

**University of São Paulo
“Luiz de Queiroz” College of Agriculture**

**Enrichment of genotyping panels for the genomic selection of special traits in
broiler chicken**

Mayara Salvian

Thesis presented to obtain the degree of Doctor in
Science. Area: Animal Science and Pastures

**Piracicaba
2020**

Mayara Salvian
Agronomist Engineer

**Enrichment of genotyping panels for the genomic selection of special traits in broiler
chicken**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:

Prof. Dr. **GERSON BARRETO MOURÃO**

Thesis presented to obtain the degree of Doctor in
Science. Area: Animal Science and Pastures

Piracicaba
2020

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Salvian, Mayara

Enrichment of genotyping panels for the genomic selection of special traits in broiler chicken / Mayara Salvian. -- versão revisada de acordo com a resolução CoPGr 6018 de 2011. -- Piracicaba, 2020.

66 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Frango de corte 2. Seleção genômica 3. Valor genético 4. SNP 5. Sequenciamento I. Título

DEDICATION

I dedicate this work to my loving parents Marlene Medeiros da Silva Salvian (*in memoriam*) and Antonio Arquimedes Salvian (*in memoriam*) whose love, education, support and words of encouragement made me the person I am today. Although they are no longer of this world, their memories continue to regulate my life.

ACKNOWLEDGMENTS

Above ground I would like to say thank you to my parents and my family, whose value to me only grows with age. I have no words to express an entire life full of good moments and memories. Although, my parents are no longer in this world, they kept me going on. This work would not have been possible without their input.

I wish to express my sincere appreciation to my supervisor, Dr. Gerson Barreto Mourão. He expertly guided me since my graduate education and who encouraged me to be professional and do the right thing, even when the road got tough. Without his wisdom, patience and persistent help, the goal of this project would not have been realized. Thank you for pushing me further than I thought I could go.

I am grateful to have had the opportunity to work under Dr. Matt Spangler. It was a pleasure to meet and work with this remarkable professional. I also want to thank his lab team and the UNL support, specially the Animal Science Department for all the opportunities. There is no place like Nebraska.

I also want to thank Dr. Luiz L. Coutinho and Dr. Mônica C. Ledur for the paterniship and for trusting in me. Your collaboration was fundamental to the conclusion of this work.

Thank you to all the institutions for the support and for making it possible to finish my Doctorate at University of São Paulo – ESALQ, Embrapa Suínos e Aves and University of Nebraska – Lincoln (UNL). A special acknowledgement to CNPq for supporting me with a scholarship (processes 206489/2017-0 and 142202/2019-3) and also for making my internship abroad at UNL possible.

My appreciation also extends to my co-workers, Amanda, Brayan, Dayse, Eula, Fabrício, Fátima, Felipe, Giovanni, Greg, Izally, Juliana, Leonardo, Luciana, Luiz, Paola, Sofia and Vamilton. I made life long friends on this team and I hope we will work together in the future. Thanks for all the moments!

I take this opportunity to record my sincere thanks to all my friends who were always by my side. Making my days more lighter and helped me in this interesting journey. Thanks Aline, Ândrea, Bruna, Cris, Elisa, Fernanda, Felipe, Dimas, Gabriela, Gabriel, Julia, Jackson, Luiz, Mileni, Priscila, Thaís, Renata and Rita. I love you!

Finally, I also place on record, my sense of gratitude to the Wittmann family (Eric, Grandma, Grandpa, Linda, Matt and Natalie). A special thanks to you Andrew, my love. I am fully indebted to you and I have no valuable words to express my gratitude for all the moments we spent together. My heart is still full of the affection received from every person. I love you guys!

EPIGRAPH

"First, think. Second, dream. Third, believe. And finally, dare".

Walt Disney

"Courage doesn't always roar. Sometimes courage is the little voice at the end of the day that says, 'I'll try again tomorrow'".

Mary Anne Radmacher

SUMMARY

RESUMO.....	9
ABSTRACT	10
LIST OF ABBREVIATIONS	11
1. INTRODUCTION.....	13
REFERENCES.....	15
2. ESTIMATION OF BREEDING VALUES USING SPARSE SNP PANELS AND IMPUTED WHOLE-GENOME SEQUENCE DATA IN BROILER CHICKENS.....	19
ABSTRACT	19
2.1. INTRODUCTION	19
2.2. MATERIALS AND METHODS.....	21
2.2.1 POPULATION AND PHENOTYPES.....	21
2.2.2 GENOTYPING	22
2.2.3 IMPUTATION.....	22
2.2.4 PREDICTION	23
2.2.5 ASSESSMENT OF ACCURACY AND BIAS	25
2.3. RESULTS	26
2.3.1 DESCRIPTIVE RESULTS	26
2.3.2 PREDICTION	27
2.4. DISCUSSION	32
2.4.1 MAF DISTRIBUTION	32
2.4.2 HERITABILITY FOR PEDIGREE-BASED AND GENOMIC MODELS.....	33
2.4.3 CORRELATION BETWEEN EBV AND GEBV	34
2.4.4 REGRESSION COEFFICIENTS.....	35
2.4.5 PREDICTIVE ABILITY	36
2.5. CONCLUSIONS.....	37
REFERENCES.....	38
3. GENOMIC PREDICTION USING HIGH DENSITY-PANEL AND IMPUTED WHOLE-GENOME SEQUENCE DATA WITH DIFFERENT GENOMIC RELATIONSHIP MATRICES IN BROILER CHICKENS.....	45
ABSTRACT	45
3.1. INTRODUCTION	45
3.2. MATERIALS AND METHODS.....	47

3.2.1	POPULATION AND PHENOTYPES	47
3.2.2	GENOTYPING	48
3.2.3	IMPUTATION.....	49
3.2.4	PREDICTION	50
3.2.5	ASSESSMENT OF ACCURACY AND BIAS	51
3.3.	RESULTS	52
3.3.1	IMPUTATION ACCURACY	52
3.3.2	MAF DISTRIBUTION	52
3.3.3	DESCRIPTIVE RESULTS	52
3.3.4	PREDICTION	53
3.4.	DISCUSSION	56
3.4.1.	GENOMIC HERITABILITY	57
3.4.2	CORRELATION BETWEEN DGV.....	58
3.4.3	REGRESSION COEFFICIENTS.....	58
3.4.5	PREDICTIVE ABILITY	59
3.5.	CONCLUSION	60
	REFERENCES.....	60

RESUMO

Enriquecimento de painéis de genotipagem para a seleção genômica de características especiais em frango de corte

O melhoramento genético modificou consideravelmente a produção de frango no Brasil e no mundo. No entanto, o intensivo processo de seleção ao longo dos anos trouxe consequências negativas em aves, como por exemplo, o aumento na deposição de gordura abdominal nos animais, resultando em dificuldades de processamento e depreciação do produto final. Nos últimos anos, os avanços tecnológicos nas áreas de genética molecular e bioinformática fizeram com que a seleção genômica (SG) com o uso de marcadores moleculares (*Single Nucleotide Polymorphisms* - SNP), e mais recentemente o sequenciamento completo do genoma (*Whole-Genomic Sequencing* - WGS), se tornasse uma importante ferramenta para aumentar o ganho genético no melhoramento animal, especialmente para características complexas e de difícil mensuração. Os objetivos deste trabalho foram estimar os valores genéticos e comparar as predições genômicas utilizando provenientes de um painel de SNP de alta densidade (HD - 600K) e de dados do sequenciamento completo do genoma (WGS), por meio de diferentes densidades de marcadores. Foram utilizadas informações de órgãos (coração, fígado, moela e pulmões) e carcaça (peito, coxa, sobrecoxa) de 2.000 aves provenientes da população referência TT pertencente ao Programa de Melhoramento Genético de Aves da EMBRAPA Suínos e Aves. Posteriormente, as predições genômicas foram realizadas utilizando os modelos PBLUP (*Pedigree-Based BLUP*), ssGBLUP (*single-step Genomic BLUP*) e BayesC em várias densidades de SNP e variantes imputadas a partir da sequência do genoma completo. As predições genômicas foram melhores quando as informações genômicas foram adicionadas nas análises. No entanto, nossos resultados não mostraram nenhum benefício no uso de dados WGS em comparação aos dados do HD quando as abordagens ssGBLUP ou BayesC foram aplicadas. Além disso, o uso de um painel de baixa densidade (~74.000 SNPs) pode fornecer resultados significativos a um baixo custo.

Palavras-chave: Frango de corte, Seleção genômica, Valor genético genômico, SNP, Sequenciamento

ABSTRACT

Enrichment of genotyping panels for the genomic selection of special traits in broiler chicken

Traditional animal breeding programs have considerably modified chicken production in Brazil. However, the intensive selection process over the years brought negative consequences in poultry production, such as increased of the abdominal fat deposition, resulting in difficulties in the industrial processing and depreciation of the final product. In recent years, technological advances in molecular genetics and bioinformatics fields have made genomic selection (GS), using molecular markers (Single Nucleotide Polymorphisms - SNP), and more recently the whole-genome sequencing (WGS), an important tool to increase the genetic gain in animal breeding, especially for complex traits and traits which are difficult to measure. The aims of this work were to estimate the genetic values and compare the genomic predictions using a high-density SNP panel (HD - 600K) and whole-genome sequencing (WGS) dataset through different marker densities. Organs (heart, liver, gizzard and lungs) and carcass (breast, thigh, drumstick) information of 2,000 animals derived from a TT broiler line belonging to the Animal Breeding Program from Embrapa Swine and Poultry were used in further analysis. Subsequently, genomic predictions were performed using pedigree-based BLUP (PBLUP), single-step genomic BLUP (ssGBLUP) and BayesC models using various densities of SNP and variants imputed from whole-genome sequence. Genomic predictions were better when the genomic information was added in the analyses. However, our results showed no benefit of using WGS data compared to HD array data when ssGBLUP or BayesC approaches were applied. Besides that, the use of array data with lower densities (~74,000 SNPs) can provide significant results at a low cost.

Keywords: Broiler chicken, Genomic selection, Genomic breeding value, SNP, Sequencing

LIST OF ABBREVIATIONS

PBLUP	Pedigree-based BLUP
ssGBLUP	Single-step Genomic BLUP
LD	Linkage Disequilibrium
SNP	Single Nucleotide Polymorphism
EBV	Estimated Breeding Values
DGV	Direct Genomic Value
HD	High Density
WGS	Whole-Genome Sequencing
QTL	Quantitative Trait Locus
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	Lung weight (g)
BRST	Breast weight (g)
THG	Thigh weight (g)
DRM	Drumstick weight (g)
BW42	Body Weight at 42 days of age (g)

1. INTRODUCTION

Chicken is considered a cheap source of protein due to the short interval between generations and the large numbers of progenies. The broiler meat production is forecast to 102.9 million tons in 2020 being USA, China and Brazil the largest world broiler producers (USDA, 2020). The traditional selection process based on phenotypic and pedigree information has been responsible for achieve the desired progress in poultry breeding programs especially for trait with high heritabilities such as body weight.

Despite many positive results obtained with traditional selection, this technique has limitations, due to the fact that the phenotypic information is considered an imperfect predictor of the breeding value, especially in cases where the negative associations between genes are not taken into account (Dekkers; Hospital 2002). In chickens, for example, traits with low heritability or sex-limited like egg production and disease resistance, which a greater number of information is required to achieve a high accuracy, this selection process is not effective as expected (Wolc, 2014). Moreover, the intense selection process over the years has also brought some negative consequences for poultry production including the abdominal fat increase and metabolic disorders (Campos et al., 2009; Leng et al., 2016).

The discovery of DNA structure in the 1950s has made with a third source of information was added in the prediction analyzes, solving problems imposed by traditional selection and improving the genetic architecture understanding, as well as which genes are involved in the expression of the trait. Additionally in livestock production, chicken was the first specie to have the genome sequenced (Hillier et al., 2004) followed by bovine (Zimin et al., 2009), sheep (Archibald et al., 2010), swine (Groenen et al. al., 2012) and equine (Orlando et al., 2013).

Over the years genomic technologies have been allowed estimating the animal breeding value based on genomic sequences becoming animal selection more accurate and efficient (Ober et al., 2012). The genomic selection (GS) introduced by Meuwissen et al. (2001) made possible to select animals accurately at an early stage of life (Liu et al., 2014b) and also led to the incorporating marker and sequence information to complement pedigree and phenotypic information (Hickey et al., 2017).

This approach has been commonly applied in dairy cattle breeding increasing the genetic gain per year and reducing the generation interval. However, in broilers, which already

have a shorter generation interval the most benefit to use it is increase the accuracy of genomic selection, especially for traits that are difficult to measure and improve (Stock; Reents, 2013; Liu et al., 2014).

The accuracy of genomic prediction can be influenced by many factors including heritability, genetic architecture, extend of linkage disequilibrium between SNPs and QTLs, the population size and the statistical method applied for the genomic prediction. Linear and non-linear methods can be used in genomic prediction, Genomic-best linear unbiased prediction (GBLUP) is an example of linear method, while Bayesian approach is an example of non-linear method (Iheshiulor et al., 2017).

The main difference between these approaches is related to distribution of SNP effects prior, GBLUP assume normal distribution for all SNP effects, while a non-normality distribution is assumed by Bayesian methods (Meuwissen et al., 2001; Chen et al., 2014). However, despite the variations between these approaches, some results showed no differences in genomic prediction using Bayes or GBLUP approach (Hayes et al., 2009; VanRaden et al., 2009; Ober et al., 2012; Heidaritabar et al., 2016).

Furthermore, according to Goddard (2009) the density of the SNP panel also has a significant effect on the GEBV prediction, because the SNPs number distributed throughout the genome also increases the probability that each QTL is in high linkage disequilibrium with at least one marker. However, it is unclear what would be the correct density of the SNP panel to achieve the desired result.

Due to sequencing costs reduction and the uncertainty regarding the correct SNP panel density to be used, some researchers are sequencing the animals whole genome and using this information to estimate the GEBV both in real data (Ober et al., 2012) as well as in simulated data (Meuwissen, Goddard, 2010; Druet et al., 2014; MacLeod et al., 2016). It is expected that the data obtained by sequencing the whole genome include the causal mutations underlying the QTL, which allow estimating the trait QTL effect regardless of linkage disequilibrium between the SNPs and QTL (van Binsbergen, et al., 2015). Thus, the SNP variance with low allelic frequency (low MAF) that are not in high LD with causal variants but which explain a part of the trait genetic variance may be used, increasing the GEBV accuracy (Heidaritabar et al., 2016).

The aim of this study was to improve the genetic understanding of organ and carcass traits in broiler chicken by using information from SNPs arrays as well as whole-genome sequencing (WGS) to be applied in genomic selection (GS). A Brazilian broiler TT population developed in Brazil by Chicken Breeding Program of EMBRAPA Swine and Poultry was used to run the analysis. All the results obtained in this thesis were carried out in agreement between “Luiz de Queiroz” College of Agriculture (University of São Paulo), EMBRAPA Swine and Poultry and University of Nebraska – Lincoln.

References

ARCHIBALD, A. L.; COCKETT, N. E.; DALRYMPLE, B. P.; FARAUT, T.; KIJAS, J. W.; MADDOX, J. F.; MCEWAN, J. C.; HUTTON ODDY, V.; RAADSMA, H. W.; WADE, C.; WANG, J.; WANG, W.; WANG, J. The sheep genome reference sequence: a work in progress. *Animal Genetics*, v. 41, 449-453, 2010.

CAMPOS, R. L. R.; NONES, K.; LEDUR, M. C.; MOURA, A. S. A. M.; PINTO, L. F. B.; AMBO, M.; BOSCHIERO, C.; RUY, D. C.; BARON, E. E.; NINOV, K.; ALTENHOFEN, C. A. B.; SILVA, R. A. M. S.; ROSÁRIO, M. F.; BURT, D. W.; COUTINHO, L. L. Quantitative trait loci associated with fatness in a broiler–layer cross. *Animal Genetics*, v. 40, p. 729-736, 2009.

CHEN, L.; LI, C.; SARGOLZAE, M.; SCHENKEL, F. Impact of Genotype Imputation on the Performance of GBLUP and Bayesian Methods for Genomic Prediction. *PLOS ONE*, v. 9, e101544, 2014.

DEKKERS, J. C. M.; HOSPITAL, F. The Use of Molecular Genetics in the Improvement of Agricultural Populations. *Nature Reviews Genetics*, v. 3, n. January, p. 22–32, 2002.

DRUET, T.; MACLEOD, I. M.; HAYES, B. J. Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity*, v. 112, p. 39–47, 2014.

GROENEN, M. A. M.; ARCHIBALD, A. L.; UENISHI, H.; TUGGLE, C. K.; TAKEUCHI, Y.; ROTHSCHILD, M. F.; et al., Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491, 393-398, 2012.

GODDARD, M. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*, v. 136, p. 245 - 257, 2009.

HAYES, B. J.; BOWMAN, P. J.; CHAMBERLAIN, A. J. GODDARD, M. E. Genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science*, v. 92, n. 2, p. 433–443, 2009.

HEIDARITABAR, M.; CALUS, M. P. L.; MEGENS, H-J.; VEREIJKEN, A.; GROENEN, M. A. M.; BASTIAANSEN, J. W. M. Accuracy of genomic prediction using imputed whole-genome sequence data in white layers. *Journal of Animal Breeding and Genetics*, v. 133, p. 167-169, 2016.

HICKEY, J. M.; CHIURUGWI, T.; MACKAY, I.; POWELL, W. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nature genetics*, v. 49, n. 9, 2017.

HILLIER, L.W.; MILLER, W.; BIRNEY, E.; WARREN, W.; HARDISON, R. C.; PONTING, C. P.; BORK, P.; BURT, D.W.; GROENEN, M.A.; DELANY, M.E. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, 432 (7018):695–716, 2004.

IHESHIULOR, O. O. M.; WOOLLIAMS, J. A.; SVENDSEN, M.; SOLBERG, T.; MEUWISSEN, T. H. E. Simultaneous fitting of genomic-BLUP and Bayes-C components in a genomic prediction model. *Genetics Selection Evolution*, v. 49, 2017.

LENG, L.; ZHANG, H. DONG, J. Q.; WANG, Z. P.; ZHANG, X. Y.; WANG, S. Z.; CAO, Z. P.; LI, Y. M.; LI, H. Selection against abdominal fat percentage may increase intramuscular fat content in broilers. *Journal of Animal Breeding and Genetics*, v. 133, p. 422-428, 2016.

LIU, T.; QU, H.; LUO, C.; SHU, D.; WANG, J.; LUND, M. S.; SU, G. Accuracy of genomic prediction for growth and carcass traits in Chinese triple-yellow chickens. *BMC Genetics*, v.15, p. 110, 2014.

MACLEOD, M. I.; BOWMAN, P. J.; VANDER JAGT, C. J.; HAILE-MARIAM, M.; KEMPER, K. E.; CHAMBERLAIN, A. J.; SCHROOTEN, C.; HAYES, B. J.; GODDARD, M. E. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics*, v. 17, p. 144, 2016.

MEUWISSEN, T.; GODDARD, M. Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics*, v. 185, p. 623 - 631, 2010.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genomewide dense marker maps. *Genetics*, v. 157, p. 1819–1829, 2001.

OBER, U.; AYROLES, J. F.; STONE, E. A.; RICHARDS, S.; ZHU, D.; GIBBS, R. A.; STRICKER, C.; GIANOLA, D.; SCHLATHER, M.; MACKAY, T. F. C.; SIMIANER, H. Using whole-genome sequence data to predict quantitative trait phenotypes in *Drosophila melanogaster*. *PLoS Genetics*, v. 8, e1002685, 2012.

ORLANDO, L.; GINOLHAC, A.; ZHANG, G.; FROESE, D.; ALBRECHTSEN, A.; STILLER, M.; SCHUBERT, M.; CAPPELLINI, E.; PETERSEN, B.; MOLTKE, I.; JOHNSON, P. L. F.; FUMAGALLI, M.; VILSTRUP, J. T.; RAGHAVAN, M.; KORNELIUSSEN, T.; MALASPINAS, A. S.; et al. Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, v. 499, 74-78, 2013.

STOCK, K.; REENTS, R. Genomic selection: Status in different species and challenges for breeding. *Reproduction in Domestic Animals*, v. 48, p. 2–10, 2013.

UNITED STATES DEPARTMENT OF AGRICULTURE (USDA). Livestock and Poultry: World Markets and Trade. Available at: <https://downloads.usda.library.cornell.edu/usda-esmis/files/73666448x/sb397r25n/0z709c25b/livestock_poultry.pdf>. Accessed: 01 Apr 2020.

VAN BINSBERGEN, R.; CALUS, M. P. L.; BINK, M. C. A. M.; VAN EEUWIJK, F. A.; SCHROOTEN C.; VEERKAMP, R. F. Genomic prediction using imputed whole-genome sequence data in Holstein Friesian cattle. *Genetics Selection Evolution*, v. 47, p. 71, 2015.

VANRADEN, P.M.; VAN TASSELL, C.P.; WIGGANS, G.R.; SONSTEGARD, T.S.; SCHNABEL, R.D.; TAYLOR, J.F.; SCHENKEL, F.S. Invited review: reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, v.92, p.16-24, 2009.

WOLC, A. Understanding genomic selection in poultry breeding. *World's Poultry Science Journal*, v. 70, n. June, p. 309–314, 2014.

2. ESTIMATION OF BREEDING VALUES USING SPARSE SNP PANELS AND IMPUTED WHOLE-GENOME SEQUENCE DATA IN BROILER CHICKENS

ABSTRACT

Traditionally, breeding values have been 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, thus genomic prediction based on high density panel (HD) or even whole-genome sequencing (WGS) data is feasible. The aim of this study was to compare the rank of estimated breeding values (EBV) for organ (heart, liver, lungs and gizzard) and carcass (breast, thigh and drumstick) weight traits in a broiler population using pedigree-based BLUP (PBLUP) and single-step genomic BLUP (ssGBLUP) models using various densities of SNP and variants imputed from whole-genome sequence. For both PBLUP and ssGBLUP, heritability estimates varied from low (LUN) to high (HRT, LIV, GIZ, BRST, THG and DRUM). Regression coefficients values of EBV on GEBV were similar for both the HD and WGS sets of SNPs, ranging from 0.87 to 0.99 across scenarios. Results show no benefit of using WGS data compared to HD array data using ssGBLUP. Therefore, the uses of array data with lower densities can provide significant results at a low cost. Our results suggest that the use of at least 74,122 SNPs (20%) can be effective to provide considerable results.

Keywords: Genomic prediction; High density panel; Whole-genomic sequence; Imputation; Broiler chicken

2.1. Introduction

Traditionally, breeding values have been estimated based on phenotypic and pedigree information by pedigree-based BLUP (PBLUP) using the numerator relationship (**A**) matrix (Henderson, 1984). 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) (Misztal et al., 2013) 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**) (Legarra et al., 2009). 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 (e.g., Howard et al., 2018). 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 (Meuwissen et al., 2016). However, the use of high density (HD) panels in forming a genomic relationship matrix has not been shown to provide significant improvements in accuracy (Misztal et al., 2013). 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 (Ober et al., 2012) as well as in simulated data (Meuwissen and Goddard, 2010; Druet et al., 2014; MacLeod et al., 2016). 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 (van Binsbergen et al., 2015).

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 (Meuwissen and Goddard, 2010; Druet et al., 2014; MacLeod et al., 2014). Interestingly, this has not always been the case in real data using cattle and chickens (e.g., van Binsbergen et al., 2015; Heidaritabar et al., 2016; MacLeod et al., 2016).

Although the expectation is 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 and degree of bias of estimated breeding values (EBV) for organs (heart, liver, lungs

and gizzard) and carcass (breast, thigh and drumstick) traits in a broiler population using PBLUP and ssGBLUP by means of various densities of SNP (high-density panel – HD) and variants imputed from whole-genome sequence (WGS) data.

2.2. MATERIALS AND 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.

2.2.1 Population and phenotypes

The chicken population used in this study was derived from a TT broiler line belonging to the Animal Breeding Program from Embrapa Swine and Poultry. 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 (Nones et al., 2012; Venturini et al., 2014). The TT reference population is a broiler population developed for genomic studies in 2008 from the crossing 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 from five hatches. Matings between relatives were avoided to improve the genetic variability as described by Marchesi et al. (2018).

A total of 1,453 animals (703 males and 750 females) were slaughtered at 42 days of age after six hours of fasting and the body weight at 42 days of age (BW42) were recorded. Blood samples from each animal were collected for DNA extraction and the eviscerated carcass was cooled. After six hour 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. (2014).

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).

2.2.2 Genotyping

Blood samples of each animal (1,453) were used to extract DNA using PureLink® Genomic DNA (Invitrogen, Carlsbad, CA, USA) kit and quantified using 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. (2013).

Axiom™ Analysis Suite (Affymetrix®) software was used to filter based on DishQC parameter, and then PLINK v.1.9 software (Purcell et al., 2007) 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 (Moreira et al., 2018).

2.2.3 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 crosses to obtain 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 (Boschiero et al., 2018) and (Moreira et al., 2018).

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 (Browning and Browning, 2008) 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 changing of detecting segregating SNPs be 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; McLaren et al., 2016) using galGal5 as reference genome.

The sequence variants selected to include 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 control of gene expression (Moreira et al., 2018). 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 consists of 69% of intergenic regions of the genome, 16% of downstream, 14% upstream and 1% of UTR3' and UTR5', respectively. Then, from those non-common variants 370,608 SNPs were randomly selected (10 times) to compose the final WGS dataset and ensure a random representation of the entire genome.

2.2.4 Prediction

Variant Subsets

Seven subsets (0.5, 1, 5, 10, 20, 40 and 80% of SNP) were randomly selected from the full HD set to determine the impact, in terms of EBV rank and bias, of using reduced subsets of SNP to inform relationships among individuals. Imputed variants from WGS were also used and mimicked the number of SNP chosen for the subsets mentioned above. In both data sets (HD and imputed WGS) the SNP selection process was repeated ten times in each scenario. Results are the average of the 10 replicates of randomly selecting subsets.

Breeding value estimation

Estimated breeding values for the weight of each organ [heart (HRT), liver (LIV), gizzard (GIZ), lung (LUN)] and carcass trait [breast (BRT), thigh (THG) and drumstick (DRM)] and BW42 were predicted using the BLUPF90 family of programs (Misztal et al., 2018) 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 hens and 260 roosters. A bivariate model with BW42 as an anchor trait has been chosen to assess genetic interactions between traits with BW42 and minimize the bias associate with them. Additionally, the use of bivariate model has greater or at least similar power compared to univariate models (Rovadoscki et al., 2018). For both methods the following bivariate animal model was used:

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

where y_1 and y_2 are the vector of observation for each carcass trait (y_1) and BW42 (y_2); X_1 and X_2 are the design matrix for fixed effects; b_1 and b_2 are the vector of fixed effects (sex and hatch) for the first and second trait, respectively; Z_1 and Z_2 are the design matrix 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 effect with a distribution $\sim N(0, I\sigma_e^2)$, where I is an identity matrix and σ_e^2 is the residual variance. The additive genetics effects were assumed to be normally distributed as $u \sim N(0, A\sigma_u^2)$ for PBLUP and $u \sim N(0, H\sigma_u^2)$ for ssGBLUP.

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. (2010) as:

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

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

$$\mathbf{G} = 0.95\mathbf{G}_w + 0.05\mathbf{A}_{22}$$

where: \mathbf{A}_{22} is the pedigree-based relationship matrix for genotyped animals only; \mathbf{G}_w is the genomic matrix obtained following (VanRaden, 2008; Heidaritabar et al., 2016):

$$\mathbf{G}_w = \frac{\mathbf{M}\mathbf{M}'}{2 \sum p_i(1 - p_i)}$$

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

2.2.5 Assessment of accuracy and bias

Spearman correlations between EBV from PBLUP and EBV 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.

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 (Legarra et al., 2008):

$$r = cor(EBV, y^*)$$

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. Moreover, the regression coefficients of EBV on GEBV in each scenario were calculated to evaluate the degree of similarity between the predictions.

2.3. RESULTS

2.3.1 Descriptive results

The descriptive statistics of the remaining data are presented in Table 1. The estimates of variance component and heritability for organ and carcass traits from PBLUP and ssGBLUP using both HD and WGS density panels are given in Tables 2. Small differences in those estimates were observed between the methods used. Except for HRT, the estimation of heritability was lower when the genomic information was used. The standard errors were also lower when the 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 weights (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 weights (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

¹HRT=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), environmental 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 trait of broiler chicken using PBLUP, HD and WGS dataset.

Trait ¹	σ_a^2	σ_e^2	σ_p^2	h^2
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)

¹HRT=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).

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 were not (Figure 1).

2.3.2 Prediction

Correlations between predicted breeding values using the different datasets and different genomic prediction methods 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 scenarios. For HRT, a slightly overestimation of the breeding values was observed ranging from 0.87 to 0.89 when less than 80% of SNPs were used in the HD dataset.

The same pattern was observed using imputed sequence data which the regression coefficients were also overestimated (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 data set did not result in a better estimation of regression coefficients 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 trait with lower heritabilities (e.g. LUN), which means that traits with low heritability requires more records to achieve higher predictive abilities as traits with high heritabilities (Table 5). Compared to PBLUP, the predictive ability of ssGBLUP was higher from 0.5% of SNPs, excepted for LUN, which the predictive ability of PBLUP was higher than ssGBLUP when 0.5% and 1% of SNPs were used.

Table 3. Regression coefficient ($b_{EBV,GEV}$) and standard error (in brackets) for organ traits in HD panel and WGS data in each scenario (SNP percentage).

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 Panel								
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)

¹HRT=Heart; LIV=Liver; GIZ=Gizzard; LUN=Lungs; BW42=Body weight at 42 days of age.

Table 4. Regression coefficient ($b_{EBV,GEV}$) and standard error (in brackets) for carcass traits in HD panel and WGS data in each scenario (SNP percentage).

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 Panel						
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)

¹BRST=Breast; DRM=Drumstick; THG =Thigh; BW42=Body weight at 42 days of age.

Table 5. Predictive abilities with HD panel and WGS datasets in each scenario (SNP percentage).

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	0.359	0.351	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	0.324	0.326	0.375	0.399	0.361	0.211	0.359	0.358	0.361	0.363	0.320	0.368	0.333
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	0.324	0.326	0.375	0.399	0.361	0.211	0.359	0.358	0.361	0.363	0.320	0.368	0.333

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

2.4. DISCUSSION

In the present study we had chosen select SNPs randomly to compose the different panel densities to investigate carcass and organ traits in broiler chicken. Although the markers were not equally spaced during the selection process, they were present in at least one chromosome across all genome.

2.4.1 MAF distribution

The distribution of MAF for the HD array was uniform while the MAF distribution for WGS variants retained for further analyses were not (Figure 1). Unlike other studies (van Binsbergen et al., 2015; Heidaritabar et al., 2016) the variants used in the current study did not show a U-shaped MAF distribution for WGS data. In accordance with Ni et al. (2017), 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 sequence 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 of the population.

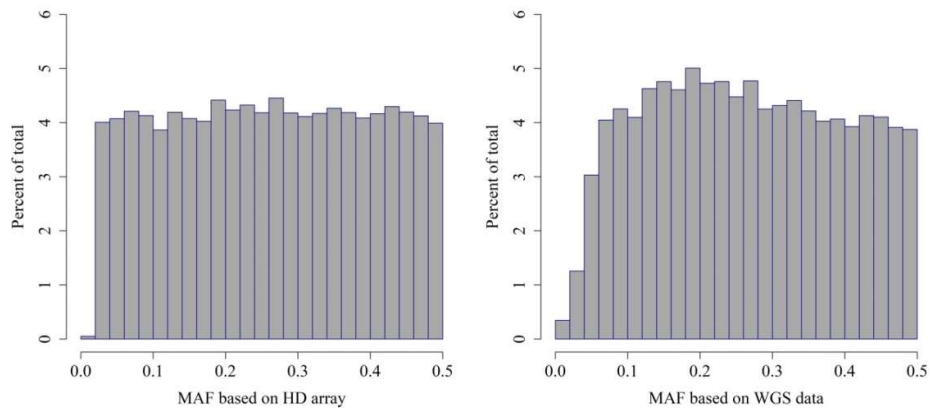


Figure 1. Distribution of minor allele frequency (MAF) for high density (HD) array data and whole-genome sequencing (WGS) data, after post-imputation filtering.

2.4.2 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, respectively. 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 traits used in this study. Using the same population (Embrapa TT), Venturini et al. (2014) reported similar pedigree-based heritability estimates for LIV (0.33 ± 0.07), GIZ (0.44 ± 0.08), BRST (0.37 ± 0.07) and DRM (0.35 ± 0.07) to those reported herein. However, the heritability estimate found in this study for THG was higher (0.44 ± 0.08) than the estimate in Venturini et al. (2014) (0.29 ± 0.06).

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 Argentão et al. (2002), Rance et al. (2002) and Gaya et al. (2006). In a study with a male broiler line, Gaya et al. (2006) 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. (2002) reported heritability estimates for HRT (0.30 ± 0.08), LIV (0.08 ± 0.06), GIZ (0.52 ± 0.10).

The heritability estimates for LUN in broiler chicken are not common in the literature. Using a F_2 experimental population Ledur et al. (2006) reported similar pedigree-based heritability estimates for LUN (0.10) than the result reported herein. Although LUN is not considered an economically important trait, it has been related to pulmonary hypertension (e.g. ascites). Heritability estimation for ascites have been reported by several authors (Moghadam et al., 2001; Deeb et al., 2002; Pakdel et al., 2002; Ledur et al., 2006; Pavlidis et al., 2007; Wideman et al. 2013). The use of a multi-trait model may be responsible for the higher heritabilities estimates differences found in this study compared to the literature since multi-traits models uses additional genetic information from link with other traits (Zhang et al., 2017).

Heritability was also estimated using the genomic matrix instead of numerator relationship matrix which resulted in relatively small differences between the estimates (Table 2). Usually the genomic heritability is smaller than heritability estimate using only the pedigree

and phenotypic information (Kim et al., 2017). Few studies have used ssGBLUP method for genomic heritability estimation for traits used in this study.

Based on the strong relationship between prediction and the heritability of the trait, wherein traits with higher heritability are more accurate comparing with traits with low heritability our results have been shown that HRT, LIV, GIZ, BRST, THG and DRM can be used as a selection criterion in this population.

2.4.3 Correlation between EBV and GEBV

Across all traits, EBV estimated with at least 0.5% SNP was highly correlated with EBV estimated from the complete HD (minimum correlation 0.94) and the minimum average correlation between 0.5% markers and PBLUP was 0.89. Indeed, lower correlations were observed when a smaller number of SNPs 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. Although, the same pattern has been noted with WGS, a slightly improvement in the correlation value (minimum correlation 0.95) was acquired with this dataset.

Comparing the correlations between predicted breeding values using different genotype datasets our results show no difference when 10% or 100% of SNPs (mean correlation 0.99) were used in the analyses which suggests that the use of evenly-spaced lower-density panel could provide a very similar ranking of EBV at a potentially lower cost, as proposed by Habier et al (2009). When applied to Japanese black cattle, Ogawa et al. (2014) suggested that using at least 4,000 equally-spaced SNPs should be enough to ranking the animals genetically for carcass weight and marbling score.

Using 50K chip, Rolf et. al (2010) pointed out that 2,500 ~ 10,000 SNPs distributed throughout the genome could address a robustly G matrix estimation for feed efficiency in Angus cattle. Thus, a lower density panel in genomic prediction analysis could be enough to generate an accurate genomic relationship matrix in genomic prediction analysis, agreeing with our findings.

2.4.4 Regression coefficients

The slope of the regression coefficients of EBV on GEBV quantifies the bias in the variance of the estimated breeding value for each scenario (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 scenarios. The regression coefficients increased as the proportion of SNP increased reaching a plateau between 5 and 10% of SNPs.

Overall, all the regression coefficients 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 that means the variance of breeding values were overestimated. However, according to Tsuruta et al. (2011) deviations of $\pm 15\%$ from unity are acceptable. Wherefore, based on these coefficients this study suggests that ssGBLUP approach is an effective method to improve the genetic prediction in broiler chicken.

Using a pure layer line, Yan et al. (2017) used the bias to compare PBLUP and ssGBLUP prediction models. These authors used the regression coefficients of phenotypes corrected for fixed effects on predicted (G)EBV and reported that ssGBLUP approach was less bias than PBLUP. On the other hand, also in layer lines, Heidaritabar et al. (2016) reported high regression coefficients (greater than 1) when PBLUP or GBLUP were used in 60K SNP panel and sequence data, indicating an underestimation of the breeding values variance.

Although, “Beavis effect” phenomenon can be pointed out as one of the reasons to cause the GEBV bias since the true SNPs effects tend to be smaller than the SNP effect reported by the trait. This effect probably was not the responsible to cause bias in this population since the SNPs used herein were selected randomly, the estimates were regressed towards the mean, thus minimizing the “Beavis effect” (Goddard, Hayes, 2009).

Bias differences among the methods may be explained by directional selection (Vitezica et al., 2011). In the present study, a 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 (Nones et al., 2012; Venturini et al., 2014).

2.4.5 Predictive ability

Predictive abilities across all traits are reported in Table 5. Compared to PBLUP, the predictive ability assessed by ssGBLUP was higher for major traits when at least 5% of SNPs were used. Although BRST have presented similar predictive ability values the slight difference between both prediction methods might be explain by the fact that not all causal variants are captured during the genomic prediction for this trait.

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 increase the number of markers, most of them are in incomplete LD with causal mutation. 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 responsible to improve the genomic prediction (Al Kaladeh et al., 2019).

Many researches have used reduced SNPs density as a solution for genotyping costs are available in the literature (e.g. Habier et al., 2009; Rolf et al., 2010; Wellmann et al., 2013; Ogawa et al., 2014; Li et al. 2018). While estimates using a genomic relationship matrix appears to be better than pedigree relationship matrix, our results show no difference in genomic prediction when a reduced number of SNPs were used to fit the genomic relationship matrix, indicating that at least 10% of SNP panel (~37,000 SNPs) can be used in genomic evaluation.

Agreeing with our findings, Su et al. (2012), Zhang et al. (2018), and Boldt et al. (2018) concluded that the use of different percentage of SNP panel in genomic prediction did not improve the genomic prediction as expected, so a reduced number of SNPs can be used.

Simulated data has shown an increase in genomic prediction accuracy when the causal mutations were included in the analyses (Meuwissen and Goddard, 2010; Druet et al., 2014; MacLeod et al., 2014). 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 (Ober et al., 2012; Van Binsbergen et al., 2015; Heidaritabar et al., 2016; MacLeod et al., 2016; VanRaden et al., 2017; Frischknecht et al., 2017; Al Kalaldehy et al., 2019).

The infinitesimal model used herein (ssGBLUP), whereby the markers are assumed to come from a normal distribution with 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. (2011) suggested that the increase of genomic prediction accuracy will be smaller when the trait is highly polygenic especially in a small reference population. Further, 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 (VanRaden et al., 2017).

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 (MacLeod et al., 2014). Thus, in highly selected population, almost the totally genetic variance can be explained by the SNPs genetic variance as result of the relationship between individuals (VanRaden et al., 2009).

Despite the imputation accuracy does not be the principal objective of the present work, it can help to explain why sequence data was not present a superior predictive ability comparing with 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 responsible to reduce the imputation accuracy since those errors may decrease the accuracy of prediction and interfere in the detection of causal mutations (Veerkamp et al.; 2016).

2.5. CONCLUSIONS

In this study, we investigated different density marker panels and methods for prediction of genomic breeding values 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 20% of SNP (~74,122) can provide decent genetic evaluations. Therefore, the use of lower-density arrays, if at a lower cost, could be used to rank individual based on genetic merit. Furthermore, the results also demonstrated no benefit of using WGS data compared to HD array data using ssGBLUP. The use of different weighted genomic matrix may also improve the predictive ability when ssGBLUP approach is used.

REFERENCES

AGUILAR, I.; MISZTAL, I.; JOHNSON, D.L.; LEGARRA, A.; TSURUTA, S.; LAWLOR, T.J. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score¹. *Journal of Dairy Science*, v. 93, p. 743–752, 2010.

AL KALALDEH, M.; GIBSON, J.; DUIJVESTIEN, N.; DAETWYLER, H.D.; MACLEOD, I.; MOGHADDAR, N.; LEE, S.H.; VAN DER WERF, J.H.J. Using imputed whole-genome sequence data to improve the accuracy of genomic prediction for parasite resistance in Australian sheep. *Genetics Selection Evolution*, 51, 32, 2019.

ARGENTÃO, C.; FILHO, T.M.; MARQUES, J.L.B.; SOUZA, E.M.; ELER, J.P.; FERRAZ, J.B.S. Genetic and phenotypic parameters of growth and carcass traits of a male line of broilers raised in tropical conditions 7th World Congress on Genetics Applied to Livestock Production, 2002.

BOLDT, R.; KEELE, J. ; KUEHN, L.; MCDANELD, T.; SMITH, T.; ENNS, R. 326 Comparison of Genomic Relationship Matrices Using Differing Number of SNP in Pooled DNA Analyses. *Journal of Animal Science*, v. 96, p. 124, 2018.

BOSCHIERO, C.; MOREIRA, G.C.M.; GHEYAS, A.A.; GODOY, T.F.; GASPARIN, G.; MARIANI, P.D.S.C.; PADUAN, M.; CESAR, A.S.M.; LEDUR, M.C.; COUTINHO, L.L. Genome-wide characterization of genetic variants and putative regions under selection in meat and egg-type chicken lines. *BMC Genomics*, 19, n. 83, 2018.

BROWNING, B.L.; BROWNING, S.R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *American Journal of Human Genetics*, v. 84, p. 210-223, 2008.

CLARK, S.A.; HICKEY, J.M.; VAN DER WERF, J.H.J. Different models of genetic variation and their effect on genomic evaluation. *Genetics Selection Evolution*, 43, 18, 2011.

DEEB N., SHLOSBERG, A.; CAHANER, A. Genotype-by-environment interaction with broiler genotypes differing in growth rate. 4. Association between responses to heat stress and to cold-induced ascites. *Poultry Science*, v.81, p.1454-1462, 2002.

DRUET, T.; MACLEOD, I.M.; HAYES, B.J. Toward genomic prediction from whole-genome sequence data: Impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity*, v.112, p.39-47, 2014.

FORNARI, M. B.; ZANELLA, R.; IBELLI, A. M. G.; FERNANDES, L. T.; CANTÃO, M. E.; THOMAZ-SOCCOL, V.; et al. Unraveling the associations of osteoprotegerin gene with production traits in a paternal broiler line. *Springerplus*, v.3, p. 1-8, 2014.

FRISCHKNECHT, M.; MEUWISSEN, T.H.E.; BAPST, B.; SEEFRIED, F.R.; FLURY, C.; GARRICK, D.; SIGNER-HASLER, H.; STRICKER, C.; BIEBER, A.; FRIES, R.; RUSS, I.; SÖLKNER, J.; BAGNATO, A.; GREDLER-GRANDL, B. Short communication: Genomic prediction using imputed whole-genome sequence variants in Brown Swiss Cattle. *Journal of Dairy Science*, v.101, p.1292-1296, 2018.

GAYA, L.G.; FERRAZ, J.B.S.; REZENDE, F.M.; MOURÃO, G.B.; MATTOS, E.C.; ELER, J.P.; FILHO, T.M. Heritability and genetic correlation estimates for performance and carcass and body composition traits in a male broiler line. *Poultry Science*, v.85, p. 837-843, 2006.

GODDARD, M.E.; HAYES, B.J. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, v.10, p.381-391, 2009.

HABIER, D.; FERNANDO, R.L.; DEKKERS, J.C.M. Genomic selection using low-density marker panels. *Genetics*, v.182, n.1, p.343-353, 2009.

HEIDARITABAR, M.; CALUS, M.P.L.; MEGENS, H.J.; VEREIJKEN, A.; GROENEN, M.A.M.; BASTIAANSEN, J.W.M. Accuracy of genomic prediction using imputed whole-genome sequence data in white layers. *Journal of Animal Breeding and Genetics*, v.133, p.167-179, 2016.

HENDERSON, C.R. *Applications of linear models in animal breeding*, 1984.

HOWARD, J.T.; RATHJE, T.A.; BRUNS, C.E.; WILSON-WELLS, D.F.; KACHMAN, S.D.; SPANGLER, M. L. The impact of truncating data on the predictive ability for single-step genomic best linear unbiased prediction. *Journal of Animal Breeding and Genetics*, v.135, p.251-262, 2018.

KIM, H.; GRUENEBERG, A.; VAZQUEZ, A.I.; HSU, S.; DE LOS CAMPOS, G. Will Big Data Close the Missing Heritability Gap? *Genetics*, v.207, p.1135-1145, 2017.

KRANIS, A., GHEYAS, A.A.; BOSCHIERO, C.; TURNER, F.; YU, L.; SMITH, S.; TALBOT, R.; PIRANI, A.; BREW, F.; KAISER, P.; HOCKING, P. M.; FIFE, M.; SALMON, N.; FULTON, J.; STROM, T.M.; HABERER, G.; WEIGEND, S.; PREISINGER, R.; GHOLAMI, M.; QANBARI, S.; SIMIANER, H.; WATSON, K.A.; WOOLLIAMS, J.A.; BURT, D.W. Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics*, 14, n.59, 2013.

LEDUR, M.C.; MELO, C.M.R.; NONES, K.; ZANELLA, E. L.; NONOV, K.; BONASSE, C. A.; JAENISCH, F.R.F.; MOURA, A.S.A.M.; COUTINHO, L.L.; SCHMIDT, G.S. Genetic and phenotypic parameters for organs, body and carcass weights and haematocrit value in a broiler x layer cross resource population. In: 8th World Congress on Genetics Applied to Livestock Production, 2006.

LEGARRA, A.; ROBERT-GRANIÉ, C.; MANFREDI, E.; ELSEN, J.M. Performance of genomic selection in mice. *Genetics*, v.180, p.611-618, 2008.

LEGARRA, A., AGUILAR, I.; MISZTAL, I. A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science*, v.92, p.4656-4663, 2009.

LI, B.; ZHANG, N.; WANG, Y.; GEORGE, A.W.; REVERTER, A.; LI, Y. Genomic Prediction of Breeding Values Using a Subset of SNPs Identified by Three Machine Learning Methods. *Frontiers Genetics*, 9, 237, 2018.

MACLEOD, I.M.; HAYES, B.J.; GODDARD, M.E. The effects of demography and long-term selection on the accuracy of genomic prediction with sequence data. *Genetics*, v.198, p.1671-1684 2014.

MACLEOD, I.M.; BOWMAN, P.J.; JAGT, C.J.V.; HAILE-MARIAM, M.; KEMPER, K.E.; CHAMBERLAIN, A.J.; SCHROOTEN, C.; HAYES, B.J.; GODDARD M.E. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics*, 17, n.144, 2016.

MARCHESI, J. A. P.; BUZANSKAS, M. E.; CANTÃO, M. E.; IBELLI A. M. G.; PEIXOTO, J. O.; JOAQUIM L. B.; et al. Relationship of runs of homozygosity with adaptive and production traits in a paternal broiler line. *Animal*, v. 12, p. 1126-1134, 2018.

MCLAREN, W.; GIL, L.; HUNT, S.E.; RIAT, H.S.; RITCHIE, G.R.S.; THORMANN, A.; FLICEK, P.; CUNNINGHAM, F. The Ensembl Variant Effect Predictor. *Genome Biology*, 17, n. 122, 2016.

MEUWISSEN, T.; GODDARD, M. Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics*, v.185, p.623-631, 2010.

MEUWISSEN, T.; HAYES, B.; GODDARD, M. Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers*, v.6, p.6-14, 2016.

MISZTAL, I.; AGGREY, S.E.; MUIR, W.M. Experiences with a single-step genome evaluation. *Poultry Science*, v.92, p.2530-2534, 2013.

MISZTAL, I., S. Tsuruta, D. Lourenco, AGUILAR, I.; LEGARRA, A., Z. Vitezica. 2018. Manual for BLUPF90 family of programs.

MISZTAL, I.; TSURUTA, S.; AGUILAR, I.; LEGARRA, A.; VANRADEN, P.M.; LAWLOR, T. J. Methods to approximate reliabilities in single-step genomic evaluation. *Journal of Dairy Science*, v.96, p.647-654, 2013.

MOGHADAM, H.K.; MCMILLAN, I.; CHAMBERS, J.R.; JULIAN, R.J. Estimation of genetic parameters for ascites syndrome in broiler chickens. *Poultry Science*, v. 7, p. 844-848, 2001.

MOREIRA, G.C.M.; BOSCHIERO, C.; CESAR, A.S.M.; REECY, J.M.; GODOY, T.F.; TREVISOLI, P.A.; CANTÃO, M.E.; LEDUR, M.C.; IBELLI, A.M.G.; PEIXOTO, J.O.; MOURA, A.S.A.M.T.; GARRICK, D.; COUTINHO, L.L. A genome-wide association study reveals novel genomic regions and positional candidate genes for fat deposition in broiler chickens. *BMC Genomics*, 19, v.374, 2018.

NI, G.; CAVERO, 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. *Genetics Selection Evolution*, v.49, 2017.

NONES, K.; LEDUR, M.C.; ZANELLA, E.L.; KLEIN, C.; PINTO, L.F.B.; MOURA, A.S.A.M.T.; RUY, D.C.; BARON, E.E.; AMBO, M.; CAMPOS, R.L.R.; BOSCHIERO, C.; BURT, D.W.; COUTINHO, L.L. Quantitative trait loci associated with chemical composition of the chicken carcass. *Animal Genetics*, v. 43, p.570-576, 2012.

OBBER, U.; AYROLES, J.F.; STONE, E.A.; RICHARDS, S.; ZHU, D.; GIBBS, R.A.; STRICKER, C.; GIANOLA, D.; SCHLATHER, M.; MACKAY, T.F.C.; SIMIANER, H. Using whole-genome sequence data to predict quantitative trait phenotypes in *Drosophila melanogaster*. *PLoS Genetics*, 8, 5, 2012.

OGAWA, S., HIROKAZU, M.; TANIGUCHI, Y.; WATANABE, T.; NISHIMURA, S.; SUGIMOTO, Y.; IWASAKI, H. 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 Genetics*, v.15, n.15, 2014.

PAKDEL, A.; VAN ARENDONK, J.A.M.; VEREIJKEN, A.L.J.; BOVENHUIS, H. Genetic and phenotypic correlations for ascites related traits in broilers. 7th World Congress on Genetics Applied to Livestock Production, 2002.

PAVLIDIS H.O., BALOG, J.M.; STAMPS, L.K.; HUGHES JR., J.D.; HUFF, W.E.; ANTHONY, N.B. Divergent Selection for ascites incidence in chickens. *Poultry Science*, v.86, p.1517-2529, 2007.

PURCELL, S. M.; NEALE, B.M.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M.A.; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, P. I.; DALY, M.J.; SHAM, P. C. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics*, v.83, p.559-575, 2007.

RANCE, K.A.; MCENTEE, G.M.; MCDEVITT, R.M. Genetic and phenotypic relationships between and within support and demand tissues in a single line of broiler chicken. *British Poultry Science*, v.43, 2002.

ROLF, M.M.; TAYLOR, J.F.; SCHNABEL, R.D.; MCKAY, S.D.; MCCLURE, M.C.; NORTHCUTT, S.L.; KERLEY, M.S.; WEABER, R.L. Impact of reduced marker set estimation of genomic relationship matrices on genomic selection for feed efficiency in Angus cattle. *BMC Genetics*, v. 11, n.24, 2010.

ROVADOSCKI, G.A.; PERTILE, S.F.N.; ALVARENGA, A.B.; CESAR, A.S.M.; PÉRTILLE, F.; PETRINI, J.; FRANZO, V.; SOARES, W.V.B.; MOROTA, G.; SPANGLER, M.L.; PINTO, L.F.B.; CARVALHO, G.G.P.; LANNA, D.P.D.; COUTINHO, L.L.; MOURÃO, G.B. Estimates of genomic heritability and genome-wide association study for fatty acids profile in Santa Inês sheep. *BMC Genomics*, 19, 375, 2018.

SU, G.; BRØNDUM, R.F.; MA, P.; GULDBRANDTSEN, B.; AAMAND, G.P.; 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. *Journal of Dairy Science*, v.95, p.4657-5665, 2012.

TSURUTA, S.; MISZTAL, I.; AGUILAR, I.; LAWLOR, T.J. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. *Journal of Dairy Science*, v.94, p.4198-4204, 2011.

VAN BINSBERGEN, R.; CALUS, M.P.L.; BINK, M.C.A.M.; VAN EEUWIJK, F.A.; SCHROOTEN, C.; VEERKAMP, R.F. Genomic prediction using imputed whole-genome sequence data in Holstein Friesian cattle. *Genetics Selection Evolution*, v.47, n.71, 2015.

VANRADEN, P.M. Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, v.91, p.4414-4423, 2008.

VANRADEN, P.M.; VAN TASSELL, C.P.; WIGGANS, G.R.; SONSTEGARD, T.S.; SCHNABEL, R.D.; TAYLOR, J.F.; SCHENKEL, F.S. Invited review: reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, v.92, p.16-24, 2009.

VANRADEN, P.M.; TOOKER, M.E.; O'CONNELL, J.R.; COLE, J.B.; BICKHART, D.M. Selecting sequence variants to improve genomic predictions for dairy cattle. *Genetics Selection Evolution*, v.49, n.32, 2017.

VEERKAMP, R.F.; BOUWMAN, A.C.; SCHROOTEN, C.; CALUS, M.P.L. Genomic prediction using preselected DNA variants from a GWAS with whole-genome sequence data in Holstein–Friesian cattle. *Genetics Selection Evolution*, 48, n.95, 2016.

VENTURINI, G.C.; CRUZ, V.A.R.; ROSA, J.O.; BALDI, F.; EL FARO, L.; LEDUR, M.C.; PEIXOTO, J.O.; MUNARI, D.P. Genetic and phenotypic parameters of carcass and organ traits of broiler chickens. *Genetic and Molecular Research*, v.13, 2014.

VITEZICA, Z.G.; AGUILAR, I.; MISZTAL, I.; LEGARRA, A. Bias in genomic predictions for populations under selection. *Genetic Research Cambridge*, v.93, p. 357-366, 2011.

WELLMANN, R.; PREU, S.; THOLEN, E.; HEINKEL, J.; WIMMERS, K.; BENNEWITZ, J. Genomic selection using low density marker panels with application to a sire line in pigs. *Genetics Selection Evolution*, 45, 28, 2013.

WIDEMAN, R.F.; RHOADS, D.D.; ERF, G.F.; ANTHONY, N.B. Pulmonary arterial hypertension (ascites syndrome) in broilers: A review. *Poultry Science*, v. 92, p. 64-83, 2013.

YAN, Y.; WU, G. ; LIU, A.; SUN, C. ; HAN, W.; LI, G.; YANG, N. Genomic prediction in a nuclear population of layers using single-step models. *Poultry Science*, v. 97, p. 397-402, 2017.

ZHANG, C.; KEMP, R.A.; STOTHARD, P.; WANG, Z.; BODDICKER, N.; KRIVUSHIN, K.; DEKKERS, J.; PLASTOW, G. Genomic evaluation of feed efficiency component traits in Duroc pigs using 80K, 650K and whole-genome sequence variants. *Genetics Selection Evolution*, v. 50, n. 14 2018.

ZHANG, X.; TSURUTA, S.; ANDONOV, S.; LOURENCO, D.A.L.; SAPP, R.L.; WANG, C.; MISZTAL, I. Relationships among mortality, performance, and disorder traits in broiler chickens: a genetic and genomic approach. *Poultry Science*, v. 97, p. 1511-1518, 2017.

3. GENOMIC PREDICTION USING HIGH DENSITY-PANEL AND IMPUTED WHOLE-GENOME SEQUENCE DATA WITH DIFFERENT GENOMIC RELATIONSHIP MATRICES IN BROILER CHICKENS

ABSTRACT

Availability of high-density panels (HD) of SNP markers has brought promises to solve the problems imposed by the traditional selection process and improve the genomic predictions enable the selection of young animals allowing the implementation of genomic selection (GS) in breeding programs. However, it is unclear what would be the optimal density of the SNP panel to achieve high estimated breeding values (EBV) accuracies with minimal genotyping costs. The aims of this study were to estimate direct genomic values (DGV) and compare results from genomic prediction analyses in a Brazilian broiler chicken using both high-density (HD) panel and imputed whole-genome sequence (WGS) data performed with BayesC model. Estimated genomic heritability for organ and carcass traits varied from low (LUN) to moderate (HRT, LIV, GIZ, LUN, BRST and DRUM). Comparing the datasets available, the highest predictive abilities were obtained from 10%, 20%, 40%, 80% and 100% of SNPs when the HD panel was used. Our findings show that there is no significant increase in prediction accuracy over the WGS file compared to HD panel using BayesC. Potential array data with lower densities (~74,122 SNPs) can provide significant results at a low cost.

Keywords: Genomic prediction; High density panel; Whole-genomic sequence; imputation; Broiler chicken

3.1. INTRODUCTION

Availability of high-density panels (HD) of SNP markers has brought promises to solve the problems imposed by the traditional selection process and improve the genomic predictions enable the selection of young animals allowing the implementation of genomic selection (GS) in breeding programs. Increasing the density of SNP panel increases the probability that any QTL is in perfect linkage disequilibrium (LD) with a single nucleotide polymorphism (SNP) marker (Meuwissen et al., 2016). However, it is unclear what would be the optimal density of the SNP panel to achieve high estimated breeding values (EBV) accuracies with minimal genotyping costs.

When the whole-genome sequence (WGS) is used in genomic predictions, it is expected that higher accuracies could be achieved due to the supposed inclusion of mutations underlying the QTL, which allow estimating the trait QTL effect regardless of LD between the SNPs and QTL (Druet et al., 2014; Van Binsbergen et al., 2015). Despite the cost of genotyping and genome resequencing have fallen, it is still relatively expensive to use a large number of animals to acquire high genomic predictions accuracy. A less expensive solution to reduce the costs, increase the number of animals and at the same time improve the genomic prediction accuracy is to impute density panels (low or high) to WGS.

In livestock production, many studies have been developed using HD array data, WGS data and also imputed data. Based on a simulated study Meuwissen and Goddard (2010) stated that the sequence data prediction accuracy were higher (5-10%) than predictions based on dense markers data. Nevertheless, this increase in accuracy has not been observed in a simulated study performed by Pérez-Enciso et al. (2015). Despite the genetic background, these authors demonstrated that the prediction accuracy did not increase when WGS was used compared to HD data. In real data, neither Ober et al. (2012) who worked with *Drosophila melanogaster* nor Calus et al. (2016) in dairy cattle found positive results using WGS compared to SNP array.

Even when imputed data is used minimal increase in accuracy was obtained. For instance, in cattle the prediction accuracy reported by Hayes et al. (2014) improved 2% when WGS and imputed data were compared to SNP array. In white layer chickens, Heidari et al. (2016) hardly improved genomic prediction accuracy (~1%) when the imputed WGS was used compared to 60K SNP panel. Similar results were found by Ni et al. (2017) which little or no substantial results were obtained in genomic prediction accuracy when using WGS data compared to HD data.

Regardless of the genomic data set used, genomic prediction accuracy can be influenced by many factors, such as the trait heritability, the nature of fixed effects, the extent of additive genetic relationships between individuals and selection candidates, the relationship between animals of reference population with target animals, the size of population, the density of the SNP panel and the statistical method applied for the GEBVs estimation (Goddard, 2009; B.J. Hayes et al., 2009; Daetwyler et al., 2010; Hayes et al., 2010; de los Campos et al., 2013; Weng et al., 2016).

Appropriate genomic prediction approaches have been chosen to take advantage of genomic information. In practice, genomic best linear unbiased prediction (GBLUP) (VanRaden, 2008) and Bayesian methods are used to perform the genomic predictions. The difference between these approaches is related to distribution of SNP effects prior, GBLUP assume normal distribution for all SNP effects, while a non-normality distribution is assumed by Bayesian methods (Meuwissen et al., 2001; Chen et al., 2014). However, despite the variations between these approaches, some results showed no differences in genomic prediction using Bayes or GBLUP approach (Hayes et al., 2009; VanRaden et al., 2009; Ober et al., 2012; Heidaritabar et al., 2016).

The aims of this study were to estimate direct genomic values (DGV) and compare results from genomic prediction analyses in a Brazilian broiler chicken using both high-density (HD) panel and imputed whole-genome sequence (WGS) data performed with BayesC model.

3.2. MATERIALS AND METHODS

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.

3.2.1 Population and phenotypes

The chicken population used in this study was derived from a TT broiler line belonging to the Animal Breeding Program from Embrapa Swine and Poultry. Since 1992, multi-trait selection has been applied in this line, mainly focused on traits such as body weight, feed conversion, carcass weights and cuts yield, fertility, hatchability and to reduce abdominal fat and metabolic syndromes (Nones et al., 2012; Venturini et al., 2014). The TT Reference population is a broiler population developed in 2008 from the crossing between 92 females with 20 males in a hierarchical scheme (1 male: 5 females) producing approximately 1,500 chickens from five hatches.

A total of 1,453 animals (703 males and 750 females) were slaughtered at 42 days of age after six hours of fasting and body weights at 42 days of age (BW42) were recorded. Blood

samples from each animal were collected for DNA extraction and the eviscerated carcass was cooled. After six hours of cooling (4°C) the carcass (breast, drumstick and thigh) and organs (heart, liver, gizzard and lung) were weighed. Descriptive statistics of 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).

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 weights (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 weights (g)						
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	46.11	113.60	464.40	14.85

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

3.2.2 Genotyping

Blood samples of each animal (1,453) were used to extract DNA using PureLink® Genomic DNA (Invitrogen, Carlsbad, CA, USA) kit and quantified using 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 (Affymetrix, Inc. Santa Clara, CA, USA). This genotyping array was developed using segregating SNPs identified in chicken populations, including four commercial broilers (meat-type chicken lines), as described by Kranis et al. (2013).

Axiom™ Analysis Suite (Affymetrix®) software was used to filter based on DishQC parameter, and then PLINK v.1.9 software (Purcell et al., 2007) was used to perform quality control analysis and genotype calling. Samples with DishQC of ≥ 0.82 and call rate of $\geq 90\%$ were kept. 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 -value $< 10^{-7}$) leaving 370,608 SNPs for further analysis (Moreira et al., 2018).

3.2.3 Imputation

Data from whole-genome sequencing (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 crosses to obtain 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 Boschiero et al. (2018) and Moreira et al. (2018).

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 600K to WGS was performed using BEAGLE 4.1 software (Browning and Browning, 2008) with 20 iterations. BEAGLE does not use pedigree information; thus genotypes are predicted based on linkage disequilibrium information using a Hidden Markov Model (HMM) process.

Imputation accuracy was assessed using the validation subset approach. Individuals used in WGS ($n=84$) were randomly split into 14 subsets with 6 animals per group and each group was used as a 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 squared correlation between observed and predicted variants.

After imputation, a quality control was applied to select the sequence variants considering MAF greater than 0.015 and imputation accuracy equal to or greater than 0.95 (e.g., $r^2 \geq 0.95$) leaving 1,421,371 SNPs for further analysis. Furthermore, SNPs were classified into five classes by Variant Effect Predictor (VEP) software (version vep-93.4; McLaren et al., 2016) using galGal5 as a reference genome.

The sequence variants selected to include in further analysis included 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 control of gene expression (Moreira

et al., 2018). Variants in common between the HD and WGS sets were removed leaving 1,095,053 SNPs to compose the WGS dataset, which consists of 69% of intergenic regions of the genome, 16% of downstream, 14% upstream and 1% of UTR3' and UTR5', respectively. Then, from those non-common variants 370,608 SNPs were randomly selected to compose the final WGS dataset.

3.2.4 Prediction

Variant subsets

Seven subsets (0.5, 1, 5, 10, 20, 40 and 80% of SNP) were selected from the full HD panel to determine the impact of using reduced subsets of SNP on genomic prediction. These subsets were chosen based on SNP effects, such that 80% of SNP represented the 80% of SNPs with largest effect and so on. Imputed variants from WGS were also used and mimicked the number of SNP chosen for the subsets mentioned above.

Direct Genomic Values

Direct genomic values (DGV) of each trait [heart (HRT), liver (LIV), gizzard (GIZ), lungs (LUN), breast (BRT), thigh (THG) and drumstick (DRM)] were predicted using GenSel software package (Garrick and Fernando, 2013). The following statistical BayesC model was used (Habier et al., 2011; Kizilkaya et al., 2010):

$$y = Xb + \sum_{j=1}^K z_j a_j + e$$

where y is the vector of observation ($n \times 1$); X is the design matrix for fixed effects; b is the vector of fixed effects which included sex and hatch; K is the number of SNPs; z_j is a vector of genotypes ($n \times 1$) at SNP j based on the number of B alleles ($-10, 0, +10$); a_j is the additive effect of SNP j ; e is the vector of random error effects with a distribution $\sim N(0, I\sigma_e^2)$, where I is an identity matrix and σ_e^2 is the residual variance. The prior for u depends on the variance, σ_u^2 and the prior probability (π) that a SNP has zero effect:

$$u|\pi, \sigma_u^2 = \begin{cases} 0 & \text{with probability } \pi \\ \sim N(0, \sigma_u^2) & \text{with probability } (1 - \pi). \end{cases}$$

Although BayesC model does not assume a known variance, all SNPs effect have a common variance which follows a scaled inverse χ^2 prior with ν_u degrees of freedom and S_a^2 scale parameter. Thus, all SNPs contribute towards the prediction of DGV. The ssGBLUP model was used to estimate the genetic and residual variances for each trait and these values were used as priors to run the BayesC model (Rolf et al., 2015; Heidaritabar et al., 2016).

The π value was assumed according to SNPs subsets to avoid fitting more SNPs than the number of animals in a given iteration. We obtained 41,000 Markov Chain Monte Carlo (MCMC) samples and the first 1,000 samples discarded as burn-in. A map file was used to position the markers split into 1 Mb windows.

3.2.5 Assessment of accuracy and bias

Spearman correlation coefficients for direct genomic values (DGV) between each subset (0.5%, 1%, 5%, 10%, 20%, 40%, 80% and 100% of SNPs) were calculated to determine the impact of using reduced subsets of SNP on the rank of animals. All the subsets were selected based on the SNP effect from the full SNP panel.

Predictive ability was defined as the correlation (r) between DGV and phenotypes corrected for fixed effects (y^*) for animals in the validation population for each trait (Legarra et al., 2008):

$$r = cor(DGV, y^*)$$

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. Moreover, the regression coefficients of the adjusted phenotype on DGV in each scenario were used to measure the degree of similarity between the predictions. A regression coefficient equal to one indicates no bias, whereas values greater than 1 or lower than 1 indicates that DGV is under- or over dispersed, respectively.

3.3. RESULTS

3.3.1 Imputation accuracy

The average of imputation accuracy from the HD panel to whole sequence imputation estimated by Beagle and assessed using the validation subset approach was 0.84 and ranged from 0.79 to 0.88 across all subsets.

3.3.2 MAF distribution

The distribution of MAF for the HD array was uniform, whereas the MAF distribution for WGS variants retained for further analyses were not, it shows a lightly right skewed (Figure 1). After applying the threshold in both HD panel (MAF < 0.02) and WGS data (MAF > 0.015), the average MAF was 0.26.

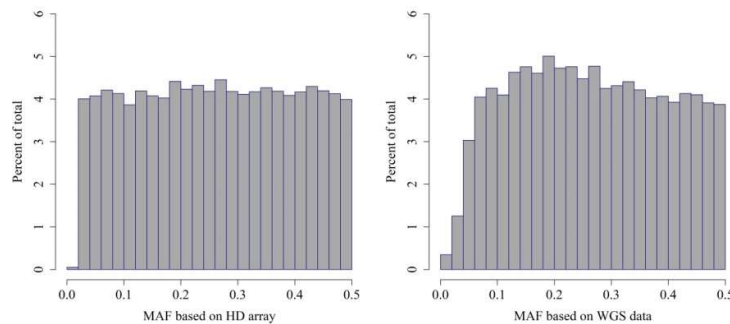


Figure 1. Distribution of minor allele frequency (MAF) for high density (HD) array data and whole-genome sequencing (WGS) data, after post-imputation filtering and before selecting the variants.

3.3.3 Descriptive results

The descriptive statistics of the remaining data are presented in Table 1. The estimates of variance component and heritability for organ and carcass traits obtained through BayesC model using both HD and WGS density panels are given in Tables 2. Despite the selection of variants, the estimation heritabilities were slightly lower when the WGS dataset were used, due to the reduction of genetic variance estimation. In general, small differences in these estimates were observed between the dataset used (HD panel and WGS dataset).

High heritability was estimated for GIZ, with values greater than 0.40 for both datasets. Moderate estimates were obtained for HRT, LIV, BRST, TGH and DRM, varying from 0.28 to 0.33, whereas LUN had low estimation of heritability (0.12), regardless the dataset used.

Table 2. Additive genetic variance (σ_a^2), residual variance (σ_e^2), phenotypic variance (σ_p^2) and genomic posterior mean heritability estimates (h^2), for organ and carcass trait of broiler chicken using BayesC (HD panel).

Trait ¹	σ_a^2	σ_e^2	σ_p^2	h^2
HD Panel				
Organ weight				
HRT	0.83	1.76	2.59	0.32
LIV	20.34	46.72	67.07	0.30
GIZ	13.85	18.92	32.73	0.42
LUN	0.74	5.26	5.99	0.12
Carcass weight				
BRST	947.69	2099.08	3046.77	0.31
DRM	146.95	299.98	446.93	0.33
THG	384.11	884.28	1268.39	0.30
WGS Dataset				
Organ weight				
HRT	0.77	1.81	2.59	0.30
LIV	19.52	47.42	66.94	0.29
GIZ	13.26	18.48	32.75	0.40
LUN	0.72	5.28	6.00	0.12
Carcass weight				
BRST	921.94	2123.82	3045.76	0.30
DRM	141.03	305.98	447.01	0.31
THG	361.43	907.60	1269.03	0.28

¹HRT=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).

3.3.4 Prediction

Regression coefficients were similar for both the HD and WGS sets of SNPs, ranging from 0.97 to 16.39 across scenarios (Table 3). For DRUM, a slightly underestimation of the breeding values was observed when the full HD panel was used (0.970). The same pattern was observed using WGS dataset (0.998). All regression coefficients were underestimation when especially lower SNP density (0.5%, 1%, 5%, 10%, 20%) were used.

However, better regression coefficients were obtained when at least 40% of SNPs were used in the analyses, except for LUN and BRST, which regression coefficients lower than 1.3 were obtained when 80% and 100% of SNPs were used. Except for BRST, the regression coefficients close to 1 were acquired only with 100% of SNP.

Regarding the predictive ability, traits with higher heritabilities (e.g. GIZ and THG) showed higher predictive ability than trait with lower heritabilities (e.g. LUN), which suggest that traits with low heritability requires more records to achieve higher predictive abilities as traits with high heritabilities. Compared to WGS dataset, the predictive ability obtained by HD panel was slightly higher from 10% of SNPs, but no significant improvement was achieved in the predictive ability when more than 20% of SNPs were used (Table 4). In general, predictive abilities were greater for subsets of SNP with largest estimated effect. Predictive ability across all traits improved when the number of SNPs with major effects rose.

Table 3. Regression coefficients (by*, DGV) and standard errors (in brackets) for organ and carcass traits in HD panel and WGS data in each scenario (SNP percentage).

HD panel							
SNP (%)	HRT ¹	LIV ¹	GIZ ¹	LUN ¹	BRST ¹	DRM ¹	THG ¹
0.5	7.60 (1.219)	6.04 (1.011)	7.39 (0.9722)	16.39 (4.237)	10.90 (1.459)	9.14 (1.111)	8.07 (1.261)
1	5.53 (0.839)	4.54 (0.693)	5.50 (0.652)	11.73 (2.845)	7.65 (0.978)	6.49 (0.75)	5.64 (0.834)
5	2.71 (0.376)	2.59 (0.333)	2.59 (0.267)	5.32 (1.193)	3.56 (0.430)	2.93 (0.328)	2.54 (0.347)
10	2.14 (0.286)	2.08 (0.258)	1.96 (0.196)	3.73 (0.826)	2.71 (0.321)	2.23 (0.245)	1.94 (0.257)
20	1.72 (0.225)	1.68 (0.204)	1.56 (0.153)	2.59 (0.575)	2.08 (0.243)	1.74 (0.193)	1.54 (0.199)
40	1.38 (0.180)	1.35 (0.162)	1.27 (0.123)	1.80 (0.400)	1.65 (0.189)	1.39 (0.154)	1.24 (0.157)
80	1.15 (0.148)	1.10 (0.132)	1.06 (0.102)	1.26 (0.285)	1.32 (0.151)	1.12 (0.126)	1.01 (0.127)
100	1.10 (0.142)	1.05 (0.126)	1.02 (0.098)	1.14 (0.261)	1.24 (0.143)	1.06 (0.120)	0.97 (0.121)
WGS data							
0.5	8.09 (1.356)	5.91 (1.046)	7.23 (0.934)	12.73 (3.902)	9.31 (1.441)	9.17 (1.161)	8.95 (1.464)
1	5.90 (0.943)	4.48 (0.718)	5.24 (0.636)	9.94 (2.691)	6.91 (0.993)	6.60 (0.789)	6.11 (0.933)
5	3.01 (0.428)	2.63 (0.351)	2.52 (0.275)	5.10 (1.174)	3.49 (0.436)	3.08 (0.341)	2.87 (0.386)
10	2.35 (0.323)	2.13 (0.270)	1.94 (0.202)	3.51 (0.804)	2.67 (0.323)	2.34 (0.256)	2.16 (0.283)
20	1.85 (0.249)	1.72 (0.211)	1.56 (0.157)	2.47 (0.561)	2.09 (0.246)	1.83 (0.199)	1.67 (0.213)
40	1.47 (0.196)	1.36 (0.166)	1.28 (0.127)	1.72 (0.389)	1.64 (0.188)	1.43 (0.157)	1.31 (0.164)
80	1.20 (0.159)	1.10 (0.133)	1.07 (0.105)	1.19 (0.275)	1.31 (0.149)	1.13 (0.127)	1.05 (0.130)
100	1.14 (0.151)	1.04 (0.126)	1.02 (0.100)	1.08 (0.251)	1.24 (0.141)	1.07 (0.121)	0.99 (0.123)

¹HRT=Heart weight (g); LIV=Liver weight (g); GIZ=Gizzard weight (g); LUN=Lungs weight (g); BRST=Breast weight (g); DRUM=Drumstick weight (g); THG =Thigh weight (g).

Table 4. Predictive abilities of HD panel and WGS data in each scenario (SNP percentage).

HD panel							
SNP (%)	HRT ¹	LIV ¹	GIZ ¹	LUN ¹	BRST ¹	DRM ¹	THG ¹
0.5	0.289	0.286	0.381	0.182	0.338	0.346	0.320
1	0.302	0.311	0.407	0.193	0.351	0.365	0.335
5	0.329	0.354	0.441	0.208	0.363	0.393	0.357
10	0.338	0.362	0.448	0.211	0.368	0.400	0.365
20	0.341	0.368	0.451	0.213	0.370	0.400	0.370
40	0.344	0.369	0.453	0.213	0.373	0.399	0.376
80	0.347	0.370	0.453	0.211	0.374	0.396	0.378
100	0.347	0.370	0.453	0.210	0.374	0.395	0.379
WGS data							
0.5	0.294	0.271	0.371	0.169	0.303	0.326	0.304
1	0.311	0.295	0.389	0.185	0.323	0.351	0.321
5	0.332	0.343	0.417	0.205	0.355	0.389	0.354
10	0.337	0.355	0.430	0.208	0.363	0.397	0.360
20	0.339	0.364	0.437	0.209	0.369	0.399	0.367
40	0.340	0.365	0.441	0.210	0.373	0.399	0.374
80	0.342	0.367	0.442	0.208	0.375	0.396	0.376
100	0.343	0.367	0.442	0.206	0.375	0.394	0.377

¹HRT=Heart weight (g); LIV=Liver weight (g); GIZ=Gizzard weight (g); LUN=Lungs weight (g); BRST=Breast weight (g); DRM=Drumstick weight (g); THG =Thigh weight (g).

3.4. DISCUSSION

In this study we investigated if the selection of subsets of major SNP effects, such as those described above, will improve the genomic selection by estimating the predictive ability of DGV for organs (heart, liver, lungs and gizzard) and carcass (breast, thigh and drumstick) traits with HD array and sequence data using BayesC model in broiler chickens.

Major SNPs identified were evaluated based on their effects estimated in training population and by re-estimation their effects in the validation set. Genomic prediction methods might be able to estimate the effect of a large number of SNPs (p) from a smaller number of phenotypic information. Due to the $p \gg n$ problem, the availability of HD data set or sequence data used to perform the genomic evaluation can lead to a poor estimation of SNP effects, an error in the estimation of the causal mutations effect and a considerable part of causal mutations with large effect might be distributed over multiple SNPs (Heidaritabar et al., 2016). Approaches like BayesC have been developed to solve $p \gg n$ problem by Markov

Chain Monte Carlo (MCMC), the idea is retaining the causal mutation by the regressing toward zero the false-positive or uninformative SNP effects.

3.4.1. Genomic heritability

The estimates of variance components and heritability for carcass and organ traits obtained through BayesC are provided in Table 2. The heritability estimates varied from low (LUN) to moderate (HRT, LIV, LUN, BRST and DRM) and high (GIZ).

Pedigree-based estimates of heritability have been reported in the literature for most of traits used in this study. Using the same population (Embrapa TT), Venturini et al. (2014) reported similar pedigree-based heritability estimates for LIV (0.33 ± 0.07) and GIZ (0.44 ± 0.08) and THG (0.29 ± 0.06) to these reported herein. However, the genomic heritability estimate found in this study for BRST and DRUM were lower (0.31 , 0.33 and 0.30) than the estimate in Venturini et al. (2014), BRST (0.37 ± 0.07) and DRM (0.35 ± 0.07).

THG and DRUM 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 Argentão et al. (2002), Rance et al. (2002) and Gaya et al. (2006). In a study with a male broiler line, Gaya et al. (2006) 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. (2002) reported heritability estimates for HRT (0.30 ± 0.08), LIV (0.08 ± 0.06), GIZ (0.52 ± 0.10).

The heritability estimates for LUN in broiler chicken are not common in the literature. Using a F_2 experimental population Ledur et al. (2006) reported similar pedigree-based heritability estimates for LUN (0.10) to this reported herein. Although LUN is not considered an economically important trait, it has been related to pulmonary hypertension (e.g. ascites). Heritability estimation for ascites have been reported by several authors (Moghadam et al., 2001; Deeb et al., 2002; Pakdel et al., 2002; Ledur et al., 2006; Pavlidis et al., 2007; Wideman et al. 2013).

3.4.2 Correlation between DGV

Eight differing subsets were assessed in HD and WGS data set (results not shown). The DGV correlation for both data sets increase with increase of the percentage of SNPs. Across all traits, DGV estimated with at least 5% of SNPs were highly correlated (>0.94) with DGV estimated from the complete HD array. Indeed, lower correlations (>0.80) were observed when a smaller number of SNPs sets (0.5% and 1%) were used. Although, the same pattern has been noted with WGS, no significant improvement in the correlation value was acquired with this dataset. Comparing the correlations between DGV using different genotype datasets our results show no difference when 20% or 100% of SNPs were used in the analyses which suggests that the use of evenly-spaced lower-density panel could provide a very similar ranking of EBV at a potentially lower cost, as proposed by Habier et al (2009).

3.4.3 Regression coefficients

DGV bias in each scenario was calculated as the regression of adjusted phenotype on DGV (Table 3). A regression coefficient (slope) equal to one indicates no bias. In practice this means that a suitable genetic merit prediction method has been applied (Vitezica et al., 2011). Our results showed higher regression coefficients in HD data set for HRT, THG and DRM whereas for the other traits (LIV, GIZ, LUN and BRST) showed lower values when the same data set was used. In both data sets, DRM and LUN showed the lowest and the highest regression coefficient, respectively.

Even though the regression values decrease with the increase of SNP density, showing better values for 80% and 100% of SNPs for all traits, the regression coefficients were significantly underestimation in the DGV predictions in lowest SNP densities, regardless the data set used. A possible explanation for this bias underestimation may be related to the fact that those SNPs are not explaining a large fraction of variation or they are not causative. In addition, the non-re-estimated of the SNPs effect after the SNP selection can also be a reason for the underestimation in the DGV predictions.

3.4.5 Predictive ability

Predictive ability of DGV was evaluated based on the correlation between the DGV and phenotypes corrected for fixed effects (Table 4) in the validation set. In the both data set the GIZ showed the highest prediction ability value range from 0.37 to 0.45, while LUN had the lowest value range from 0.16 to 0.21. Our findings showed there is no clearly significant increase in prediction accuracy over the imputed file compared to HD dataset regardless of the SNP percentage used.

Simulated data has shown an increase in genomic prediction accuracy when the causal mutations were included in the analyses (Meuwissen and Goddard, 2010; Druet et al., 2014; MacLeod et al., 2014). Contrary to those findings, our study showed no significant increase in prediction accuracy over the imputed file compared to HD dataset. Indeed, lower or no significant benefits had been achieved in predictive ability gain using sequence data comparing with SNP array (Ober et al., 2012; Van Binsbergen et al., 2015; Heidaritabar et al., 2016; MacLeod et al., 2016; VanRaden et al., 2017; Frischknecht et al., 2017).

Many reasons might be responsible to explain why not any increase in predictive ability was observed in those results. In a simulated study, Clark et al. (2011) suggested that the increase of genomic prediction accuracy will be smaller when the trait is highly polygenic especially in a small reference population. Further, 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 (VanRaden et al., 2017).

Another possible reason is related to the population structure. When a small effective population size suffer selection process over the years no significant results in accuracy are obtained resulting in a high level of LD (MacLeod et al., 2014). Thus, in this type of population, almost the totally genetic variance can be explained by the SNPs genetic variance as result of the relationship between individuals, which leads to a low extension of LD (VanRaden et al., 2009). Hence, a large number of SNPs are required to reach the genomic prediction accuracy (Wray et al., 2007).

One important factor that is not usually taken into account is the WGS full information. In livestock studies the use of WGS data is commonly focused on SNPs, but, in

theory, this data also include all DNA variants, such as, InDels and copy number variations (CNV) which are important role to gene expression and phenotype variation (McCarroll et al., 2006; Redon et al., 2006) and therefore, may be influence in the prediction ability. Moreover, select variants and use appropriate prediction models are needed to increase the predictive ability since the density of WGS data is higher compared to HD array.

3.5. CONCLUSION

In this study, we investigated different density marker panels for prediction of genomic breeding values in a broiler chicken population. Our results demonstrate that there is no difference when 20% or 100% of SNPs were used to inform kinship in the prediction of direct genomic breeding values, but currently, suggest that at least 20% of SNP (~74,122 SNPs) can provide consistent genetic evaluations.

Therefore, the use of lower-density arrays, if at a lower cost, could be used to rank individual based on genetic merit. In addition, no significant improvements in genomic prediction accuracy were noticed when imputation to WGS data were used in broiler chickens comparing with the predictions based on HD array. Increasing the number of sequenced animals or reducing the relationship between animals in reference population may help to improve the predictions when WGS is available.

REFERENCES

ARGENTÃO, C.; FILHO, T.M.; MARQUES, J.L.B.; SOUZA, E.M.; ELER, J.P.; FERRAZ, J.B.S. Genetic and phenotypic parameters of growth and carcass traits of a male line of broilers raised in tropical conditions. 7th World Congress on Genetics Applied to Livestock Production, 2002.

BOSCHIERO, C.; MOREIRA, G.C.M.; GHEYAS, A.A.; GODOY, T.F.; GASPARIN, G.; MARIANI, P.D.S.C.; PADUAN, M.; CESAR, A.S.M.; LEDUR, M.C.; COUTINHO, L.L. Genome-wide characterization of genetic variants and putative regions under selection in meat and egg-type chicken lines. *BMC Genomics*, 19, n. 83, 2018.

BROWNING, B.L.; BROWNING, S.R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *American Journal of Human Genetics*, v. 84, p. 210-223, 2008.

CALUS, M.P.L.; BOUWMAN, A.C.; SCHROOTEN, C.; VEERKAMP, R.F. Efficient genomic prediction based on whole-genome sequence data using split-and-merge Bayesian variable selection. *Genetic Selection Evolution*, v.48, n.49, 2016.

CHEN, L.; LI, C.; SARGOLZAEI, M.; SCHENKEL, F. Impact of genotype imputation on the performance of GBLUP and Bayesian methods for genomic prediction. *PLoS One*, v.9, 2014.

CLARK, S.A.; HICKEY, J.M.; VAN DER WERF, J.H.J. Different models of genetic variation and their effect on genomic evaluation. *Genetics Selection Evolution*, 43, 18, 2011.

DAETWYLER, H.D.; PONG-WONG, R.; VILLANUEVA, B.; WOOLLIAMS, J.A. The impact of genetic architecture on genome-wide evaluation methods. *Genetics*, v.185, p.1021-1031, 2010.

DE LOS CAMPOS, G.; HICKEY, J.M.; PONG-WONG, R.; DAETWYLER, H.D.; CALUS, M.P.L. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, v.193, p.327-345, 2013.

DEEB N., SHLOSBERG, A.; CAHANER, A. Genotype-by-environment interaction with broiler genotypes differing in growth rate. 4. Association between responses to heat stress and to cold-induced ascites. *Poultry Science*, v.81, p.1454-1462, 2002.

DRUET, T.; MACLEOD, I.M.; HAYES, B.J. Toward genomic prediction from whole-genome sequence data: Impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity*, v.112, p.39-47, 2014.

FRISCHKNECHT, M.; MEUWISSEN, T.H.E.; BAPST, B.; SEEFRIED, F.R.; FLURY, C.; GARRICK, D.; SIGNER-HASLER, H.; STRICKER, C.; BIEBER, A.; FRIES, R.; RUSS, I.; SÖLKNER, J.; BAGNATO, A.; GREDLER-GRANDL, B. Short communication: Genomic prediction using imputed whole-genome sequence variants in Brown Swiss Cattle. *Journal of Dairy Science*, v.101, p.1292-1296, 2018.

GARRICK, D.J.; FERNANDO, R.L. Implementing a QTL detection study (GWAS) using genomic prediction methodology. *Methods in Molecular Biology*, v. 1019, 2013.

GAYA, L.G.; FERRAZ, J.B.S.; REZENDE, F.M.; MOURÃO, G.B.; MATTOS, E.C.; ELER, J.P.; FILHO, T.M. Heritability and genetic correlation estimates for performance and carcass and body composition traits in a male broiler line. *Poultry Science*, v.85, p. 837-843, 2006.

GODDARD, M. Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica*, v.136, n.245-257, 2009.

HABIER, D.; FERNANDO, R.L.; DEKKERS, J.C.M. Genomic selection using low-density marker panels. *Genetics*, v.182, n.1, p.343-353, 2009.

HABIER, D.; FERNANDO, R.L.; KIZILKAYA, K.; GARRICK, D.J. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics*, v.12, n.186, 2011.

HAYES, B.J.; PRYCE, J.; CHAMBERLAIN, A.J.; BOWMAN, P.J.; GODDARD, M.E. Genetic architecture of complex traits and accuracy of genomic Prediction: Coat colour, Milk-fat percentage, and type in holstein cattle as contrasting model traits. *PLoS Genetics*, v.6, 2010.

HAYES, B.J.; BOWMAN, P.J.; CHAMBERLAIN, A.J.; GODDARD, M.E. Invited review: Genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science*, v.92, p.433-443, 2009.

HAYES, B.J.; MACLEOD, I.M.; DAETWYLER, H.D.; BOWMAN, P.J.; CHAMBERLIAN, A.J.; VANDER JAGT, C.J.; CAPITAN, A.; PAUSCH, A., STOTHARD, P.; LIAO, X.; SCHROOTEN, C.; MULLAART, E.; FRIES, R.; GULDBRANDTSEN, B.; LUND, M.S.; BOICHARD, D.A.; VEERKAMP, R.F.; VANTASSELL, C.P.; GREGLER, B., DRUET, T.; BAGNATO, A.; VILKKI, J.; DEKONING, D.J.; SANTUS, E.; GODDARD, M.E. Genomic prediction from whole genome sequence in livestock: the 1000 bull genomes project. In: *Proceedings of the 10th World Congress on Genetics Applied to Livestock Production*. Vancouver (Canada), 2014.

HEIDARITABAR, M.; CALUS, M.P.L.; MEGENS, H.J.; VEREIJKEN, A.; GROENEN, M.A.M.; BASTIAANSEN, J.W.M. Accuracy of genomic prediction using imputed whole-genome sequence data in white layers. *Journal of Animal Breeding and Genetics*, v.133, p.167–179, 2016.

KIZILKAYA, K.; FERNANDO, R.L.; GARRICK, D.J. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *Journal of Animal Science*, v.88, p.544-551, 2010.

KRANIS, A., GHEYAS, A.A.; BOSCHIERO, C.; TURNER, F.; YU, L.; SMITH, S.; TALBOT, R.; PIRANI, A.; BREW, F.; KAISER, P.; HOCKING, P. M.; FIFE, M.; SALMON, N.; FULTON, J.; STROM, T.M.; HABERER, G.; WEIGEND, S.; PREISINGER, R.; GHOLAMI, M.; QANBARI, S.; SIMIANER, H.; WATSON, K.A.; WOOLLIAMS, J.A.; BURT, D.W. Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics*, 14, n.59, 2013.

LEDUR, M.C.; MELO, C.M.R.; NONES, K.; ZANELLA, E. L.; NONOV, K.; BONASSE, C. A.; JAENISCH, F.R.F.; MOURA, A.S.A.M.; COUTINHO, L.L.; Schmidt, G.S. Genetic and phenotypic parameters for organs, body and carcass weights and haematocrit value in a broiler x layer cross resource population. In: 8th World Congress on Genetics Applied to Livestock Production, 2006.

LEGARRA, A.; ROBERT-GRANIÉ, C.; MANFREDI, E.; ELSEN, J.M. Performance of genomic selection in mice. *Genetics*, v.180, p.611-618, 2008.

MACLEOD, I.M.; HAYES, B.J.; GODDARD, M.E. The effects of demography and long-term selection on the accuracy of genomic prediction with sequence data. *Genetics*, v.198, p.1671-1684 2014.

MACLEOD, I.M.; BOWMAN, P.J.; JAGT, C.J.V.; HAILE-MARIAM, M.; KEMPER, K.E.; CHAMBERLAIN, A.J.; SCHROOTEN, C.; HAYES, B.J.; GODDARD M.E. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics*, 17, n.144, 2016.

MCCARROLL, S. A.; HADNOTT, T.N.; PERRY, G.H.; SABETI, P.C.; ZODY, M.C.; BARRETT, J.C.; DALLAIRE, S.; GABRIEL, S.B.; LEE, C.; DALY, M.J.; ALTSHULER, D.M. Common deletion polymorphisms in the human genome. *Nature Genetetic*, v.38, p.86-92, 2006.

MCLAREN, W.; GIL, L.; HUNT, S.E.; RIAT, H.S.; RITCHIE, G.R.S.; THORMANN, A.; FLICEK, P.; CUNNINGHAM, F. The Ensembl Variant Effect Predictor. *Genome Biology*, 17, n. 122, 2016.

MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, v.157, p.1819-1829, 2001.

MEUWISSEN, T.; GODDARD, M. Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics*, v.185, p.623-631, 2010.

MEUWISSEN, T.; HAYES, B.; GODDARD, M. Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers*, v.6, p.6-14, 2016.

MOGHADAM, H.K.; MCMILLAN, I.; CHAMBERS, J.R.; JULIAN, R.J. Estimation of genetic parameters for ascites syndrome in broiler chickens. *Poultry Science*, v. 7, p. 844-848, 2001.

MOREIRA, G.C.M.; BOSCHIERO, C.; CESAR, A.S.M.; REECY, J.M.; GODOY, T.F.; TREVISOLI, P.A.; CANTÃO, M.E.; LEDUR, M.C.; IBELLI, A.M.G.; PEIXOTO, J.O.; MOURA, A.S.A.M.T.; GARRICK, D.; COUTINHO, L.L. 2018. A genome-wide association study reveals novel genomic regions and positional candidate genes for fat deposition in broiler chickens. *BMC Genomics*, 19, v.374, 2018.

NI, G.; CAVERO, 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. *Genetics Selection Evolution*, v.49, 2017.

NONES, K.; LEDUR, M.C.; ZANELLA, E.L.; KLEIN, C.; PINTO, L.F.B.; MOURA, A.S.A.M.T.; RUY, D.C.; BARON, E.E.; AMBO, M.; CAMPOS, R.L.R.; BOSCHIERO, C.; BURT, D.W.; COUTINHO, L.L. Quantitative trait loci associated with chemical composition of the chicken carcass. *Animal Genetics*, v. 43, p.570-576, 2012.

OBBER, U.; AYROLES, J.F.; STONE, E.A.; RICHARDS, S.; ZHU, D.; GIBBS, R.A.; STRICKER, C.; GIANOLA, D.; SCHLATHER, M.; MACKAY, T.F.C.; SIMIANER, H. Using whole-genome sequence data to predict quantitative trait phenotypes in *Drosophila melanogaster*. *PLoS Genetics*, 8, 5, 2012.

PAKDEL, A.; VAN ARENDONK, J.A.M.; VEREIJKEN, A.L.J.; BOVENHUIS, H. Genetic and phenotypic correlations for ascites related traits in broilers. 7th World Congress on Genetics Applied to Livestock Production, 2002.

PAVLIDIS H.O., BALOG, J.M.; STAMPS, L.K.; HUGHES JR., J.D.; HUFF, W.E.; ANTHONY, N.B. Divergent Selection for ascites incidence in chickens. *Poultry Science*, v.86, p.1517-2529, 2007.

PÉREZ-ENCISO, M.; RINCÓN, J.C.; LEGARRA, A. Sequence- vs. chip-assisted genomic selection: Accurate biological information is advised. *Genetic Selection Evolution*, v.47, n.43, 2015.

RANCE, K.A.; MCENTEE, G.M.; MCDEVITT, R.M. Genetic and phenotypic relationships between and within support and demand tissues in a single line of broiler chicken. *British Poultry Science*, v.43, 2002.

REDON, R.; ISHIKAWA, S.; FITCH, K.R.; FEUK, L.; PERRY, G.H.; ANDREWS, T.D.; FIEGLER, H.; SHAPER, M.H.; CARSON, A.R.; CHEN, W.; CHO, E.K.; DALLAIRE, S.; FREEMAN, J.L.; GONZÁLEZ, J.R.; GRATACÒS, M.; HUANG, J.; KALAITZOPOULOS, D.; KOMURA, D.;

MACDONALD, J.R.; MARSHALL, C.R.; MEI, R.; MONTGOMERY, L.; NISHIMURA, K.; OKAMURA, K.; SHEN, F.; SOMERVILLE, M.J.; TCHINDA, J.; VALSESIA, A.; WOODWARK, C.; YANG, F.; ZHANG, J.; ZERJAL, T.; ZHANG, J.; ARMENGOL, L.; CONRAD, D.F.; ESTIVILL, X.; TYLER-SMITH, C.; CARTER, N.P.; ABURATANI, H.; LEE, C.; JONES, K.W.; SCHERER, S.W.; HURLES, M.E. Global variation in copy number in the human genome. *Nature*, v.444, p.444-454, 2006.

ROLF, M.M.; GARRIK, D.J.; FOUNTAIN, T.; RAMEY, H.R.; WEABER, R.L.; DECKER, J.E.; POLLAK, E.L.; SCHNABEL, R.D.; TAYLOR, J.F. Comparison of Bayesian models to estimate direct genomic values in multi-breed commercial beef cattle. *Genetic Selection Evolution*, v.47, n.23, 2015.

VANRADEN, P.M. Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, v.91, p.4414-4423, 2008.

VANRADEN, P.M.; VAN TASSELL, C.P.; WIGGANS, G.R.; SONSTEGARD, T.S.; SCHNABEL, R.D.; TAYLOR, J.F.; SCHENKEL, F.S. Invited review: reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, v.92, p.16-24, 2009.

VANRADEN, P.M.; Tooker, M.E.; O'Connell, J.R.; Cole, J.B.; Bickhart, D.M. Selecting sequence variants to improve genomic predictions for dairy cattle. *Genetics Selection Evolution*, v.49, n.32, 2017.

VENTURINI, G.C.; CRUZ, V.A.R.; ROSA, J.O.; BALDI, F.; EL FARO, L.; LEDUR, M.C.; PEIXOTO, J.O.; MUNARI, D.P. Genetic and phenotypic parameters of carcass and organ traits of broiler chickens. *Genetic and Molecular Research*, v.13, 2014.

VITEZICA, Z.G.; AGUILAR, I.; MISZTAL, I.; LEGARRA, A. 2011. Bias in genomic predictions for populations under selection. *Genetic Research Cambridge*, v.93, p. 357-366, 2011.

WENG, Z.; WOLC, A.; SHEN, X.; FERNANDO, R.L.; DEKKERS, J.C.M.; ARANGO, J.; SETTAR, P.; FULTON, J.E.; O'SULLIVAN, N.P.; GARRICK, D.J. Effects of number of training generations on genomic prediction for various traits in a layer chicken population. *Genetic Selection Evolution*, v.48, n.22, 2016.

WIDEMAN, R.F.; RHOADS, D.D.; ERF, G.F.; ANTHONY, N.B. Pulmonary arterial hypertension (ascites syndrome) in broilers: A review. *Poultry Science*, v. 92, p. 64-83, 2013.

WRAY, N. R.; GODDARD, M.E.; VISSCHER, P.M. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Research*, v.17, p.1520-1528, 2007.

ZHANG, X.; TSURUTA, S.; ANDONOV, S.; LOURENCO, D.A.L.; SAPP, R.L.; WANG, C.; MISZTAL, I. Relationships among mortality, performance, and disorder traits in broiler chickens: a genetic and genomic approach. *Poultry Science*, v. 97, p. 1511-1518, 2017.