Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Accuracy of genotype imputation with different low density panels in Braford and Hereford cattle

M. L. Piccoli[123], J. Braccini Neto[1], F. F. Cardoso[4], M. Sargolzaei[35] and F. S. Schenkel[3].
[1] Universidade Federal do Rio Grande do Sul, Departamento de Zootecnia, Porto Alegre, Brazil.
[2] GenSys Consultores Associados S/S, Porto Alegre, Brazil.
[3] Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, Ontario, Canada.
[4] Embrapa Southern Region Animal Husbandry, Bagé, Brazil.
[5] The Semex Alliance, Guelph, Canada.

ABSTRACT: The main objective of this research was to test alternative low density SNP panels to impute Illumina 50K SNP panel genotypes in Braford and Hereford cattle. Genotypes from 3,768 Hereford, Braford and Nellore animals were used for testing imputation from low density SNP panels (3K, 6K, 8K, 15K and 20K) to the Illumina 50K SNP panel, under four different scenarios: including or not Nellore genotypes in the reference population in combination with the use or not of pedigree information. There were no significant differences in imputation accuracy among these four scenarios within each panel. However, significant differences between panels were found. The best accuracy was given by a customized 15K SNP panel, with an overall genotype concordance rate of 0.977, with 93.3% of the animals imputed with a concordance rate above 0.95. The concordance rates for the other SNP panels were 0.872, 0.952, 0.957 and 0.958 for 3K, 6K, 8K and 20K SNP panel, respectively. Therefore, in the Braford/Hereford population considered in this study, all the alternative panels denser than 3K could be used for imputing to the 50K SNP panel with an overall high imputation accuracy. However, the best results were obtained with the customized 15K SNP instead of the alternative commercial panels. The use of Nellore sire genotypes and pedigree information did not increase accuracy of imputation in this population.

Keywords: imputation, Braford, low density panel, SNP.

Introduction

Breeding animals for traits of economic importance has been practiced over the years based on the phenotypic information and relationships among individuals. Recent advances in DNA analysis, led to the complete sequencing of several species, including cattle (The Bovine Genome Sequencing and Analysis Consortium, 2009). With the development of the genomic science new technologies have emerged. For instance, different panels for genotyping SNPs (*Single Nucleotide Polymorphisms*) are available, such as the Illumina BeadChip BovineHD (Illumina Inc., San Diego, USA), that enables genotyping 777K SNPs in a single chip. These new genotyping technologies have stimulated the development of new research areas, such as the genotype imputation technology.

In general, research has demonstrated greater genetic progress in breeding programs if genomic predictions of genetic merit are used in genetic evaluations (Aguilar et al. (2010); Brito et al. (2011)). This greater progress is associated with a shorter generation interval, increase in the selection intensity and accuracy (Van Raden (2008); Goddard et al. (2010)).

Procedures for imputation of genotypes, a technique that refers to prediction of ungenotyped SNP genotypes, have been the subject of recent studies (Sargolzaei et al. (2010); Druet et al. (2010); Zhang and Druet (2010)). The main goal of this technology is impute high-density SNP panel genotypes from lower density panels. It could enable greater use of genotyping by farmers, since the cost of low density panels are more affordable for the cattle industry. The main objective of this research was to test alternative low density SNP panels to impute Illumina 50K SNP panel (Illumina Inc., San Diego, USA) genotypes in Braford and Hereford cattle.

Material and Methods

Data: Data was from the Conexão Delta G's genetic improvement program - Hereford and Braford (Zebu x Hereford) cattle, containing approximately 520,000 animals from 97 farms located in the South, Southeast, Midwest and Northeast regions of Brazil. There were 683 Hereford and 2997 Braford animals genotyped for either the 50K Illumina SNP panel (n=3,550) or the HD Illumina SNP panel (777K, n=130) from 17 farms located in the South of Brazil. There were also 88 Nellore bulls from the Paint Program genotyped with the HD SNP panel. Overall in the genotype dataset there were 62, 73, and 88 Hereford, Braford, and Nellore sires; 151 and 705 Hereford and Braford cows; 370 and 1,748 Hereford and Braford young bulls and 100 and 471 Hereford and Braford heifers, respectively. Only 21% of the genotyped animals had their sire also genotyped and 2% of the genotyped animals had their dam genotyped in the dataset.

Data editing: The animals genotyped with the HD

panel had their panel information masked to the same SNPs contained in the 50K SNP panel. After that, all animals had 49,345 SNPs. The SNP quality control included GC score (>=0.15), Call Rate (>=0.90), Hardy-Weinberg Equilibrium ($P>=10^{-6}$) and autosomal chromosome. The individual sample quality control considered GC Score (>=0.15), Call Rate (>=0.90), heterozygosity deviation (limit of ± 3 SD), repeated sampling and paternity errors. After the quality control, 3,698 samples and 43,248 SNP were retained in the study.

Reference and imputation sets: The dataset was divided in two sets. The reference population included all animals, except those that were born in 2011, which were assigned to the imputation population. This splitting of the dataset resulted in 2,735 reference and 963 imputation animals. For animals in the imputation population, 3K, 6K, 8K, 15K and 20K low density panels were created. The 15K SNP panel was created based on the 8K SNP panel by expanding it with SNPs selected based on minor allele frequency (MAF>0.23), linkage disequilibrium (LD<0.088) and preferably located evenly spaced between 2 SNPs in the 8K SNP panel. All the other panels were commercial panels: 3K and 6K Illumina panels (Illumina Inc., San Diego, USA), 8K and 20K GGP (GeneSeek genomic profiler) panels (Gene Seek Inc., Lincoln, USA).

Imputation scenarios: Animals born in 2011 were imputed in four different scenarios: including Nellore genotypes in the reference population and either including pedigree information (ne-p) or not including pedigree information (ne-np); not including Nellore genotypes in the reference population and either including pedigree information (nne-p) or not including pedigree information (nne-np). Imputation in all the scenarios was carried out by FImpute software (Sargolzaei et al. (2011)).

Results and Discussion

There were not significant differences in concordance rate among the four scenarios within each panel (Figure 1). However, there were substantial differences in concordance rate between panels (Table 1 and Figure 1). It was expected that including Nellore genotypes in the reference population could increase imputation accuracy. The reason was that the imputation population included mostly Braford animals (about 82%) that have in their breed composition between 15% and 75% of zebu breeds, including the Nellore breed. However, negligible gains were found for all alternative low density panels imputed to 50K when Nellore genotypes were used (Figure 1). The use of all pedigree information was also not important for any of the low density panels. The likely reason for this is related to the fact that approximately 50% of the animals in pedigree, as well as the genotyped animals, had unknown sire (offspring from groups of multiple sires) and only 21% and 2% the genotyped animals had their sires and dams genotyped, respectively.

The best imputation results were obtained using the 15K SNP panel (concordance rate equal to 0.977) in all scenarios, while the concordance rate using the 3K panel across all the scenarios was about 0.872. The 6K, 8K and 20K panels showed a concordance rate of 0.952, 0.957 and 0.958 over all scenarios, respectively (Table 1, Figure 1). In general, the highest was the SNP density in the low density panel, the best was the imputation accuracy to the 50K panel, except for the 20K panel. However, this panel was developed for imputation to 777K panel and only contains 7,033 SNPs in common with the 50k SNP panel after quality control. This resulted in very similar results to those from the 8K SNP panel, which is mostly nested in the 50K panel. Wang et al. (2012) in Angus, Dassonneville et al. (2012) in Blonde d'Aquitaine, Huang et al. (2012) in Hereford, and Chud (2014) in Canchim cattle tested the accuracy of imputation using different densities of markers and reported results generally in favor of a denser low density panel in beef cattle. Similar results were found in this research, where, with the exception of the 3K SNP panel, all the other low density panels tested could be used to accurately impute to 50K SNP panel, but with the best results found for the 15K panel.

For all panels and in the different scenarios, 100% of the genotypes were imputed. The concordance rate per chromosome was always above 0.93, except for the 3K panel (data not shown). For the 15K panel, the lowest concordance rate was on chromosome 28 (0.970) and the highest value on chromosome 1 (0.980).

Concordance rates greater than 0.95 may be expected to result in relatively small losses in accuracy of genomic prediction (Sargolzaei et al. (2010)). Over the four scenarios, the 15K SNP panel showed the highest percentage of animals with concordance rate between 0.95 to 1.00 (=93.3%), whereas for the 3K SNP panel this percentage was 6.5%. For the 6K, 8K and 20K panels, these percentages were 61.9%, 68.4% and 69.7% respectively (Table 1).

Conclusions

In the Braford/Hereford population considered in this study, all the alternative panels denser than 3K could be used for imputing to the 50K SNP panel with an overall high imputation accuracy. However, the best results were obtained with the customized 15K SNP panel instead of alternative commercial panels. The use of Nellore sire genotypes and pedigree

information did not increase accuracy of imputation in this population.

Literature Cited

Aguilar, I., Misztal, I., Johnson, D. L. et al. (2010) J. Dairy Sci., 93: 743-752.

Brito, F. V., Braccini Neto J., Sargolzaei, M. et al. (2011). BMC Genet., 12:80.

Chud, T. C. S. (2014). M.Sc. Thesis Universidade Estadual Paulista., Jaboticabal, Brazil.

Dassonneville, R., Fritz, S., Ducrocq, V. et al. (2012) J. Dairy Sci., 95:4136-4140.

Druet, T., Schrooten, C., Roos, A. P. W. (2010). J. Dairy Sci., 93: 5443–5454.

Goddard, M. E., Hayes, B. J., Meuwissen, T. H. E. (2010). Proc. 9th WCGALP.

Huang, Y., Maltecca, C., Cassady, J. P. et al. (2012). J. Anim. Sci., 90:4203-4208.

Sargolzaei, M., Schenkel, F. S., Chesnais, J. (2010). Dairy Cattle Breeding and Genetics Committee Meeting.

Sargolzaei, M. Chesnais, J. P., Schenkel, F. S. (2011). J. Anim. Sci., 89, E-Suppl. 1.

Van Raden, P.M. (2008). J. Dairy Sci., 91: 4414–4423.

Wang, H., Woodward, B., Bauck, S. et al. (2012). Livest. Sci.

Zhang, Z., Druet, T. (2010). J. Dairy Sci., 93: 5487–5494.

Figure 1. Concordance rate for imputation from alternative low-density panels to 50K SNP panel under 4 scenarios: (ne-p) - using Nellore genotypes in the reference population and considering pedigree information; (nne-p) - not using Nellore genotypes in the reference population and considering pedigree information; (ne-np) - using Nellore genotypes in the reference population and not using pedigree information; (nne-np) - not using Nellore genotypes in the reference population and not using pedigree information.

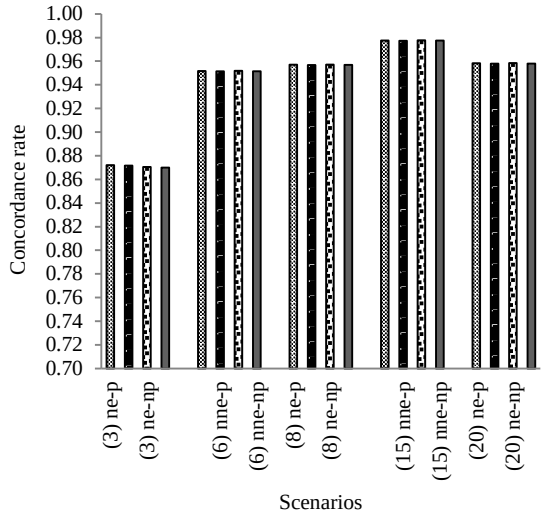

Table 1. Statistics for genotype imputation from alternative low-density panels (3K, 6K, 8K, 15K and 20K) to the 50K SNP panel over all scenarios (including or not including Nellore genotypes in the reference population x using or not using pedigree information).

Correct call range	Animals		Incorrect call (%)	Concor-dance rate
	No.	(%)		
		3K-50K		
0.95 - 1.00	63	6.5	2.968	0.970
0.90 - 0.95	292	30.3	7.663	0.923
0.85 - 0.90	278	28.8	12.358	0.876
0.80 - 0.85	208	21.6	17.245	0.828
0.75 - 0.80	90	9.3	21.804	0.782
0.70 - 0.75	22	2.3	26.830	0.732
0.65 - 0.70	8	0.8	31.662	0.683
0.60 - 0.65	4	0.4	36.668	0.634
Mean			12.848	0.872
		6K-50K		
0.95 - 1.00	596	61.9	2.856	0.971
0.90 - 0.95	304	31.6	6.893	0.931
0.85 - 0.90	47	4.9	11.962	0.880
0.80 - 0.85	11	1.1	17.483	0.825
0.75 - 0.80	4	0.4	22.509	0.775
0.70 - 0.75	1	0.1	25.473	0.745
Mean			4.844	0.952
		8K-50K		
0.95 - 1.00	659	68.4	2.715	0.973
0.90 - 0.95	260	27.0	6.728	0.933
0.85 - 0.90	32	3.3	12.031	0.880
0.80 - 0.85	8	0.9	17.038	0.830
0.75 - 0.80	4	0.4	22.284	0.777
Mean			4.313	0.957
		15K-50K		
0.95 - 1.00	899	93.3	1.337	0.987
0.90 - 0.95	56	5.8	3.554	0.964
0.85 - 0.90	8	0.9	7.026	0.930
0.80 - 0.85	1	0.1	10.801	0.892
Mean			2.262	0.977
		20K-50K		
0.95 - 1.00	671	69.7	2.662	0.973
0.90 - 0.95	256	26.5	6.738	0.933
0.85 - 0.90	24	2.5	12.397	0.876
0.80 - 0.85	8	0.9	16.871	0.831
0.75 - 0.80	4	0.4	21.980	0.780
Mean			4.191	0.958