IMPUTATION ACCURACY MEASUREMENT AND POST-IMPUTATION QUALITY IN IMPUTED SNP GENOTYPES FOR DAIRY CATTLE

K.M. Tiplady and R.G. Sherlock

LIC, Private Bag 3016, Hamilton 3240, New Zealand

SUMMARY

Imputation of genotypes is a cost-effective method for generating genotypes for un-typed loci and allows data from different genotyping panels and platforms to be combined. Accuracy of imputation can be defined in a number of ways to distinguish well-imputed from poorly-imputed SNP. The aims of this study were to compare different measures of imputation accuracy in low density panel data and determine how well the estimated allelic R^2 (AR^2) measure reported by BEAGLE performs across minor allele frequency (MAF) as a post-imputation filtering tool. Genotypes for 28,793 New Zealand mixed-breed dairy cows from a low density BeadChip (n=16,512 SNP) were used in the study. For 17,593 animals, 9,166 SNP were masked and imputed using version 4.0 of BEAGLE software. Imputation accuracy for SNP with MAF $\geq$ 0.005 was high, but was variable for low MAF (< 0.005) SNP. Genotypic concordance was not informative for low MAF SNP and was poorly correlated with AR^2 for low MAF SNP. Other imputation accuracy measures (genotypic correlation, minor allele sensitivity and imputation quality score) were informative for low MAF SNP and were highly correlated with AR^2 across all MAF classifications (r > 0.81). Results showed that post-imputation filtering based on AR^2 is an effective approach for removing poorly-imputed SNP, including those of low MAF.

INTRODUCTION

Genotype imputation increases the power of existing data by providing predicted genotypes for loci that have not been directly assayed. It allows data from different genotyping platforms to be combined and makes additional variants available for analysis without the cost of actually genotyping them. Compared to using a smaller set of only true genotypes, the additional power from imputed genotypes can provide better signal in genome wide association studies (Khatkar *et al.* 2013) and better estimates of direct genetic values (Khatkar *et al.* 2012; Weigel *et al.* 2010). However, incorrectly imputed genotypes can add noise and compromise an analysis (Weigel *et al.* 2010; Chen *et al.* 2014). Imputation correctness has been evaluated based on a number of accuracy metrics in previous studies (Khatkar *et al.* 2013; Calus *et al.* 2014), each providing a different way to distinguish well-imputed from poorly-imputed SNP. This differentiation can be particularly problematic for low minor allele frequency (MAF) SNP where accurate imputation is more difficult and sensitive to genotype calling errors (Lin *et al.* 2010; Calus *et al.* 2014). Also, some measures of accuracy are highly dependent on MAF and can give misleading results for low MAF SNP (Lin *et al.* 2010; Hickey *et al.* 2012). In this study, accuracy of imputation was examined for genotypes from New Zealand (NZ) progeny test dairy herds which were genotyped on a custom GGP-LD BeadChip. Imputing low MAF SNP well is important within this context because these custom SNP chip panels are often updated with new loci, many of which are low MAF, with a requirement for these to be imputed through the historically genotyped population. Generating imputation accuracy metrics requires a comparison set of true and imputed genotypes, and this is often obtained by selecting a subset of animals as a validation set. For this validation subset, a set of SNP of interest are masked and then imputed. In practical applications of imputation where a complete "truth set" is unavailable, pedigree relationships can sometimes be used to infer true genotypes and the level of imputation accuracy. However, a generally-available post-imputation quality measure which is not dependent on having a "truth set" and is reliable across MAF is desirable. Browning and Browning (2009) outline a post-imputation estimate of imputation accuracy, the estimated allelic R^2 (AR^2) which is not dependent on allele frequency or having a "truth set" of genotypes. The aims of this study were to compare the AR^2 reported by BEAGLE (Browning and Browning 2009) to a number of different imputation accuracy metrics derived from comparing true with imputed genotypes, and determine how well the AR^2 performs across MAF as a post-imputation filtering tool.

MATERIALS AND METHODS

Genotypes from New Zealand (NZ) progeny test dairy herds (Holstein-Friesian, Jersey and crossbreed) were obtained from a custom version of the GGP-LD BeadChip with 20,183 SNP. After removing animals with a call rate < 0.95 and any SNP that were non-autosomal or had a call rate < 0.9, 19,143 SNP for each of 28,793 animals were included in the study.

Imputation reference. Reference animals were selected as those with progeny in the wider population (11,062 females; 138 males). Average pedigree relationships between reference animals were 0.034 (sd=0.031). Monomorphic SNP were removed and missing SNP were imputed using version 4.0 of BEAGLE (Browning and Browning 2009) with default parameters. This resulted in an imputation reference of 16,512 SNP for 11,200 animals.

Imputation target. Genotypes for 17,593 animals not included in the imputation reference were included in the imputation target population. Of the target population, 38.4% had at least 1 parent in the reference, and the average pedigree relationship between reference and target animals was 0.033 (sd=0.029). Of the 16,512 SNP in the imputation reference, 9,166 were masked to leave only the SNP in common with an earlier version of the GGP-LD BeadChip. Imputation was carried out using version 4.0 of BEAGLE with default parameters. True and imputed genotypes were compared for 9,166 masked SNP on 17,593 animals.

Imputation accuracy. Imputation accuracy was assessed according to 4 measures: Genotypic concordance (GCONC; proportion of genotype calls where the true genotype matches the most likely imputed genotype), genotypic correlation (GCORR; correlation between observed and imputed number of copies of the alternate allele), minor allele sensitivity (MAS; proportion of times a minor allele is correctly called when it is present, analogous to non-reference sensitivity) and imputation quality score (IQS; concordance adjusted for chance agreement) as defined by Lin *et al.* (2010).

Post-imputation quality. Post-imputation quality was assessed using the AR^2 calculated by BEAGLE. This is an estimate of the squared correlation between the allele dosage of the most likely imputed genotype and the allele dosage of the true genotype. The true genotype is unknown but the allelic R^2 is estimated from the distribution of imputed posterior genotype probabilities.

MAF classifications. SNP were grouped by frequency of the minor allele in the reference.

RESULTS AND DISCUSSION

Table 1 summarises imputation accuracy as measured by GCONC, GCORR, MAS and IQS, and the AR^2 reported by BEAGLE. For SNP with MAF < 0.005, GCORR, MAS and IQS all indicated measures of accuracy ≤ 0.462, whereas GCONC indicated a high accuracy (0.999). Also, a decrease in GCONC was observed with increasing MAF, but an increase in accuracy was observed when measured by GCORR, MAS and IQS. This is because GCONC is dependent on MAF, and demonstrates that measuring accuracy based on GCONC can be misleading for low MAF SNP, as outlined by Calus *et al.* (2014). Mean AR^2 values also increased with MAF and were particularly low (0.188) for SNP with MAF < 0.005. Imputation accuracy levels were high (≥ 0.864) when MAF ≥ 0.005 based on all 4 measures considered in this study. The MAF at which SNP are accurately imputed would be expected to increase as the size of the imputation reference decreases.

Table 1. Mean imputation accuracy (GCONC, GCORR, MAS, IQS) and post-imputation quality (AR^2) for SNP classified by MAF.

MAF classification	N	GCONC	GCORR	MAS	IQS	AR^2
< 0.005	1218	0.999	0.462	0.198	0.217	0.188
0.005-0.01	130	0.998	0.906	0.864	0.887	0.774
0.01-0.05	557	0.995	0.951	0.930	0.947	0.873
≥ 0.05	7261	0.974	0.965	0.974	0.951	0.910
All	9166	0.979	0.928	0.867	0.852	0.810

Correlations between imputation accuracy measures and AR^2 are shown in Table 2. GCONC was poorly correlated with AR^2 for SNP with MAF < 0.005. Other imputation accuracy measures (GCORR, MAS, IQS) were highly correlated (≥ 0.812) with AR^2 across all minor allele frequencies. High correlations between these accuracy measures and AR^2 suggest that AR^2 may be a good tool for screening SNP post-imputation.

Table 2. Correlations between AR^2 and imputation accuracy (GCONC, GCORR, MAS, IQS) classified by MAF.

MAF classification	N	GCONC	GCORR	MAS	IQS
< 0.005	1218	-0.069	0.851	0.908	0.903
0.005-0.01	130	0.763	0.824	0.852	0.900
0.01-0.05	557	0.644	0.837	0.846	0.882
≥ 0.05	7261	0.888	0.925	0.812	0.947
All	9166	-0.077	0.927	0.972	0.974

Figure 1 shows the distribution of GCORR values prior to and post filtering based on an AR^2 threshold of 0.7. Prior to filtering imputed genotypes, GCORR values were highly variable, in particular for SNPs with MAF < 0.005 (Figure 1a). After filtering, the variation in GCORR values was significantly reduced, particularly for SNP with MAF < 0.005 (Figure 1b). In total, 1191 SNP were removed, most of which were SNP with MAF < 0.005. Results for MAS and IQS were similar (not presented here).

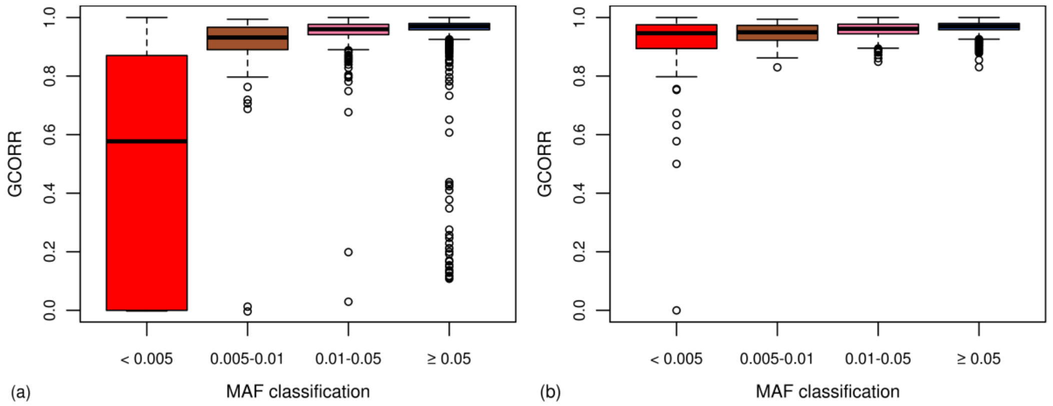

Figure 1. Distribution of genotypic correlation (GCORR) (a) prior to filtering and (b) post filtering based on an AR^2 threshold of 0.7.

Browning and Browning (2009) demonstrated that at high SNP density, AR^2 is a good metric for estimating imputation accuracy without dependence on allele frequency. Kelly *et al.* (2013) also showed that in a population of composite tropical cattle, AR^2 was an effective measure for identifying a large number of poorly-imputed SNP when imputing from Illumina BovineSNP50 to Illumina BovineHD SNP panels. Table 3 summarises mis-classifications of SNP in this study that resulted when a post-imputation filter of $AR^2 > 0.7$ was used to predict SNP that had been imputed well according to each of the accuracy measures GCORR, MAS and IQS. For each measure, well-imputed SNP are defined as those where the measure was > 0.7. False negative (FN) SNP were defined as those with an $AR^2 \leq 0.7$ but an imputation accuracy > 0.7. False positive (FP) SNP were defined as those with an $AR^2 > 0.7$ but an imputation accuracy ≤ 0.7. Low FN rates (≤ 3.77%) were observed for SNP with MAF ≥ 0.01, but were higher for SNP with MAF < 0.01 (5.83-22.88%). Very low FP rates (≤ 0.9%) were observed for SNP with MAF < 0.005 and were all zero for SNP with MAF ≥ 0.005. These results confirm that post-imputation filtering based on AR^2 is an effective approach for removing poorly-imputed SNP, including those of low MAF.

Table 3. Percentage of false positive (FP) and false negative (FN) SNP for imputation accuracy measures (GCORR, MAS, IQS) based on an AR^2 threshold of 0.7.

	GCORR		MAS		IQS	
MAF classification	FN	FP	FN	FP	FN	FP
< 0.005	22.88	0.85	5.83	0.90	9.52	0.41
0.005-0.01	20.93	0	17.69	0	16.15	0
0.01-0.05	3.77	0	3.41	0	3.41	0
≥ 0.05	0.25	0	0.36	0	0.15	0
All	2.35	0.06	1.52	0.12	1.82	0.05

CONCLUSION

Genotypic concordance was not informative for low MAF SNP and was poorly correlated with AR^2 values reported by BEAGLE for low MAF SNP. Other imputation accuracy measures examined (GCORR, MAS, IQS) were informative for SNP across all minor allele frequencies. These measures were also highly correlated with each other and with post-imputation AR^2 values. Post-imputation filtering based on an AR^2 threshold of 0.7 was shown to be an effective way of removing poorly-imputed SNP for imputed genotypes from a population of NZ dairy cows genotyped on a low density panel.

REFERENCES

Browning B.L. and Browning S.R. (2009) *Am J Hum Genet* **84**: 210.

Calus M.P., Bouwman A.C., Hickey J.M., Veerkamp R.F. and Mulder H.A. (2014) *Animal* **8**: 1743.

Chen L., Li C., Sargolzaei M. and Schenkel F. (2014) *PLoS ONE,* **9(7)**: e101544.

Hickey J.M., Crossa J., Babu R. and de los Campos G. (2012) *Crop Science* **52:** 654.

Kelly M., Fortes M.R.S. and Moore S.S. (2013) *Proc. Assoc. Advmt. Anim. Breed Genet.* **20**: 550-553.

Khatkar M.S., Moser G., Hayes B.J. and Raadsma H.W. (2012) *BMC Genomics* **13**: 538.

Khatkar M.S., Thomson P.C. and Raadsma H.W. (2013) *Proc. Assoc. Advmt. Anim. Breed Genet.* **20**: 554-557.

Lin P., Hartz S. M., Zhang Z., Saccone S.F., Wang J., *et al.* (2010) *PLoS ONE* **5(3)**: e9697.

Weigel K.A., de Los Campos G. and Vazquez A., *et al.* (2010) *J. Dairy Sci.* **93**: 5423.