Animal Breeding and Genetics

J. Anim. Breed. Genet. ISSN 0931-2668

ORIGINAL ARTICLE

Within- and across-breed imputation of high-density genotypes in dairy and beef cattle from medium- and low-density genotypes

D.P. Berry[1], M.C. McClure[2] & M.P. Mullen[3]

1 Animal & Grassland Research and Innovation Centre, Cork, Ireland
2 Irish Cattle Breeding Federation, Cork, Ireland
3 Animal & Grassland Research and Innovation Centre, Galway, Ireland

Keywords

Illumina; Beagle; impute; genotype; genomic selection.

Correspondence

D.P. Berry, Animal & Grassland Research and Innovation Center, Teagasc, Moorepark, Fermoy, Co. Cork, Ireland.
Tel: +353 25 42386;
Fax: +353 25 42340;
E-mail: donagh.berry@teagasc.ie

Received: 14 May 2013;
accepted: 5 November 2013

Summary

The objective of this study was to evaluate, using three different genotype density panels, the accuracy of imputation from lower- to higher-density genotypes in dairy and beef cattle. High-density genotypes consisting of 777 962 single-nucleotide polymorphisms (SNP) were available on 3122 animals comprised of 269, 196, 710, 234, 719, 730 and 264 Angus, Belgian Blue, Charolais, Hereford, Holstein-Friesian, Limousin and Simmental bulls, respectively. Three different genotype densities were generated: low density (LD; 6501 autosomal SNPs), medium density (50K; 47 770 autosomal SNPs) and high density (HD; 735 151 autosomal SNPs). Imputation from lower- to higher-density genotype platforms was undertaken within and across breeds exploiting population-wide linkage disequilibrium. The mean allele concordance rate per breed from LD to HD when undertaken using a single breed or multiple breed reference population varied from 0.956 to 0.974 and from 0.947 to 0.967, respectively. The mean allele concordance rate per breed from 50K to HD when undertaken using a single breed or multiple breed reference population varied from 0.987 to 0.994 and from 0.987 to 0.993, respectively. The accuracy of imputation was generally greater when the reference population was solely comprised of the breed to be imputed compared to when the reference population comprised of multiple breeds, although the impact was less when imputing from 50K to HD compared to imputing from LD.

Introduction

Genomic selection (Meuwissen *et al.* 2001) exploiting genome-wide information on a large population of animals is the method of genetic evaluations in many dairy (Hayes *et al.* 2009) and some beef (Saatchi *et al.* 2011, 2012) populations. The accuracy of the genomic predictions is a function of the size of the population of animals with both phenotypes and genotypes. Greater prediction accuracy is achievable with larger reference populations (Daetwyler *et al.* 2008), although the relationships among its individuals and between the reference population and candidate individuals are also important (Pszczola *et al.* 2012). There is nonetheless a cost to genotyping large populations of animals especially for higher-density genotypes. This cost, however, could be reduced by genotyping using a lower-density (i.e. lower cost) genotype panel and imputing to a higher density. Imputation nonetheless still requires a population of animals genotyped on the higher-density genotype panel. Imputation has been documented to be accurate within dairy (Weigel *et al.* 2009; Berry & Kearney 2011; Mulder *et al.* 2012) and beef cattle

  doi:10.1111/jbg.12067

(Dassonneville *et al.* 2012; Huang *et al.* 2012). The aforementioned studies, however, have primarily evaluated imputation from low- to medium-density genotype panels, although a few studies on imputation to high-density genotype panels also exist (Erbe *et al.* 2012; Pausch *et al.* 2013; VanRaden *et al.* 2013). The cost of acquiring higher-density genotypes could potentially be further reduced if the reference population of animals genotyped on the higher density could be generated from multiple breeds. Nevertheless, there is little information on the usefulness of across-breed imputation in cattle (Brøndum *et al.* 2012), especially genetically diverse breeds such as beef and dairy breeds.

The objective therefore of the present study was to evaluate the accuracy of imputation from lower-density genotyping panels to higher-density genotyping panels in dairy and beef cattle using a single breed reference population or multibreed reference population. The results from this study will be useful in determining the accuracy of imputation to very high-density panels (>700 000 single-nucleotide polymorphisms) as well as the contribution to accuracy of imputation by exploiting high-density genotype information from multiple breeds.

Materials and methods

Genotype data

Illumina (http://www.illumina.com) high-density (HD) genotypes (777 962 single-nucleotide polymorphisms; SNP) were available on 3122 dairy and beef bulls. The SNP positions were based on the UMD 3.1 genome build (Zimin *et al.* 2009). The number of bulls per breed was 269, 196, 710, 234, 719, 730, and 264 for Aberdeen Angus, Belgian Blue, Charolais, Hereford, Holstein-Friesian, Limousin and Simmental, respectively. Mendelian inconsistencies were used to validate animal identification through parentage assessment but also to discard autosomal SNPs that did not adhere to Mendelian inheritance. Only autosomal SNPs with a known UMD 3.1 genomic location were retained for this analysis.

As well as the HD panel described above, two alternative SNP density panels were generated to represent the Illumina Bovine50 BeadChip (50K; Matukumalli *et al.* 2009) and Illumina low-density (LD; Boichard *et al.* 2012) genotyping panel. A total of 47 770 of the autosomal SNPs on both the HD panel and 50K genotyping were retained. A total of 6501 of the autosomal SNPs on both the HD and LD panels were retained. The 6501 autosomal SNPs were also in common between the LD and 50K panel.

Imputation

Animals were partitioned into either a reference or a validation population to test the accuracy of imputation. All animals, irrespective of breed, born after 2005 (n = 698) were assumed to represent the validation bulls; all other bulls were included in the reference population. Three separate analyses were undertaken as follows: (i) imputation from LD to HD, (ii) imputation from 50K to HD and (iii) imputation of LD to 50K. In all analyses, the full complement of higher-density genotypes was retained in the reference animals. Genotypes were masked in the validation animals to represent the lower-density panel.

Imputation to the higher-density genotypes was undertaken for each chromosome separately using the freely available software Beagle version 3.1.0 (Browning & Browning 2007, 2009); Beagle exploits population-wide linkage disequilibrium in the imputation process. The default of ten iterations was used in all scenarios. Imputation was undertaken within and across breeds. In all analyses, the same animals were included in the validation population. However, when the analysis was within breed, only the animals of that breed were included in the reference population. Furthermore, the accuracy of imputation of a single breed validation population was also calculated using simply the modal genotype (i.e. the most frequent genotype) of the single breed reference population.

Several statistics were calculated to compare the accuracy of imputation in the validation population of animals: (i) genotype concordance rate defined as the average proportion of correctly imputed genotypes within SNP or within animal, (ii) allele concordance rate defined as the average proportion of correctly imputed alleles within SNP or within animal; in this instance, a genotype imputed to be heterozygote but was truly homozygote was assumed to have one correct allele imputed and (iii) the correlation between the actual and imputed genotypes. In all instances, the accuracy of imputation was calculated by including also the actual genotypes used in the imputation process. This was to generate results that are therefore applicable in the real-life situation.

Results and discussion

The number of SNPs per chromosome used in the three alternative genotype density panels is displayed

Table 1 Number of single-nucleotide polymorphisms for the high-density (HD), medium-density (50K) and low-density (LD) genotyping panels for each chromosome (BTA)

BTA	HD	50k	LD	BTA	HD	50k	LD
1	46487	3126	391	16	24173	1538	205
2	40050	2548	340	17	22263	1440	188
3	35568	2272	305	18	19383	1246	175
4	34974	2353	302	19	18903	1270	178
5	34834	2044	300	20	21486	1404	204
6	35513	2371	306	21	21171	1311	183
7	33162	2137	281	22	18030	1190	164
8	33523	2177	293	23	15212	973	148
9	31056	1897	271	24	18616	1206	175
10	30443	1971	264	25	12928	902	134
11	32010	2053	274	26	15239	1009	145
12	26122	1597	225	27	13148	892	137
13	23590	1662	211	28	13034	885	126
14	24775	1683	219	29	14707	963	133
15	24751	1580	224				

in Table 1. Summary statistics for the accuracy of within-breed and across-breed imputation across the different genotyping platforms are summarized in Tables 2 and 3, respectively. The allele concordance rate was always greater than the genotype concordance rate; the difference between the two statistics per individual varied from 0.001 to 0.124 (within-breed imputation) and 0.001 to 0.139 (across-breed imputation). Furthermore, the variation among individuals in mean imputation accuracy was lower for the allele concordance rate. Berry & Kearney (2011) when imputing from 2730 SNPs to the 50K panel in Holstein-Friesian dairy cattle also documented a greater average, and lower variation, in allele concordance rate compared to genotype concordance rate. Allele concordance rate is arguably more informative for evaluating the accuracy of imputation for use in genomic selection as most genomic selection algorithms assume only additive allele effects. Hickey *et al.* (2012), however, strongly advocated the use of the correlation between actual and imputed genotypes as the best measure of imputation accuracy as it was less sensitive to minor allele frequency compared to concordance rates. The correlation between the true and imputed genotypes was in between the allele and genotype concordance rates. The mean accuracy of imputation of heterozygous SNPs was lower than the mean imputation accuracy of homozygous SNPs (Table 4). Furthermore, the accuracy of imputation (irrespective of statistic used) declined as the minor allele frequency increased (Figure 1) which is consistent with other studies (Berry & Kearney 2011). There was no difference in mean imputation accuracy at the ends of each chromosome (i.e. peripheral 50 SNPs) compared to the rest of the chromosome.

The mean accuracy of imputation (for all three accuracy statistics) per chromosome was similar, although variation in imputation accuracy did exist across the genome. Mean allele concordance rate per SNP for the across-breed imputation from LD to HD is in Figure 2; mean allele concordance rate per SNP for the across-breed imputation from 50K to HD is in Figure 3. Concordance rate by genomic location for the within-breed analysis was very similar. Several obvious genomic regions existed where imputation accuracy was low and was relatively consistent irrespective of whether the LD or 50K was the lower-density panel. The low accuracy of imputation in such regions could be due to a multitude of factors including (i) incorrect annotation of the genomic position of the SNP, (ii) recombination hot spots located in the vicinity, (iii) incorrect genotype calling during the laboratory genotyping process and (iv) a greater level of heterozygosity in these regions which subsequently

Table 2 Genotype and allele concordance rate (standard deviation in parenthesis) as well as the correlation (*r*) between the actual and imputed genotypes for alterative scenarios of within-breed imputation across low-density (LD), medium-density (50K) and high-density (HD) genotyping panels in each of the seven different breeds. Also included is the number of animals included in the reference (*R*) and validation (*V*) population

			LD to 50K			LD to HD			50K to HD		
Breed	*R*	*V*	Genotype	Allele	*r*	Genotype	Allele	*r*	Genotype	Allele	*r*
AA	195	74	0.950 (0.029)	0.974 (0.015)	0.962	0.936 (0.035)	0.967 (0.018)	0.951	0.983 (0.013)	0.992 (0.006)	0.988
BB	140	56	0.935 (0.030)	0.967 (0.016)	0.950	0.916 (0.040)	0.956 (0.021)	0.933	0.974 (0.014)	0.987 (0.007)	0.980
CH	526	184	0.953 (0.028)	0.976 (0.015)	0.964	0.948 (0.032)	0.973 (0.017)	0.960	0.987 (0.01)	0.994 (0.005)	0.990
HE	189	45	0.961 (0.022)	0.980 (0.012)	0.970	0.949 (0.029)	0.974 (0.015)	0.960	0.988 (0.009)	0.994 (0.005)	0.991
HF	688	31	0.929 (0.055)	0.963 (0.030)	0.943	0.922 (0.066)	0.959 (0.036)	0.937	0.977 (0.026)	0.988 (0.013)	0.982
LM	506	224	0.946 (0.026)	0.973 (0.013)	0.959	0.942 (0.030)	0.970 (0.016)	0.955	0.986 (0.008)	0.993 (0.004)	0.989
SI	180	84	0.935 (0.034)	0.967 (0.018)	0.951	0.923 (0.043)	0.960 (0.023)	0.940	0.977 (0.015)	0.988 (0.008)	0.983

AA = Aberdeen Angus; BB = Belgian Blue; CH = Charolais; HE = Hereford; HF = Holstein-Friesian; LM = Limousin; SI = Simmental.

Table 3 Genotype and allele concordance rate (standard deviation in parenthesis) as well as the correlation (*r*) between the actual and imputed genotypes for alterative scenarios of across-breed imputation across low-density (LD), medium-density (50K) and high-density (HD) genotyping panels in the entire data set or in different breeds. Also included is the number of animals (N) included in the validation population; the number of animals in the reference population was always 2424

		LD to 50K			LD to HD			50K to HD		
Breed	N	Genotype	Allele	*r*	Genotype	Allele	*r*	Genotype	Allele	*r*
All	698	0.925 (0.038)	0.961 (0.021)	0.942	0.922 (0.045)	0.960 (0.024)	0.938	0.982 (0.012)	0.991 (0.006)	0.987
AA	74	0.925 (0.037)	0.961 (0.020)	0.942	0.919 (0.036)	0.958 (0.019)	0.940	0.984 (0.010)	0.992 (0.005)	0.988
BB	56	0.912 (0.037)	0.954 (0.02)	0.931	0.890 (0.041)	0.940 (0.021)	0.918	0.974 (0.013)	0.987 (0.007)	0.981
CH	184	0.937 (0.033)	0.968 (0.017)	0.952	0.927 (0.034)	0.961 (0.018)	0.948	0.986 (0.009)	0.993 (0.005)	0.990
HE	45	0.935 (0.033)	0.966 (0.018)	0.949	0.928 (0.031)	0.961 (0.017)	0.949	0.987 (0.008)	0.993 (0.004)	0.990
HF	31	0.912 (0.062)	0.954 (0.034)	0.928	0.894 (0.072)	0.942 (0.039)	0.920	0.975 (0.026)	0.987 (0.013)	0.981
LM	224	0.926 (0.033)	0.962 (0.018)	0.943	0.919 (0.033)	0.956 (0.018)	0.941	0.983 (0.009)	0.991 (0.004)	0.987
SI	84	0.905 (0.044)	0.951 (0.023)	0.927	0.901 (0.053)	0.949 (0.028)	0.922	0.974 (0.016)	0.987 (0.008)	0.981

AA = Aberdeen Angus; BB = Belgian Blue; CH = Charolais; HE = Hereford; HF = Holstein-Friesian; LM = Limousin; SI = Simmental.

affect imputation accuracy (Table 4). Variation across the bovine genome in imputation accuracy has been reported elsewhere (Erbe *et al.* 2012; Pausch *et al.* 2013). The location of most genomic regions that deemed to be poorly imputed in the present study was very similar to the regions documented by Erbe *et al.* (2012) who imputed from 50K to HD in a population of Holstein-Friesian and Jersey dairy cattle and Pausch *et al.* (2013) in a population of 797 Fleckvieh animals. Erbe *et al.* (2012) reported that 1231 of the HD SNPs in their population had a genotype concordance rate of <0.80, while the equivalent statistic in the present study when evaluating the accuracy of across-breed imputation from 50K to HD was 2234 SNPs. Pausch *et al.* (2013) analysing linkage disequilibrium patterns in high-density genotypes of Fleckvieh cattle suggested that 5039 SNPs may actually be incorrectly positioned on the Illumina HD SNP manifest. Although, on average, the imputation accuracy across all SNPs was very good in the present study, reduced imputation accuracy for individual SNPs could have serious implications for genomic selection or genome-wide association algorithms if these genomic regions harbour polymorphisms with large effects on the phenotype(s) under investigation.

Accuracy of imputation across different genotype densities

Irrespective of whether using allele concordance rate, genotype concordance rate or the correlation between actual or imputed genotype to depict accuracy, the accuracy of imputation was, on average, greatest when imputing from 50K to HD and was poorest when imputing from LD to HD (Tables 2 and 3). The minor allele frequency for the LD, 50K and HD genotype panel across all animals in the present study was 0.39, 0.24 and 0.25, respectively. On an individual animal basis, the mean accuracy of imputation from 50K to HD was always superior to the mean individual accuracy of imputation from LD to either 50K or HD. The same conclusion was evident irrespective of whether the imputation was undertaken within or across breed. The accuracy of imputation of HD genotypes was up to 20 percentage units better for some individual animals when imputing from 50K genotypes compared to imputing from LD. Although, on average, the accuracy of imputation from LD to 50K was slightly better than the accuracy of imputation from LD to HD, the individual animal imputation accuracy from LD to HD was better than the individual animal imputation accuracy from LD to 50K for 18% (within-breed imputation) to 40% (across-breed imputation) of the individuals.

Table 4 Proportion of genotypes correctly imputed across the different genotype platform imputation scenarios when the true genotype is homozygous or heterozygous and the imputation is undertaken within breed (Within) or across breeds (Across)

	Homozygotes		Heterozygotes	
Genotype platforms	Within	Across	Within	Across
LD to 50K	0.962	0.944	0.907	0.879
LD to HD	0.955	0.939	0.900	0.882
50K to HD	0.989	0.987	0.972	0.972

The imputation accuracies in the present study are consistent with those documented in most other populations of dairy (Weigel *et al.* 2009; Berry & Kearney 2011; Erbe *et al.* 2012) and beef cattle (Dassonneville *et al.* 2012) using a range of different genotype densities and imputation algorithms. To our knowledge,

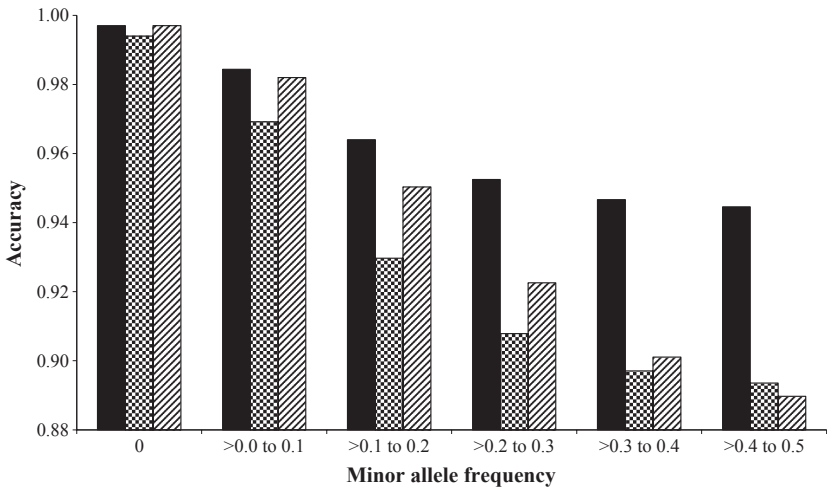

Figure 1 Accuracy of imputation from low- to high-density genotypes represented by allele concordance rate (shaded bars), genotype concordance rate (checked bars) and correlation between actual and imputed genotypes (striped bars) across different minor allele frequency categories.

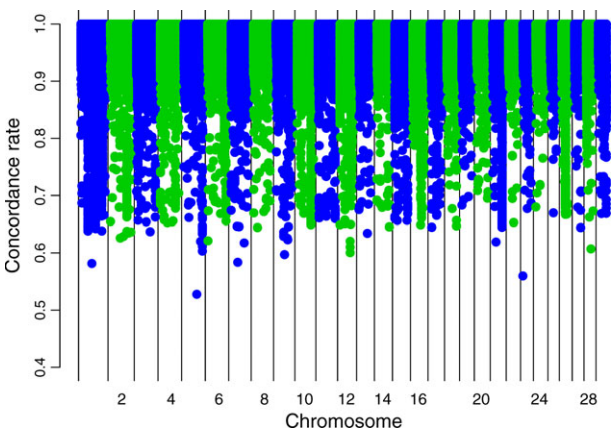

Figure 2 Individual single-nucleotide polymorphism mean allele concordance rate when imputing from the low-density to high-density genotype platform using a multiple breed reference population.

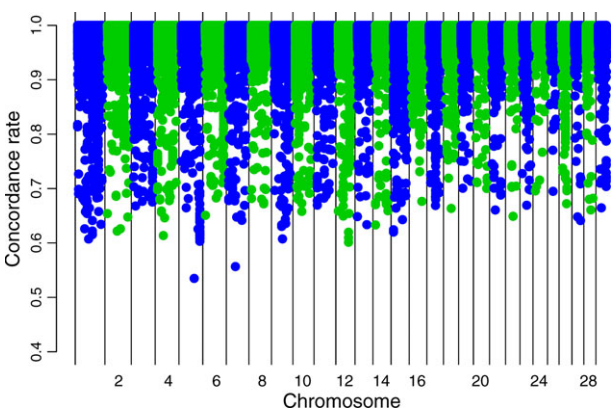

Figure 3 Individual single-nucleotide polymorphism mean allele concordance rate when imputing from the medium-density to high-density genotype platform using a multiple breed reference population.

the present study is the only study that compared, using exactly the same population, the imputation accuracy across these three commercially available genotype platforms. Although difficult to compare studies because of differences in population structure (e.g. relationships between reference and validation animals) and study design (e.g. reference population size and SNPs evaluated), comparing studies that imputed from 50K to HD (Brøndum *et al.* 2012; Erbe *et al.* 2012; Khatkar *et al.* 2012) to studies that imputed from low density to 50K (Weigel *et al.* 2009; Berry & Kearney 2011) suggest that, corroborating the present study, the accuracy of imputation for the latter scenario was lower.

Irrespective of whether imputation was undertaken within breed or across breed, the proportion of correlated imputed homozygous SNPs was always poorest when imputing from LD to HD and was always greatest when imputing from 50K to HD (Table 4).

Accuracy of imputation within or across breeds

Mean imputation accuracy per breed was always superior when undertaken within breed compared to undertaken across breed with the exception of the 50K to HD imputation scenario when undertaken in Angus and Belgian Blue cattle, although the difference was minuscule; allele concordance rate was 0.0003–0.0004 superior for the Angus and Belgian Blue animals, respectively, when undertaken across breed. A contributing factor to the different trend in Angus animals may be due to smaller-sized population included in the analyses, although small populations were also used for the Hereford and Simmental.

The Belgian Blue is a relatively recent breed formed from the crossing of local Belgian breeds with the Shorthorn breeds. It is therefore likely that haplotypes present in other breeds included in the present study may still exist within the Belgian Blue breed, and thus, imputation in the Belgian Blue (or other such breeds and composites) may indeed benefit from across-breed imputation. Moreover, on an individual animal basis, mean imputation accuracy when imputing from LD to 50K was always superior when undertaken within breed except for two animals (i.e. 0.3% of the data); when imputing from LD to HD, individual animal imputation accuracy was always superior when undertaken within breed compared to across breed for all animals with the exception of four animals. Although, on average, imputation accuracy from 50K to HD was superior when undertaken within breed, for 14% of the validation animals, the opposite was true.

These results therefore suggest that in this population at least and in the scenarios investigated (including the imputation algorithms used), there is no benefit of exploiting higher-density genotypes from multiple breeds for imputation across breeds. This is likely due to a lack of linkage disequilibrium phase between SNPs across breeds, and this hypothesis was substantiated here by the difference between across breed and within breed being largest from LD to 50K or HD but almost negligible when imputing from 50K to HD. Imputation algorithms require that the haplotype to be imputed from the lower density also exists in the higher-density panel. Haplotypes, especially over a larger genomic region, are likely to differ between many breeds. The linkage disequilibrium phase among breeds between adjacent SNPs in the 50K is likely to be greater than between SNPs on the LD because of the greater marker density in the former. This therefore suggests that there may indeed be some benefit of across-breed imputation from HD to sequence data, especially if only a few animals are sequenced and with low genome coverage.

Despite the differences in reference population sizes of the breeds, there was no obvious breed differences in mean imputation accuracy when imputation was undertaken within or across breeds; the reference population size of the Holstein-Friesian population was 688 compared to 140 for Belgian Blues. When the reference population included a single breed, the mean genotype and allele concordance rate per individual in the other breeds, when imputing from LD to HD, was 0.617 and 0.777, respectively; if the imputed genotypes were simply assumed to equal the modal genotype of the reference population, the respective values were 0.592 and 0.768. No obvious imputation accuracy difference existed when alternative breeds in the reference and validation population were evaluated. Therefore, poor imputation accuracy is achieved if the breeds to be imputed are not included in the reference population with the higher-density genotypes.

Few studies have evaluated the usefulness of multiple cattle breeds in the imputation process, and to our knowledge, no study has included both dairy and beef breeds in the reference population for imputation. Brøndum *et al.* (2012) concluded that combining the three populations of Danish, Swedish and Finnish Red dairy cattle improved the accuracy of imputation. Including Holstein-Friesian animals in the reference population only increased the accuracy of imputation in the Danish Red population. Imputation accuracy, however, in the Holstein population was reduced when the reference population included the Red breeds. The Red breeds, nevertheless, are unlikely to be genetically very diverse. Using a relatively small population of four different sheep breeds (Border Leicester, White Faced Suffolk, Poll Dorset and Merino), Hayes *et al.* (2011) reported a lower accuracy of imputation from low-density (1000–5000 SNPs) to medium-density (48 640 SNPs) genotypes when the reference population comprised of multiple breeds compared to a reference population with just the breed of animals being imputed.

Impact of relationships between reference and validation animals

The accuracy of imputation is known to be influenced by the relationships between the reference animals and the validation animals (Berry & Kearney 2011). When imputation was undertaken across breeds, having the genotype of the sire only (not the dam or maternal grandsire) in the reference data set increased the genotype concordance rate by 0.049, 0.042 and 0.010 units when imputing from LD to HD, LD to 50K and 50K to LD, respectively, compared to when neither the sire nor maternal grandsire genotypes were in the reference population. Relative to when just the sire genotype was included in the reference population, having also the genotype of the maternal grandsire increased the genotype concordance rate by 0.020, 0.017 and 0.005 when imputing from LD to HD, LD to 50K and 50K to LD, respectively. These results therefore corroborate other studies that emphasized the importance of having the higher-density genotypes on back-pedigree to improve the accuracy of imputation (Druet *et al.* 2010; Zhang & Druet 2010; Berry

& Kearney 2011). The present study, however, indicates that the influence of genotyped back-pedigree diminishes as the SNP density of lower-density genotype panel increases (i.e. LD to 50K).

Conclusions

Imputation accuracy from the medium-density genotype panel (50K) to the HD panel was superior to that of imputation from lower-density genotype panels. On average, the accuracy of imputation was very high. There was, on average, no benefit in imputation accuracy from exploiting a multibreed reference population, and in most instances, the accuracy of imputation was reduced when imputation was undertaken using multiple breeds as opposed to a single breed in the reference population. This is likely due to a lack of consistent linkage disequilibrium phases between SNPs across different breeds.

Acknowledgements

Financial support of the Irish Department of Agriculture Research Stimulus Fund (RSF-06-0353; RSF-06-0428; 11/SF/311), Science Foundation Ireland (09/IN.1/B2642) and the Irish dairy and beef industry are gratefully acknowledged.

References

Berry D.P., Kearney J.F. (2011) Imputation of genotypes from low-to high-density genotyping platforms and implications for genomic selection. *Anim*, **5**, 1162–1169.

Boichard D., Chung H., Dassonneville R., David X., Eggen A., Fritz S., Gietzen K.J., Hayes B.J., Lawley C.T., Sonstegard T.S., Van Tassell C.P., VanRaden P.M., Viaud-Martinez K.A., Wiggans G.R., the Bovine LD Consortium. (2012) Design of a bovine low-density SNP array optimised for imputation. *PLoS One*, **7**, e34130.

Brøndum R.F., Ma P., Lund M.S., Su G. (2012) Genotype imputation within and across Nordic cattle breeds. *J. Dairy Sci.*, **95**, 6795–6800.

Browning S.R., Browning B.L. (2007) Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. *Am. J. Hum. Genet.*, **81**, 1084–1097.

Browning B.L., Browning S.R. (2009) A unified approach to genotype imputation and haplotype phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.*, **84**, 210–223.

Daetwyler H.D., Villanueva B., Woolliams J.A. (2008) Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One*, **3**, e3395.

Dassonneville R., Fritz S., Ducrocq V., Boichard D. (2012) Imputation performance of 3 low-density marker panels in beef and dairy cattle. *J. Dairy Sci.*, **95**, 4136–4140.

Druet T., Schrooten C., de Roos A.P.W. (2010) Imputation of genotypes from different single nucleotide polymorphism panels in dairy cattle. *J. Dairy Sci.*, **93**, 5443–5454.

Erbe M., Hayes B.J., Matukumalli L.K., Goswami S., Bowman P.J., Reich C.M., Mason B.A., Goddard M.E. (2012) Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J. Dairy Sci.*, **95**, 4114–4129.

Hayes B.J., Bowman P.J., Chamberlain A.J., Goddard M.E. (2009) Invited review: genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.*, **92**, 433–443.

Hayes B.J., Bowman P.J., Daetwyler H.D., Kijas J.W., van der Werf J.H.J. (2011) Accuracy of genotype imputation in sheep breeds. *Anim. Genet.*, **43**, 72–80.

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

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

Khatkar M.S., Moser G., Hayes B.J., Raadsma H.W. (2012) Strategies and utility of imputed SNP genotypes for genomic analysis in dairy cattle. *BMC Genomics*, **13**, 538.

Matukumalli L.K., Lawley C.T., Schnabel R.D., Taylor J.F., Allan M.F., Heaton M.P., O'Connell J., Moore S.S., Smith T.P.L., Sonstegard T.S., Van Tassell C.P. (2009) Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One*, **4**, e5350.

Meuwissen T.H.E., Hayes B.J., Goddard M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, **157**, 1819–1829.

Mulder H.A., Calus M.P.L., Druet T., Schrooten C. (2012) Imputation of genotypes with low-density chips and its effect on reliability of direct genomic values in Dutch Holstein cattle. *J. Dairy Sci.*, **95**, 876–889.

Pausch H., Aigner B., Emmerling R., Edel C., Götz K.-U., Fries R. (2013) Imputation of high-density genotypes in the Fleckvieh cattle population. *Genet. Sel. Evol.*, **45**, 3.

Pszczola M., Strabel T., Mulder H.A., Calus M.P.L. (2012) Reliability of direct genomic values for animals with different relationships within and to the reference population. *J. Dairy Sci.*, **95**, 389–400.

Saatchi M., McClure M.C., McKay S.D., Rolf M.M., Kim J., Decker J.E., Taxis T.M., Chapple R.H., Ramey H.R.,

Northcutt S.L., Bauck S., Woodward B., Dekkers J.C.M., Fernando R.L., Schnabel R.D., Garrick D.J., Taylor J.F. (2011) Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genet Sel Evol*, **43**, 40.

Saatchi M., Schnabel R.D., Rolf M.M., Taylor J.F., Garrick D.J. (2012) Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle. *Genet. Sel. Evol.*, **44**, 38.

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

Weigel K.A., de los Campos G., González-Recio O., Naya H., Wu X.L., Long N., Rosa G.J.M., Gianola D. (2009) Predictive ability of direct genomic values for lifetime net merit of Holstein sires using selected subsets of single nucleotide polymorphism markers. *J. Dairy Sci.*, **92**, 5248–5257.

Zhang Z., Druet T. (2010) Marker imputation with low-density marker panels in Dutch Holstein cattle. *J. Dairy Sci.*, **93**, 5487–5494.

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

