Using 1K panel in 3 French pig breeds for genomic selection: Accuracy of Imputation and estimation of genomic breeding values using 1K SNP panel, designed for several breeds in French pig populations

Céline Carillier-Jacquin, Alban Bouquet, Yann Labrune, Pauline Brenaut, Juliette Riquet, Catherine Larzul

Using 1K SNP panel for genomic selection in 3 French pig breeds:
Accuracy of Imputation and estimation of genomic breeding values using 1K SNP panel, designed for several breeds in French pig populations

C. Carillier-Jacquin[1], A. Bouquet[2], Y. Labrune[1], P. Brenaut[2], J. Riquet[1], C. Larzul[1]

[1] GenPhySE, Université de Toulouse, INRA, INPT, ENVT, 31320, Castanet-Tolosan, France
celine.carillier-jacquin@inra.fr (Corresponding Author)
[2] IFIP-Institut du Porc, 35651 Le Rheu Cedex, France

Introduction

The current cost of medium density SNP chips is a limit to the development of genomic selection in pig populations (Badke et al., 2014; Wellmann et al., 2013). To reduce the cost of genotyping, a low density (LD) SNP chip was designed in 2016 and has been used in routine. This LD panel of around 1100 SNP was optimized for imputation accuracy in the French Landrace (Land) pig population using equally spaced SNP with minor allele frequency (MAF) larger than 0.2. In the present study, we proposed to adapt the panel to two other major French pig breeds *i.e.* Large White (LW) and Pietrain (PI) lines. Imputation accuracy as well as the impact on genomic estimated breeding value (GEBV) were estimated in the three breeds using this new SNP chip design.

Material and methods

Data

Datasets composed of 2 822 Landrace, 619 Large White and 984 Pietrain genotyped using Porcine SNP60 Illumina BeadChip (Illumina, San Diego, CA) and 902 Landrace and 442 Large White using GeneSeek Genomic Profiler HD 80k (GeneSeek, Licoln, NE). Landrace and Large White animals were mostly breeding animals and candidates whereas Pietrain pigs were part of a specific experimental design (i.e. 96 sires and around 10 sons per sire). After quality control (Hardy Weinberg equilibrium, MAF estimated in each population > 0.05, call rate > 0.98), a total of 26 492 SNP in common between the two medium density (MD) panels were conserved.

The low density panel was defined selecting two types of SNP among the 26 492: 467 SNP, were selected for parentage assignation or as polymorphisms of interest for several traits (Christophe Audebert, Genes Diffusion, personal communication), and 658 SNP were selected in order to have equidistant localization on autosomes (based on the SNP position on the Sus Scrofa v11.1 draft sequence (www.ncbi.nlm.nih.gov)), excluding the ends of the chromosomes (1 Mb on each side), and with a MAF larger than 0.2 in each of the three French pig populations.

Imputation

Imputation was carried out separately in each breed. To study imputation quality, for each breed a subset of animals was selected as a reference population (2 651 Landrace, 950 Large White and 888 Pietrain), the last ones (111 Large White and 1 073 Landrace pigs born in the second semester of 2016, and 96 Pietrain (1 son randomly chosen per sire) being considered as candidate animals genotyped with the LD panel. For candidate animals, genotypes at the 25 436 SNP not included in the LD panel were imputed based on MD genotypes of the

reference populations. FImpute software (Sargolzaei et al., 2011) was used to impute the candidate genotypes based on pedigree and SNP information. Imputation accuracy was evaluated using the Pearson correlations between imputed and real genotypes per animal and per SNP and the imputation error rate estimated as:

$$\text{error rates} = \frac{S\text{err}}{\text{nb}} \quad (1)$$

where *S*err is the sum of imputation errors for the SNP considered and nb is the number of genotyped animals.

Genomic breeding value prediction

Genomic evaluations were conducted separately in each breed using single-step genomic BLUP (GBLUP) method with blupf90 program (Misztal et al., 2002). Traits studied were reproduction traits for Landrace and Large White female lines, respectively number of piglets born alive (NBA), number of piglets weaned by the sow (NW), average (AWB) and standard deviation (SDWB) of weight of piglet at birth. For Pietrain population, genomic evaluations were performed on production traits recorded in test station or at slaughter house, *i.e.* average daily gain (ADG), feed intake (FI), backfat depth (BD), loin muscle depth (LMD), percentage of lean meat (PLM), dressing yield (DY) and ham ultimate pH (pH). The maximum number of records used were 223 394 for Landrace, 450 685 for Large White coming from the official genetic evaluation of July, 19th 2016 and 749 for Pietrain pigs.

Models used were the same than the ones used for official genetic evaluations in those populations:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{(Wp)} + \mathbf{(Tl)} + \mathbf{(Ps)} + \mathbf{e} \quad (2)$$

where **y** was the vector of performances, **β** the fixed effects depending on the trait considered (herd*batch number*sex and halothane for BD and LMD, herd*batch number, halothane and weight at the beginning of control for ADG and FI, herd*batch number, halothane and carcass weight for PLM, DY and pH; or herd*batch number at delivery and parity of the dam for reproduction trait (Bouquet et al., 2014)), **p** the random effect of permanent environment of the dam specific for reproduction traits, **l** the random effects depending on the trait considered (litter for BD and LMD, growing pen for ADG and FI, breed of the sire of piglets for NW, or age of piglets at birth weighting (0 or 1 day) for AWB and SDWB), **s** the random effect of the sire of piglets specific for reproduction traits, **u** the breeding values and **e** a vector of random normal errors. Genomic breeding values **u** were normally distributed with $Var(\mathbf{u}) = \mathbf{H}\sigma_u^2$, where **H** was the genetic relationship matrix combining information of SNP markers and pedigree (Legarra et al., 2009) derived from allele frequencies considered in each population.

The impact of genomic evaluation using imputed genotypes was evaluated as the correlation between GEBVs estimated with imputed genotypes and the ones estimated with MD genotypes for the candidate animals (1 073 Landrace,111 Large White and 96 Pietrain pigs).

Results and discussion

Imputation quality

Average imputation error rates across chromosomes were 0.06 for Landrace, 0.07 for Large White and 0.09 for Pietrain populations. Figure 1 shows the imputation error rate estimated per chromosome in each population. It ranges from 0.03 for chromosome 1 in Landrace breed

to 0.13 for chromosome 16 in Pietrain population. As in Xiang et al. (2015) imputations were the least accurate for chromosomes with the lowest average linkage disequilibrium in each population (i.e. chromosomes 17 and 18 in Landrace and Large White, chromosomes 10, 11, 12 and 16 in Pietrain). The best results were obtained for the Landrace population which could be explained by the larger number of animals genotyped in MD compared to the other populations. Similar trends were obtained for correlations between imputed and real genotypes estimated per SNP which ranged from 0.53 for chromosome 17 to 0.84 for chromosome 1 in Landrace population. Correlations between imputed and real genotypes estimated per animal over the whole genome were larger than per SNP i.e. 0.93, 0.92 and 0.88 in Landrace, Large White and Pietrain populations respectively.

Imputation accuracy obtained in this study was lower than in studies in which a 8K-10K panel was used for a variety of breeds (Xiang et al., 2015), (Badke et al., 2014) (Gualdrón Duarte et al., 2013). It was consistent with other studies having considered very low density SNP panels in pigs (Cleveland and Hickey, 2013; Huang et al., 2012).

Genomic evaluation

Table 1 presents correlations between GEBV estimated using imputed genotypes on the one hand and on the 26 492 SNP in common between the two original MD panels on the other hand for the validation animals in Landrace and Large White population. These correlations ranged from 0.89 for SDWB for Landrace to 0.97 for Large White. It was better than the ones obtained for Pietrain population from 0.81 for LMD to 0.97 for PLM (see Table 2) which could be explained by the lower number of performance records used in this population. In maternal breeds, correlations between GEBV obtained were similar for females and for male candidates. The highest impact on GEBV was estimated on animals with fewer relatives genotyped with MD chips. The accuracy of genomic selection using imputed genotypes was correlated to the relatedness between training and validation animals (see additional table 3 in Landrace) as in Badke et al. (2014).

These correlations between GEBV based on original and imputed genotypes were better than expected considering the quality of imputation found in this study. Indeed the decrease in genomic evaluation accuracy using imputed genotypes was similar to the decrease estimated in Badke et al. (2014)' study with better imputation quality.

Conclusion

A design for a customised low density chip (i.e. 1 100 SNP) was proposed for the three French pig populations Landrace, Large White and Pietrain. Imputation quality considering error rates and correlations between imputed and true genotypes was similar to other studies considering the same LD panel size. It differed in the three populations which could be explained by the difference in number of animals genotyped and in relatedness between animals genotyped with LD panel and those genotyped with MD one. Considering the small size of the LD panel used in this study, imputation accuracy obtained is sufficient to use imputed genotypes from LD panel in genomic evaluations. In practice, genotyping candidates with LD chip could be a solution to select future sires at low cost. Imputation quality using the LD panel was not good for animals with no parent genotyped with MD chip. It could be considered to genotype sires of these animals with MD panel and dams with LD chip to improve imputation accuracy.

List of References

Badke, Y.M., Bates, R.O., Ernst, C.W., Fix, J., Steibel, J.P., 2014. Accuracy of Estimation of Genomic Breeding Values in Pigs Using Low-Density Genotypes and Imputation. G3 Genes Genomes Genet. 4, 623–631.

Bouquet, A., Delaunay, I., Merour, I., Bellec, T., Ligonesche, B., Sourdioux, M., Bidanel, J.P., Canario, L., 2014. Conséquences d'une sélection sur l'homogénéité du poids des porcelets à la naissance sur la productivité numérique des truies Large White et Landrace Français. Presented at the Journées Recherche Porcine, 46, 19-24.

Cleveland, M.A., Hickey, J.M., 2013. Practical implementation of cost-effective genomic selection in commercial pig breeding using imputation. J. Anim. Sci. 91, 3583–3592.

Gualdrón Duarte, J.L., Bates, R.O., Ernst, C.W., Raney, N.E., Cantet, R.J., Steibel, J.P., 2013. Genotype imputation accuracy in a F2 pig population using high density and low density SNP panels. BMC Genet. 14, 38.

Huang, Y., Hickey, J.M., Cleveland, M.A., Maltecca, C., 2012. Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. Genet. Sel. Evol. 44, 25.

Legarra, A., Aguilar, I., Misztal, I., 2009. A relationship matrix including full pedigree and genomic information. J Dairy Sci 92, 4656–4663.

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., Lee, D.H., 2002. BLUPF90 and related programs (BGF90). Presented at the 7th World Congress in Genetics Applied to Livestock Production, 19-23 August 2002, Montpellier, France.

Sargolzaei, M., Chesnais, J.P., Schenkel, F.S., 2011. FImpute - An efficient imputation algorithm for dairy cattle populations. J. Dairy Sci. 94: E-Suppl. 1, 421.

Wellmann, R., Preuß, S., Tholen, E., Heinkel, J., Wimmers, K., Bennewitz, J., 2013. Genomic selection using low density marker panels with application to a sire line in pigs. Genet. Sel. Evol. 45, 28.

Xiang, T., Ma, P., Ostersen, T., Legarra, A., Christensen, O.F., 2015. Imputation of genotypes in Danish purebred and two-way crossbred pigs using low-density panels. Genet. Sel. Evol. 47, 54.

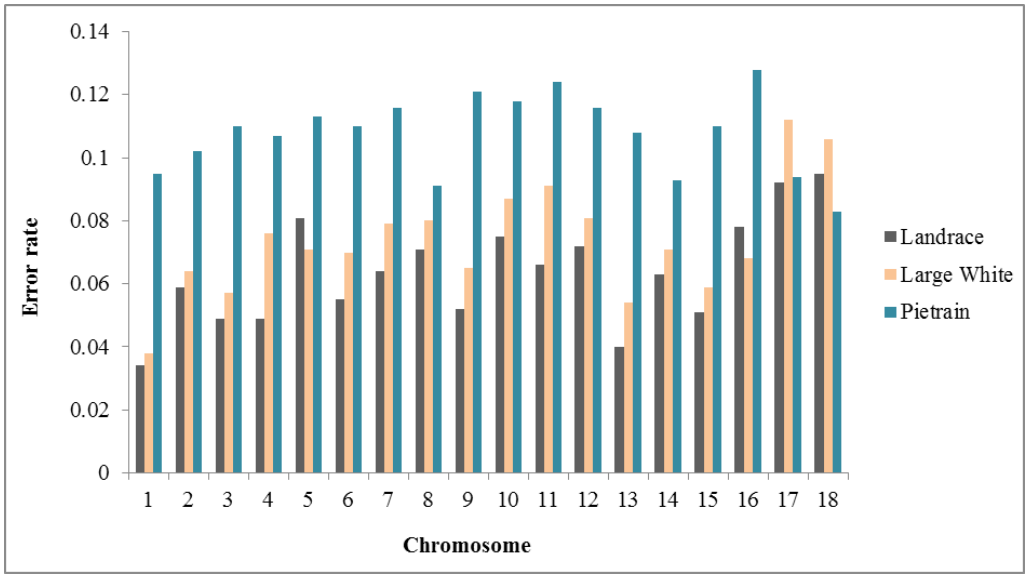

Figure 1. Imputation error rate per chromosome in the three French pig population Landrace, Large White and Pietrain.

Table 1. Correlations between GEBV estimated using imputed genotypes on the one hand and original MD genotypes on the other hand in Large White and Landrace populations

	Landrace	LW
number of piglets born alive	0.96	0.94
number of piglets weaned by the saw	0.95	0.94
average weight of piglets at birth	0.93	0.96
standard deviation of the piglets weight at birth	0.89	0.97

Table 2. Correlations between GEBV estimated using imputed genotypes on the one hand and original MD genotypes on the other hand in Pietrain population

Trait	Correlation
Average daily gain	0.95
Feed intake	0.95
Backfat depth	0.89
Loin muscle depth	0.81
Percentage of lean meat	0.97
Dressing yield	0.96
Meat pH	0.96

Table 3. Correlations between GEBV estimated using imputed genotypes on the one hand and original MD genotypes on the other hand in Landrace population for animals with 1, 2 or no parent genotyped

Number of parents genotyped	No	1	2
number of piglets born alive	0.82	0.96	0.99
number of piglets weaned by the saw	0.80	0.95	0.99
average weight of piglets at birth	0.67	0.93	0.99
standard deviation of the piglets weight at birth	0.79	0.91	0.99