

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Papers and Publications in Animal
Science

Animal Science Department

2023

Estimation of Breeding Values Using Different Densities of Snp to Inform Kinship in Broiler Chickens

Mayara Salvian

Gabriel Costa Monteiro Moreira

Ândrea Plotzki Reis

Brayan Dias Dauria

Fabrcio Pilonetto

See next page for additional authors

Follow this and additional works at: <https://digitalcommons.unl.edu/animalscifacpub>



Part of the [Genetics and Genomics Commons](#), and the [Meat Science Commons](#)

This Article is brought to you for free and open access by the Animal Science Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Papers and Publications in Animal Science by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Mayara Salvian, Gabriel Costa Monteiro Moreira, Ândrea Plotzki Reis, Brayan Dias Dauria, Fabrício Pilonetto, Izally Carvalho Gervásio, Mônica Corrêa Ledur, Luiz Lehmann Coutinho, Matthew L. Spangler, and Gerson Barreto Mourao

Estimation of Breeding Values Using Different Densities of Snp to Inform Kinship in Broiler Chickens

Mayara Salvian

Universidade de São Paulo Escola Superior de Agricultura Luiz de Queiroz <https://orcid.org/0000-0002-8336-4082>

Gabriel Costa Monteiro Moreira

Universidade de São Paulo Escola Superior de Agricultura Luiz de Queiroz

Ândrea Plotzki Reis

Universidade Federal de Vicosa

Brayan Dias Dauria

Universidade de São Paulo Escola Superior de Agricultura Luiz de Queiroz

Fabrcio Pilonetto

Universidade de São Paulo Escola Superior de Agricultura Luiz de Queiroz

Izally Carvalho Gervásio

Universidade de São Paulo Escola Superior de Agricultura Luiz de Queiroz

Mônica Corrêa Ledur

Empresa Brasileira de Pesquisa Agropecuária Suínos e Aves

Luiz Lehmann Coutinho

Universidade de São Paulo Escola Superior de Agricultura Luiz de Queiroz

Matthew Lee Spangler

University of Nebraska-Lincoln

Gerson Barreto Mourão (✉ gbmourao@usp.br)

<https://orcid.org/0000-0002-0990-4108>

Research article

Keywords: Genomic prediction, high density panel, whole-genomic sequence, imputation, broiler chicken

DOI: <https://doi.org/10.21203/rs.3.rs-32429/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Published in *Livestock Science* 267 (2023) 105124

doi:10.1016/j.livsci.2022.105124

Abstract

Background: Traditionally, breeding values are estimated based on phenotypic and pedigree information using the numerator relationship (**A**) matrix. With the availability of genomic information, genome-wide markers can be included in the estimation of breeding values through genomic kinship. However, the density of genomic information used can impact the cost of implementation. The aim of this study was to compare the rank, accuracy, and bias of estimated breeding values (EBV) for organs [heart (HRT), liver (LIV), gizzard (GIZ), lungs (LUN)] and carcass [breast (BRST), drumstick (DRM) and thigh (THG)] weight traits in a broiler population using pedigree-based BLUP (PBLUP) and single-step genomic BLUP (ssGBLUP) methods using various densities of SNP and variants imputed from whole-genome sequence (WGS).

Results: For both PBLUP and ssGBLUP, heritability estimates varied from low (LUN) to high (HRT, LIV, GIZ, BRST, DRM and THG). Regression coefficients values of EBV on genomic estimated breeding values (GEBV) were similar for both the high density (HD) and WGS sets of SNPs, ranging from 0.87 to 0.99 across scenarios.

Conclusion: Results show no benefit of using WGS data compared to HD array data using an unweighted ssGBLUP. Our results suggest that 10% of the content of the HD array can yield unbiased and accurate EBV.

Background

Traditionally, breeding values are estimated based on phenotypic and pedigree information by pedigree-based BLUP (PBLUP) using the numerator relationship (**A**) matrix [1]. With the advent of genomic selection and the availability of dense SNP arrays, genomic information has been included in the estimation of breeding values. Currently, many genetic evaluation systems have implemented a single-step genomic BLUP (ssGBLUP) [2] approach that makes use of genomic, phenotypic, and pedigree data simultaneously. This approach combines the **A** matrix with the genomic relationship matrix (**G**) into a single kinship matrix (**H**) [3]. The benefit of this approach is in the ability to account for Mendelian inheritance information and thus a more accurate prediction of breeding values can be obtained as compared with PBLUP.

Despite the reduction in the cost of genotyping, it still represents a non-trivial cost. Consequently, the ability to optimize the cost of implementing genomic selection and the rate of genetic gain from having done so is of interest. One potential way to do this is to reduce the proportion of animals genotyped in a strategic manner [4]. Another option is to simply reduce the density of the marker panel used. Theoretically, denser SNP panels lead to an increased probability that any QTL (Quantitative Trait Loci) is in perfect linkage disequilibrium (LD) with a SNP [5]. However, the use of high density (HD) SNP panels in forming a genomic relationship matrix has not been shown to provide significant improvements in accuracy [2]. Despite numerous studies, it is unclear what the optimal density of a SNP panel would be to achieve increased estimated breeding value (EBV) accuracies with minimal genotyping costs.

Recently, efforts have been allocated to whole-genome sequencing (WGS) and using this information to estimate EBV both in real data [6] as well as in simulated data [7–9]. Thus, it is expected that data obtained by sequencing the whole genome include the causal mutations underlying the QTL, which would enable estimating the trait QTL effect regardless of LD between the SNPs and QTL [10]. Performing WGS at moderate to high-depths for every animal in a population would be cost prohibitive to many if not all livestock breeding programs. A less expensive solution would be to genotype individuals with less expensive SNP panels and impute sequence variants throughout the population by only sequencing targeted individuals. Simulated data has shown an increase in genomic prediction accuracy when the causal mutations were included in the analyses [7, 8, 11]. Interestingly, this has not always been the case in real data using cattle and chickens [9, 10, 12].

Although the expectation that genomic selection using HD panels and even WGS data increase prediction accuracy in chickens for traits that are difficult or costly to measure, it is unclear what marker density is sufficient. Therefore, the aim of this study was to compare the rank, accuracy, and degree of bias of estimated breeding values (EBV), as well as, genomic estimated breeding values (GEBV) for organs (heart, liver, lungs and gizzard) and carcass (breast, thigh and drumstick) traits in a broiler population using PBLUP and ssGBLUP with various densities of SNP and variants imputed from whole-genome sequence.

Methods

All experimental protocols related to animals in this study were performed in agreement with the resolution number 010/2012 approved by the Embrapa Swine and Poultry Ethics Committee on Animal Utilization to ensure compliance with international guidelines for animal welfare.

Population And Phenotypes

The chicken population used in this study was derived from a TT broiler line belonging to the Poultry Breeding Program from the Embrapa Swine and Poultry National Research Center. Since 1992, multi-trait selection has been applied in this line, mainly focused on traits such as body weight, feed conversion, carcass weights and yield, fertility, hatchability, and to reduce abdominal fat and metabolic syndromes [13, 14]. The TT reference population is a broiler population developed for genomic studies in 2007 from the mating between 92 females (one from each female family) with

20 males (one from each male family) in a hierarchical scheme (1 male: 5 females) producing approximately 1,500 chickens of both sexes from five hatches. Matings between relatives were avoided to improve the genetic variability as described by Marchesi et al. [15].

A total of 1,453 animals (703 males and 750 females) were weighted at 42 days of age (BW42) after six hours of fasting and euthanized by cervical dislocation followed by bleeding, according to the approval of the Embrapa Swine and Poultry Ethical Committee of Animal Use (CEUA), under protocol number 010/2012. Blood samples from each animal were collected for DNA extraction and the eviscerated carcass was cooled. After four hours of cooling (4 °C), the carcass (breast, drumstick, and thigh) and organs (heart, liver, gizzard and lung) were weighed. More details about the rearing condition and phenotypes measurements are available in Fornari et al. [16].

Descriptive statistics for the carcass and organ traits involved in the study (Table 1) were obtained through the PROC MEANS procedure of SAS® (SAS 9.4, SAS Institute).

Genotypes

Blood samples of each animal (1,453) were used to extract DNA using PureLink® Genomic DNA (Invitrogen, Carlsbad, CA, USA) kit and quantified with Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). After extraction, the diluted genomic DNA was prepared following Affymetrix protocol to perform the genotyping analysis using 600K Affymetrix Axiom Genotyping Array (HD) (Affymetrix, Inc. Santa Clara, CA, USA). This genotyping array was developed using segregating SNPs identified in chicken populations, including four commercial broiler lines, as described by Kranis et al. [17].

Axiom™ Analysis Suite (Affymetrix®) software was used to filter based on DishQC parameter, and then PLINK v.1.9 software [18] was used to perform quality control analysis and genotype calling. Samples that exhibited DishQC of ≥ 0.82 and call rate of $\geq 90\%$ were kept. In order to select markers with high quality, a SNP quality control was applied for removing SNP with call rate lower than 98%, MAF lower than 2% and significant deviations from HWE ($p\text{-value} < 10^{-7}$) leaving 370,608 SNP for further analysis [19].

Imputation

Data from WGS were obtained using the Illumina HiSeq2500® System (Illumina, Inc., San Diego, EUA) with coverage of 10X for 84 animals from Brazilian broiler and layer lines; 14 of those were randomly selected from the 20 males used in the matings to obtain the TT reference population. These data were aligned to Build 5 of the chicken reference genome (Gallus_gallus-5.0) with BWA (version GCA_000002315.3). The read alignment, as well as variant calling and quality control, were performed following the same pipeline adopted by [20] and [19].

After filtering, 12,577,770 SNPs remained in the set of 84 animals sequenced and were used as the reference dataset to impute the HD array to sequence data. Imputation from HD to WGS was performed using BEAGLE 4.1 software [21] with 20 iterations. Imputation accuracy was assessed using the validation subset approach. Sequenced individuals ($n = 84$) were randomly divided into 14 subsets with 6 animals per group and each group was used as validation set once. The imputation process was carried out again for each validation subset masking the SNPs from HD, and then the imputed values for the validation set were compared to their observed values from sequence. Imputation accuracy was defined as the average of squared correlation between observed and predicted variants. The accuracy of imputation was 0.84.

After imputation, a quality control was applied to select the sequence variants with a MAF greater than 0.015 and imputation accuracy equal to or greater than 0.95 (e.g., $r^2 \geq 0.95$) which left 1,421,371 SNPs for further analysis. By using this MAF it is expected that the chance of detecting segregating SNPs is greater, thus it would reduce the cost of genotyping of non-segregating selected SNPs. Furthermore, SNPs were classified into five classes by Variant Effect Predictor (VEP) software (version vep-93.4; [22]) using galGal5 as reference genome. The sequence variants selected to be included in further analysis were UTR3', UTR5', downstream, upstream and intergenic regions of the genome. Genetic variants annotated in those regions were considered potentially functional and thus could have a role in the regulation of the phenotype or even be responsible for controlling gene expression [20]. To ensure that the WGS dataset had the same number of variants as the HD array set, the common genetic variants between those data sets were removed from WGS data, leaving only the non-common variants. After the selection of non-common variants, 1,095,053 SNPs remained to compose the WGS dataset, which consisted of 69% of intergenic regions of the genome, 16% of downstream, 14% upstream and 1% of UTR3' and UTR5', respectively. From the non-common variants, 370,608 SNPs were randomly selected (replicated 10 times) to compose the final WGS dataset and ensure a random representation of the entire genome.

Prediction

Correlations between EBV using PBLUP and ssGBLUP with different SNP subsets were high and ranged from 0.88 to 0.94 for organ traits and from 0.92 to 0.95 for carcass traits. Regression coefficients were similar for both the HD and WGS sets of SNPs, ranging from 0.87 to 0.99 across subsets. For HRT, GEBV tended to be overinflated with regression coefficients ranging from 0.87 to 0.89 when less than 80% of SNPs were used from the HD dataset. The same pattern was observed using imputed sequence data in which the regression coefficients were also less than 1

(Tables 3 and 4). The regression coefficients increased as the proportion of SNP increased reaching a plateau between 5 and 10% of SNP. However, the use of WGS variants did not result in regression coefficients closer to 1 compared to HD panel, except for LUN.

Regarding the predictive ability, traits with higher heritabilities (e.g. GIZ and DRM) showed higher predictive ability than traits with lower heritabilities (e.g. LUN) (Table 5). Compared to PBLUP, the predictive ability of ssGBLUP was greater when at least 5% of SNPs were used (HD and WGS sets), excepted for THG and BW42, where the predictive ability of ssGBLUP was greater than PBLUP when at least 10% of SNPs were used in HD panel for the evaluated traits. However, for THG and BW42 the predictive ability of PBLUP was greater than ssGBLUP when WGS dataset was used.

Table 3

Regression coefficient ($b_{EBV,GEBV}$) and standard error (in brackets) for organ traits using different SNP subsets from the high-density panel (HD Panel) and whole genome sequence (WGS Data).

HD Panel								
SNP (%)	HRT ¹	BW42 ¹	LIV ¹	BW42 ¹	GIZ ¹	BW42 ¹	LUN ¹	BW42 ¹
0.5	0.87 (0.009)	0.94 (0.007)	0.92 (0.010)	0.94 (0.009)	0.93 (0.009)	0.94 (0.009)	0.97 (0.012)	0.94 (0.009)
1	0.88 (0.009)	0.96 (0.007)	0.93 (0.009)	0.95 (0.008)	0.93 (0.008)	0.95 (0.008)	0.98 (0.010)	0.95 (0.008)
5	0.89 (0.008)	0.97 (0.006)	0.93 (0.008)	0.95 (0.007)	0.94 (0.007)	0.95 (0.007)	0.98 (0.009)	0.95 (0.007)
10	0.89 (0.008)	0.97 (0.006)	0.93 (0.008)	0.95 (0.007)	0.94 (0.007)	0.95 (0.007)	0.98 (0.009)	0.95 (0.007)
20	0.89 (0.008)	0.97 (0.006)	0.93 (0.008)	0.95 (0.007)	0.94 (0.007)	0.95 (0.007)	0.98 (0.009)	0.95 (0.007)
40	0.89 (0.008)	0.97 (0.006)	0.93 (0.008)	0.95 (0.007)	0.94 (0.007)	0.95 (0.007)	0.98 (0.009)	0.95 (0.007)
80	0.89 (0.008)	0.97 (0.006)	0.93 (0.008)	0.95 (0.007)	0.94 (0.007)	0.95 (0.007)	0.98 (0.009)	0.95 (0.007)
100	0.89 (0.008)	0.97 (0.006)	0.93 (0.008)	0.95 (0.007)	0.94 (0.007)	0.95 (0.007)	0.98 (0.009)	0.95 (0.007)
WGS Dataset								
0.5	0.87 (0.010)	0.90 (0.007)	0.92 (0.010)	0.94 (0.009)	0.93 (0.009)	0.94 (0.009)	0.98 (0.012)	0.94 (0.009)
1	0.88 (0.009)	0.91 (0.007)	0.92 (0.009)	0.94 (0.008)	0.93 (0.008)	0.94 (0.008)	0.98 (0.011)	0.94 (0.008)
5	0.89 (0.008)	0.93 (0.006)	0.93 (0.009)	0.95 (0.008)	0.94 (0.008)	0.95 (0.008)	0.99 (0.010)	0.95 (0.008)
10	0.89 (0.008)	0.93 (0.006)	0.93 (0.009)	0.95 (0.008)	0.94 (0.008)	0.95 (0.008)	0.99 (0.010)	0.95 (0.008)
20	0.89 (0.008)	0.93 (0.006)	0.93 (0.009)	0.95 (0.008)	0.94 (0.008)	0.95 (0.007)	0.99 (0.009)	0.95 (0.008)
40	0.89 (0.008)	0.93 (0.006)	0.93 (0.008)	0.95 (0.007)	0.94 (0.008)	0.95 (0.007)	0.99 (0.009)	0.95 (0.008)
80	0.89 (0.008)	0.93 (0.006)	0.93 (0.008)	0.95 (0.007)	0.94 (0.008)	0.95 (0.007)	0.99 (0.009)	0.95 (0.008)
100	0.89 (0.008)	0.93 (0.006)	0.93 (0.008)	0.95 (0.007)	0.94 (0.008)	0.95 (0.007)	0.99 (0.009)	0.95 (0.008)
¹ HRT=Heart; LIV = Liver; GIZ = Gizzard; LUN = Lungs; BW42 = Body weight at 42 days of age.								

Table 4

Regression coefficient ($b_{EBV,GEBV}$) and standard error (in brackets) for carcass traits using different SNP subsets from the high-density panel (HD Panel) and whole genome sequence (WGS Data).

HD Panel						
SNP (%)	BRST ¹	BW42 ¹	DRM ¹	BW42 ¹	THG ¹	BW42 ¹
0.5	0.93 (0.009)	0.94 (0.009)	0.92 (0.009)	0.94 (0.009)	0.93 (0.009)	0.94 (0.009)
1	0.94 (0.008)	0.95 (0.008)	0.93 (0.008)	0.94 (0.008)	0.93 (0.008)	0.95 (0.008)
5	0.95 (0.007)	0.95 (0.007)	0.94 (0.008)	0.95 (0.007)	0.94 (0.007)	0.95 (0.007)
10	0.95 (0.007)	0.95 (0.007)	0.94 (0.007)	0.95 (0.007)	0.94 (0.007)	0.95 (0.007)
20	0.95 (0.007)	0.95 (0.007)	0.94 (0.007)	0.95 (0.007)	0.94 (0.007)	0.95 (0.007)
40	0.95 (0.007)	0.95 (0.007)	0.94 (0.007)	0.95 (0.007)	0.94 (0.007)	0.95 (0.007)
80	0.95 (0.007)	0.95 (0.007)	0.94 (0.007)	0.95 (0.007)	0.94 (0.007)	0.95 (0.007)
100	0.95 (0.007)	0.95 (0.007)	0.94 (0.007)	0.95 (0.007)	0.94 (0.007)	0.95 (0.007)
WGS Dataset						
0.5	0.93 (0.009)	0.94 (0.009)	0.93 (0.009)	0.94 (0.009)	0.93 (0.009)	0.94 (0.009)
1	0.93 (0.008)	0.94 (0.008)	0.93 (0.009)	0.94 (0.008)	0.93 (0.009)	0.94 (0.008)
5	0.93 (0.008)	0.95 (0.008)	0.93 (0.008)	0.95 (0.008)	0.94 (0.008)	0.95 (0.008)
10	0.94 (0.007)	0.95 (0.008)	0.94 (0.008)	0.95 (0.008)	0.94 (0.008)	0.95 (0.008)
20	0.94 (0.007)	0.95 (0.008)	0.94 (0.008)	0.95 (0.008)	0.94 (0.008)	0.95 (0.008)
40	0.94 (0.007)	0.95 (0.008)	0.94 (0.008)	0.95 (0.008)	0.94 (0.008)	0.95 (0.008)
80	0.94 (0.007)	0.95 (0.008)	0.94 (0.008)	0.95 (0.008)	0.94 (0.008)	0.95 (0.008)
100	0.94 (0.007)	0.95 (0.008)	0.94 (0.008)	0.95 (0.008)	0.94 (0.008)	0.95 (0.008)
¹ BRST=Breast; DRM = Drumstick; THG = Thigh; BW42 = Body weight at 42 days of age.						

Table 5

Predictive abilities with different SNP subsets from the high-density panel (HD Panel) and whole genome sequence (WGS Data).

HD Panel														
SNP (%)	HRT ¹	BW42 ¹	LIV ¹	BW42 ¹	GIZ ¹	BW42 ¹	LUN ¹	BW42 ¹	BRST ¹	BW42 ¹	DRM ¹	BW42 ¹	THG ¹	BW42 ¹
0.5	0.335	0.325	0.316	0.356	0.401	0.340	0.208	0.334	0.336	0.328	0.369	0.353	0.346	0.352
1	0.342	0.329	0.333	0.368	0.410	0.350	0.219	0.345	0.347	0.338	0.375	0.366	0.359	0.367
5	0.347	0.334	0.349	0.385	0.426	0.369	0.223	0.363	0.360	0.352	0.391	0.380	0.365	0.377
10	0.349	0.332	0.343	0.381	0.425	0.363	0.225	0.360	0.360	0.352	0.391	0.381	0.371	0.379
20	0.350	0.332	0.347	0.382	0.426	0.365	0.224	0.361	0.360	0.352	0.392	0.382	0.370	0.380
40	0.351	0.332	0.346	0.383	0.426	0.365	0.224	0.362	0.360	0.352	0.392	0.383	0.371	0.380
80	0.350	0.332	0.346	0.383	0.427	0.366	0.224	0.362	0.360	0.352	0.392	0.383	0.371	0.380
100	0.350	0.332	0.347	0.383	0.427	0.365	0.224	0.362	0.360	0.352	0.392	0.383	0.371	0.380
PBLUP	0.305	0.333	0.326	0.387	0.399	0.361	0.211	0.360	0.358	0.361	0.363	0.372	0.368	0.381
WGS Data														
0.5	0.329	0.323	0.313	0.355	0.397	0.339	0.199	0.320	0.345	0.328	0.360	0.347	0.342	0.351
1	0.332	0.327	0.331	0.369	0.412	0.351	0.200	0.353	0.350	0.336	0.374	0.366	0.351	0.367
5	0.344	0.329	0.338	0.376	0.416	0.355	0.210	0.360	0.359	0.343	0.385	0.372	0.357	0.370
10	0.344	0.330	0.339	0.378	0.418	0.358	0.220	0.362	0.360	0.344	0.385	0.373	0.358	0.372
20	0.345	0.330	0.339	0.378	0.419	0.358	0.220	0.362	0.361	0.346	0.386	0.373	0.358	0.372
40	0.346	0.330	0.339	0.378	0.420	0.359	0.220	0.362	0.361	0.346	0.386	0.374	0.358	0.373
80	0.346	0.330	0.340	0.378	0.419	0.359	0.220	0.362	0.362	0.346	0.386	0.374	0.359	0.373
100	0.346	0.330	0.340	0.379	0.421	0.359	0.220	0.362	0.361	0.346	0.387	0.374	0.359	0.373
PBLUP	0.305	0.333	0.326	0.387	0.399	0.361	0.211	0.360	0.358	0.361	0.363	0.372	0.368	0.381

¹HRT=Heart; LIV = Liver; GIZ = Gizzard; LUN = Lungs; BRST = Breast; DRM = Drumstick; THG = Thigh; BW42 = Body weight at 42 days of age.

Breeding Value Estimation

Estimated breeding values for the weight of each organ [heart (HRT), liver (LIV), gizzard (GIZ), lungs (LUN)] and carcass trait [breast (BRST), thigh (THG) and drumstick (DRM)], and BW42 were predicted using the BLUPF90 family of programs [23] using three approaches: 1) Pedigree based BLUP (PBLUP), 2) Single-step genomic BLUP (ssGBLUP) using subsets from the HD panel, and 3) ssGBLUP using subsets from WGS. The pedigree used consisted of 2,130 animals, 430 dams and 260 sires. A series of bivariate animal models were fitted with BW42 and each organ and carcass trait. The inclusion of BW42 was to account for the inherent (co)variance between bird weight and the weight of organs or carcass traits. For both methods, the following bivariate animal model was used:

$$\begin{bmatrix} \text{\textit{varvec}}y_1 \\ \text{\textit{varvec}}y_2 \end{bmatrix} = \begin{bmatrix} \text{\textit{varvec}}X_1 & 0 \\ 0 & \text{\textit{varvec}}X_2 \end{bmatrix} \begin{bmatrix} \text{\textit{varvec}}b_1 \\ \text{\textit{varvec}}b_2 \end{bmatrix} + \begin{bmatrix} \text{\textit{varvec}}Z_1 & 0 \\ 0 & \text{\textit{varvec}}Z_2 \end{bmatrix} \begin{bmatrix} \text{\textit{varvec}}u_1 \\ \text{\textit{varvec}}u_2 \end{bmatrix} + \begin{bmatrix} \text{\textit{varvec}}e_1 \\ \text{\textit{varvec}}e_2 \end{bmatrix}$$

where y_1 and y_2 are the vector of observation for each evaluated trait (y_1) and BW42 (y_2); X_1 and X_2 are the design matrices for fixed effects; b_1 and b_2 are the vectors of fixed effects (sex and hatch) for the first and second trait, respectively; Z_1 and Z_2 are the design matrices for random effects; u_1 and u_2 are the vector of random additive genetic effects; e_1 and e_2 are the vector of random error effects. The vector of genetic effects, $u = [u_1, u_2]$, was assumed to be multivariate normal distributed with mean 0 and variance $\text{\textit{varvec}}A^{-1} \otimes \mathbf{F}$ and $\text{\textit{varvec}}H^{-1} \otimes \mathbf{F}$ for PBLUP and ssGBLUP, respectively, where \otimes is the Kronecker product and \mathbf{F} is the additive genetic (co)variance matrix of trait 1 and 2. The vector of residuals, $e = [e_1, e_2]$, was assumed to be multivariate normal distributed with mean 0 and variance $\text{\textit{varvec}}I \otimes \text{\textit{varvec}}R$, where $\text{\textit{varvec}}I$ was an identity matrix and $\text{\textit{varvec}}R$ was the residual (co)variance matrix.

The **H** matrix combines information from numerator relationship matrix (**A**) and genomic matrix (**G**). The inverse of **H** was calculated following the approach of Aguilar et al. [24] as:

$$\text{varvec}H^{-1} = \text{varvec}A^{-1} \begin{bmatrix} 0 & 0 \\ 0 & \text{varvec}G^{-1} - \text{varvec}A_{22}^{-1} \end{bmatrix}$$

where \mathbf{A}^{-1} is the inverse of a numerator relationship matrix; \mathbf{G}^{-1} is the inverse of a blended genomic matrix; and \mathbf{A}_{22}^{-1} is the inverse of a pedigree-based relationship matrix for genotyped animals only. The **G** blended matrix was obtained as follows:

$$\text{varvec}G = 0.95\text{varvec}G_{\text{varvec}w} + 0.05\text{varvec}A_{22}$$

where: $\text{varvec}A_{22}$ is the pedigree-based relationship matrix for genotyped animals only; G_w is the genomic matrix obtained following [12, 25]:

$$\text{varvec}G_{\text{varvec}w} = \frac{\text{varvec}M\text{varvec}M}{2 \sum p_i(1 - p_i)}$$

where: *M* is the SNP matrix, coded as 0, 1 or 2, and p_i is the allelic frequency for the i^{th} SNP.

Assessment Of Accuracy And Bias

Spearman correlations between EBV from PBLUP and GEBV for ssGBLUP using genotypes from eight subsets (0.5%, 1%, 5%, 10%, 20%, 40%, 80% and 100% of SNPs) from both the HD and WGS imputed set were calculated to determine the impact of using reduced subsets of SNP on EBV rank.

Approximately one-third of the animals had their phenotypes masked and were chosen to be in the validation set. These animals were randomly selected, and three subsets were created to ensure that all the animals were in the validation set once. Predictive ability of EBV, from PBLUP and ssGBLUP, was defined as the correlation (*r*) between EBV and phenotypes corrected for fixed effects (y^*) for animals in the validation set for each trait [26]:

$$r = \text{cor}(\text{EBV}, y^*)$$

Moreover, the regression coefficients of EBV on GEBV in each scenario were calculated to evaluate the degree of inflation/deflation of GEBV.

Results

Descriptive statistics

The descriptive statistics of phenotypic data are presented in Table 1. Imputation accuracy estimated by Beagle and assessed using the validation subset approach was 0.84 and ranged from 0.79 to 0.88. After filtering, the distribution of MAF for the HD array was uniform while the MAF distribution for WGS variants retained for further analyses was not (Fig. 1). The estimates of variance components and heritability for organ and carcass traits from PBLUP and ssGBLUP using both HD and WGS subsets are given in Table 2. Small numerical differences in parameter estimates were observed between PBLUP and ssGBLUP. Heritability estimates and their standard errors were lower when genomic information was used.

Table 1

Number of observations (N), mean, standard deviation (SD), minimum (MIN) maximum (MAX) and coefficient of variation (CV) values of carcass and organ traits of broiler chickens.

Trait ¹	N	Mean	SD	Min	Max	CV, %
Organ weight (g)						
HRT	1421	12.34	2.15	6.30	19.70	17.35
LIV	1422	52.34	8.73	25.40	82.40	16.68
GIZ	1423	32.00	6.04	17.80	56.10	18.86
LUN	1430	15.31	3.06	6.60	24.60	19.98
Carcass weight (g)						
BW42	1452	2223.86	260.24	988	2971.00	11.70
BRST	1426	500.76	63.48	211.30	710.80	12.68
DRM	1421	205.87	31.24	86.20	306.60	15.17
THG	1427	310.49	46.15	113.60	464.40	14.85
¹ HRT=Heart; LIV = Liver; GIZ = Gizzard; LUN = Lungs; BW42 = Body weight at 42 days of age; BRST = Breast; DRM = Drumstick; THG = Thigh.						

Table 2

Additive genetic variance (σ_a^2), residual variance (σ_e^2), phenotypic variance (σ_p^2) and heritability estimates (h^2), with their respective standard errors (in brackets) for organ and carcass traits of broiler chickens using PBLUP and ssGBLUP from the high-density (HD) panel and whole genome sequence (WGS) Dataset.

Trait ¹	$\sqrt{\text{varvec}\sigma_a^2\text{varvec}}$	$\sqrt{\text{varvec}\sigma_e^2\text{varvec}}$	$\sqrt{\text{varvec}\sigma_p^2\text{varvec}}$	$\sqrt{\text{varvec}h^2\text{varvec}}$
PBLUP				
Organ weight				
HRT	1.01 (0.24)	1.77 (0.16)	2.78 (0.14)	0.37 (0.07)
LIV	23.49 (5.53)	44.93 (3.68)	68.43 (3.37)	0.34 (0.07)
GIZ	15.42 (3.22)	17.33 (1.95)	32.75 (1.82)	0.47 (0.07)
LUN	1.09 (0.34)	5.10 (0.29)	6.19 (0.26)	0.17 (0.05)
Carcass weight				
BRST	1306.2 (291.47)	1854.3 (182.81)	3160.5 (167.51)	0.41 (0.07)
DRM	205.37 (47.73)	266.45 (28.76)	471.82 (26.10)	0.43 (0.08)
THG	522.31 (120.08)	801.08 (76.23)	1323.4 (69.35)	0.39 (0.07)
HD Panel				
Organ weight				
HRT	1.12 (0.18)	1.71 (0.11)	2.83 (0.13)	0.39 (0.04)
LIV	22.37 (3.89)	45.86 (2.60)	68.23 (3.15)	0.33 (0.04)
GIZ	14.12 (2.05)	18.32 (1.15)	32.44 (1.60)	0.43 (0.05)
LUN	0.97 (0.26)	5.20 (0.25)	6.17 (0.25)	0.16 (0.04)
Carcass weight				
BRST	1046.1 (181.66)	2054.40 (118.67)	3100.50 (145.27)	0.34 (0.05)
DRM	162.49 (27.50)	296.13 (17.47)	458.62 (21.77)	0.35 (0.05)
THG	431.62 (75.76)	870.23 (49.93)	1301.90 (60.78)	0.33 (0.05)
WGS Dataset				
Organ weight				
HRT	1.04 (0.17)	1.77 (0.10)	2.81 (0.13)	0.37 (0.04)
LIV	21.52 (3.78)	46.58 (2.56)	68.11 (3.13)	0.31 (0.04)
GIZ	13.54 (1.97)	18.78 (1.13)	32.33 (1.58)	0.42 (0.04)
LUN	0.92 (0.25)	5.24 (0.25)	6.16 (0.25)	0.15 (0.04)
Carcass weight				
BRST	1009.7 (173.80)	2073.2 (115.28)	3082.90 (143.17)	0.33 (0.05)
DRM	153.64 (26.20)	302.69 (17.08)	456.33 (21.40)	0.34 (0.05)
THG	395.15 (70.96)	895.75 (48.83)	1290.70 (59.10)	0.31 (0.05)
¹ HRT=heart weight (g); LIV = liver weight (g); GIZ = gizzard weight (g); LUN = lung weight (g); BRST = breast weight (g); DRM = drumstick weight (g); THG = thigh weight (g).				

Prediction

Correlations between EBV using PBLUP and ssGBLUP with different SNP subsets were high and ranged from 0.88 to 0.94 for organ traits and from 0.92 to 0.95 for carcass traits. Regression coefficients were similar for both the HD and WGS sets of SNPs, ranging from 0.87 to 0.99 across subsets. For HRT, GEBV tended to be overinflated with regression coefficients ranging from 0.87 to 0.89 when less than 80% of SNPs were used from the HD dataset. The same pattern was observed using imputed sequence data in which the regression coefficients were also less than 1

(Tables 3 and 4). The regression coefficients increased as the proportion of SNP increased reaching a plateau between 5 and 10% of SNP. However, the use of WGS variants did not result in regression coefficients closer to 1 compared to HD panel, except for LUN.

Regarding the predictive ability, traits with higher heritabilities (e.g. GIZ and DRM) showed higher predictive ability than traits with lower heritabilities (e.g. LUN) (Table 5). Compared to PBLUP, the predictive ability of ssGBLUP was greater when at least 5% of SNPs were used (HD and WGS sets), excepted for THG and BW42, where the predictive ability of ssGBLUP was greater than PBLUP when at least 10% of SNPs were used in HD panel for the evaluated traits. However, for THG and BW42 the predictive ability of PBLUP was greater than ssGBLUP when WGS dataset was used.

Discussion

In the present study we chose SNPs at random to be part of reduced subsets to investigate the genomic prediction of carcass and organ traits in broiler chickens. Although the markers were not equally spaced during the selection process, they were present in at least one chromosome across all genome.

MAF distribution

The distribution of MAF for the HD array was uniform while the MAF distribution for WGS variants retained for further analyses was not (Figure 1). Unlike other studies [10, 12], the variants used in the current study did not show a U-shaped MAF distribution for WGS data. In accordance with Ni et al. [27], which also found a non-U-shaped MAF distribution for sequence data in layer chickens, this distribution in WGS data may have occurred due to two possible reasons. First, some of the rare SNPs in the sequenced animals were removed during the imputation process as a result of poor imputation accuracy of SNPs with low MAF. Second, these same rare SNPs were not available in all animals in the population.

Heritability for pedigree-based and genomic models

Estimates of variance components and heritability for carcass and organ traits obtained through PBLUP and ssGBLUP are provided in Table 2. The heritability estimates varied from low (LUN) to moderate (HRT, LIV and THG) and high (GIZ, BRST and DRM) and the standard errors associated with those estimates were low.

Pedigree-based heritability estimates have been reported in the literature for most of the traits used in this study. Using the same population (Embrapa TT), Venturini et al. [14] reported similar pedigree-based heritability estimates for LIV (0.33 ± 0.07), GIZ (0.44 ± 0.08) and BRST (0.37 ± 0.07) to those reported herein. However, the heritability estimate found in this study for DRM (0.43 ± 0.08) and THG (0.39 ± 0.07) was than the estimate in Venturini et al. [14] (0.35 ± 0.07 for DRM and 0.29 ± 0.06 for THG). THG and DRM are commonly analyzed together as a leg trait, so heritability estimates for those traits are scarce in the literature. Heritability estimates for leg in chicken were reported by Argento et al. [28] (0.34), Rance et al. [29] (0.48 ± 0.07) and Gaya et al. [30] (0.33 ± 0.03). In a study with a male broiler line, Gaya et al. [30] reported heritability estimates for HRT (0.38 ± 0.04), LIV (0.25 ± 0.03), GIZ (0.39 ± 0.04) and BRST (0.33 ± 0.03). Rance et al. [29] reported heritability estimates for HRT (0.30 ± 0.08), LIV (0.08 ± 0.06), GIZ (0.52 ± 0.10) and BRST (0.59 ± 0.08).

The heritability estimates for LUN in broiler chickens are not common in the literature. Using an F_2 experimental population, Ledur et al. [31] reported similar pedigree-based heritability estimates for LUN (0.10) to the result reported herein. Although LUN is not considered an economically important trait, it has been related to pulmonary hypertension (e.g. ascites). Heritability estimates for ascites have been reported by several authors [32–36]. The use of a multi-trait model may be responsible for higher estimates of heritability reported in this study compared to the literature since multi-trait models use additional genetic information from links with other traits [37].

Heritability was also estimated using the **H** matrix instead of the numerator relationship (**A**) matrix which resulted in relatively small differences between the estimates (Table 2). In general, the genomic heritability is smaller than heritability estimated using only the pedigree and phenotypic information [38].

Correlation between EBV and GEBV

Across all traits, EBV estimated with at least 0.5% of SNP were highly correlated with EBV estimated from the complete HD (minimum correlation of 0.94) and the minimum average correlation between 0.5% of SNP and PBLUP was 0.89. Indeed, lower correlations were observed when a smaller number of SNP sets were used, but correlations between predicted breeding values were higher when the genomic matrix was incorporated in the analyses, regardless the SNP set selected. A similar pattern was observed with WGS subsets.

Comparing the correlations between EBV using different SNP subsets showed no differences when 10% or 100% of SNPs (mean correlation 0.99) were used in the analyses which suggests that the use of an evenly-spaced lower-density panel could provide a very similar ranking of EBV at a potentially lower cost, as proposed by Habier et al. [39]. When applied to Japanese black cattle, Ogawa et al. [40] suggested that using at least 4,000 equally spaced SNPs was sufficient for genetic prediction for carcass weight and marbling score. Using a 50K chip, Rolf et al. [41] reported that between 2,500 and 10,000 SNPs distributed throughout the genome could be used to form a **G** matrix to accurately predict EBV for feed

efficiency in Angus cattle. The current study supports the use of reduced subsets of SNP, demonstrated in other species, for genomic prediction in broiler chickens.

Regression coefficients

The regression coefficients of EBV on GEBV quantifies the bias in the variance of the estimated breeding value for each SNP subset (Tables 3 and 4). Regression coefficients were similar for both the HD and WGS sets of SNPs, ranging from 0.82 to 0.99 across subsets. The regression coefficients increased as the proportion of SNP increased reaching a plateau between 5 and 10% of SNPs. Overall, all the regression coefficient values were the same when HD or WGS was used. In practice, regression coefficient equal to one indicates no bias. Except for LUN, our results showed regression coefficients lower than one for both HD and WGS sets, which means the variance of breeding values were overestimated. However, according to Tsuruta et al. [42] deviations of $\pm 15\%$ from unity are acceptable.

Using a pure layer line, Yan et al. [43] used bias to compare PBLUP and ssGBLUP prediction methods. These authors used the regression coefficients of phenotypes corrected for fixed effects on predicted (G)EBV and reported that EBVs from ssGBLUP were less biased than those from PBLUP. Also, in layer lines, Heidaritabar et al. [12] reported high regression coefficients (greater than 1) when PBLUP or GBLUP methods were used. According to these authors, regardless the incorporation of genotypes from a 60K SNP panel and sequence data the regression coefficients remained greater than 1, indicating an underestimation of the breeding value variance.

Bias differences among the methods may be explained by directional selection [44]. In the present study, multi-trait selection has been applied in this line, mainly focused on traits such as body weight, feed conversion, carcass weights and yield, fertility, hatchability, and to reduce abdominal fat and metabolic syndromes [13, 14].

Predictive ability

Predictive abilities across all traits are reported in Table 5. Compared to PBLUP, the predictive ability assessed by ssGBLUP was higher for most of the traits when at least 5% of SNPs were used. The incorporation of sequencing variants is generally thought to have the potential to improve predictive abilities, since it is expected that a high proportion of genetic variation may be explained when a high-density panel or even sequencing data are used. Although WGS increases the number of markers, most of them are in incomplete LD with causal mutations. Variants in incomplete LD with causal mutations limited the increase of prediction abilities, thus the use of variants in strong LD with causal mutations could be useful in improving the accuracy of genomic prediction [45].

The concept of reducing SNP density as a solution for genotyping costs has been widely reported [39–41, 46, 47]. While accuracy estimates using a genomic relationship matrix appear to be better than those from a pedigree relationship matrix, our results show no difference in genomic prediction accuracy when a reduced number of SNPs were used to fit the genomic relationship matrix, indicating that at least 10% of SNP from the HD panel (~37,000 SNPs) can be used in genomic evaluation. A previous study with dairy cattle [48] has shown small genomic prediction gains with the increased marker density from medium density (~54,000K) to HD (~777,000K). Agreeing with our findings, similar results were also obtained in pigs by Zhang et al. [37] which increasing marker density (80K, 650K and WGS) had a little or no advantage in genomic prediction for feed intake.

Simulated data has shown an increase in genomic prediction accuracy when the causal mutations were included in the analyses [7, 8, 11]. Contrary to those findings, our study showed no significant increase in prediction accuracy when using WGS variants as opposed to SNP from the HD. Other authors have also observed lower or no significant benefits in predictive ability gain using sequence data comparing with SNP arrays [6, 9, 10, 12, 45, 49, 50].

The infinitesimal model used herein (ssGBLUP), whereby all markers are assumed to have an effect and a common variance, showed no significant increase in prediction accuracy using WGS variants as compared to the HD markers. In a simulated study, Clark et al. [51] suggested that the increase of genomic prediction accuracy would be small when the trait is highly polygenic especially in a small reference population. Furthermore, false positives, including sequencing, alignment and calling errors, which are not included in simulated analysis but are present in real data, can also be responsible for these results [49].

Another possible reason is related to the population structure. When a small effective population size undergoes selection for an extended period of time no significant gains in prediction accuracy are obtained regardless of using HD panel or WGS dataset [11]. Thus, in highly selected population, almost all of the genetic variance can be explained by the SNPs genetic variance as result of the relationship between individuals [52].

Although imputation accuracy was not the main objective of the present study, it can help to explain why sequence data did not show a superior predictive ability compared with the HD panel. In our study, the average of imputation accuracy assessed using the validation subset approach was 0.84. Although this value is suitable, possible errors in the genomic map might be responsible for reducing the imputation accuracy since those errors may decrease the accuracy of prediction and interfere in the detection of causal mutations [53].

Conclusions

In the current study, we investigated different SNP panel densities to inform kinship for single-step genomic BLUP in a broiler population. Our results show no difference when 10% or 100% of SNPs were used to inform kinship in the prediction of breeding values, suggesting that at least 10% of SNP (~ 37,060) can provide decent genetic evaluations. Therefore, the use of low-density arrays, if at a low cost, could be applied to reliably rank individuals based on genetic merit. Furthermore, the results also demonstrated no benefit of using WGS data compared to HD array data using unweighted ssGBLUP. The use of a weighted genomic matrix may prove useful to exploit causal variants from whole-genome sequence data.

Abbreviations

PBLUP
Pedigree-based BLUP
ssGBLUP
Single-step Genomic BLUP
LD
Linkage Disequilibrium
SNP
Single Nucleotide Polymorphism
EBV
Estimated Breeding Values
GEBV
Genomic Estimated Breeding Values
HD
High Density
WGS
Whole-Genome Sequencing
QTL
Quantitative Trait Loci
MAF
Minor Allele Frequency
HWE
Hardy-Weinberg Equilibrium
VEP
Variant Effect Predictor
HRT
Heart weight (g)
LIV
Liver weight (g)
GIZ
Gizzard weight (g)
LUN
Lungs weight (g)
BRST
Breast weight (g)
THG
Thigh weight (g)
DRM
Drumstick weight (g)
BW42
Body Weight at 42 days of age (g)

Declarations

Ethics approval and consent to participate

All experimental protocols related to animal experimentation in this study were performed in agreement with the resolution number 010/2012 approved by the Embrapa Swine and Poultry Ethics Committee on Animal Utilization to ensure compliance with international guidelines for animal

welfare.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are public and included in this published article. The datasets used and/or analyzed during the current study (genotypes and phenotypes) are available from the corresponding author on reasonable request.

Competing interests

Not applicable.

Funding

This study was supported by the Brazilian Agricultural Research Corporation – Embrapa (project number 01.11.07.002.04.02) and by the thematic project (2014/08704–0) from São Paulo Research Foundation (FAPESP). The TT Reference Population was funded by the National Council of Scientific and Technological Development (CNPq) grant number 481755/2007–1 from the Brazilian Government. G.C.M. Moreira received fellowships from FAPESP, grants 14/21380–9 (in cooperation agreement with CAPES), and 16/00569–1. M. Salvian received a fellowship from CNPq, grants 206489/2017-0. L.L. Coutinho and M.C. Ledur are recipients of productivity fellowship from CNPq.

Authors' contributions

MS, GCMM, MCL, LLC, MLS and GBM conceived the idea of this research and participated in the experimental design. MS, GCMM and APR performed data analysis. MS drafted the manuscript. MS, GCMM, APR, MCL, LLC, MLS and GBM collaborated with interpretation, discussion and writing of the manuscript. MCL, LLC, MLS and GBM provided the experimental environment, phenotype and data analysis support. All authors have read and approved the final manuscript.

Acknowledgments

The authors would like to thank Embrapa Swine & Poultry for providing the genotype, phenotype and pedigree data. We also thank Animal Biotechnology Laboratory at University of São Paulo (ESALQ-USP) for sequencing and for genotyping the animals. The authors would like to acknowledge the collaborative efforts among Embrapa, University of São Paulo, and University of Nebraska – Lincoln; especially the Animal Science Department and the Holland Computer Center (HCC) for providing resources to perform the analysis.

Authors' information

Not applicable.

References

1. Henderson CR. Applications of linear models in animal breeding. 1984.
2. Misztal I, Aggrey SE, Muir WM. Experiences with a single-step genome evaluation. *Poult Sci*. 2013.
3. Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. *J Dairy Sci*. 2009.
4. Howard JT, Rathje TA, Bruns CE, Wilson-Wells DF, Kachman SD, Spangler ML. The impact of truncating data on the predictive ability for single-step genomic best linear unbiased prediction. *J Anim Breed Genet = Zeitschrift fur Tierzucht und Zuchtungsbiologie*. 2018.
5. Meuwissen T, Hayes B, Goddard M. Genomic selection: A paradigm shift in animal breeding. *Anim Front*. 2016;6 x:6. doi:10.2527/af.2016-0002.
6. Ober U, Ayroles JF, Stone EA, Richards S, Zhu D, Gibbs RA, et al. Using whole-genome sequence data to predict quantitative trait phenotypes in *Drosophila melanogaster*. *PLoS Genet*. 2012.
7. Meuwissen T, Goddard M. Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics*. 2010.
8. Druet T, Macleod IM, Hayes BJ. Toward genomic prediction from whole-genome sequence data: Impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity (Edinb)*. 2014.
9. MacLeod IM, Bowman PJ, Vander Jagt CJ, Haile-Mariam M, Kemper KE, Chamberlain AJ, et al. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics*. 2016.
10. Van Binsbergen R, Calus MPL, Bink MCAM, Van Eeuwijk FA, Schrooten C, Veerkamp RF. Genomic prediction using imputed whole-genome sequence data in Holstein Friesian cattle. *Genet Sel Evol*. 2015.

11. MacLeod IM, Hayes BJ, Goddard ME. The effects of demography and long-term selection on the accuracy of genomic prediction with sequence data. *Genetics*. 2014.
12. Heidaritabar M, Calus MPL, Megens HJ, Vereijken A, Groenen MAM, Bastiaansen JWM. Accuracy of genomic prediction using imputed whole-genome sequence data in white layers. *J Anim Breed Genet*. 2016;133:167–79.
13. Nones K, Ledur MC, Zanella EL, Klein C, Pinto LFB, Moura ASAMT, et al. Quantitative trait loci associated with chemical composition of the chicken carcass. *Anim Genet*. 2012.
14. Venturini GC, Cruz VAR, Rosa JO, Baldi F, El Faro L, Ledur MC, et al. Genetic and phenotypic parameters of carcass and organ traits of broiler chickens. *Genet Mol Res*. 2014.
15. Marchesi JAP, Buzanskas ME, Cantão ME, Ibelli AMG, Peixoto JO, Joaquim LB, et al. Relationship of runs of homozygosity with adaptive and production traits in a paternal broiler line. *Animal*. 2018;12:1126–34.
16. Fornari MB, Zanella R, Ibelli AMG, Fernandes LT, Cantão ME, Thomaz-Soccol V, et al. Unraveling the associations of osteoprotegerin gene with production traits in a paternal broiler line. *Springerplus*. 2014;3:1–8.
17. Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, Smith S, et al. Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics*. 2013.
18. Purcell SM, Neale BM, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet*. 2007.
19. Moreira GCM, Boschiero C, Cesar ASM, Reecy JM, Godoy TF, Trevisoli PA, et al. A genome-wide association study reveals novel genomic regions and positional candidate genes for fat deposition in broiler chickens. *BMC Genomics*. 2018.
20. Boschiero C, Moreira GCM, Gheyas AA, Godoy TF, Gasparin G, Mariani PDSC, et al. Genome-wide characterization of genetic variants and putative regions under selection in meat and egg-type chicken lines. *BMC Genomics*. 2018.
21. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet*. 2008.
22. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016.
23. Misztal I, Tsuruta S, Lourenco D, Aguilar I, Legarra A, Vitezica Z. Manual for BLUPF90 family of programs. 2018.
24. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score¹. *J Dairy Sci*. 2010.
25. VanRaden PM. Efficient Methods to Compute Genomic Predictions. *J Dairy Sci*. 2008.
26. Legarra A, Robert-Granié C, Manfredi E, Elsen JM. Performance of genomic selection in mice. *Genetics*. 2008.
27. Ni G, Caverio D, Fangmann A, Erbe M, Simianer H. Whole-genome sequence-based genomic prediction in laying chickens with different genomic relationship matrices to account for genetic architecture. *Genet Sel Evol*. 2017;49:1–14.
28. Argentão C, Filho TM, Marques JLB, Souza EM, Eler JP, Ferraz JBS. Genetic and phenotypic parameters of growth and carcass traits of a male line of broilers raised in tropical conditions. *7th World Congr Genet Appl to Livest Prod*. 2002;;2–5.
29. Rance KA, McEntee GM, McDevitt RM. Genetic and phenotypic relationships between and within support and demand tissues in a single line of broiler chicken. *Br Poult Sci*. 2002.
30. Gaya LG, Ferraz JBS, Rezende FM, Mourão GB, Mattos EC, Eler JP, et al. Heritability and genetic correlation estimates for performance and carcass and body composition traits in a male broiler line. *Poult Sci*. 2006.
31. M. C. Ledur, C.M.R. Melo, K. Nones EL, Zanella, K. Ninov, C.A. Bonassi, F.R.F. Jaenisch, A.S.A.M.T. Moura, L.L. Coutinho G, Schmidt. S. 8th World Congress on Genetics Applied to Livestock Production, August 13-18, 2006, Belo Horizonte, MG, Brasil. *Genetics*. 2006;5:8–9.
32. Moghadam HK, McMillan I, Chambers JR, Julian RJ. Estimation of genetic parameters for ascites syndrome in broiler chickens. *Poult Sci*. 2001;80:844–8. doi:10.1093/ps/80.7.844.
33. Deeb N, Cahaner A. Genotype-by-environment interaction with broiler genotypes differing in growth rate. 3. Growth rate and water consumption of broiler progeny from weight-selected versus nonselected parents under normal and high ambient temperatures. *Poult Sci*. 2002;81:293–301. doi:10.1093/ps/81.3.293.
34. Pakdel A, van Arendonk JAM, Vereijken ALJ, Bovenhuis H. Genetic parameters of ascites-related traits in broilers: effect of cold and normal temperature conditions. *Br Poult Sci*. 2005;46:35–42. doi:10.1080/00071660400023938.
35. Pavlidis HO, Balog JM, Stamps LK, Hughes JD, Huff WE, Anthony NB. Divergent selection for ascites incidence in chickens. *Poult Sci*. 2007;86:2517–29. doi:10.3382/ps.2007-00134.
36. Wideman RF, Rhoads DD, Erf GF, Anthony NB. Pulmonary arterial hypertension (ascites syndrome) in broilers: A review. *Poult Sci*. 2013;92:64–83. doi:10.3382/ps.2012-02745.
37. Zhang X, Tsuruta S, Andonov S, Lourenco DAL, Sapp RL, Wang C, et al. Genetics and genomics: Relationships among mortality, performance, and disorder traits in broiler chickens: A genetic and genomic approach. *Poult Sci*. 2018;97:1511–8.

38. Kim H, Grueneberg A, Vazquez AI, Hsu S, de Los Campos G. Will Big Data Close the Missing Heritability Gap? *Genetics*. 2017.
39. Habier D, Fernando RL, Dekkers JCM. Genomic selection using low-density marker panels. *Genetics*. 2009;182:343–53.
40. Ogawa S, Matsuda H, Taniguchi Y, Watanabe T, Nishimura S, Sugimoto Y, et al. Effects of single nucleotide polymorphism marker density on degree of genetic variance explained and genomic evaluation for carcass traits in Japanese Black beef cattle. *BMC Genet*. 2014;15:1–13.
41. Rolf MM, Taylor JF, Schnabel RD, McKay SD, McClure MC, Northcutt SL, et al. Impact of reduced marker set estimation of genomic relationship matrices on genomic selection for feed efficiency in Angus cattle. *BMC Genet*. 2010;11.
42. Tsuruta S, Misztal I, Aguilar I, Lawlor TJ. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. *J Dairy Sci*. 2011.
43. Yan G, Qiao R, Zhang F, Xin W, Xiao S, Huang T, et al. Imputation-Based Whole-Genome Sequence Association Study Rediscovered the Missing QTL for Lumbar Number in Sutan Pigs. *Sci Rep*. 2017.
44. Vitezica ZG, Aguilar I, Misztal I, Legarra A. Bias in genomic predictions for populations under selection. *Genet Res (Camb)*. 2011;93:357–66.
45. Al Kalaldehy M, Gibson J, Duijvesteijn N, Daetwyler HD, MacLeod I, Moghaddar N, et al. Using imputed whole-genome sequence data to improve the accuracy of genomic prediction for parasite resistance in Australian sheep. *Genet Sel Evol*. 2019;51:32. doi:10.1186/s12711-019-0476-4.
46. Wellmann R, Preuß S, Tholen E, Heinkel J, Wimmers K, Bennewitz J. Genomic selection using low density marker panels with application to a sire line in pigs. *Genet Sel Evol*. 2013;45:1–11.
47. Li B, Zhang N, Wang YG, George AW, Reverter A, Li Y. Genomic prediction of breeding values using a subset of SNPs identified by three machine learning methods. *Front Genet*. 2018;9 JUL:1–20.
48. Su G, Brøndum RF, Ma P, Guldbrandtsen B, Aamand GP, Lund MS. 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*. 2012;95:4657–65.
49. VanRaden PM, Tooker ME, O'Connell JR, Cole JB, Bickhart DM. Selecting sequence variants to improve genomic predictions for dairy cattle. *Genet Sel Evol*. 2017.
50. Frischknecht M, Meuwissen THE, Bapst B, Seefried FR, Flury C, Garrick D, et al. Short communication: Genomic prediction using imputed whole-genome sequence variants in Brown Swiss Cattle. *J Dairy Sci*. 2018;101:1292–6. doi:10.3168/jds.2017-12890.
51. Clark SA, Hickey JM, Van Der Werf JHJ. Different models of genetic variation and their effect on genomic evaluation. *Genet Sel Evol*. 2011.
52. VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, et al. Invited review: reliability of genomic predictions for North American Holstein bulls. *J Dairy Sci*. 2009.
53. Veerkamp RF, Bouwman AC, Schrooten C, Calus MPL. Genomic prediction using preselected DNA variants from a GWAS with whole-genome sequence data in Holstein-Friesian cattle. *Genet Sel Evol*. 2016;48:1–14.

Figures

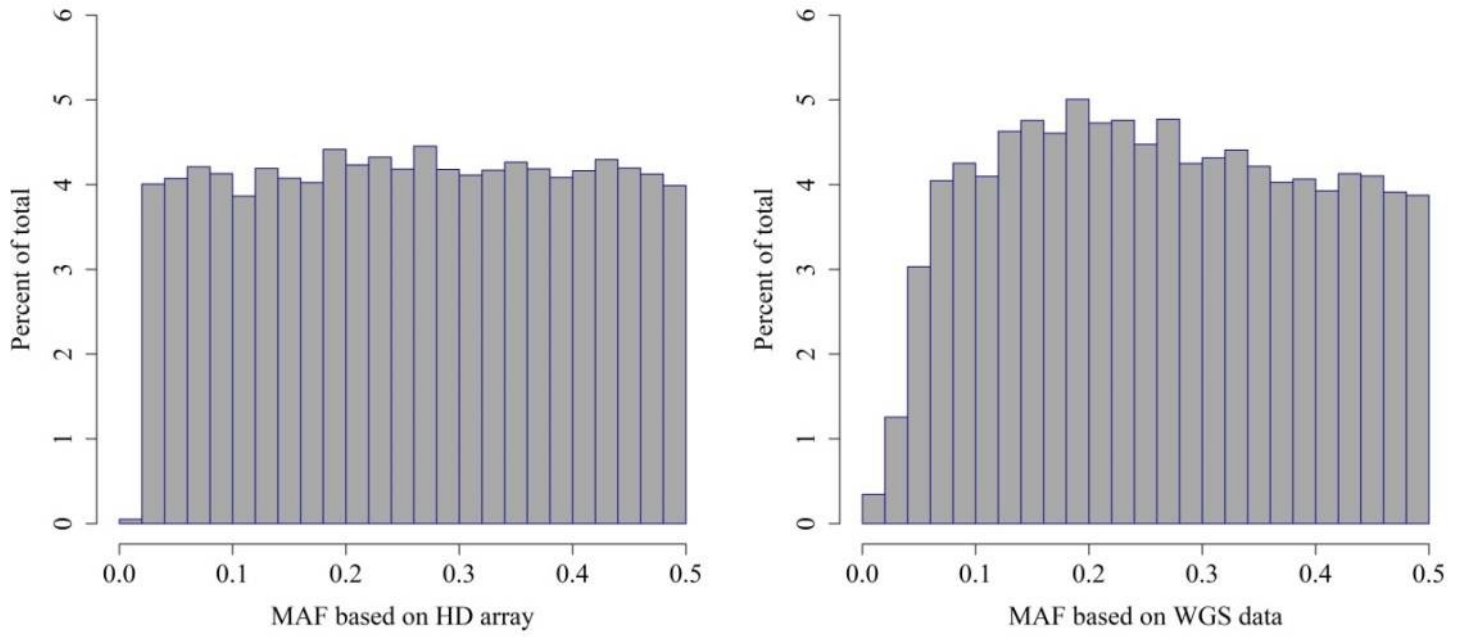


Figure 1
 Distribution of minor allele frequency (MAF) for high density (HD) array data and whole-genome sequencing (WGS) data post-imputation filtering.