Imputation Efficiency with Different Low Density Chips in French Dairy and Beef Breeds

Romain Dassonneville [1,2], Sébastien Fritz [3], Didier Boichard[1] and Vincent Ducrocq[1]
[1]INRA, UMR 1313 Génétique Animale et Biologie Intégrative, 78 352 Jouy-en-Josas, France
[2]Institut de l'Elevage, 149, rue de Bercy, 75 595 Paris Cedex 12 France
[3]Union nationale des Coopératives d'Elevage et d'Insémination Animale (UNCEIA), 149, rue de Bercy, 75 595 Paris Cedex 12 France
e-mail: romain.dassonneville@jouy.inra.fr

Abstract

Low density chips are appealing alternative tools in order to reduce genotyping costs. Such a chip is already commercially available. The best way to use low density chips is to impute data to a more dense coverage such as the standard 50K genotype. Two alternative in silico chips are presented and include markers selected to optimise Minor Allele Frequency and spacing. The objective of this study was to compare imputation accuracy of these custom low density chips with the commercially available 3K chip. Three French dairy and beef breeds were studied: Holstein, Montbéliarde and Blonde d'Aquitaine with respectively 4,037, 1,219 and 991 50K-genotypes. Markers were masked for the validation population in order to mimic a low density genotype. Imputation was realised with *Beagle* software. 95% to 99% of alleles were correctly imputed depending on the breed and the low density chip. Custom low density chip gave better results. The gain to use 6K chip was found to be even higher for beef breeds such as Blonde d'Aquitaine. A low density chip with 6,000 markers is a valuable genotyping tool that is suitable for both dairy and beef breeds. Such a tool can be used for pre-selection of young animals or large screening of the female population.

1. Introduction

Genomic selection requires estimation of effects for SNP (Single Nucleotide Polymorphism) covering the whole genome at a sufficient density. For this purpose, the Bovine 50K chip from Illumina is widely used. The number of genotyped animals increases every year. But genotyping a large fraction of the female population requires to drastically reduce genotyping costs. For this purpose, another genotyping alternative is offered by low density chips. The Golden Gate Bovine 3K chip (GG 3K) was developed in 2009 by Illumina with a cheaper technology (Golden Gate). Nowadays, other low density chips can be considered at a similar reduced cost. This low density chips could be especially developed to maximize imputation accuracy to reconstruct genotypes of the widely used Bovine 50K chip. With low density chips, markers density is reduced while the genome is still entirely covered. Imputation offers the possibility to predict –impute– a dense (50K) genotype based on data from a low density chip (Zhang and Druet, 2010). Genomic breeding values (GEBV) can be obtained from calculations based on imputed genotypes (Dassonneville and Brøndum *et al.*, 2011). The purpose of this study is to compare imputation accuracy of different chips: the commercially available chip and 2 custom low density chips of various markers density (3K and 6K).

2. Materials and Method

Data

Three breeds were chosen for this study: the international Holstein breed, the French Montbéliarde dairy breed, and the Blonde d'Aquitaine beef breed.

Reference population

The reference population of dairy breeds (Holstein and Montbéliarde) included progeny-tested bulls across several generations (4,037 and 1,219 respectively). The validation population was defined through a cut-off date such that 25% of the bulls of the reference population forming the validation population were born after that date. Most of the male ancestors of the animals

included in the validation population were genotyped and included in the training population. The reference population of the Blonde d'Aquitaine beef breed included 961 young bulls and their 30 sires. 240 young bulls were randomly selected to form the validation population. Among the ancestors of individuals included in the validation population, only the sire was in the training population.

Low density chips

The GG 3K chip is based on a different genotyping technology (Golden Gate) contrary to other Bovine SNP chips (HD, 50K) which are based on Infinium technology. The 2 other chips studied here were only created in silico. They can be considered as based on Infinium technology (since markers were chosen among Bovine 50K SNP).

Criteria to select markers to include

Markers had to be included in the Bovine SNP50 versions 1 and 2, and in the Bovine HD chip from Illumina; markers had to have a known position on btau 4.0 and UMD3 assemblies and the position had to be consistent between the 2 assemblies (less than 10 Mb apart); call rates were above 0.98 and no technical problem was observed in the sample of gentoyped animals at INRA; markers were checked for Hardy-Weinberg equilibrium (q value > 0.01). Once all of these criteria were checked, markers were chosen in order to maximize the Minor Allele Frequency (MAF). For the custom 3K chip, for each Mb, the SNP with the highest average MAF over the 3 main French dairy breeds (Holstein, Normande and Montbéliarde) was kept. For the custom 6K chip, first the best SNP was selected for dairy breeds (as for the custom 3K), then the SNP with the highest average MAF over 8 dairy and beef breeds (Blonde d'Aquitaine, Brown Swiss, Charolaise, Holstein, Limousine, Normande, Montbéliarde and Maine-Anjou) was added. Finally, a few more SNP were added to cover every half Mb and to ensure a better coverage of chromosome extremities. For these 2 custom low density chips, both MAF and spacing were optimised.

Table 1. Number of markers including in the 3 low density chips.

Low density chip	Number of markers
GoldenGate Bovine3K	2900
French custom 3K	2929
French custom 6K	6144

The GG 3K chip includes 2900 SNP (Table 1) among which 2635 are kept after quality control as described above. For the custom chips, including 2929 and 6144 SNP, only markers on sexual chromosomes were discarded. These SNP were added in order to be able to use the low density chip for sexing. Markers used for parentage were also included.

Simulating low density chips

The SNPs included in the low density chips are all included in the Bovine 50K chip. To mimic the low density chip, marker genotypes of animals of the validation population were obtained by erasing the remaining marker information from the 50K.

Imputation method

Imputation was performed using the software *Beagle* (Browing and Browning, 2007) version 3.2 based on population linkage disequilibrium.

Imputation error rate calculation

The number of errors was counted as 0 when the imputed and observed marker genotypes were identical, 1 if the real marker genotype was homozygous and the imputed genotype was heterozygous (or vice versa), and 2 if real and imputed marker types were opposite homozygous. The error rate was calculated as the total number of errors divided by twice the number of imputed loci.

3. Results

In figure 1, based on error rate calculation, the percentage of alleles correctly imputed is presented, for the 3 different breeds, for the 3 different low density chips.

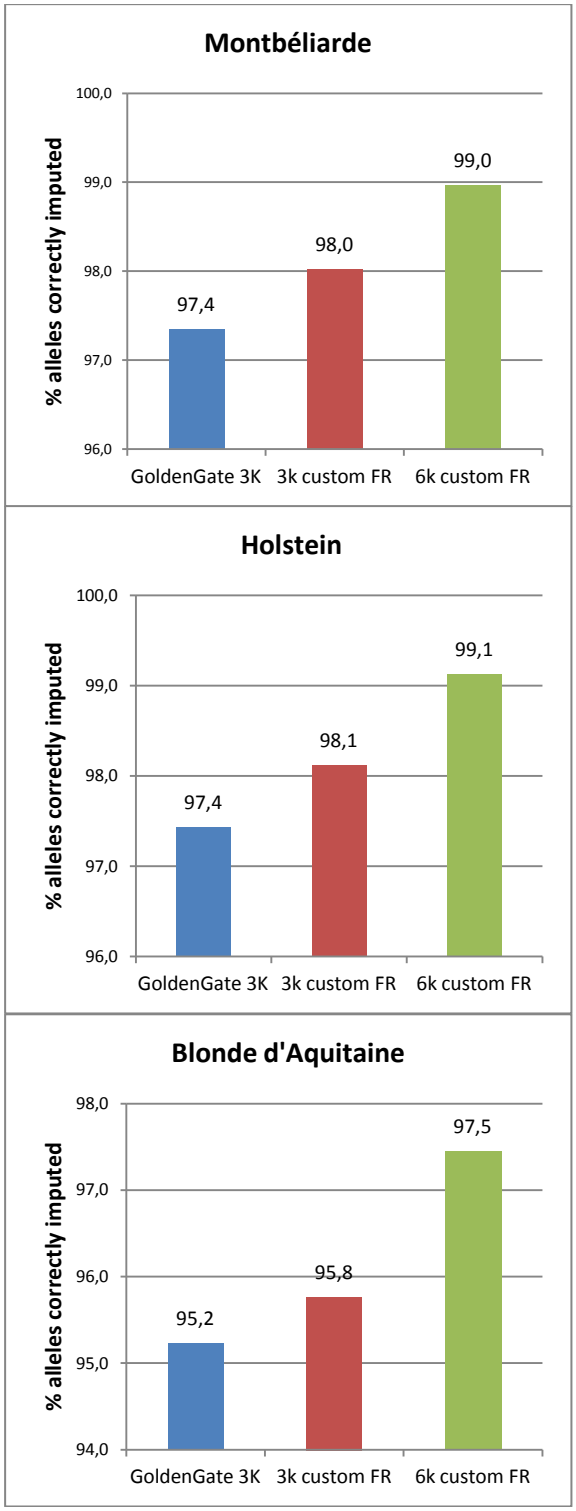

Figure 1. Imputation accuracy (fraction of masked alleles correctly imputed) for the 3 low density chips for the 3 breeds.

4. Discussion

In the Montbéliarde breed, 97% of the masked alleles were correctly imputed from the GG 3K chip. This result is high enough to implement genotyping with low density chips. However it can be improved with other chips. The French in silico custom 3K chip was optimised for French dairy breeds including Montbéliarde. For this reason, imputation accuracy (98% of alleles correctly imputed) was improved compared to GG3K. With the 6K chip, both markers density and MAF optimisation were improved, resulting in higher imputation accuracy (99%).

Holstein was one of the 3 breeds involved in the choice of markers for the GG 3K. For this reason, one could expect similar results between the two 3K chips. But the custom 3K chip gives better results (1.9% error rate instead of 2.6%). One possible explanation is the lower number of effective markers. Another one is the Golden Gate constraints in the choice of markers, limiting the possible optimization on MAF and spacing. With no surprises, results were better with 6K chip (less than 1% error rate).

Imputation accuracy results for the French beef breed Blonde d'Aquitaine for the two 3K chips were similar. Imputation accuracy was lower compared to the two other breeds, maybe due to smaller and different reference population. Still a slight gain can be obtained with the French custom 3K chip (95.8% vs. 95.2%) whereas MAF of that breed were not taken into account in any of these 2 chips. Maybe this is due to a better optimisation of MAF and spacing as reported above. The biggest gain for the 6K chip compared to 3K chips was obtained with the beef breed (2.5% error rate vs. 4-5%) since MAF of that breed and other French beef breeds were included in the analysis to select markers.

5. Conclusion

Imputation using *Beagle* software was efficient to reconstruct a dense - 50K - genotype from low density chips data. Accuracy ranges between 95 and 99%.

Low density chips are appealing alternative tools that reduce genotyping costs. This could allow to genotype more animals. They can be

used for pre-selection of young animals. It is even more interesting in the perspective of using female information in genomic selection.

The existing Golden Gate Bovine 3K chip presents satisfactory results. However, other choices of markers are possible for low density chips in order to optimise MAF and spacing for various breeds so that imputation is more accurate. The use of the 6K chip is even more interesting for the French beef breed Blonde d'Aquitaine, dividing by 2 the error rate for instance. The 6K chip always give better imputation results for all the breeds even for dairy breeds such as Holstein.

A low density chip with 6,000 markers is an interesting genotyping tool that is suitable for dairy and beef breeds.

Acknowledgments

Agence Nationale de la Recherche and APISGENE are gratefully acknowledged for funding the AMASGEN project. Blonde d'Aquitaine data were obtained from the Qualvigene project.

References

Browning, S.R. & Browning, B.L. 2007. Rapid and accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *The American Journal of Human Genetics 81,*1084-1097.

Dassonneville, R., Brøndum, R.F., Druet, T., Fritz, S., Guillaume, F., Guldbrandtsen, B., Lund, M.S., Ducrocq, V. & Su, G. 2011. Impact of imputing markers from a low density chip on the reliability of genomic breeding values in Holstein populations. *J. Dairy Sci. 94:7*, 3679-3686.

Illumina website:
http://www.illumina.com/Documents/products/datasheets/datasheet_bovine3k.pdf

Zhang, Z. & Druet, T. 2010. Marker imputation with low-density marker panels in Dutch Holstein cattle. *J. Dairy Sci. 93:11*, 5487-5494.