed from the 3K to the Illumina 50K chip using 171 reference animals of the Gyr breed. This scenario was used in the Gyr population genotyped with the GGP35IND chip, showing a low mean MAF of 0.2644 and highest allelic error rate (3.58%). However, Utsunomiya et al. (2019) reported a MAF $>$ 0.35 for GGP35IND in four Zebu breeds. We obtained a lower MAF for the GGP35IND chip, possibly because we only used the Gyr breed for the imputations.

In a study on Swedish cattle, Ma et al. (2013) reported that the CR was high when the MAF was low, while the correlation coefficient was low. The CR metric does not consider correct random filling, which is preferred for markers with low MAF. Hickey et al. (2012) highlighted that some imputation errors for a marker with low MAF can substantially reduce the correlation coefficient for this marker.

Another possible explanation for the low imputation accuracy in the LD9K_35IND_GIR scenario might be the design of the GGP35IND chip, which was customized based on the analysis of genotyped Nellore animals (69.32%), with a small percentage (7.0%) of SNPs from Gyr animals (Ferraz et al. 2018), i.e., the SNPs customized based on Nellore animals may be monomorphic for the Gyr breed. Although the population was smaller compared with the other scenarios, other imputation studies using similar populations achieved a greater imputation accuracy (Boison et al. 2015).

The LD9K_35IND_GIR_NEL scenario, including the Nellore reference population, did not improve the accuracy of imputation, i.e., the inclusion of other animal studies with Zebu populations Gyr did not improve the imputation accuracy, even when the commercial indicus chip was tested. Different factors influence the accuracy of imputation, such as the presence of monomorphic SNPs, low linkage disequilibrium, and a small reference population, reducing the accuracy of imputations (Ventura et al. 2016).

Furthermore, the chips for *B. indicus* use genotyped animals of the Nellore breed as a reference because this breed is more representative in Brazil and has a higher percentage of SNPs related to the selection objective (meat production – beef cattle). This fact results in low accuracy when breeds with different selection objectives, such as dairy and dual-purpose breeds, are used for imputation. In the present study, we found that CS is the best measure of imputation accuracy for the different scenarios. According to Ma et al. (2013), the correlation coefficient better captures the difference between imputation accuracy and random filling than CR.

In the present study, we tested a customized chip to determine the imputation accuracy in the Gyr breed. Although the design of the customized chip used genotype information from the same population as the imputed chip, it helps us to choose the correct panel for standardization of the data of the Gyr population. According to Carvalheiro et al. (2014), commercial and customized chips cannot be adequately compared. However, the authors highlight the importance of using

specific information of the populations to design LD chips. In addition, the performance of different marker density panels (such as LD or HD) and different genotyping platforms needs to be investigated as they present differences, with some showing high determination bias in relation to taurine breeds and others showing differences in the indicus breeds (Utsunomiya et al. 2014). The highest accuracy using the three metrics analyzed (CS, CR, and allelic error rate) was observed for the LD9K_30 K scenario, which was the most accurate for the Gyr population studied. In addition, it showed a statistically significant difference with LD9K_35IND_GIR and LD9K_35IND_GIR_NEL scenarios (p-value $<$ 2.2E-16).

Imputation in the Gyr breed was accurate for the GGP Bovine LDv4 chip. The scenarios that included the Nellore breed did not increase the imputation accuracy of the chip. The scenarios using the GGP indicus chip were less accurate for the Gyr population; however, it is necessary to study more Gyr populations and different Zebu breeds, as well as to test the efficiency of the chip. The findings suggest the need for greater representation of SNPs containing genotype data from different Zebu breeds in the design of new *B. indicus* chips.



## Acknowledgements

The authors wish to thank Associação Brasileira de Criadores de Gir Leiteiro (ABCGIL), Asociación Colombiana de Criadores de Ganado Cebu (ASOCEBU), and the Nelore Qualitas breeding program, Brazil, for providing the data used in this study. We also acknowledge the financial support of Fundação de Apoio a Pesquisa do Estado de São Paulo (FAPESP, 2015/06686-7) e (FAPESP, 2017/10630-2), Agence Universitaire de la Francophonie (PRISA, 1140/2018), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, 304047/2017-1). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

## References

Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S., and Lawlor, T.J. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. **93**: 743–752. doi:10.3168/jds.2009-2730. PMID:20105546.

Asocebu. 2020. Asociación Colombiana de Criadores de Ganado Cebú. Available at https://www.asocebu.com/index.php/razas/gyr [accessed 10 August 2020].

Berry, D.P., Mcclure, M.C., and Mullen, M.P. 2014. Within- and across-breed imputation of high-density genotypes in dairy and beef cattle from medium- and low-density genotypes. J. Anim. Breed. Genet. **131**: 165–172. doi:10.1111/jbg.12067. PMID:24906026.

Boichard, D., Chung, H., Dassonneville, R., David, X., Eggen, A., Fritz, S., et al. 2012. Design of a bovine low-density snp array optimized for imputation. PLoS ONE, **7**: e34130. doi:10.1371/journal.pone.0034130. PMID:22470530.

Boison, S.A., Santos, D.J.A., Utsunomiya, A.H.T., Carvalheiro, R., Neves, H.H.R., O'Brien, A.M.P., et al. 2015. Strategies for single nucleotide polymorphism (SNP) genotyping to enhance genotype imputation in Gyr (*Bos indicus*) dairy cattle: Comparison of commercially available SNP chips. J. Dairy Sci. **98**: 4969–4989. doi:10.3168/jds.2014-9213. PMID:25958293.

Bolormaa, S., Gore, K., Van Der Werf, J.H.J., Hayes, B.J., and Daetwyler, H.D. 2015. Design of a low-density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy. Anim. Genet. **46**: 544–556. doi:10.1111/age.12340. PMID:26360638.

Carvalheiro, R., Boison, S.A., Neves, H.H.R., Sargolzaei, M., Schenkel, F.S., Utsunomiya, Y.T., et al. 2014. Accuracy of genotype imputation in Nelore cattle. Genet. Sel. Evol. **46**: 1–11. doi:10.1186/s12711-014-0069-1, doi:10.1186/1297-9686-46-1.

Chud, T.C.S., Ventura, R.V., Schenkel, F.S., Carvalheiro, R., Buzanskas, M.E., Rosa, J.O., et al. 2015. Strategies for genotype imputation in composite beef cattle. BMC Genet. **16**(1): 10. doi:10.1186/s12863-015-0251-7. PMID:25652552.

Druet, T., Schrooten, C., and de Roos, A.P.W. 2010. Imputation of genotypes from different single nucleotide polymorphism panels in dairy cattle. J. Dairy Sci. **93**: 5443–5454. doi:10.3168/jds.2010-3255. PMID:20965360.

Ferraz, J.B.S., Wu, X., Li, H., Xu, J., Ferretti, R., Simpson, B., et al. 2018. Design of a low-density SNP chip for *Bos indicus*: GGP indicus technical characterization and imputation accuracy to higher density SNP genotypes. *In* Proceedings of the 10th World Congress on Genetics Applied to Livestock Production. pp. 3–7. Available from http://www.wcgalp.org/system/files/proceedings/2018/design-low-density-snp-chip-bos-indicus-ggp-indicus-technical-characterization-and-imputation.pdf.

Ghoreishifar, S.M., Moradi-Shahrbabak, H., Moradi-Shahrbabak, M., Nicolazzi, E.L., Williams, J.L., Iamartino, D., and Nejati-Javaremi, A. 2018. Accuracy of imputation of single-nucleotide polymorphism marker genotypes for water buffaloes (*Bubalus bubalis*) using different reference population sizes and imputation tools. Livest. Sci. **216**: 174–182. doi:10.1016/j.livsci.2018.08.009.

Hickey, J.M., Crossa, J., Babu, R., and de los Campos, G. 2012. Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. Crop Sci. **52**: 654–663. doi:10.2135/cropsci2011.07.0358.

Howie, B., Marchini, J., and Stephens, M. 2011. Genotype imputation with thousands of genomes. G3 Genes, Genomes, Genet. **1**: 457–470. doi:10.1534/g3.111.001198.

Iqbal, N., Liu, X., Yang, T., Huang, Z., Hanif, Q., Asif, M., et al. 2019. Genomic variants identified from whole-genome resequencing of indicine cattle breeds from Pakistan. PLoS ONE, **14**: e0215065. doi:10.1371/journal.pone.0215065. PMID:30973947.

Khatkar, M.S., Moser, G., Hayes, B.J., and Raadsma, H.W. 2012. Strategies and utility of imputed SNP genotypes for genomic analysis in dairy cattle. BMC Genomics, **13**: 538. doi:10.1186/1471-2164-13-538. PMID:23043356.

Larmer, S.G., Sargolzaei, M., and Schenkel, F.S. 2014. Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across Canadian dairy breeds. J. Dairy Sci. **97**: 3128–3141. doi:10.3168/jds.2013-6826. PMID:24582440.

Ma, P., Brøndum, R.F., Zhang, Q., Lund, M.S., and Su, G. 2013. Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish Red Cattle. J. Dairy Sci. **96**: 4666–4677. doi:10.3168/jds.2012-6316. PMID:23684022.

Montaldo, H.H., Casas, E., Ferraz, J.B.S., Vega-Murillo, V.E., and Román-Ponce, S.I. 2012. Opportunities and challenges from the use of genomic selection for beef cattle breeding in Latin America. Anim. Front. **2**: 23–29. doi:10.2527/af.2011-0029.

Oliveira Júnior, G.A., Chud, T.C.S., Ventura, R.V., Garrick, D.J., Cole, J.B., Munari, D.P., et al. 2017. Genotype imputation in a tropical crossbred dairy cattle population. J. Dairy Sci. **100**: 9623–9634. doi:10.3168/jds.2017-12732. PMID:28987572.

Piccoli, M.L., Braccini, J., Cardoso, F.F., Sargolzaei, M., Larmer, S.G., and Schenkel, F.S. 2014. Accuracy of genome-wide imputation in Braford and Hereford beef cattle. BMC Genet. **15**: 1. doi:10.1186/s12863-014-0157-9. PMID:24387126.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. **81**: 559–575. doi:10.1086/519795. PMID:17701901.

Sargolzaei, M., Chesnais, J.P., and Schenkel, F.S. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics, **15**: 478. doi:10.1186/1471-2164-15-478. PMID:24935670.

Toro Ospina, A.M., Silva Faria, R.A., Vercesi Filho, A.E., Cyrillo, J.N., dos, S.G., Zerlotti Mercadante, M.E., et al. 2020. Genome-wide identification of runs of homozygosity islands in the Gyr breed (*Bos indicus*). Reprod. Dom. Anim. **55**: 333–342. doi:10.1111/rda.13639.

Utsunomiya, Y.T., Bomba, L., Lucente, G., Colli, L., Negrini, R., Lenstra, J.A., et al. 2014. Revisiting AFLP fingerprinting for an unbiased assessment of genetic structure and differentiation of taurine and zebu cattle. BMC Genet. **15**(1): 47. doi:10.1186/1471-2156-15-47. PMID:24739206.

Utsunomiya, Y.T., Milanesi, M., Fortes, M.R.S., Porto-Neto, L.R., Utsunomiya, A.T.H., Silva, M.V.G.B., et al. 2019. Genomic clues of the evolutionary history of *Bos indicus* cattle. Anim. Genet. **5 0**: 557–568. doi:10.1111/age.12836. PMID:31475748.

Ventura, R.V., Miller, S.P., Dodds, K.G., Auvray, B., Lee, M., Bixley, M., et al. 2016. Assessing accuracy of imputation using different SNP panel densities in a multi-breed sheep population. Genet. Sel. Evol. **48**: 1–20. doi:10.1186/s12711-016-0244-7.

Villalobos-Cortés, A., Martínez, A., Vega-Pla, J.L., Landi, V., Quiroz, J., Marques, J.R., and Delgado, J.V. 2015. Genetic relationships among five zebu breeds naturalized in America accessed with molecular markers. Ital. J. Anim. Sci. **14**: 3280. doi:10.4081/ijas.2015.3280.