Tropical Animal Health and Production (2022) 54:90
https://doi.org/10.1007/s11250-022-03102-0

REGULAR ARTICLES

Assessing accuracy of genotype imputation in the Afrikaner and Brahman cattle breeds of South Africa

S. Mdyogolo[1,2] · M. D. MacNeil[1,2,3] · F. W. C. Neser[2] · M. M. Scholtz[1,2] · M. L. Makgahlela[1,2]

Received: 13 July 2021 / Accepted: 1 February 2022

Abstract

Imputation may be used to rescue genomic data from animals that would otherwise be eliminated due to a lower than desired call rate. The aim of this study was to compare the accuracy of genotype imputation for Afrikaner, Brahman, and Brangus cattle of South Africa using within- and multiple-breed reference populations. A total of 373, 309, and 101 Afrikaner, Brahman, and Brangus cattle, respectively, were genotyped using the GeneSeek Genomic Profiler 150 K panel that contained 141,746 markers. Markers with MAF ≤ 0.02 and call rates ≤ 0.95 or that deviated from Hardy Weinberg Equilibrium frequency with a probability of ≤ 0.0001 were excluded from the data as were animals with a call rate ≤ 0.90. The remaining data included 99,086 SNPs and 360 Afrikaner, 75,291 SNPs and 288 animals Brahman, and 97,897 SNPs and 99 Brangus animals. A total of 7986, 7002, and 7000 SNP from 50 Afrikaner and Brahman and 30 Brangus cattle, respectively, were masked and then imputed using BEAGLE v3 and FImpute v2. The within-breed imputation yielded accuracies ranging from 89.9 to 96.6% for the three breeds. The multiple-breed imputation yielded corresponding accuracies from 69.21 to 88.35%. The results showed that population homogeneity and numerical representation for within and across breed strategies, respectively, are crucial components for improving imputation accuracies.

Keywords Beagle · FImpute · Minor allele frequency · Reference population

Introduction

Genomic selection (GS) and genome-wide association studies (GWAS) require the use of large numbers of animals (at least ≥ 1000) that have been genotyped at a very large number of loci (Calus et al., 2013; Lashmar et al., 2019). Although the cost of high-density panels and genotype-by-sequencing has decreased in recent years with the prospect of decreasing further in the future, per animal cost of genotyping on high-density panels remains high (Lashmar et al., 2019; Gebrehiwot et al., 2021). The high cost of genotyping on high-density panels hinders genetic progress in most breeding programs, particularly in developing countries. However, imputation may be used to increase the density of genotypes for animals that have been assessed with low-density panels (Friedrich et al., 2018). Genotype imputation is used to infer missing genotypes of animals assayed with low-density panels from those in a reference population that was genotyped using a high-density panel (Marchini et al., 2007; Li et al., 2009).

In developing countries, having a genotyped reference population of adequate size can be problematic, and this is particularly true for indigenous breeds. An alternative to increasing the number of animals in the reference population of the target breed is to combine data sets from related populations (Lund et al., 2014). Rowan et al. (2019) found that a large multi-breed reference population significantly increased imputation accuracy compared with a within-breed reference panel. However, Korkuć et al. (2019) argue that when imputing a small population and a large reference panel is not available and that a smaller reference panel should be used without including a different related breed.

Various approaches exist for imputing genotypes (Boison et al., 2015). Currently, it is difficult to generate high-density genotypes or whole genome sequence data for many

✉ S. Mdyogolo
sinebongo@gmail.com

[1] Department of Animal Breeding and Genetics, Agricultural Research Council, Irene, South Africa

[2] Department of Animal, Wildlife and Grassland Sciences, University of the Free State, Bloemfontein, South Africa

[3] Delta G, Miles City, MT, USA

Published online: 08 February 2022

individuals due to the high cost of genotyping. Fortunately, pedigree information for the majority of individuals may be readily available. The use of pedigree data can be especially valuable in identifying rare genetic variants (Cohen et al., 2004; Leigh et al., 2008; Manolio et al., 2008; Manolio et al., 2009; Cirulli and Goldstein, 2010; Ott et al., 2011; Cheung et al., 2013). However, in some breeding programs, incomplete or nonexistent pedigree data hinders its use in imputation.

The genetic architecture and molecular mechanisms underlying the adaptation of the South African indigenous breeds (e.g., Afrikaner) have not been thoroughly investigated (Lashmar et al., 2019). The same can be said for *Bos indicus*- (e.g., Brahman) and *Bos indicus*-derived (Brangus) cattle in South Africa. Despite cost challenges, there have been efforts towards accumulating a relatively modest number of high-density genotypes for these breeds in South Africa through the South African Beef Genomics Program. To date, except for Bonsmara, these efforts have fallen short of the data requirements for GS and GWAS. The need for reduced cost of genotyping and an increased number of animals with high-density genotypes provide motivation for maximizing the usefulness of the available samples.

A large proportion of SNP assays failing on an individual DNA sample may be indicative of a poor-quality DNA sample, which could lead to aberrant genotype calling. Samples with low genotyping efficiency, after first removing markers which have a low genotype call rate across samples, should be eliminated from further analysis as they might indicate a poorer-quality DNA sample (Turner et al., 2011). In developing countries, this editing of the data further reduces the typically already too small number of genotyped animals. Genotype assays that failed on many samples are poor assays and are likely to result in spurious data and it is recommended that these markers also be removed from the analysis (Turner et al., 2011). This editing can compromise the distance between markers, thus further compromising the analysis. Therefore, the aim of this study was to evaluate the use of imputation to rescue missing genotypes for the Afrikaner, Brahman, and Brangus breeds using within-breed reference populations and alternatively a multi-breed reference population.

Materials and methods

Pedigrees for the Afrikaner, Brahman, and Brangus cattle were obtained from the respective breed societies. Influential animals were identified using PEDIG (Boichard, 2002) from the marginal genetic contributions to their respective populations. A total of 373, 309, and 101 animals were found to represent the Afrikaner, Brahman, and Brangus breeds, respectively. These animals were genotyped using the GeneSeek Genomic Profiler (GGP or 150 K) consisting of 141,746 SNP markers. Sampling and genotyping procedures were approved by the Agricultural Research Council of South Africa ethics committee (APIC18/03). Quality control was uniform for all breeds. Using Plink v1.9 (Chang et al., 2015), MAF, call rates for markers and animals, and genotype frequency were determined for each breed. Markers with MAF ≤ 0.02 and call rates ≤ 0.95 or that deviated from Hardy Weinberg Equilibrium frequency with a probability of ≤ 0.0001 were excluded as were animals with a call rate ≤ 0.90. After these edits, 99,086, 97,987, and 75,291 SNP markers and 360, 288, and 99 animals remained in the data for the Afrikaner, Brahman, and Brangus breeds, respectively.

For each breed, a random set of animals was selected and a random portion of the genotypes of these animals were masked. For Afrikaner and Brahman, 7986 and 7002 SNPs from 50 randomly selected animals were masked. For Brangus, 60,696 SNPs of the 97,897 SNPs, from 30 randomly selected animals were masked.

Within breeds, the masked genotypes were imputed using BEAGLE (Browning et al., 2018) and FImpute (Sargolzaei et al., 2014). BEAGLE phases and imputes missing genotypes based on the similarity of haplotypes (Browning and Browning, 2009). The program restricts the hidden Markov model (HMM) calculations to clusters of markers that are genotyped in the study population, which reduces memory requirements and computation time (Browning and Browning, 2016). FImpute uses an overlapping sliding window approach to efficiently exploit relationships or haplotype similarities between individuals in the study and reference population (Sargolzaei et al., 2014). The program starts with long windows to capture similarities between close relatives followed by short windows to capture information from more distant relatives (Sargolzaei et al., 2014). FImpute also uses pedigree information if it is available.

A multi-breed reference population that was created by pooling genotypes from the three breeds included a total of 617 animals (Afrikaner = 310, Brahman = 238, and Brangus = 69) animals. This reference population used for multi-breed genotype imputation contained 60,364 SNPs, after combined quality control, that were common to the three breeds. To test this reference population, study populations in which 7986 SNPs of the 99,086 SNP from 50 randomly selected Afrikaner cattle, 7002 SNPs of the 97,987 SNPs from 50 randomly selected Brahman cattle and 20,000 SNPs of the 75,291 from 30 randomly selected Brangus cattle were masked. The genotypes at the masked loci were again imputed using both BEAGLE (Browning et al., 2018) and FImpute (Sargolzaei et al., 2014).

A multidimensional scaling (MDS) cluster analysis of genotyped individuals based on genome-wide identity by state (IBS) pairwise distances was conducted using Plink

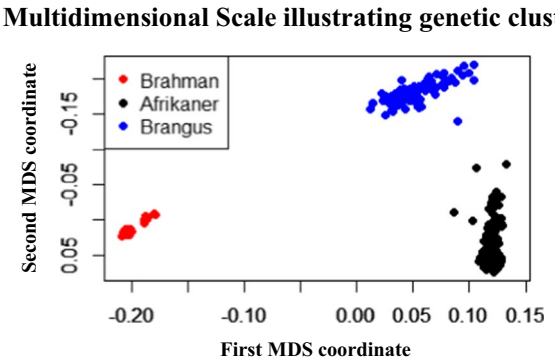

Fig. 1 Multidimensional scale (MDS) clustering illustration of the genetic distinctiveness of the Afrikaner, Brahman, and Brangus cattle breeds

v1.9 (Chang et al., 2015). The accuracy of imputation was assessed by the concordance rate (CR) between actual markers masked as missing and imputed genotypes using SnpSift (Cingolani et al., 2012). Lastly, the influence of MAF and LD on CR for imputed data (within- and multi-breeds) was assessed with simple linear regression and plotted.

Results and discussion

Population cluster, minor allele frequency, and linkage disequilibrium

The Afrikaner (*Bos taurus africanus*) and Brahman (*Bos indicus*) breeds of cattle are well adapted to the harsh environmental conditions of South Africa and are extensively used for domestic beef production. Although bred for similar purposes, the Afrikaner and Brahman cattle breeds are characterized by different historical origins in South Africa. The Afrikaner is a Sanga-type breed, combining indicine and African taurine ancestry, and is genetically distinct from European *Bos taurus* breeds (Makina et al., 2016). The Brahman is a *Bos indicus* breed originating from backcrossing of Zebu cattle (e.g., Ongole, Gir, Guzerat, and Krishna Valley cattle) adapted to the subtropical environmental conditions of south Asia with European taurine breeds in America (Koufariotis et al., 2018; Low et al., 2020). Afrikaner cattle are believed to be the product of southward migration from Eurasia along the west coast of the continent via the horn of Africa (Hanotte et al., 2002). In contrast, the Brahman cattle were first imported from America directly into South Africa in the mid-1950s (http://southafrica.co.za/brahman-cattle.html).

Brangus cattle were used in this study together with cattle of the Afrikaner and Brahman breeds with the view that this breed could potentially increase the accuracy of imputation for the other breeds when a multi-breed reference population strategy served as a base. Brangus was developed to combine the superior traits of Angus and Brahman cattle. Thus, they are a stabilized composite breed with 3/8 Brahman and 5/8 Angus genetics (http://www.thebeefsite.com/breeds/beef/43/brangus/). Like Brahman, the initial development of Brangus was in the United States of America (USA) at a United States Department of Agriculture research station at Jeanerette, Louisiana (USDA Yearbook of Agriculture, 1935).

The separation of the Afrikaner, Brahman, and Brangus breeds was apparent as shown by MDS clustering (Fig. 1). The Afrikaner and Brangus had similar values for the first MDS coordinate while Brahman was substantially distant from the other breeds. Afrikaner and Brahman had similar values on the second MDS coordinate, with Brangus being distant. To some degree, the development of these

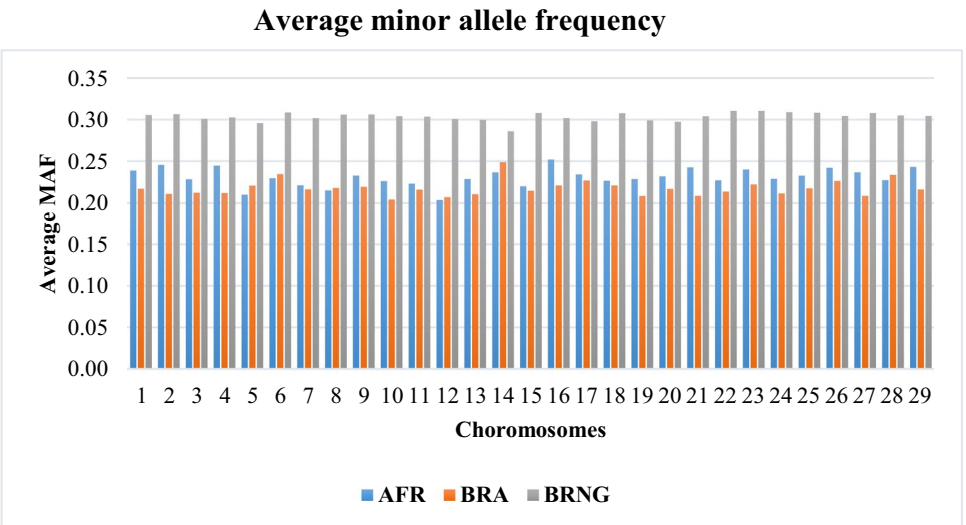

Fig. 2 Average minor allele frequency (MAF) across 29 autosomes for the Afrikaner (AFR), Brahman (BRA) and Brangus (BRNG) breeds of cattle

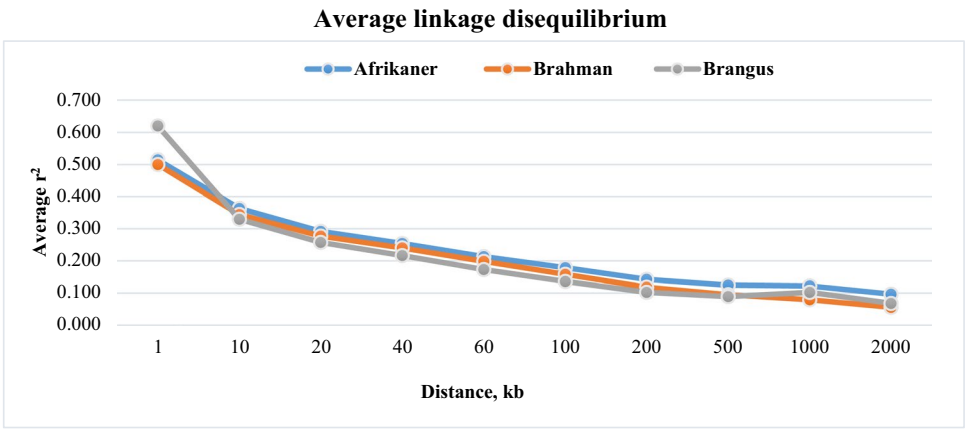

Fig. 3 Average linkage disequilibrium (r2) for each pair of SNP alleles within a window of 10 000 kb estimated for marker spacing from of 1 to 2 000 kb

three breeds constitutes varying combinations of indicine and taurine genetics. It might be speculated that the separation of Brahman from Afrikaner and Brangus along the X-axis would be attributable to differences in the proportion of indicine genetics. Likewise, it could be speculated that the separation of Brangus from Afrikaner and Brahman along the Y-axis might reflect differences in heterozygosity. The size and compactness of the MDS clusters of animals might be interpreted as indicating the degree of within-breed genetic homogeneity, with the Brahman evidently being the most homogenous of the three breeds.

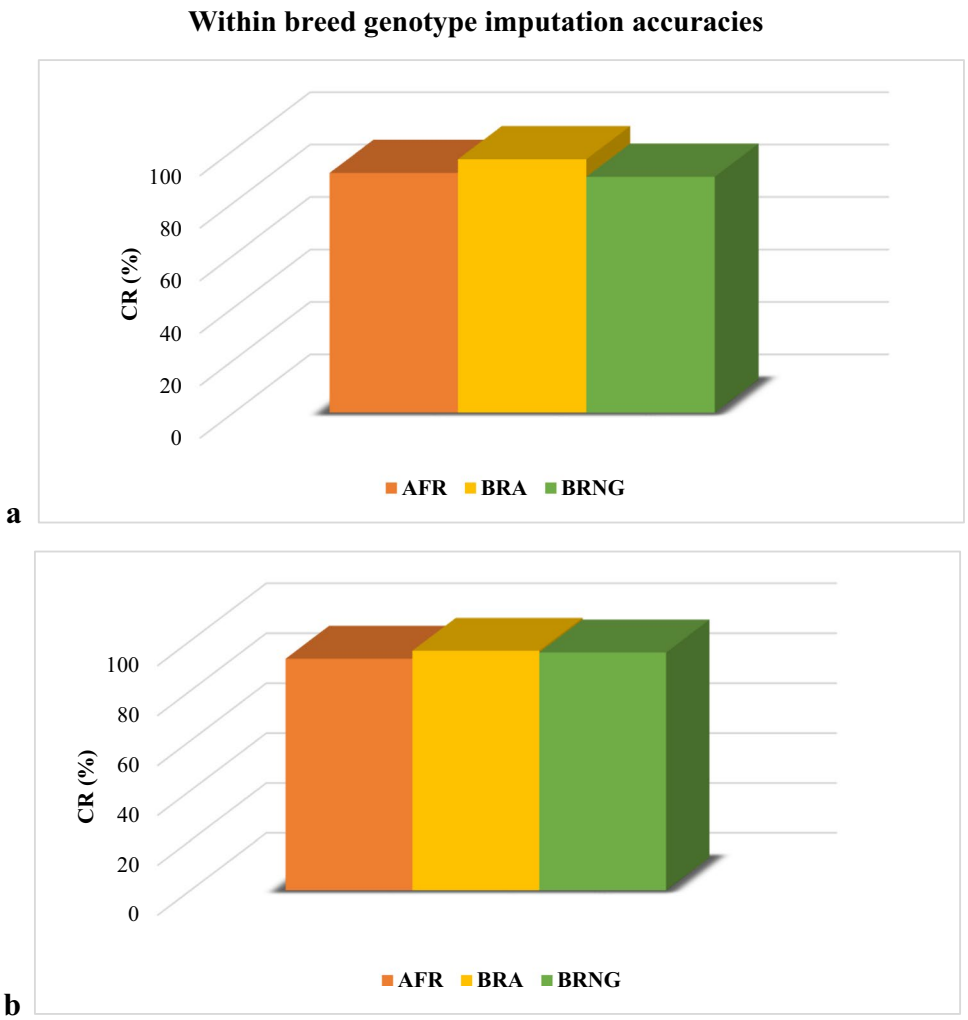

Fig. 4 Concordance rate (%) used to measure within breed genotype imputation accuracies for Afrikaner (AFR), Brahman (BRA) and Brangus (BRNG) utilizing **a.** BEAGLE & **b.** FImpute

The average MAF for Afrikaner, Brahman, and Brangus was 0.23, 0.22, and 0.30, respectively (Fig. 2). Minor allele frequency ranged from 0.20 to 0.31 for majority of the chromosomes implying that the effect of low MAF on the overall LD estimates for each breed should be negligible as previously suggested by Sargolzaei et al. (2008). The number of markers with low MAF (<0.10) were 20,319, 10,075, and 3077 for Afrikaner, Brahman, and Brangus, respectively. The observation of lower MAF in Afrikaner and Brahman may be partly reflective of the SNP panel discovery breeds being mainly of European taurine origin. Although also displaying a similar percentage of markers with low MAF, the Brahman breed may be better characterized by the GeneSeek Genomic Profiler than the Afrikaner due to the SNP array containing some *Bos indicus* markers and the use of taurine breeds in the initial development of Brahman. The high MAF of Brangus and smaller number of markers with low MAF may reflect its more recent formation as an indicine-taurine composite breed in comparison to Afrikaner and Brahman. Previous research has shown MAF for these breeds as 0.25 for Afrikaner (Makina et al., 2015), 0.24 for Brangus (He et al., 2018), and 0.27 for Brahman (Farah et al., 2018).

Unlike European breeds (e.g., *Bos taurus* breeds), indigenous and *Bos indicus* breeds were underrepresented in the SNP panel discovery breeds. Thus, in South Africa investigations were conducted to test the viability of existing SNP assays to study indigenous cattle of South Africa (Qwabe et al., 2013; Lashmar et al., 2019). Observations from these initial studies showed lower minor allele frequencies (MAF) and fewer informative SNPs for the indigenous breeds (Qwabe et al., 2013; Zwane et al., 2016; Lashmar et al., 2018). Gebrehiwot et al. (2021) concluded that a new SNP panel was needed to accommodate crossbred African dairy cattle and possibly other indigenous cattle. Although still in its infancy, genomic research has yielded promising results in identifying genes associated with production, adaptation, and reproduction in different breeds of South African cattle (Wang et al., 2015; Makina et al., 2016; Zwane et al., 2019).

Although the observed MAF for the three breeds is expected to have minor effects on LD, they may have considerable effects on imputation accuracies. Some imputation programs have been found to perform poorly for low MAF SNP (Shi et al., 2019). Shi et al. (2019) explained the poor performance of some programs as being due to

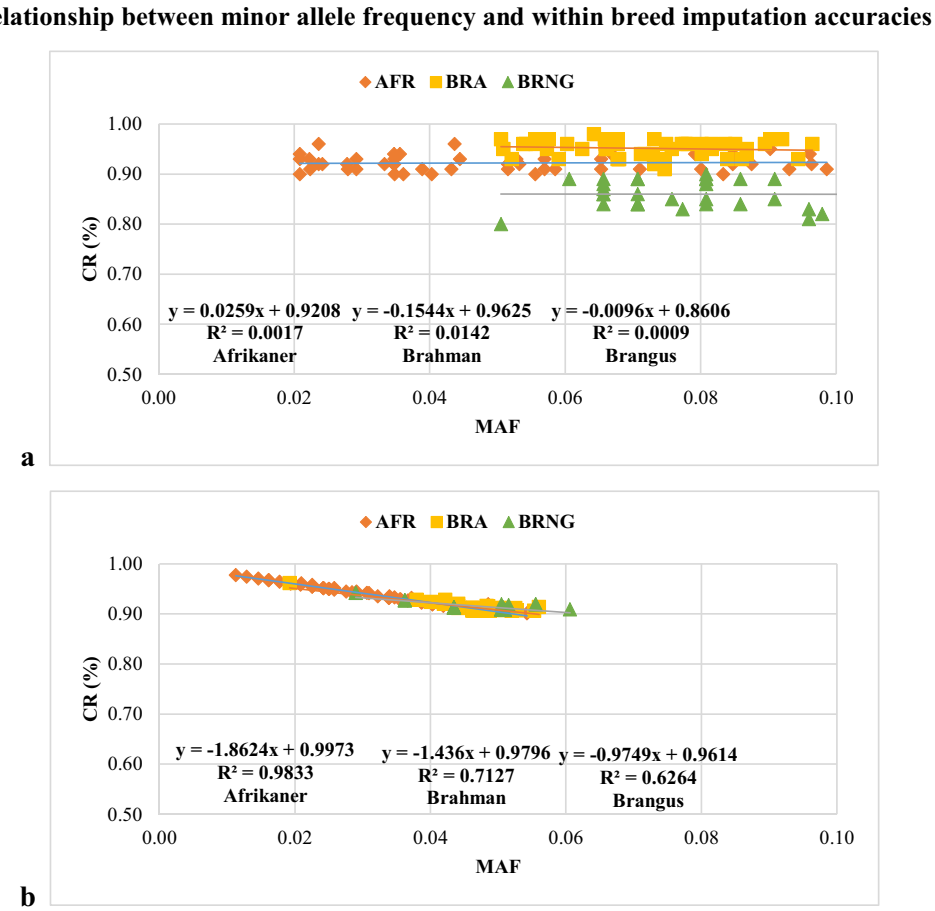

Fig. 5 Linear regression displaying the relationship between MAF and CR within breed data imputed with **a.** BEAGLE & **b.** FImpute

reference population bias. Thus, the importance of the composition of reference population coupled with the use of a suitable imputation program is emphasized to achieve relatively high accuracy imputation. In the current study, it was hypothesized that MAF effects may be expected for a breed that is not numerically well represented in the reference population (e.g., Brangus) or a breed that is less genetically diverse (e.g., Brahman).

The squared correlation between SNP alleles at different loci indicated greater initial LD in Brangus than in Afrikaner and Brahman with observed LD at a marker spacing of 1 kb being greater in Brangus (0.62) than in Afrikaner (0.51) and Brahman (0.50) (Fig. 3). However, a sharp decline in LD was noted in Brangus and average LD was 0.23, 0.20, and 0.21 for the Afrikaner, Brahman, and Brangus, respectively. Zhu et al. (2013) and Rogers (2014) indicated that populations formed recently, i.e., through crossbreeding programs, are typically characterized by a steep decline in initial LD; and this is in agreement with the current result for the Brangus breed. Linkage disequilibrium for Brahman in this study was slightly greater at marker spacing of 10 kb, than estimates of 0.25 for Australian Brahman (Porto-Neto et al., 2014) and 0.21 for Brazilian Gyr (Silva et al., 2010). However, the present estimates of LD for Brahman (0.33–0.36 at an interval of 10 kb) are similar to those of other indicine breeds such as Nellore (McKay et al., 2007).

Within-breed genotype imputation

Figure 4 presents the concordance rate, defined as the percentage of correctly imputed genotypes over all study individuals. Using BEAGLE, imputation of Brahman genotypes had an average concordance rate of 96.6% while the average accuracies for Afrikaner and Brangus were 91.4% and 89.9%, respectively (Fig. 4a). Using FImpute, imputation of Brahman genotypes was 96.0% accurate while the concordance rates for Afrikaner and Brangus were 92.8% and 95.3%, respectively (Fig. 4b). Thus, using pedigree information in addition to LD (i.e., using FImpute) resulted in at least as accurate imputation as did using LD information only (i.e., using BEAGLE). However, the differences between approaches in accuracy of imputation were generally negligible. Nelson et al. (2016) and Bai et al. (2019) stated that a genetically diverse study population requires a corresponding diverse reference population. For Brahman, the least diverse of these three breeds as presented in Fig. 1,

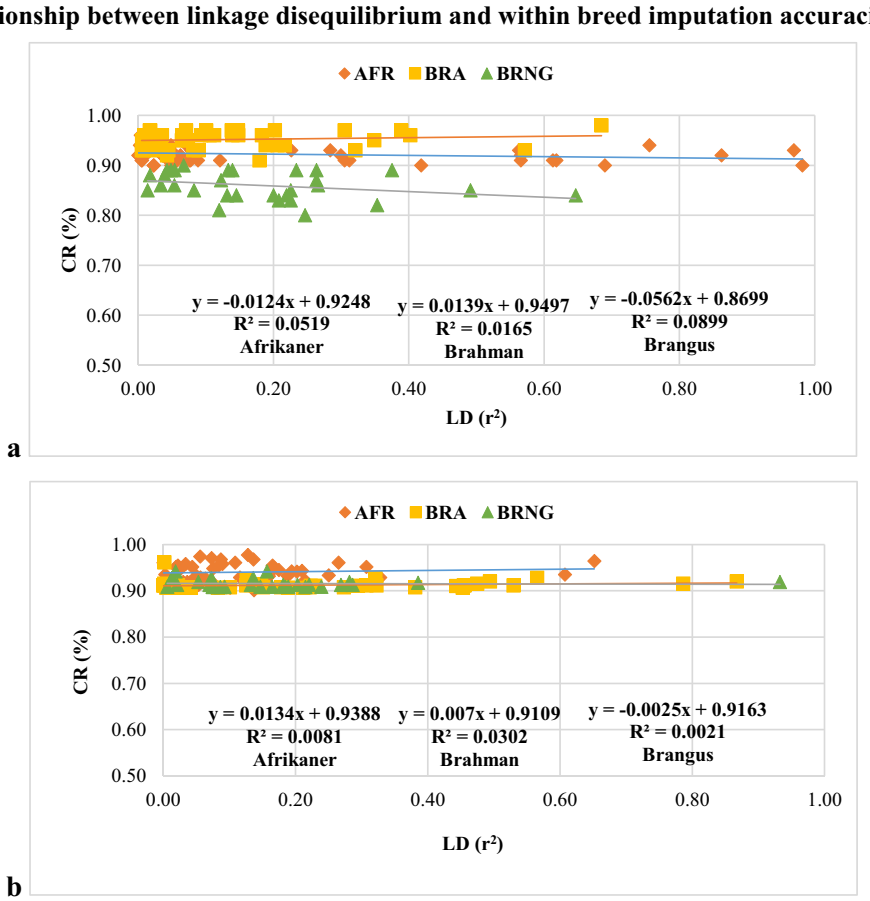

Fig. 6 Linear regression displaying the relationship between Linkage disequilibrium and concordance rate for within breed data imputed with **a.** BEAGLE & **b.** FImpute

genotype imputation accuracies were numerically greater for both BEAGLE and FImpute than for the more genetically diverse Afrikaner and Brangus. Thus, within-breed imputation appeared slightly more accurate when the breed was less genetically diverse and consequently making the animals for whom genotypes were to be imputed more closely related to the reference population (Huang et al., 2012; Bouwman et al., 2014).

The coefficient of determination from the simple linear regression (R^2) of the accuracy of imputation on MAF ranged from 0.001 to 0.014 illustrating the negligible influence of MAF on the concordance rate for within breed imputation, when using BEAGLE (Fig. 5a). However, when FImpute was used for imputation, differences in MAF had a much larger effect on the concordance rate ($R^2 = 0.626$ to 0.983) (Fig. 5b). Thus, MAF had an important influence on the accuracy of imputation when pedigree relationships were used in the prediction of missing genotypes. However, whether BEAGLE or FImpute was used for imputation, the influence of LD at 10 kb marker spacing on the accuracy of imputation was small as indicated by low coefficients of determination ($R^2 = 0.017$ to 0.090 and $R^2 = 0.002$ to 0.030, respectively) (Fig. 6a, b). These findings on family-based imputation methods agreed with findings by Liu et al. (2019). Chassier et al. (2018) found that LD and MAF had no major influence on the imputation accuracies. Also, these findings were consistent with Shi et al. (2019) in that the effect of MAF on CR depended on the choice of software used for imputation.

Imputation using a multi-breed reference population

When using BEAGLE for imputation with a multi-breed reference population composed of Afrikaner, Brahman, and Brangus, the concordance rates were 82.6%, 79.4%, and 69.2%, respectively (Fig. 7a). When using FImpute to impute with the multi-breed reference population, Afrikaner, Brahman, and Brangus had concordance rates of 88.4%, 81.7%, and 72.3%, respectively (Fig. 7b). Thus, the accuracy of imputation was reduced by more than 10% with the use of a multi-breed reference population.

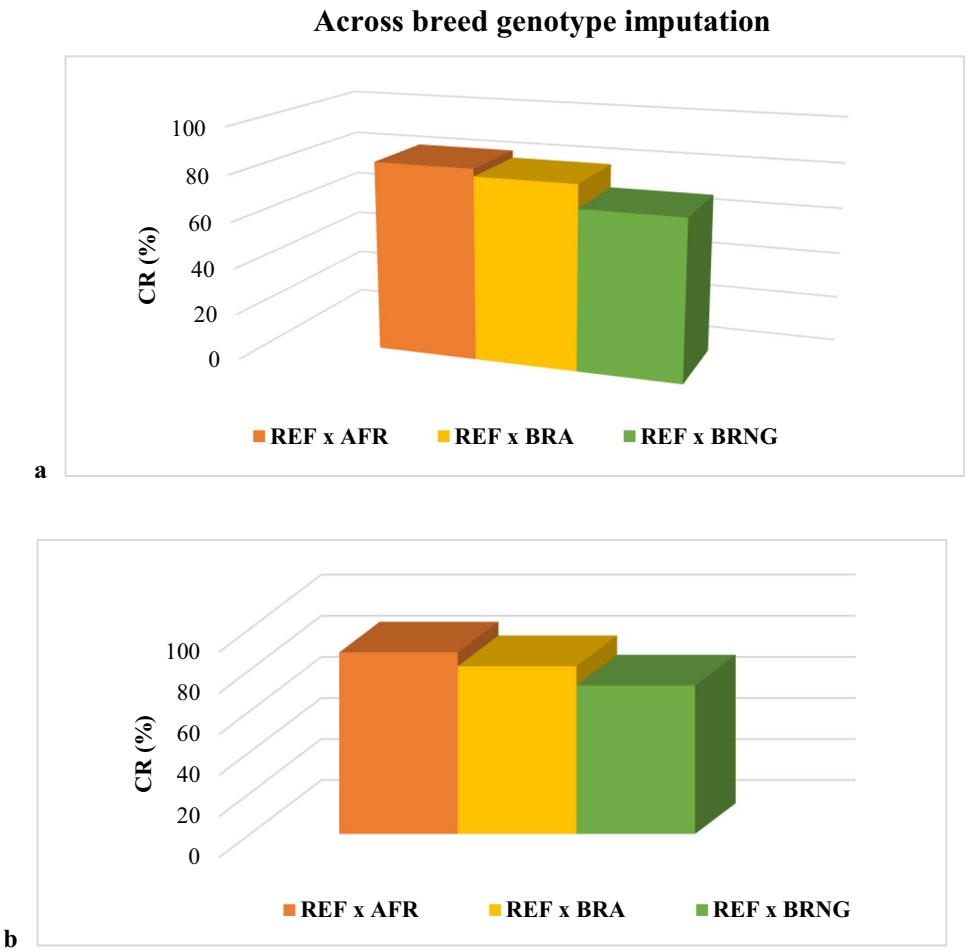

Fig. 7 Concordance rate (%) used to measure across breed genotype imputation accuracies for Afrikaner (AFR), Brahman (BRA), and Brangus (BRNG) utilizing a multibreed reference population using **a.** BEAGLE & **b.** FImpute

Breed representation in the reference population appeared to affect the observed outcomes of imputation using the multi-breed reference population. The Afrikaner had the highest representation ($n = 310$) in the reference population followed by Brahman ($n = 238$) and Brangus ($n = 69$). Of particular interest was the observation that Afrikaner, with its better representation in the reference population, had more animals with low MAF in the imputed genotypes, relative to Brahman and Brangus. Thus, with better representation of each breed, and especially Brangus, improved accuracies might be achieved.

Coefficients of determination from the simple linear regression of concordance rate on MAF ($R^2 = 0.0015$ to 0.1287) with the multi-breed reference population were similar to those that had been obtained with the breed-specific reference populations when using BEAGLE (Fig. 8a). Using FImpute, and the multi-breed reference population, the relationship between MAF and concordance rate was reduced ($R^2 = 0.1463$ to 0.7404) compared to the results obtained with breed-specific reference panels (Fig. 8b); thus, signifying the value of exploiting both LD and family information when using a multi-breed reference population. BEAGLE exploits LD between markers and ignores IBD derived from pedigree data (Browning and Browning, 2007), whereas FImpute uses a combination of pedigree and LD information. In FImpute, haplotype similarities based on IBD between individuals in reference and study populations are identified, thereby identifying rare variants (Sargolzaei et al., 2014). Luan et al. (2012) found both IBD and IBS to be important for accurate imputation because they both contributed information to the relationship among animals. It was shown that within breed, IBD was reliable and improved GS accuracies and this was in agreement with within breed imputation displayed in the current study.

The effects of LD on accuracy of imputation remain negligible with the multi-breed reference population irrespective of the method used ($R^2 = 0.0004$ to 0.0804) (Fig. 9a, b). These results confirm that of Hozé et al. (2013) wherein it was shown that LD had no major influence on the accuracy of imputed genotypes relative to the reference population size and the relationship between the reference and target population.

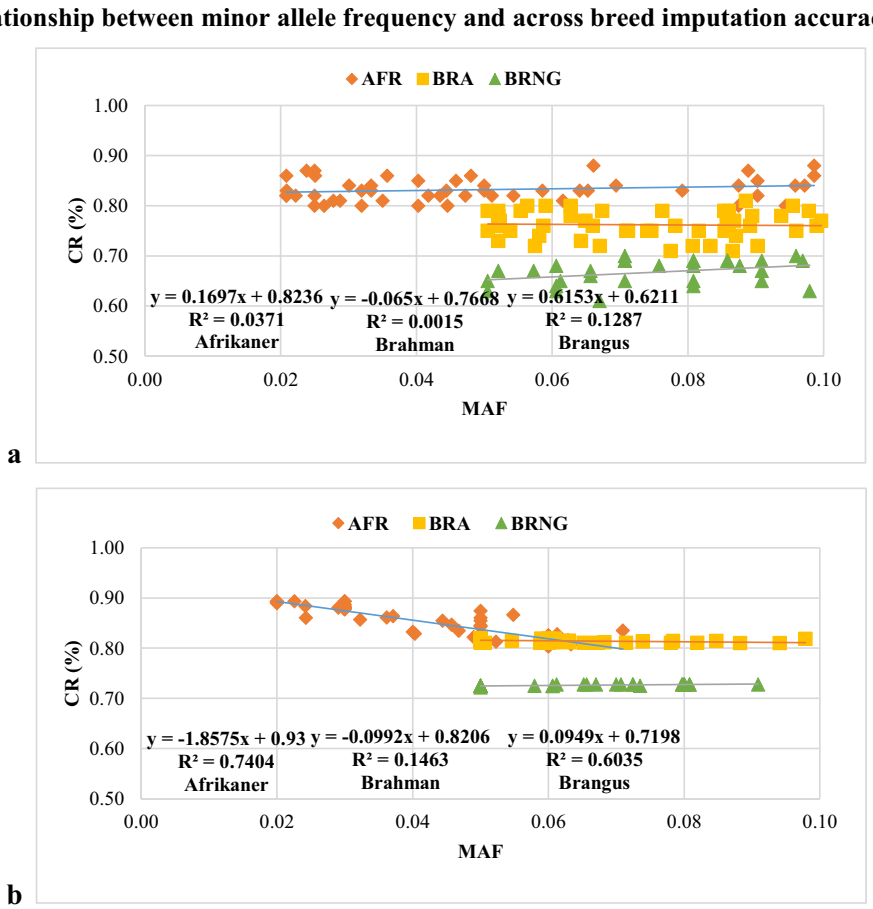

Fig. 8 Linear regression displaying the relationship between minor allele frequency and concordance rate for multi-breed reference population imputation strategy using **a.** BEAGLE **b.** FImpute

Accuracy of imputation is influenced by the size and composition of the reference population and is crucial to the accuracy of downstream GS and GWAS (e.g., Huang et al., 2012; Ullah et al., 2019). In cattle breeding programs, reference populations for genotype imputation vary, comprising of only individuals that are closely related to the individuals in the study population (e.g., Huang et al., 2012), to those composed of individuals from a variety of breeds with diverse genetic relationships to the study group (e.g., Rowan et al., 2019). For more genetically diverse breeds, the accuracy of imputation may benefit from utilizing a large and more genetically diverse reference population that includes multiple breeds (Brøndum et al., 2012; Rowan et al., 2019). However, Berry et al. (2014) stated that it is advantageous for accurate imputation to include only animals from the breed being imputed in the reference population. Thus, when making use of population-wide LD imputation, the accuracy can be positively affected by the presence of close relatives in the reference population and choice of the imputation program used to capture LD between markers. Likewise, when a large reference population is not available, Korkuć et al. (2019) recommended using a smaller same-breed reference population without including different related breeds.

Conclusion

Concordance rates were less affected by the method used for imputation, while MAF affected the accuracy of imputation more with FImpute compared to BEAGLE. Across all three breeds, the accuracy of imputation was higher using within-breed strategies and is potentially the most feasible strategy among the currently tested strategies. Based on the findings from this study, it is recommended that the South African breeding industry vigorously make use of within breed imputation to initiate the utility of genomic tools, while working towards the goal of increasing the number of genotypes for the different breeds.

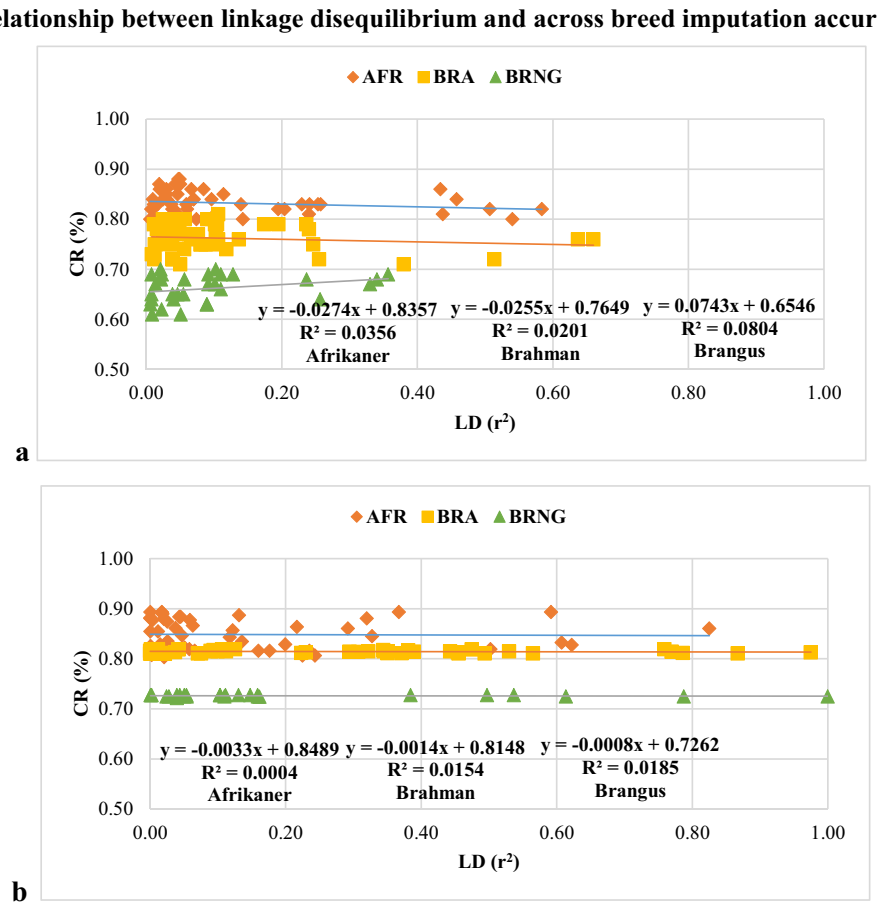

Fig. 9 Linear regression displaying the relationship between linkage disequilibrium and concordance rate for multi-breed reference population using **a.** BEAGLE **b.** FImpute

Acknowledgements SM thanks the University of the Free State (UFS), the National Research Foundation (NRF), and the Technology Innovation Agency (TIA), an implementing agency of the Department of Science and Innovation for financial support. Without the financial support of the Red Meat Research and Development SA (RMRD SA) and the Beef Genomics Program (BGP), this research would not have been possible. The breed societies granting of permission to use the data is gratefully acknowledged.

Author contribution The article was drawn from the PhD thesis of SM. SM and MLM planned the study. SM implemented analyses of the data. FWCN and MLM supervised SM. MMS and MDM provided inputs on the interpretation of the results. MDM edited the article that was originally drafted by SM. All authors read and approved the final manuscript.

Funding This project received financial support from the Department of Science and Innovation (DSI), the Red Meat Research and Development South Africa (RMRD SA), and the South African Beef Genomics Program (BGP). We would also like to thank the southern African Brahman Breeders' Society for their financial contribution towards the generation of genomic data.

Data availability Data cannot be shared with the public, due to the breed society's still participating in the Beef Genomics Program of South Africa and under the process of accumulating and preliminarily analyzing their data for their breeding program purposes.

Declarations

Conflict of interest The authors declare no competing interests.

Ethics approval Approval of the study was granted by the Animal Ethics Committee (AEC) of the Agricultural Research Council of South Africa (APIC18/03).

References

Bai, W.Y., Zhu, X.W., Cong, P.K., Zhang, X.J., Richards, J.B., and Zheng, H-F., 2019. Genotype imputation and reference panel: a systematic evaluation on haplotype size and diversity. Briefing in Bioinformatics, 1–12.

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. Journal of Animal Breeding and Genetics, 131, 165-172.

Boichard, D., 2002. Pedig: A Fortran package for pedigree analysis suited for large populations. Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France. CD-ROM communication no. 28–13

Boison, S.A., Santos, D.J.A., Utsunomiya, A.H.T., Carvalheiro, R., Neves, H.R.H., O'Brien, A.M.P., Garcia, J.F., Sölkner, J., and 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(7).

Bouwman, A.C., Hickey, J.M., Calus, M.P.L., and Veerkamp, R.F., 2014. Imputation of non-genotyped individuals based on genotyped relatives: assessing the imputation accuracy of a real case scenario in dairy cattle. Genetics Selection Evolution, 46, 6.

Brøndum, R.F., Ma, P., Lund, M.S., and Su, G., 2012. Short communication: Genotype imputation within and across Nordic cattle breeds. Journal of Dairy Science, 95, 6795-6800.

Browning, B.L., and Browning, S.R., 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. American Journal of Human Genetics, 84, 210–223.

Browning, B.L., and Browning, S.R., 2016. Genotype imputation with millions of reference samples. American Journal of Human Genetics, 98, 116–126.

Browning, B.L., Zhou, Y., and Browning, S.R., 2018. A one-penny imputed genome from next generation reference panels. American Journal of Human Genetics, 103(3), 338-348.

Browning, S.R., and Browning, B.L., 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. American Journal of Human Genetics, 81, 1084-1097.

Calus, M.P.L., de Haas, Y., and Veerkamp, R.F., 2013. Combining cow and bull reference populations to increase accuracy of genomic prediction and genome-wide association studies. Journal of Dairy Science, 96, 6703–6715.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience, 4(7).

Chassier, M., Barrey, E., Robert, C., Duluard, A., Danvy, S., and Ricard, A., 2018. Genotype imputation accuracy in multiple equine breeds from medium- to high-density genotypes. Journal of Animal Breeding and Genetics, 135, 420-431.

Cheung, C.Y.K., Thompson, E.A., and Wijsman, E.M., 2013. GIGI: An approach to effective imputation of dense genotypes on large pedigrees. American Journal of Human Genetics, 92, 504-516.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly, 6(2), 80-92.

Cirulli, E.T., and Goldstein, D.B., 2010. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nature Reviews Genetics, 11, 415–425.

Cohen, J.C., Kiss, R.S., Pertsemlidis, A., Marcel, Y.L., McPherson, R., and Hobbs, H.H., 2004. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science, 305, 869–872.

Farah, M.M., Fortes, M.R.S., Kelly, M., Porto-Neto, L.R., Meira, C.T., Carreño, L.O.D., Ricardo da Fonseca, R., and Moore, S.S., 2018. Accuracy of genomic selection predictions for hip height in Brahman cattle using different relationship matrices. Pesquisa Agropecuária Brasileira, 53(6), 717-726.

Friedrich, J., Antolın, R., Edwards, S.M., Sanchez-Molano, E., Haskell, M.J., Hickey, J.M., and Wiener, P., 2018. Accuracy of genotype imputation in Labrador retrievers. Animal Genetics, 49, 303-311.

Gebrehiwot, N.Z., Aliloo, H., Strucken, E.M., Marshall K, Al Kalaldeh, M., Missohou, A., and Gibson, J.P., 2021. Inference of ancestries and heterozygosity proportion and genotype imputation in West African cattle populations. Frontiers in Genetics, 12, 584355.

Hanotte, O., Bradley, D.G., Ochieng, J.W., Verjee, Y., Hill, E.W., and Rege, J.E.O., 2002. African pastoralism: genetic imprints of origins and migrations. Science, 296, 336-339.

He, J., Guo, Y., Xu, J., Li, H., Fuller, A., Tait Jr, R.G.T., Wu, X.L., and Bauck, S., 2018. Comparing SNP panels and statistical methods for estimating genomic breed composition of individual animals in ten cattle breeds. BMC Genetics, 19, 56.

Hozé, C., Fouilloux, M-N., Venot, E., Guillaume, F., Dassonneville, R., Fritz, S., Ducrocq, V., Phocas, F., Boichard, D., and Croiseau, P., 2013. High-density marker imputation accuracy in sixteen French cattle breeds. Genetics Selection Evolution, 45, 33.

Huang, Y., Maltecca, C., Cassady, J.P., Alexander, L.J., Snelling, W.M., and MacNeil, M.D., 2012. Effects of reduced panel, reference origin, and genetic relationship on imputation of genotypes in Hereford cattle. Journal of Animal Sciences, 90(12), 4203-4208.

Korkuć, P., Arends, D., and Brockmann, G. A., 2019. Finding the optimal imputation strategy for small cattle populations. Frontiers in Genetics, 10, 52.

Koufariotis, L., Hayes, B.J., Kelly, M., Burns, M., Lyons, R., Stothard, P., Chamberlain, A.J., and Moore, S., 2018. Sequencing the mosaic genome of Brahman cattle identifies historic and recent introgression including polled. Nature, 8, 17761.

Lashmar, S.F., Muchadeyi, F.C., and Visser, C., 2019. Genotype imputation as a cost-saving genomic strategy for South African Sanga cattle: A review. South African Journal of Animal Science, 49 (2).

Lashmar, S.F., Visser, C., Van Marle-Köster, E., and Muchadeyi, F.C., 2018. Genomic diversity and autozygosity within the SA Drakensberger beef cattle breed. Livestock Science, 212, 111-119.

Leigh, S.E.A., Foster, A.H., Whittall, R.A., Hubbart, C.S., and Humphries, S.E., 2008. Update and analysis of the University College London low density lipoprotein receptor familial hypercholesterolemia database. Annals of Human Genetics, 72, 485–498.

Li, Y., Willer, C., Sanna, S., and Abecasis, G., 2009. Genotype imputation. Annual Review of Genomics and Human Genetics, 10, 387-406.

Liu, C-T., Deng, X., Fisher, V., Heard-Costa, N., Xu, H., Zhou, Y., Vasan, R.S., and Cupples, L.A., 2019. Revisit population-based and family-based genotype imputation. Nature, 9, 1800.

Low, W.Y., Tearle, R., Liu, R., Koren, S., Rhie, A., Bickhart, D.M., Rosen, B.D., Kronenberg, Z.N., Kingan, S.B., Tseng, E., Thibaud-Nissen, F., Martin, F.J., Billis, K., Ghurye, J., Hastie, A.R., Lee, J., Pang, A.W.C., Heaton, M.P., Phillippy, A.M., Hiendleder, S., Smith, T.P.L., and Williams, J.L., 2020. Haplotype-resolved genomes provide insights into structural variation and gene content in Angus and Brahman cattle. Nature Communications, 11, 2071.

Luan, T., Woolliams, J.A., Ødegård, J., Dolezal, M., Roman-Ponce, S.I., Bagnato, A., and Meuwissen, T.H.E., 2012. The importance of identity-by-state information for the accuracy of genomic selection. Genetics Selection Evolution, 44, 28.

Lund, M.S., Su, G., Janss, L., Guldbrandtsen, B., and Rasmus Froberg Brøndum, R.F., 2014. Genomic evaluation of cattle in a multi-breed context. Livestock Science, 166, 101-110.

Makina, S.O., Taylor, J.F., van Marle-Köster, E., Muchadeyi, F.C., Makgahlela, M.L., MacNeil, M.D., and Maiwashe, A., 2015. Extent of linkage disequilibrium and effective population size in four South African Sanga cattle breeds. Frontier in Genetics, 6, 337.

Makina, S.O., Whitacre, L.K., Decker, J.E., Taylor, J.F., MacNeil, M.D., Scholtz, M.M., van Marle-Köster, E., Muchadeyi, F.C., Makgahlela, M.M., and Maiwashe, A., 2016. Insight into the genetic composition of South African Sanga cattle using SNP data from cattle breeds worldwide. Genetics Selection Evolution, 48, 88.

Manolio, T.A., Brooks, L.D., and Collins, F.S., 2008. A HapMap harvest of insights into the genetics of common disease. Journal of Clinical Investigation, 118, 1590–1605.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Augustine Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F.C., McCarroll, S.A., and Visscher, P.M., 2009. Finding the missing heritability of complex diseases. Nature, 461, 747–753.

Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P., 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. Nature Genetics, 39(7), 906-913.

McKay, S.D., Schnabel, R.D., Murdoch, B.M., Matukumalli, L.K., Aerts, J., Coppieters, W., Crews, D., Neto, E.D., Gill, C.A., Gao, C., Mannen, H., Stothard, P., Wang, Z., Van Tassell, C.P., Williams, J.L., Taylor, J.F., and Moore, S.S., 2007. Whole genome linkage disequilibrium maps in cattle. BMC Genetics, 8(74).

Nelson, S.C., Stilp, A.M., Papanicolaou, G.J., Taylor, K.D., Rotter, J.I., Thornton, T.A., and Laurie, C.C., 2016. Improved imputation accuracy in Hispanic/Latino populations with larger and more diverse reference panels: applications in the Hispanic community health study/study of Latinos (HCHS/SOL). Human Molecular Genetics, 25(15), 3245-3254.

Ott, J., Kamatani, Y., and Lathrop, M., 2011. Family-based designs for genome-wide association studies. Nature Reviews Genetics, 12, 465–474.

Porto-Neto, L.R., Kijas, J.W., and Reverter, A., 2014. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. Animal Genetics, 45, 180-190.

Qwabe, S.O., Van Marle-Köster, E., Maiwashe, A., and Muchadeyi, F.C., 2013. Evaluation of the BovineSNP50 genotyping array in four South African cattle populations. South African Journal of Animal Science, 43, 64-67.

Rogers, A.R., 2014. How population growth affects linkage disequilibrium. Genetics, 197, 1329-1341.

Rowan, T.N., Hoff, J.L., Crum, T.E., Taylor, J.F., Schnabel, R.D., and Decker, J.E., 2019. A multi-breed reference panel and additional rare variation maximizes imputation accuracy in cattle. Genetics Selection Evolution, 51, 77.

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.

Sargolzaei, M., Schenkel, F.S., Jansen, G.B., and Schaeffer, L.R., 2008. Extent of linkage disequilibrium in Holstein cattle in North America. Journal of Dairy Science, 91, 2106–2117.

Shi, S., Yuan, N., Yang, M., Du, Z., Wang, J.,Xin Sheng, X., Jiayan Wu, J., and Xiao, J., 2019. Comprehensive assessment of genotype imputation performance. Human Heredity, 83,107–116.

Silva, R.C., Neves, H., Hde Rezende, S., Queiroz, A., Sena, J.A.D., and Pimentel, E.C.G., 2010. Extent of linkage disequilibrium in Brazilian Gyr dairy cattle based on genotypes of AI sires for dense SNP markers. Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany.

Turner S., Armstrong L.L., Bradford Y., Carlson C.S., Crawford D.C., Crenshaw A.T., de Andrade M, Doheny K.F., Haines J.L., Hayes G., Jarvik G., Jiang L., Kullo I.J., Li R., Ling H., Manolio T.A., Matsumoto M., McCarty C.A., McDavid A.N., Mirel D.B., Paschall J.E., Pugh E.W., Rasmussen L.V., Wilke R.A., Zuvich R.L., and Ritchie M.D., 2011. Quality control procedures for genome-wide association studies. Current Protocols in Human Genetic. Chapter 1: Unit 1.19.

Ullah, E., Mall, R., Abbas, M.M., Kunji, K., Nato Jr, A.Q., Bensmail, H., Wijsman, E.M., and Saad, M., 2019. Comparison and assessment of family and population-based genotype imputation methods in large pedigrees. Genome Research, 29, 125–134.

United States Department of Agriculture., 1935. Yearbook of Agriculture, M.S. Eisenhower (ed). Government Printing Office, Washington, DC.

Wang, M.D., Dzama, K., Hefer, C.H., and Muchadeyi, F.C., 2015. Genomic population structure and prevalence of copy number variations in South African Nguni cattle. BMC Genomics, 16, 894.

Zhu, M., Zhu, B., Wang, Y.H., Wu, Y., Xu, L., Guo, L.P., Yuan, Z.R., Zhang, L.P., Gao, X., Gao, H.J., Xu, S.Z., and Li, J.Y., 2013.

Linkage disequilibrium estimation of Chinese beef Simmental cattle using high-density SNP panels. Asian Australasian Journal of Animal Sciences, 26(6), 772-779.

Zwane, A.A., Maiwashe, A., Makgahlela, M.L., Choudhury, A., Taylor, J.F., and Van Marle-Köster, E., 2016. Genome-wide identification of breed-informative single-nucleotide polymorphisms in three South African indigenous cattle breeds. South African Journal of Animal Science, 46, 302-312.

Zwane, A.A., Schnabel, R.D., Hoff, J., Choudhury, A., Makgahlela, M.L., Maiwashe, A., van Marle-Köster, E., and Taylor, J.F., 2019. Genome-wide SNP discovery in indigenous cattle breeds of South Africa. Frontiers in Genetics, 10, 273.