Interest of using imputation for genomic evaluation in layer chicken

Florian Herry, David Picard Druet, Frédéric Herault, Amandine Varenne, Thierry Burlot, Pascale Le Roy, Sophie Allais

IMPUTATION FOR GENOMIC EVALUATION IN LAYERS

Interest of using imputation for genomic evaluation in layer chicken

Florian Herry,[*,†] David Picard Druet,[†] Frédéric Hérault,[†] Amandine Varenne,[*] Thierry Burlot,[*] Pascale Le Roy,[†] and Sophie Allais[†,1]

[*]*NOVOGEN, Mauguerand 22800 Le Foeil, France;* †*PEGASE, INRA, Agrocampus Ouest, 16 Le Clos 35590 Saint-Gilles, France*

[1]Corresponding author: sophie.allais@agrocampus-ouest.fr

GENETICS AND GENOMICS

ABSTRACT

With the availability of the 600K Affymetrix® Axiom® high-density (HD) single nucleotide polymorphism (SNP) chip, genomic selection has been implemented in broiler and layer chicken. However, the cost of this SNP chip is too high to genotype all selection candidates. A solution is to develop low density SNP chip, at a lower price, and to impute all missing markers. But to routinely implement this solution, the impact of imputation on genomic evaluation accuracy must be studied. It is also interesting to study the consequences of the use of low density SNP chips on genomic evaluation accuracy. In this perspective, the interest of using imputation in genomic selection was studied in a pure layer line.

Two low density SNP chip design were compared: an equidistant (EQ) methodology and a methodology based on linkage disequilibrium (LD). Egg weight, egg shell color, egg shell strength and albumen height were evaluated with single-step GBLUP methodology. The impact of imputation errors or the absence of imputation on the ranking of the male selection candidates was assessed with a genomic evaluation based on ancestry. Thus, genomic estimated breeding values (GEBV), with imputed HD genotypes or low density genotypes, were compared to GEBV obtained with the HD SNP chip. The relative accuracy of GEBV was also investigated by considering as reference GEBV estimated on offspring.

A limited reordering of the breeders, selected on a multi-trait index, was observed. Spearman correlations between GEBV on HD genotypes and GEBV on low density genotypes (with or without imputation) were always higher than 0.94 with more than 3K SNPs. For the genetically closer top 150 individuals for a specific trait, with imputation, the reordering was reduced with correlation higher than 0.94 with more than 3K SNPs. Without imputation the correlations remained below 0.85 with less than 3K and 16K SNPs for EQ and LD methodology, respectively. The differences in GEBV correlations between both methodologies never were significant. The conclusions were the same for all studied traits.

Key words: Genomic selection, layer chicken, low density panel, imputation accuracy, genomic evaluation accuracy

INTRODUCTION

The availability of single nucleotide polymorphisms (SNP) enabled the development of high-throughput genotyping technologies leading to the use of the 600K Affymetrix® Axiom® high density (HD) genotyping array, a high-density genotyping chip developed by Kranis et al. in 2013, in layer and broiler breeding. Genomic selection as described by Meuwissen et al. (2001) has then been implemented in many livestock species with different statistical methods like genomic best linear unbiased prediction methods (GBLUP) (Legarra et al., 2009; Goddard et al., 2011) or Bayesian methods (Meuwissen et al., 2001; Xu, 2003; Habier et al., 2009). From a reference population with genotypes and phenotypes, it is possible to estimate the genomic value of the genotyped selection candidates with or without phenotype. The main objective is to choose among the selection candidates of generation *N*, the best breeders for one or more traits to produce the individuals of the generation *N+1*. In addition, compared to a genetic selection, genomic selection may increase the genetic gain through the decrease in generation interval, most particularly for species with high generation interval, through the increase in selection intensity by genotyping many selection candidates and through the increase in evaluation accuracy.

However, the high cost of such high density (HD) SNP chip is still a problem for all livestock species. To reduce the cost of genomic selection, low density SNP chips can be developed. The idea is to select a subset of markers from the HD SNP chip and to impute the genotypes at missing markers. Three main methods to select the marker panel have been developed: (1) selection of a subset of SNPs chosen at regular intervals along each chromosome taking into account or not the MAF of the selected SNPs (Habier et al., 2009; Weigel et al., 2009; Zhang et al., 2011; Cleveland & Hickey, 2013; Wang et al., 2013; Herry et al., 2018), (2) selection of a subset of SNPs having high effects on different traits of interest (Weigel et al., 2009, Zhang et al., 2011), or (3) selection of a subset of SNPs based on linkage

disequilibrium (LD) between markers (Herry et al., 2018). This latter method was studied because of the particularities of the Gallus gallus genome (International Chicken Genome Sequencing Consortium, 2004) and the particular structure of the avian linkage disequilibrium (Megens et al., 2009; Qanbari et al., 2010; Hérault et al., 2018).

Factors influencing imputation accuracy are well documented as well as the relation between imputation accuracy and genomic evaluation of the selection candidates. Theoretically, due to imputation errors, genomic evaluation accuracy with imputed genotypes is expected to be lower than a genomic evaluation done with HD genotypes. The literature confirms it for very low density SNP chip (from few SNPs to 3K SNPs) with a decrease in genomic evaluation accuracy with a decrease, sometimes limited, in imputation accuracy (Weigel et al., 2009; Weigel et al., 2010; Mulder et al., 2012; Cleveland & Hickey, 2013, Raoul et al., 2017). But concerning intermediate low density SNP chip (between 6K and 20K SNPs), other studies showed that the impact of imputation errors was very limited (Weigel et al., 2010; VanRaden et al., 2011; VanRaden et al., 2012; Moghaddar et al., 2015; Wang et al., 2016). However, few studies about the impact of imputation on genomic evaluation have been led on chickens (Wang et al, 2013).

In addition, several studies showed that for traits affected by few large QTL, genomic evaluations are more sensitive to imputation errors. This was shown by Habier et al. (2009) and Zhang et al. (2011) in simulation studies and confirmed by Chen et al. (2014) on real data. They showed, in Holstein bulls, that the accuracy of direct genomic value (DGV) for milk fat percentage, a trait affected by few large QTL, decreased by 34% via GBLUP using imputed genotypes. Conversely, they showed that the accuracy of DGV for the somatic cell score, a trait affected by many small QTL, decreased only by 15%. In layer chickens, most of studied traits are affected by many small QTL. This could indicate that genomic evaluation would not be severely impacted by imputation errors.

126 Finally, most studies investigated the impact of imputation on genomic evaluation
127 accuracy, but only few studies focused on the impact of the use of medium density SNP chip
128 (Su et al., 2012; Moghaddar et al., 2015) or low density SNP chip (Weigel et al., 2009; Harris
129 & Johnson, 2010) without imputation on genomic evaluation.
130 The main objective of a company is to select their breeders and to describe the
131 consequences on the loss of selection response and on genetic progress by investigating if the
132 ranking of their best candidates would be modified with the use of low density SNP chip.
133 Thus, focusing on four generations of a pure line of laying hens, the first objective of this
134 study was to investigate the impact of imputation errors on genomic evaluation with an
135 evaluation based on ancestry of the candidates of the second generation with true HD
136 genotyping or imputed HD genotyping. The second objective was to study the impact of a
137 direct use of low density SNP chips, without imputation, on genomic evaluation. To do so, a
138 comparison was done between the same previous genomic evaluation of the candidates based
139 on ancestry with true HD genotyping or with low density genotyping without imputation.
140 Then, to get closer to the true breeding values of the candidates, their genomic estimated
141 breeding values (GEBV) was estimated with a genomic evaluation with optimal information
142 (phenotypes on descendants). Thus, the third objective was to assess the relative accuracy of
143 genomic evaluation by comparing the GEBV of the candidates of the second generation with
144 optimal information (phenotypes on their descendants of the third and fourth generations) and
145 their GEBV based on ancestry with imputed HD genotyping. Finally, imputed HD genotyping
146 of the candidates were replaced by their low density genotyping without imputation.
147 Therefore, the fourth objective was to assess the relative accuracy of genomic evaluation of
148 the candidates without imputation.
149
150

151

152 # MATERIAL AND METHODS

153 ## *Ethics Statement*

154 All blood samples were carried out as part of the commercial and selection activities
155 of Novogen. These animals studied and the scientific investigations described herein are
156 therefore not to be considered as experimental animals per se, as defined in EU directive
157 2010/63 and subsequent national application texts. As a consequence, we did not seek ethical
158 review and approval of this study as one inclusing the use of experimental animals. All
159 animals were reared in compliance with national regulations pertaining to livestock
160 production and according to procedures approved by the French Veterinary Services.

161

162 ## *Animals*

163 All animals studied were detailed in Herry et al. (2018). They consisted in a
164 commercial pure line of Rhode Island (RI) laying hens. This line was created and selected by
165 Novogen (Plédran, France). The population studied was comprised of 21,475 chickens split in
166 four generations. Each generation was divided in three batches and a new batch was bred
167 every six months from 2010 to 2015 (Figure 1).
168 Concerning the laying hens, phenotypic data were recorded from 60 to 90 weeks of age, when
169 birds where bred in individual cages. Each data collected was associated with a laying hen.
170 There were 75,121 measures recorded for 7983 birds. Finally, the sires were bred in
171 individual cages.
172 Genomic selection was implemented in 2015 on males of this line. However, females were
173 still selected based on pedigree and performances, and not with genomic selection. Thus, this
174 study concerned male selection candidates. In addition, among the different parameters

studied and detailed in a next section, the relative accuracy of genomic selection was investigated. To calculate this relative accuracy, it is necessary to have a set of male selection candidates with information on their offspring. These male selection candidates were the 67 male breeders of the generation G1.

Genotyping

Genotyping are briefly described because detailed in Herry et al. (2018). 2370 animals were genotyped for 580,961 SNPs using the 600K Affymetrix® Axiom® HD genotyping array (Kranis et al., 2013).

Based on the fifth annotation release of Gallus gallus genome (Warren et al., 2017), these SNPs were distributed on macro-chromosomes (1 to 5), intermediate chromosomes (6 to 10), micro-chromosomes (11 to 28 and 33), one linkage group (LGE64), two sexual chromosomes Z and W, as well as a group of 3,724 SNPs with unknown location.

Genotypes were filtered through six successive steps (Table 1) including individual call rate (<95%), MAF (<0.05), SNP call rate (<95%) and Hardy-Weinberg equilibrium ($P < 10^{-4}$). SNPs with unknown location or located on sexual chromosome W were removed, as well as the animals showing pedigree incompatibilities. Most of the SNPs had to be removed because they showed zero MAF. Finally, 300,351 SNPs and 2362 individuals remained available for the analyses.

Low Density SNP Chips Design

Several low density SNP chips were previously designed in silico by selecting a subset of SNPs (Herry et al., 2018) from the HD SNP chip.

An equidistant (EQ) methodology was studied by selecting SNPs at regular physical intervals (in pb) along each chromosome. In addition, for each interval, the SNP with the

highest MAF, or the one located furthest on the left, in case of equivalent MAF, was selected. 12 low density "equi" SNP chips were designed according to this method with different SNP densities: 1K, 2K, 3K, 4K, 5K, 7.5K, 10K, 15K, 20K, 30K, 40K and 50K SNPs.

A linkage disequilibrium (LD) methodology was studied considering the particular structure of the chicken linkage disequilibrium (Robert et al., 2015). Low density SNP chips were designed using the SS4I software (Hérault et al., 2016). This software enabled to obtain clusters of SNPs according to a chosen LD threshold. For each cluster, the SNP with the highest MAF was selected and used as representative of this cluster. 9 low density "LD" SNP chips were designed with different LD thresholds: 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 and 0.8.

Imputation Accuracy

In our study, the selection candidates were the 580 sires of the second generation (G1) with simulated low density genotyping. The selection candidates were imputed from the high density genotyping of the 447 sires of the first generation (G0). These 447 individuals were the fathers or the fathers' half-brothers of the selection candidates. Thus, the selection candidates were directly related to them.

For each low density SNP chip designed, imputation accuracy of the selection candidates was previously assessed as the mean correlation between true and imputed genotypes (Herry et al., 2018). Correlations were calculated one SNP at a time for all the candidates, as suggested in Pearson's method. The mean correlation was then estimated on 300,351 correlations. The mean correlations obtained were subsequently compared for the different low density SNP chips and/or scenarios, using Student tests with type 1 error rate of 0.1%.

Measurement of Traits

Four distinct traits were studied in this paper. They are named according to Animal Trait Ontology for Livestock (Atol Ontology, 2012). From 60 to 75 weeks, egg production was recorded each day for all individuals. There were individual data. 75,121 eggs concerning 7983 birds were measured from (G0) to (G3).

One egg was collected per layer and per week, between 60 and 75 weeks, for all layers. These eggs were then transferred at Zootests (Ploufragan, France) to study egg quality traits. The first step was to measure Egg Weight (EW, in g). Then, three traits concerning egg shell color were estimated with a Minolta Chroma Meter: redness (a*), yellowness (b*) and lightness (L*) of egg shell. Egg Shell Color (ESC) was then calculated as $ESC = 100 - (L^* - a^* - b^*)$. The next step consisted in measuring Egg Shell Strength (ESS, in N) by using a compression machine to evaluate the shell static stiffness. ESS corresponded to the maximum force recorded before fracturing the shell. Finally, each egg was broken and Albumen Height (AH) was measured using a tripod.

Genomic Evaluation Strategies

EW, ESC, ESS and AH were evaluated with single-step GBLUP methodology (Legarra et al., 2009) using BLUPF90 programs (Misztal et al., 2002).

The first part aimed to investigate the impact of imputation errors on genomic evaluations (Figure 2a). To do so, a genomic evaluation based on ancestry "Anc_HD" was done using true HD genotyping of the 447 G0 sires and selection candidates (G1), and phenotypes of the first generation (G0). A second genomic evaluation based on ancestry "Anc_Imputed" was done using the same data for the 447 G0 sires and imputed HD genotyping of the selection candidates (G1) from simulated low density SNP chips previously

250 designed. For each low density SNP chip and for each trait, Spearman correlations, that
251 enabled to estimate the reordering of the selection candidates, were calculated between true
252 “Anc_HD” Genomic Estimated Breeding Value (GEBV) and “Anc_Imputed” GEBV.
253 Spearman correlations were calculated for the top 150 individuals from G1 according to each
254 trait. Spearman correlations were limited to the top 150 males to better describe the
255 consequences of imputation errors on the reordering of these individuals, and thus to better
256 describe the consequences on the loss of selection response and on genetic progress. The
257 objective was to identify the good candidates and to successfully rank them among
258 themselves. We did not focus on the ranking of the less good candidates. There were also
259 calculated for the 67 breeders from G1 having at least 10 offspring in G2.

260 Then, concerning the second objective, imputed HD genotyping of the candidates were
261 replaced by their low density genotyping without imputation, allowing to simulate the impact
262 of the direct use of the different low density SNP chips without imputation (Figure 2b). This
263 part also implied the use of low density genotyping without imputation for the reference
264 population. For each low density SNP chip and for each trait, Spearman correlations were
265 calculated between the same previous true “Anc_HD” GEBV and “Anc_Not_Imputed”
266 GEBV obtained with low density genotyping (without imputation). These correlations were
267 calculated for the same 67 breeders of G1 and the top 150 individuals from G1 according to
268 each trait.

269 The third objective was to study the attainable relative accuracy with imputation
270 (Figure 2c). To calculate this relative accuracy, it is necessary to have a set of male selection
271 candidates with information on their offspring. On one hand, males don’t have own
272 phenotypes and only a few of them have daughter records. Thus, information from them is
273 limited. On the other hand, Generation 2 had 662 genotyped females with own performances
274 and some of them with progeny records. They would provide a more reliable validation set

with GEBVs using all available information fairly close to the true breeding values. However, females were still selected based on pedigree and performances, and not with genomic selection. Thus, this study focused on male selection candidates. To get closer to the true breeding values for the males, a genomic evaluation "Full_HD" of the G1 candidates was done with all available information (phenotypes and genotypes) from (G0) to (G3). These "Full_HD" GEBV leaded to closer to the true breeding values of the G1 candidates which cannot be calculated. These "Full_HD" GEBV represented the maximum of relative accuracy attainable regarding this genomic evaluation with all information and were calculated only for the 67 G1 breeders which had at least 10 offspring in G2. Then, these "Full_HD" GEBV were compared by Pearson correlations with the previous GEBV based on ancestry "Anc_Imputed" with imputed HD genotyping of the breeders, for each simulated low density SNP chip.

Finally, imputed HD genotyping of the candidates were replaced again by their low density genotyping without imputation. The "Full_HD" GEBV of the 67 G1 breeders were compared by Pearson correlations with their GEBV obtained with low density genotyping without imputation ("Anc_Not_Imputed" GEBV). The fourth objective was thus to investigate the impact of a direct use of low density SNP chips without imputation on relative accuracy of genomic evaluation (Figure 2d).

The four traits were jointly estimated according to a classical multi-trait animal model: $Y = 1\mu + X\beta + Zu + \varepsilon$. Y is a vector of the four traits of each individual, μ is the vector of means of each trait, β is a vector of fixed effects including batches, battery and position in the battery, u is a vector of genomic breeding values and ε is a vector of random residual effects. X and Z are design matrixes relating respectively phenotypes to fixed effects and phenotypes to genomic breeding values (u). It is assumed that $u \sim N(0, H \otimes W)$ where H is the genetic relationship matrix combining SNP information and pedigree data (Legarra et al., 2009) and W is the matrix of variance and covariance of the genomic breeding values of the four traits.

Finally, $\varepsilon \sim N(0, I \otimes R)$ where I is the identity matrix and R is the matrix of residual variance and covariance of the four traits.

Software

FImpute V2.2 (Sargolzaei et al., 2014) was used to impute the selection candidates with low density genotyping to high density genotyping from the individuals of G0 with high density genotyping.

The scenario with all available information (Full_HD) was used to estimate the genetic parameters of the model. Remlf90 (Misztal et al., 2002) was used to estimate the genetic and residual variance components. Once fixed, all different genomic evaluations based on ancestry were performed with Blupf90. The variance components were compared to components estimated with a pedigree based model using all phenotypes. They were highly correlated (Picard Druet et al., 2019).

RESULTS AND DISCUSSION

Imputation Accuracy

All the results concerning imputation accuracy were presented in Herry et al. (2018) but the evolution of the mean correlations between true and imputed genotypes for the two different methodologies were recalled in Figure 3. For both methodologies, there was an increase in mean correlation with an increase in the number of SNPs on the different low density SNP chips. Better imputation accuracies were obtained with the LD methodology at an equivalent SNP density. The differences observed in mean correlation between the two methodologies were all significant. In addition, for the EQ methodology at a very low density of 1K SNPs, the mean correlation was 0.7098 indicating a quite deteriorated imputation accuracy. This corresponded to a genotyping imputation error rate of 18.5%.

These results were consistent with those found in the literature (Dassonneville et al., 2012; Carvalheiro et al., 2014) where an increase in the number of SNPs on low density SNP chip led to better imputations.

Impact of Imputation Errors

The impact of imputation errors was investigated by comparing the results of a genomic evaluation based on ancestry, with true HD genotyping or with imputed HD genotyping. Only the results for Egg Weight (EW) were shown to simplify the reading and because of the similarity of the results for the other traits.

Results For The Top 150 Individuals.

For both methodologies (Figure 4a), there was an increase in Spearman correlations between "Anc_HD" GEBV and "Anc_imputed" GEBV with an increase in SNP density. Indeed, for the LD0.05 and LD0.8 SNP chips, the mean correlations were respectively 0.8661 and 0.9931. For the 3Kequi and 20Kequi SNP chips, there were respectively 0.9045 and 0.9885. These results are in agreement with imputation accuracies obtained with the different low density SNP chips. There was an increase in mean correlation concerning the evaluations with an increase in imputation accuracy which is consistent with the literature. Moghaddar et al. (2015) showed, for Merino sheep, that the mean correlations between GEBV based on true genotypes (50K) and GEBV based on imputed genotypes (50K imputed from 12K) increased with imputation accuracies.

It was noticed that for both methodologies, with more than 5K SNPs, the mean correlations were above 0.90 indicating a re-ranking rather reduced of the best individuals for EW. However, for the 1Kequi SNP chip, the mean correlation was 0.7833 indicating a reordering quite important of the best individuals for egg weight.

348 Finally, at equivalent SNP density of 3K SNPs, the EQ methodology seemed to
349 present higher results than the LD methodology with mean GEBV correlations of respectively
350 0.9045 and 0.8661 for the 3Kequi and LD0.05. But the differences were not significant since
351 the standard errors were ±0.04 for both SNP chips. At a density of 20K SNPs, both
352 methodologies were equivalent with mean GEBV correlations of respectively 0.9885 and
353 0.9931 for the 20Kequi and LD0.8. However, as seen previously, the LD methodology
354 appeared to be better to get good imputation accuracies. Thus, higher imputation accuracies
355 with the LD methodology were not synonymous of better mean correlations between GEBV
356 compared to the EQ methodology. This could be due to the methodology itself. Indeed, Harris
357 and Johnson (2010) and Weigel et al. (2010) said that an equidistant methodology was better
358 to get good genomic evaluation results for traits controlled by many small QTL, which is the
359 case for the four traits studied. On the contrary, genomic evaluations concerning traits
360 controlled by few large QTL were more sensitive to equidistant methodology which was
361 consequently not the most appropriated methodology. Moreover, ssGBLUP methodology
362 considers a same variance for each SNP (Legarra et al., 2009) and consequently would favor
363 the EQ methodology. Finally, another reason could be due to the errors done with imputation.
364 Some imputation errors from LD SNP chips could degrade more the GEBV estimation than
365 imputation errors from equidistant SNP chips. The EQ methodology would be more robust
366 than the LD methodology in case of imputation errors.
367
368 ***Results For The Breeders.***
369 Spearman correlations between "Anc_HD" GEBV and "Anc_Imputed" GEBV were
370 also calculated for the 67 G1 breeders having at least 10 offspring in the next generation G2.
371 For both methodologies (Figure 4b), there was an increase in Spearman correlations with an
372 increase in SNP density. Indeed, for the LD0.05 and LD0.08 SNP chips, the mean GEBV

correlations were respectively 0.9777 and 0.9979. For the 3Kequi and the 20Kequi SNP chips, the results were respectively 0.9771 and 0.9972. Thus, the results were higher compared to the results for the top 150 individuals. This is due to the distribution of the 67 breeders which were not the best breeders of G1 for EW, but the best for a set of selection criteria. This was confirmed by plotting the normal distribution of HD GEBV estimated on ancestry with true HD genotyping for all G1 candidates (Figure 5). The 67 breeders (in red on the plot) were well distributed among the 580 individuals of G1 which reduced the reordering of the individuals.

The results also showed that even with a SNP density superior to 2K SNPs, good mean correlations (superior to 0.95) could be obtained indicating a very reduced re-ranking of the individuals. With only 5K SNPs imputed to the HD SNP chips, mean correlations above 0.98 could be reached.

However, with the 1Kequi SNP chip, the mean GEBV correlation was under 0.95. This decrease in correlation was also illustrated by Cleveland and Hickey (2013) in pig. They used only 450 SNPs imputed to the Illumina PorcineSNP60 BeadChip which resulted in a decrease in correlation to 0.866 (for an imputation accuracy of 0.914). Thus, by decreasing to much the SNP density, the reduced imputation accuracies can have negative consequences on genomic evaluations.

Finally, our results did not show any difference between EQ and LD methodologies.

Impact of the Absence of Imputation

Given the good results of genomic evaluations with imputed genotyping, the impact of the absence of imputation was studied. Only the results for Egg Weight (EW) were shown to simplify the reading and because of the similarity of the results for the other traits.

Results For The Top 150 Individuals.

For the top 150 individuals for both methodologies (Figure 6a), there was an increase in Spearman correlation between "Anc_HD" GEBV and "Anc_Not_Imputed" GEBV with an increase in SNP density. Indeed, the mean correlations for the 3Kequi and the 20Kequi SNP chips were respectively 0.8507 and 0.9379. For the LD0.05 and the LD0.8 SNP chips there were respectively 0.7816 and 0.8658. Zhang et al. (2011) showed in simulation studies that compared to the results of a genomic evaluation done with HD SNP chip, the results of genomic evaluations done with low density SNP chips without imputation also decreased. With an effective population size of 100, heritability of 0.5, 241 QTL, and a SNP chip of 10K markers, the relative accuracy of the GBLUP evaluation decreased from 0.88 with 5K markers to 0.69 with only 200 markers.

For both methodologies, there was a consequent decrease in mean correlations compared to the results of the genomic evaluations done with imputed HD genotyping. For the 1Kequi and the 50Kequi SNP chips, both imputed, the results were respectively 0.7833 and 0.9964. Without imputation, the results were respectively 0.6261 and 0.9503. Likewise, for the LD0.05 and the LD0.8 SNP chips with imputation, the results were respectively 0.8661 and 0.9931. Without imputation, the results decreased respectively to 0.7816 and 0.8658. Furthermore, from 20K SNPs, the results for the EQ methodology seemed to reach a mean correlation threshold of 0.95 whereas with imputation the mean correlations were above 0.99. Thus, imputations enabled to increase significantly the mean correlations, mainly for very low density SNP chips. In addition, these results indicate that the ranking of the best 150 individuals of G1 for EW obtained without imputation was quite different from the ranking obtained with HD genotyping. The lower results obtained for very low SNP density indicated that using few SNPs could not be sufficient to accurately rank individuals having very close genomes.

Finally, at equivalent SNP density, a tendency to get higher results with the EQ methodology was observed. Indeed, at 3K SNPs, the difference in mean correlation between 3Kequi and LD0.05 SNP chips was equal to 0.07. The same difference was obtained between 20Kequi and LD0.8 SNP chips. Such differences were higher than with imputation but were not significant. However, we can note that the correlations remained always below 0.90 for the top 150 individuals whatever the SNP density with the LD methodology without imputation. The differences between methodologies are consistent with the genetic determinism of the four traits as explained in the previous part (Harris and Johnson, 2010; Weigel et al., 2010). In addition, the EQ methodology enabled a covering of all chromosomes more optimal than the LD methodology (Herry et al., 2018). With the LD methodology, there were some gaps on chromosomes without SNPs selected on low density SNP chips. With the EQ methodology, the number of gaps was decreased, or at least their size was lower.

Results For The Breeders.

Spearman correlations between "Anc_HD" GEBV and "Anc_Not_Imputed" GEBV were also calculated for the 67 breeders (Figure 6b). For both methodologies, there was an increase in Spearman correlations with an increase in SNP density. At equivalent SNP density, the results for the 3Kequi and 20Kequi SNP chips were respectively 0.9484 and 0.9802. For the LD0.05 and LD0.8 the results were respectively 0.9349 and 0.9665. Compared to the results for the top 150 individuals, the results were better for the 67 breeders as shown previously in the scenario with imputation. Finally, for a SNP density higher than 3K, the mean correlations were above 0.94 for both methodologies, indicating a reordering rather reduced of the 67 breeders. In bovine, Weigel et al. (2009) showed that compared to the top 500 bulls selected from progeny testing, 306 were truly selected with 32K SNPs chosen from the Illumina BovineSNP50 Bead Chip. With 2K equally spaced SNPs, 292 bulls were

chosen. With only 500 equally spaced SNPs, 247 bulls were chosen. This illustrates that compared to the HD SNP chip, the re-ranking was limited and that even with few SNPs, the reordering of the individuals was limited.

Compared to the results obtained with imputation, there was a slight decrease in correlations with "Anc_HD" GEBV. Indeed, for the 1Kequi and 50Kequi SNP chips, the results were respectively 0.9316 (±0.0451) and 0.9983 (±0.0072) with imputation, and 0.8718 (±0.0608) and 0.9815 (±0.0238) without imputation. Likewise, for the LD0.05 and the LD0.8, the results were respectively 0.9777 (±0.0261) and 0.9979 (±0.0080) with imputation, and 0.9349 (±0.0440) and 0.9665 (±0.0318) without imputation. Thus, the differences observed for both methodologies were not significant and the results were still high whatever the SNP chip used. These results were rather different from those obtained by Aliloo et al. (2018). They showed in bovine, for 1034 individuals, that correlations between HD GEBV (on 777K genotypes) and GEBV based on imputed HD genotyping were significantly higher than without imputation. Indeed, according to their MAFI (Minor Allele Frequency within Interval) method which was the closest to our EQ methodology, using 4013 and 25,410 SNPs imputed to 777K SNPs resulted respectively in correlations of 0.9398 and 0.9927. These results decreased dramatically without imputation with correlations of respectively 0.6485 and 0.8598. Such a large decrease was not observed in our study.

Finally, the differences observed between the two methodologies were also not significant. Consequently, the simpler EQ methodology seems to be sufficient to get good genomic evaluation results for traits controlled by many small QTL, which is the case for the four traits studied.

Impact Of Imputation On Relative Accuracy Of Genomic Evaluation

The impact of imputation on the attainable relative accuracy of genomic evaluations was studied by comparing a genomic evaluation "Full_HD" of the 67 G1 breeders using all available information (phenotypes and genotypes) from generation (G0) to (G3) and GEBV of the G1 breeders based on ancestry with imputed HD genotyping ("Anc_Imputed" GEBV), for each low density SNP chip. Only the results for Egg Weight (EW) were shown to simplify the reading and because of the similarity of the results for the other traits.

It was noticed (Figure 7) for the EQ methodology a slight increase in Pearson correlations from very low density SNP chips to 20K SNPs. Indeed, for the 1Kequi and the 20Kequi SNP chips, the mean correlations were respectively 0.4472 and 0.4854. But for the LD methodology, the results were rather stable with mean correlations of respectively 0.4917 and 0.4875 for the LD0.05 and LD0.8 SNP chips. For both methodologies, the results varied slightly up to 20K SNPs. They became steady for the EQ methodology from 20K to higher SNP densities. Finally, for both methodologies, the correlations of "Anc_Imputed" GEBV with "Full_HD" GEBV were not significantly different from those obtained by comparison between true HD GEBV on ancestry and "Full_HD" GEBV. The mean correlation was 0.4848 and corresponded to a theoretical maximum value attainable. The standard error for each low density SNP chip was ± 0.11 indicating that there was no difference with the theoretical maximum value. For information purposes, the mean correlations for ESC, ESS and AH were 0.2618 ± 0.12, 0.4027 ± 0.11 and 0.4802 ± 0.11. This is consistent with the previous results showing a very slight impact of imputations errors on GEBV estimations of the 67 breeders on ascendance. For both methodologies, from a density of 5K SNPs imputed to the HD SNP chip, the mean correlations were above 0.98 between "Anc_HD" GEBV and "Anc_Imputed" GEBV. These results are also in agreement with the literature. Indeed, Harris and Johnson (2010) showed that in bovine, from 5K to 1000K SNPs, the increase in correlations between true phenotypes and predicted phenotypes was very limited (0.62 to

497 0.65). VanRaden et al. (2012) showed that, for 28 traits tested in bovine, in average, the
498 estimated genomic reliability was 61.1% with 300K SNPs and decreased to only 60.7% when
499 they used 45K SNPs. In the study of Wellman et al. (2013), 768 SNPs imputed to the Illumina
500 PorcineSNP60 BeadChip (60K SNPs) led to a negligible loss in genomic evaluation accuracy.
501 Likewise, Chen et al. (2014) estimated in bovine that the accuracy of genomic prediction with
502 observed 50K or imputed 50K (from 6K) genotypes was 0.61 for milk yield and 0.62 for
503 somatic cell score (SCS).

504 However, a decrease in relative accuracy was observed with the 1Kequi SNP chip with
505 a mean correlation of 0.4472. The highest decrease was observed for albumen height (AH)
506 where the mean correlation for the 1Kequi SNP chip was 0.4045 (±0.11) and the theoretical
507 maximum value was 0.4802. One cannot conclude about the significance of this difference
508 but this decrease was also expected because the results regarding the impact of imputation
509 accuracies showed a mean correlation of 0.9316 for the 1Kequi SNP chip. Other studies
510 showed that decreasing to much the SNP density has consequences on genomic evaluation
511 accuracies. Raoul et al. (2017) illustrated this point in Merino sheep where using only 500 or
512 250 SNPs imputed to the Illumina OvineSNP50 BeadChip resulted respectively in a decrease
513 in accuracies from 0.53 (with HD SNP chip) to 0.45 and 0.38. Wellman et al. (2013) showed
514 that 384 SNPs imputed to the Illumina PorcineSNP60 BeadChip led to a loss of 3% in
515 genomic evaluation accuracy. Likewise, Chen et al. (2014) showed that the accuracy of
516 genomic prediction decreased from 0.61 to 0.49 for milk yield and from 0.62 to 0.53 for SCS
517 with imputed 50K genotypes from 384 SNPs.

518 Consequently, we can conclude that the effects of imputation errors on GEBV relative
519 accuracies were very limited even if slightly more important for very low densities.

520

Impact of the Direct Use of Low Density SNP Chips Without Imputation on Relative Accuracy of Genomic Evaluation

The impact of the direct use of low density SNP chips on relative accuracy of genomic evaluation was studied by comparing the "Full_HD" GEBV of the G1 and GEBV of the G1 breeders on ancestry with low density genotyping without imputation ("Anc_Not_Imputed" GEBV), for each low density SNP chip. For both methodologies, only the results for Egg Weight (EW) were shown to simplify the reading and because of the similarity of the results for the other traits.

Both methodologies were rather stable with slight variations in Pearson correlations up to 20K SNPs (Figure 8). The results for the 3Kequi and 20Kequi SNP chips were respectively 0.4471 and 0.4675. For the LD0.05 and LD0.8 the correlations were respectively 0.4583 and 0.4888. However, the standard errors associated to these results were ±0.11 and the correlation between the "Full HD" GEBV and the HD GEBV based on ancestry was 0.4848. This indicates that the differences observed between each low density SNP chip, and consequently between the two methodologies, were not significant. These results are in agreement with the previous results showing a very slight impact of the absence of imputation on GEBV estimation of the 67 breeders on ascendance. However, the results for the 1Kequi was 0.4018 (±0.11). This lower but non-significant result was also expected because the correlation between "Anc_HD" GEBV and "Anc_Not_Imputed" GEBV was lower (0.8718 ±0.0608) than those obtained with higher SNP densities. This was the case for all traits studied.

The results found in the literature are contrasted. Moghaddar et al. (2015) showed in Merino sheep, that the accuracy of genomic prediction based on observed 50K genotypes was 0.446 for post-weaning weight (PWW) and 0.219 for post-weaning eye muscle depth (PW_EMD). Based on genotypes imputed from 12K to 50K genotypes, with imputation

accuracy comprised between 0.88 and 0.99, the accuracy of genomic prediction was 0.443 for PWW and 0.219 for PW_EMD. Based on observed 12K genotypes, the accuracy was 0.412 for PWW and 0.205 for PW_EMD. Thus, the results were slightly better with imputation compared to a direct use of the 12K without imputation, but in both cases, there was not a dramatic decrease in genomic prediction accuracy despite a significant gap of SNP density between HD and low density chips. Weigel et al. (2009) had a gap of SNP density closer to our but the results were rather different. The correlation between the results from progeny testing and the genomic result with a HD SNP chip of 32K was 0.612. With 300, 1K and 2K equally spaced SNPs, the results were respectively 0.253, 0.422 and 0.539. Contrary to the results of Moghaddar et al. (2015), there was a significant decrease in their results with the use of low density SNP chips without imputation. In 2010, they showed that their results were better with imputation.

Finally, for a SNP density higher than 3K, using low density SNP chips without imputation leaded to results as good as those obtained with the HD SNP chip itself.

CONCLUSIONS

This study showed a very limited reordering of the breeders, selected on a multi-traits index, with low density genotyping (with or without imputation) instead of HD genotyping. Indeed, Spearman correlations between GEBV on HD genotyping and GEBV on low density genotyping were always higher than 0.94 with more than 3K SNP. For the top 150 individuals, who are genetically closer than the breeders, the reordering was a bit more important. Thus, the correlations between GEBV with HD genotyping and GEBV with low density genotyping remained below 0.85 with less than 3K SNP with the EQ methodology and less than 16K SNP (LD0.6) with the LD methodology. The differences in GEBV

correlations between the two methodologies were never significant but seemed to indicate that the simpler EQ methodology was sufficient to obtain similar results.

Thus, using directly low density SNP chips designed with the EQ methodology with more than 5K SNPs could enable to get good results of genomic evaluation and could be a cost effective solution for genomic selection. However, only four traits were studied. These four traits were controlled by many small QTL, which explained why the equidistant methodology was more appropriated to realize genomic evaluation with ssGBLUP than the LD methodology, whereas the results on imputation accuracies were inverted. Further investigations on other traits with different genetic architectures should be conducted.

Finally, as shown by Habier et al. (2009), there could be a decrease in genomic evaluation accuracy over the generations with low density genotyping. This would require to genotype at higher density birds selected at each generation to avoid a decrease in genomic evaluation accuracy which could be prejudicial for genomic selection. In addition, in our study, only the males were genotyped but having both parents genotyped could lead to higher genomic evaluation accuracies.

ACKNOWLEDGMENTS

This research project was partly supported by the French national research agency "ANR" within the framework of project ANR-10-GENOM_BTV-015 UtOpIGe. FlH is a PhD fellow supported by the poultry breeding company Novogen as part of a CIFRE thesis between PEGASE INRA's unit, Agrocampus Ouest and Novogen.

REFERENCES

593 Aliloo, H., R. Mrode, A. M. Okeyo, G. Ni, M. E. Goddard, and J. P. Gibson. 2018. The
594 feasibility of using low-density marker panels for genotype imputation and genomic
595 prediction of crossbred dairy cattle of East Africa. J. Dairy Sci. 101:1-20.
596 Atol Ontology. INRA 2012. [http://www.atol-ontology.com]. Accessed 11 Mar 2015.
597 Carvalheiro, R., S. A. Boison, H. H. R. Neves, M. Sargolzaei, F. S. Schenkel, Y. T.
598 Utsunomiya, A. M. P. O'Brien, J. Sölkner, J. C. McEwan, C. P. Van Tassell, T. S.
599 Sonstegard, and J. F. Garcia. 2014. Accuracy of genotype imputation in Nelore cattle.
600 Genet. Sel. Evol. 46:69-79.
601 Chen, L., C. Li, M. Sargolzaei, and F. Schenkel. 2014. Impact of genotype imputation on the
602 performance of GBLUP and Bayesian methods for genomic prediction. PLoS One.
603 9:e101544.
604 Cleveland, M. A., and J. M. Hickey. 2013. Practical implementation of cost-effective genomic
605 selection in commercial pig breeding using imputation. J. Anim. Sci. 91:3583-3592.
606 Dassonneville, R., S. Fritz, V. Ducrocq, and D. Boichard. 2012. Short communication:
607 Imputation performances of 3 low density marker panels in beef and dairy cattle. J.
608 Dairy Sci. 95:4136-4140.
609 Deelen, P., M. J. Bonder, K. J. van der Velde, H. J. Westra, E. Winder, D. Hendriksen, L.
610 Franke, and M. A. Swertz. 2014. Genotype harmonizer: automatic strand alignment and
611 format conversion for genotype data integration. BMC Res. Notes. 7:901-904.
612 Goddard, M. E., B. J. Hayes, and T. H. E. Meuwissen. 2011. Using the genomic relationship
613 matrix to predict the accuracy of genomic selection. J. Anim. Breed Genet. 128:409-
614 421.
615 Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2009. Genomic selection using low-density
616 marker panels. Genetics. 182:343-353.

617 Harris, B. L., and D. L. Johnson. 2010. The impact of high density SNP chips on genomic
618 evaluation in dairy cattle. Interbull Bull. 42:40-43.
619 Hérault, F., J. Yon, F. Herry, S. Allais, and P. Le Roy. 2016. SS4I: select SNP susbset for
620 imputation. (in French). https://prodinra.inra.fr/record/375448.
621 Hérault, F., F. Herry, A. Varenne, T. Burlot, D. Picard-Druet, J. Recoquillay, C. Macé, F.
622 Fagnoul, S. Allais, and P. Le Roy. 2018. A linkage disequilibrium study in layer and
623 broiler commercial chicken populations. Proc. 11th World Congr. Genet. Appl. Livest.
624 Prod., Auckland, New-Zealand.
625 Herry, F., F. Hérault, D. Picard Druet, A. Varenne, T. Burlot, P. Le Roy, and S. Allais. 2018.
626 Design of low density SNP chips for genotype imputation in layer chicken. BMC
627 Genetics. 19:108-121.
628 International Chicken Genome Sequencing Consortium. 2004. Sequence and comparative
629 analysis of the chicken genome provide unique perspectives on vertebrate evolution.
630 Nature. 432:695-716.
631 Kranis, A., A. A. Gheyas, C. Boschiero, F. Turner, L. Yu, S. Smith, R. Talbot, A. Pirani, F.
632 Brew, P. Kaiser, P. M. Hocking, M. Fife, N. Salmon, J. Fulton, T. M. Strom, G.
633 Haberer, S. Weigend, R. Preisinger, M. Gholami, S. Qanbari, H. Simianer, K. A.
634 Watson, J. A. Woolliams, and D. W. Burtet. 2013. Development of a high density 600K
635 SNP genotyping array for chicken. BMC Genomics. 14:59-71.
636 Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and
637 genomic information. J. Dairy Sci. 92:4656-4663.
638 Megens, H. J., R. P. M. A. Crooijmans, J. W. M. Bastiaansen, H. H. D. Kerstens, A. Coster,
639 R. Jalving, A. Vereijken, P. Silva, W. M. Muir, H. H. Cheng, O. Hanotte, and M. A. M.
640 Groenen. 2009. Comparison of linkage disequilibrium and haplotype diversity on
641 macro- and microchromosomes in chicken. BMC Genetics. 10:86-96.

642 Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value
643 using genome-wide dense marker maps. Genetics. 157:1819-1829.
644 Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. H. Lee. 2002. BLUPF90 and
645 related programs (BGF90). Proc. 7th. World Congr. Genet. Appl. Livest. Prod.,
646 Montpellier, France.
647 Moghaddar, N., K. P. Gore, H. D. Daetwyler, B. J. Hayes, and J. H. J. van der Werf. 2015.
648 Accuracy of genotype imputation based on random and selected reference sets in
649 purebred and crossbred sheep populations and its effect on accuracy of genomic
650 prediction. Genet. Sel. Evol. 47:97-108.
651 Mulder, H. A., M. P. L. Calus, T. Druet, and C. Schrooten. 2012. Imputation of genotypes
652 with low-density chips and its effect on reliability of direct genomic values in Dutch
653 Holstein cattle. J. Dairy Sci. 95:876-889.
654 Picard Druet, D., A. Varenne, F. Herry, F. Hérault, S. Allais, T. Burlot and P. Le Roy. 2019.
655 Properties of genomic evaluation for egg quality traits in layers. In submission in Genet.
656 Sel. Evol.
657 Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P.
658 Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A Tool Set for
659 Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum.
660 Genet. 81:559-575.
661 Qanbari, S., M. Hansen, S. Weigend, R. Preisinger, and H. Simianer. 2010. Linkage
662 disequilibrium reveals different demographic history in egg laying chickens. BMC
663 Genetics. 11:103-115.
664 Raoul, J., A. A. Swan, and J. M. Elsen. 2017. Using a very low-density SNP panel for
665 genomic selection in a breeding program for sheep. Genet. Sel. Evol. 49:76-87.

Robert, R., F. Hérault, H. Romé, A. Varenne, H. Chapuis, A. Vignal, T. Burlot, and P. Le Roy. 2015. A linkage disequilibrium study in a layer chicken population. Proc. 9th. Eur. Symp. Poult. Genet., Tuusula, Finland.

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 15:478-489.

Su, G., R. F. Brøndum, P. Ma, B. Guldbrandsten, G. P. Aamand, and M. S. Lund. 2012. Comparison of genomic predictions using medium-density (~54,000) and high-density (~777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. J. Dairy Sci. 95:4657-4665.

VanRaden, P. M., J. R. O'Connell, G. R. Wiggans, and K. A. Weigel. 2011. Genomic evaluation with many more genotypes. Genet. Sel. Evol. 43:10-20.

VanRaden, P. M., J. D. Null, M. Sargolzaei, G. R. Wiggans, M. E. Tooker, J. B. Cole, T. S. Sonstegard, E. E. Connor, M. Winters, J. B. C. H. M. van Kaam, A. Valentini, B. J. Van Doormaal, M. A. Faust, and G. A. Doak. 2012. Genomic imputation and evaluation using high-density Holstein genotypes. J. Dairy Sci. 95:1-11.

Wang, C., D. Habier, B. L. Peiris, A. Wolc, A. Kranis, K. A. Watson, S. Avendano, D. J. Garrick, R. L. Fernando, S. J. Lamont, and J. C. M. Dekkers. 2013. Accuracy of genomic prediction using an evenly spaced, low-density single nucleotide polymorphism panel in broiler chickens. Poult. Sci. 92:1712-1723.

Wang, Y., G. Lin, C. Li, and P. Stothard. 2016. Genotype imputation methods and their effects on genomic predictions in cattle. Springer Sci. Rev. 4:79-98.

Warren, W. C., L. D. W. Hillier, C. Tomlinson, P. Minx, M. Kremitzki, T. Graves, C. Markovic, N. Bouk, K. D. Pruitt, F. Thibaud-Nissen, V. Schneider, T. A. Mansour, C. T. Brown, A. Zimin, R. Hawken, M. Abrahamsen, A. B. Pyrkosz, M. Morisson, V. Fillon, A. Vignal, W. Chow, K. Howe, J. E. Fulton, M. M. Miller, P. Lovell, C. V.

691 Mello, M. Wirthlin, A. S. Mason, R. Kuo, D. W. Burt, J. B. Dodgson, and H. H. Cheng.
692 2017. A new chicken genome assembly provides insight into avian genome structure.
693 G3. 7:109-117.
694 Weigel, K. A., G. de los Campos, O. González-Recio, H. Naya, X. L. Wu, N. Long, G. J. M.
695 Rosa, and D. Gianola. 2009. Predictive ability of direct genomic values for lifetime net
696 merit of Holstein sires using selected subsets of single nucleotide polymorphism
697 markers. J. Dairy Sci. 92:5248-5257.
698 Weigel, K. A., G. de los Campos, A. I. Vazquez, G. J. M. Rosa, D. Gianola, and C. P. van
699 Tassell. 2010. Accuracy of direct genomic values dervied from imputed single
700 nucleotide polymorphism genotypes in Jersey cattle. J. Dairy Sci. 93:5423-5435.
701 Wellmann, R., S. Preuβ, E. Tholen, J. Heinkel, K.Wimmers, and J. Bennewitz. 2013.
702 Genomic selection using low density marker panels with application to a sire line in
703 pigs. Genet. Sel. Evol. 45:28-38.
704 Xu, S. 2003. Estimating polygenic effects using markers of the entire genome. Genetics.
705 163:789-801.
706 Zhang, Z., X. Ding, J. Liu, Q. Zhang, and D. J. de Koning. 2011. Accuracy of genomic
707 prediction using low-density marker panels. J. Dairy Sci. 94:3642-3650.
708
709
710
711
712
713
714
715

Table 1. Summary of the different steps of quality control

Genotypes filtration	**RI Line**
Individual Call Rate (<95%)	8
MAF (=0)	204,122
MAF (<0.05)	54,650
SNP Call Rate (<95%)	7541
Hardy-Weinberg equilibrium ($P<10^{-4}$)	12,538
SNP with unknown location or on chromosome W	1759
Pedigree Incompatibility	0
SNP retained for analyses	**300,351**
Animals retained for analyses	**2362**

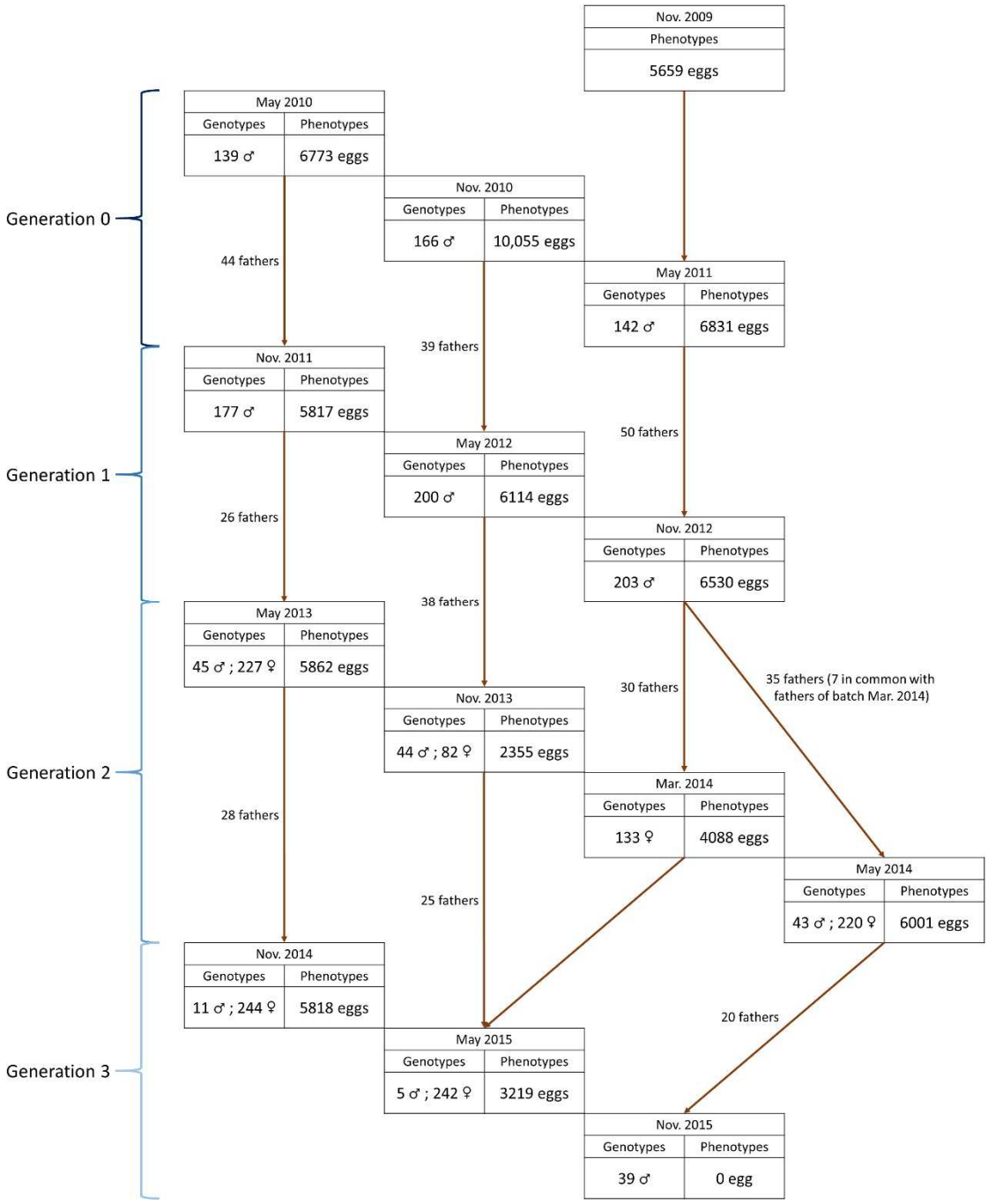


Figure 1. Population structure of the RI line

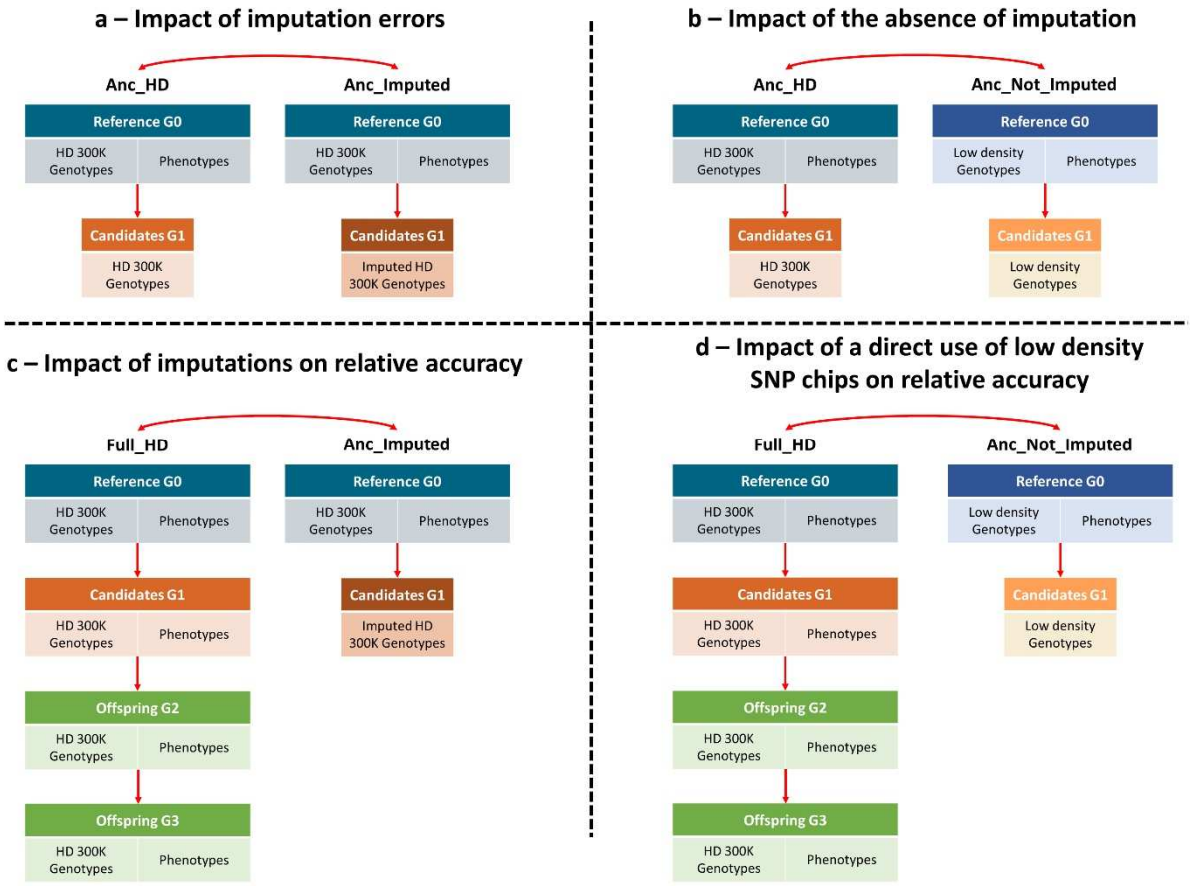


Figure 2. Summary of all different genomic evaluation strategies studied

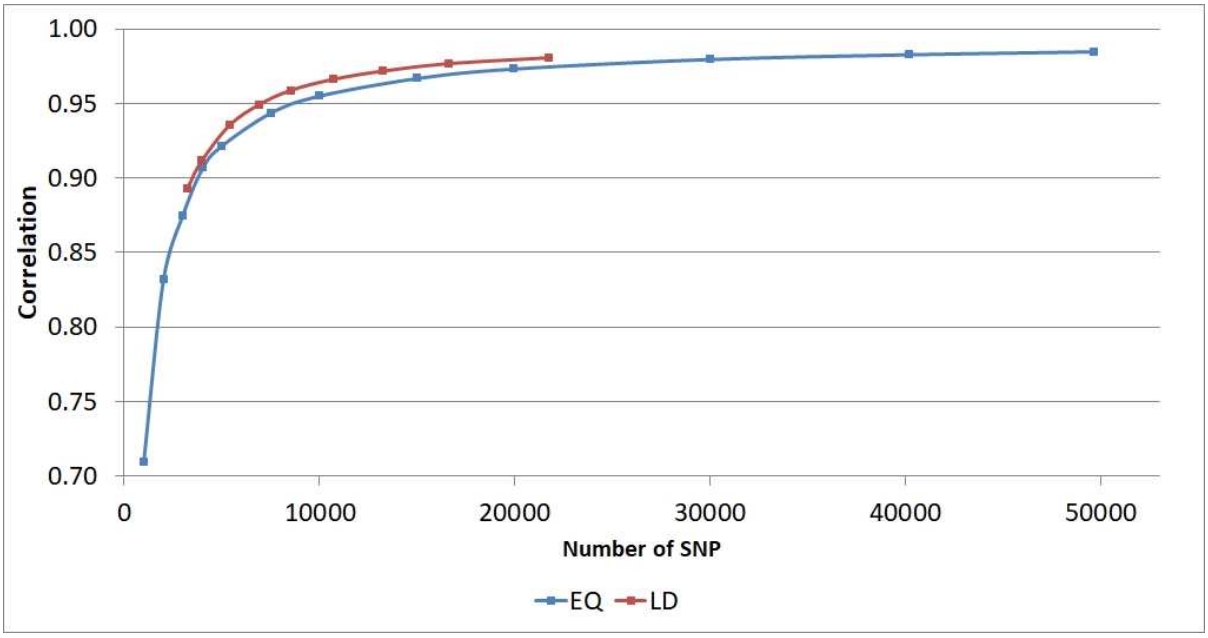
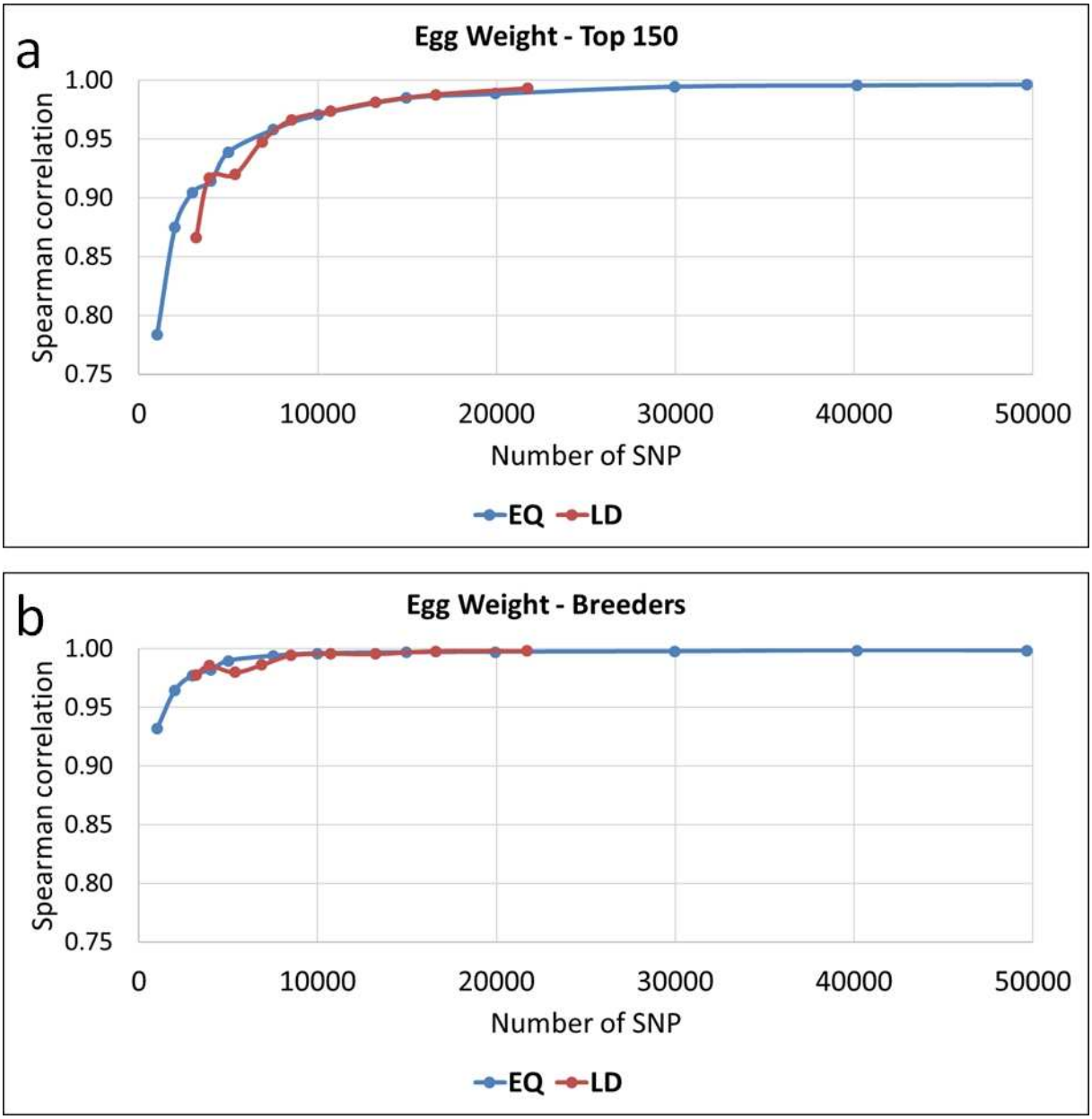


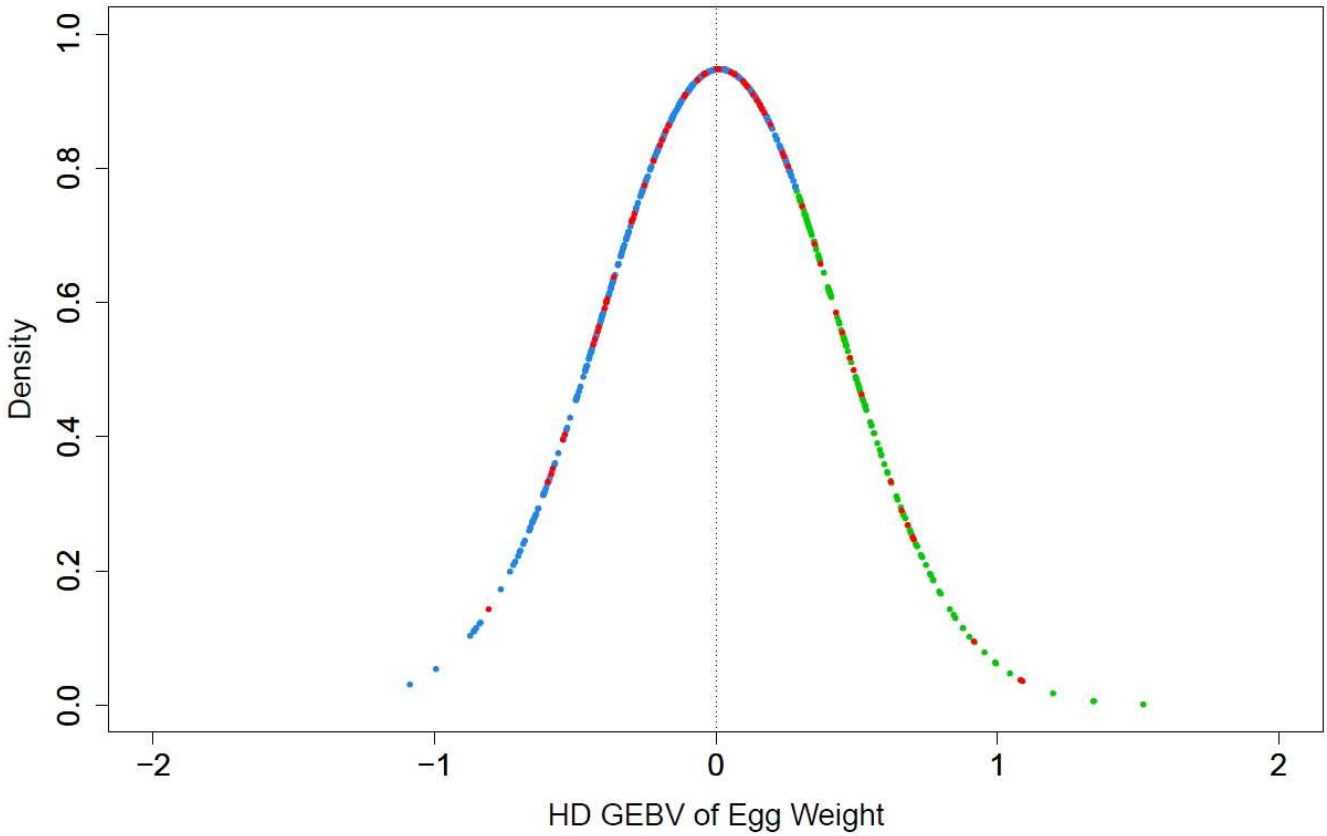
Figure 3. Mean correlations between true and imputed genotypes according to the number of SNPs on low density SNP chips for EQ and LD methodologies.



742

743 **Figure 4.** Spearman correlations between GEBV based on ancestry obtained with true HD
744 genotyping and GEBV based on ancestry obtained with imputed HD genotyping. Results are
745 shown for egg weight and for the top 150 individuals (a) or the 67 breeders (b) according to
746 the number of SNPs on low density SNP chip for both methodologies.

747



748

749 **Figure 5.** Normal distribution of all G1 selection candidates according to their HD GEBV of
750 Egg Weight estimated on ancestry with true HD genotyping. Red dots represent the 67 G1
751 breeders, green dots represent the top 150 individuals for EW, and blue dots represent the
752 other selection candidates.

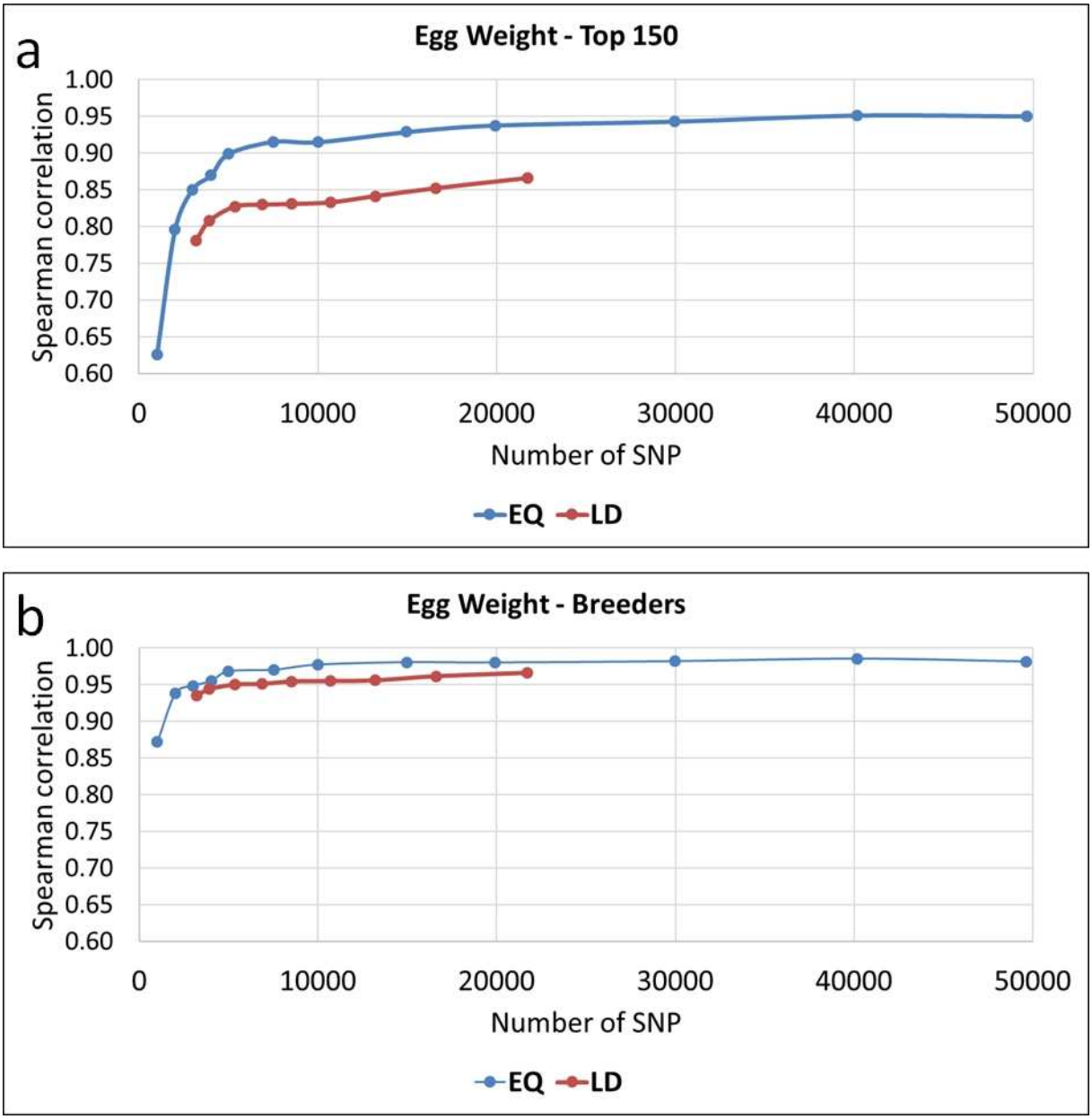


Figure 6. Spearman correlations between GEBV based on ancestry obtained with true HD genotyping and GEBV based on ancestry obtained with low density genotyping (without imputation). Results are shown for egg weight and for the top 150 individuals (a) or the 67 breeders (b) according to the number of SNPs on low density SNP chip for both methodologies.

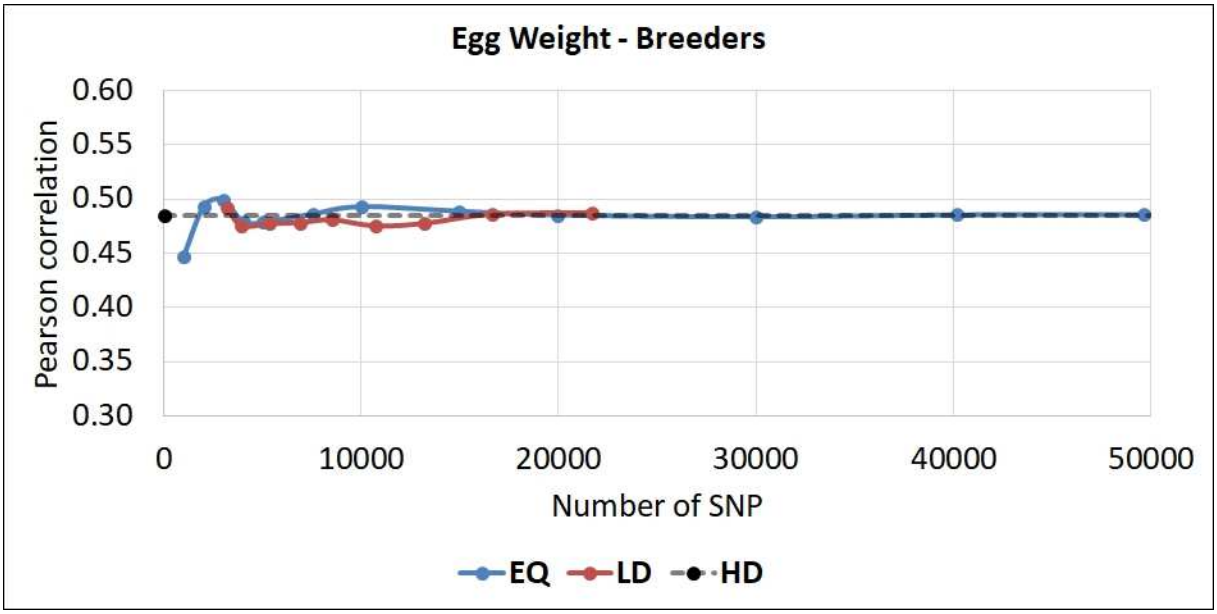


Figure 7. Pearson correlations between "Full_HD" GEBV based on offspring with true HD genotyping and GEBV based on ancestry with imputed HD genotyping. Results are shown for egg weight and for the 67 G1 breeders according to the number of SNPs on low density SNP chip for both methodologies.

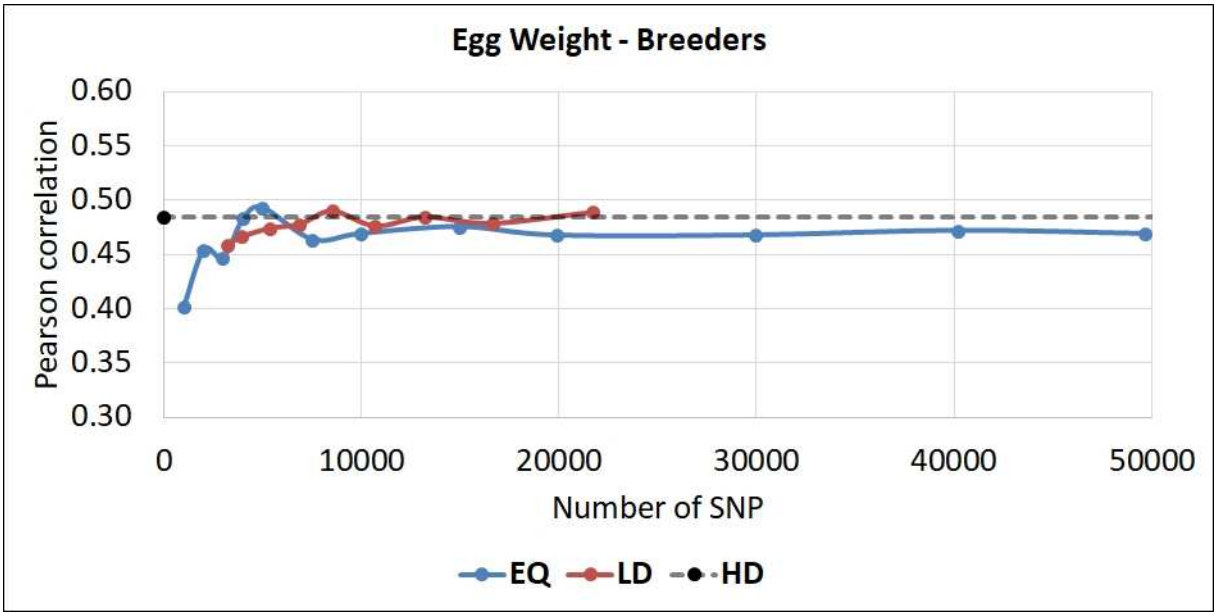


Figure 8. Pearson correlations between "Full_HD" GEBV based on offspring with true HD genotyping and GEBV based on ancestry with low density genotyping (without imputation). Results are shown for egg weight and for the 67 G1 breeders according to the number of SNPs on low density SNP chip for both methodologies.