Effect of allele frequencies, effect sizes and number of markers on prediction of quantitative traits in chickens


1 Department of Animal Science, University College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran
2 Department of Animal Sciences, University of Wisconsin, Madison, WI, USA
3 Department of Dairy Science, University of Wisconsin, Madison, WI, USA
4 Aviagen Ltd, Midlothian, UK
5 The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, UK
6 Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI, USA

Keywords
Broiler chicken; genomic best linear unbiased prediction; genome-enabled prediction; marker density; marker effect sizes; minor allele frequency; predictive ability.

Summary
The objective was to assess goodness of fit and predictive ability of subsets of single nucleotide polymorphism (SNP) markers constructed based on minor allele frequency (MAF), effect sizes and varying marker density. Target traits were body weight (BW), ultrasound measurement of breast muscle (BM) and hen house egg production (HHP) in broiler chickens. We used a 600 K Affymetrix platform with 1352 birds genotyped. The prediction method was genomic best linear unbiased prediction (GBLUP) with 354 564 single nucleotide polymorphisms (SNPs) used to derive a genomic relationship matrix (G). Predictive ability was assessed as the correlation between predicted genomic values and corrected phenotypes from a threefold cross-validation. Predictive ability was 0.27 ± 0.002 for BW, 0.33 ± 0.001 for BM and 0.20 ± 0.002 for HHP. For the three traits studied, predictive ability decreased when SNPs with a higher MAF were used to construct G. Selection of the 20% SNPs with the largest absolute effect sizes induced a predictive ability equal to that from fitting all markers together. When density of markers increased from 5 K to 20 K, predictive ability enhanced slightly. These results provide evidence that designing a low-density chip using low-frequency markers with large effect sizes may be useful for commercial usage.

Introduction
The ability to predict quantitative trait phenotypes accurately from information on genotypic variation is important for plant and animal breeding, medicine and evolutionary biology. In classical animal and plant breeding, the prediction of genetic values has been made on the basis of recorded phenotypes and pedigree information. However, the advent of high-throughput genotyping platforms for many agronomic species (e.g. Harris et al. 2011; Kranis et al. 2013) has enabled genotyping large numbers of individuals for dense panels of single nucleotide polymorphisms (SNPs) spanning the entire genome. The pedigree-based numerator relationship matrix used in best linear unbiased prediction models traditionally applied to genetic evaluation can then be replaced by some genomic relationship (G) matrix (e.g. Nejati-Javaremi et al. 1997). In animal and plant breeding, selection based on genome-based predictions of genetic values may markedly increase the rate of genetic progress (Schaeffer 2006; Hayes et al. 2009) and has been rapidly adopted in many species.
Recently, a large proportion of the ‘missing heritability’ for human height was recovered statistically by including hundreds of thousands of SNPs jointly in a linear statistical model (Makowsky et al. 2011). However, it is unclear to what extent a larger proportion of explained variance translates into an improved predictive accuracy in a future sample because an excellent goodness of fit in training data and predictive accuracy in a testing set often are in conflict. Heritability estimates can be regarded as measures of goodness of fit but increasing goodness of fit will not necessarily lead to increased predictive accuracy in future samples due to issues such as model over-fitting (Makowsky et al. 2011).

The accuracy of prediction methods based on marker data depends on the heritability of the trait, the number of loci affecting trait variation, mode of inheritance, distribution of allelic effects, distribution of allele frequencies, linkage disequilibrium (LD), size of the genome, marker density and the size and type of sample used in the statistical analysis (Makowsky et al. 2011; de los Campos et al. 2013).

It has been argued that the minor allele frequency (MAF) of SNPs in a panel may be different from that of causal variants, with causal variants having a lower minor allele frequency than common SNPs investigated to date (Yang et al. 2010). We found that 75% of genomic variance was explained by SNPs with MAF <0.20 for body weight (BW) and ultrasound measurement of breast muscle (BM), and MAF <0.30 for hen house egg production (HHP) (Abdollahi-Arpanahi et al. in press). Furthermore, while an estimate of genomic heritability may be viewed as a goodness of fit measure that is valid for the data set used to fit the model, it may not reflect quality of prediction. Different MAF bins may deliver different predictive ability and that is typically masked in a standard genomic best linear unbiased prediction (GBLUP) implementations.

Lettre (2011) and Park et al. (2011) pointed out that there is an inverse relationship between estimated marker effect size and allele frequencies, suggesting that effect sizes are estimated with low precision when MAF is small, that rare alleles have truly large effects, or both. However, using SNPs at low frequencies in a population may not be feasible for prediction because they are likely to appear as monomorphic in a finite size training sample.

Although there have been some studies on relating the marked genetic variance to various partitionings of the genome (e.g. Yang et al. 2010; Williams et al. 2013), we are not aware of reports on the relationship between genome-enabled prediction and different partitions of markers when using a high-density chip in livestock. The objective of this study was to compare the goodness of fit and predictive ability stemming from using different categories of SNPs formed according to MAF, effects sizes and marker densities when studying BW, BM and HHP in chickens. These three traits display considerable genetic variation in chicken populations and have been under intense artificial selection for a long time (Koerhuis & McKay 1996). We evaluated goodness of fit and predictive ability via a threefold cross-validation (CV) replicated five times.

Materials and methods

Data

Data were provided by Aviagen (Newbridge, Scotland). The traits studied were BW at 35 days of age, ultrasound area of BM and HHP defined as the total number of eggs laid between weeks 28 and 54, measured on 1351, 1336 and 823 birds, respectively. Phenotype records for BW and BM were precorrected for combined effects of sex, hatch week, contemporary group of parents and pen in the growing farm, whereas phenotype records for HHP were precorrected for hatch effects. Corrected phenotypes were merged with marker records on individuals that were typed using the Affymetrix® Axiom® 600 K Array Set. Details on genotyping are described by Kranis et al. (2013) and characteristics of the phenotypic data are presented in R. Abdollahi-Arpanahi, A. Pakdel, A. Nejati-Javaremi, M. Moradi Shahrbabak, A. Kranis, G. Morota, B.D. Valente, G.J.M. Rosa, D. Gianola (in preparation).

A total of 580 954 SNP marker genotypes were originally available in the data set. Markers were excluded from the analysis if departing from Hardy–Weinberg equilibrium (p < 10^{-6}) had MAF <0.01 and missing rate >0.05. Finally, 354 364 markers remained for the analysis. The small proportion of missing genotypes that remained was imputed using BEAGLE (Browning & Browning 2009). The mean MAF in our study was 0.24. Only SNPs on 28 autosomes (1–28) were included. Data edition was performed using the PLINK software (Purcell et al. 2007). Characteristics of the SNPs are shown in Table 1.

Partitioning markers into subsets

To investigate the goodness of fit and predictive ability delivered by different MAF bins, SNPs were classified into five groups of MAF (0.01–0.09, n = 70 780; 0.09–0.195, n = 71 027; 0.195–0.294, n = 71 254;
Prediction of complex traits with SNP markers

Where the frequency of base population and therefore needed to be estimated. However, recent work by Forni et al. (2011) suggests that similar results can be obtained using the allele frequencies of the current population.

Goodness of fit and predictive ability

The criteria for model comparison included goodness of fit statistics and an evaluation of predictive ability. Goodness of fit for each group of SNPs was measured by computing the correlation between predicted genomic breeding values (GEBV) and corrected phenotype values (COR) and using the Likelihood Ratio statistic (LRT) in the entire data set. LRT is calculated as twice the difference in log-likelihoods between the full ($\sigma_g^2 \neq 0$) and reduced ($\sigma_g^2 = 0$) models and it indicates ‘significance’ of the marked variance.

The predictive ability of the GBLUP model using all markers was assessed by a 3-fold CV. To assess the predictive ability of each subset of SNPs, the data set was partitioned into three parts. Solutions for all SNPs effects in the training set (twofolds) were estimated and used to predict observations in the testing set (third fold). The criterion for comparison was the correlation between predicted GEBV and corrected phenotype in testing sets. For each bin of SNPs, we ran five replicates of a 3-fold CV. All analyses were conducted using restricted maximum likelihood methods as implemented in the Genome-wide Complex Trait Analysis (GCTA) tool (Yang et al. 2011).

Results and Discussion

Goodness of fit statistics

Figure 1 displays the correlation between GEBVs and corrected phenotypes and the LRT by class of SNPs binned based on MAF using the entire data set. The goodness of fit statistics were also calculated with every training set and averaged out. As results of these two analyses were similar, only those from the entire data set are presented. The correlation and LRT tended to decrease as MAF of SNPs increased. For the three traits, goodness of fit was better when all markers were used for constructing the $G$, without evidence of over-fitting. Correlations were larger for BM, followed by BW and HHP. These results were expected and reflect the magnitudes of the heritability of each trait (Hayes et al. 2010).

Figure 2 shows the goodness of fit statistics by subset of SNPs constructed according to the absolute values of inferred effect sizes. The LRT and the correlation increased when moving from quintile one (small effect size) to quintile 5 (large effect size). It
should be noticed that LRT and the correlation resulting from quintile five for all three traits were larger than when all markers were used. Perhaps including redundant markers into the evaluation model harms the ability of discerning training data. However, as mentioned before, an excellent goodness of fit in training set can be in conflict with predictive ability in testing set.

The correlation and LRT statistics by bin of model size (marker densities) are presented in Figure 3. For BW and BM, both correlation and LRT increased slightly as more SNPs were included in the model until 40 K, when a plateau was reached indicating no additional improvement in model fit by increasing the number of markers. This may be an indication of the larger extent of LD in the chicken genome (relative to

Figure 1 Correlation between predicted genomic values and corrected phenotypes (COR) and Likelihood Ratio statistics (LRT) obtained with the entire data set using SNPs binned based on minor allele frequency (MAF) and genomic best linear unbiased prediction (GBLUP) for body weight (BW), ultra-sound of breast muscle (BM) and hen house egg production (HHP). The number of SNPs included in each MAF bin was: 0.01–0.09, n = 70 780; 0.09–0.20, n = 71 027; 0.20–0.29, n = 71 254; 0.29–0.40, n = 71 101; 0.40–0.50, n = 70 767). LRT was calculated as twice the difference in log-likelihoods between the full ($\sigma^2 = 0$) and reduced ($\sigma^2 = 0$) models.
the human), so it would seem that a chip about 10–20 K would be adequate to capture training sample variability. For HHP, gains from fitting more markers were small but a plateau was reached at about 20 K. The required value for significance in the LR test statistic is more than 3.84 (with p < 0.05) so that differences in HHP obtained with >10 K SNPs are virtually uniform. Overall, the relationship between goodness of fit and model size was different for the three traits.

Figure 2 Correlation between predicted genomic values and corrected phenotypes (COR) and Likelihood Ratio statistic (LRT) obtained in the entire data set using SNPs partitioned into quintiles of absolute marker effect sizes (from smallest, Q1, to largest, Q5) and genomic best linear unbiased prediction (GBLUP) for body weight (BW), ultra-sound of breast muscle (BM) and hen house egg production (HHP). LRT was calculated as twice the difference in log-likelihoods between the full ($\sigma_g^2 \neq 0$) and reduced ($\sigma_g^2 = 0$) models.
Predictive ability

The predictive correlations obtained using all SNPs together were $0.27 \pm 0.002$ for BW, $0.33 \pm 0.001$ for BM and $0.20 \pm 0.002$ for HHP. Using common SNPs (MAF >0.01) and a Bayesian model, Lee et al. (2008) found correlations in the range of 0.4–0.9 for some quantitative traits in mice. Using ~2.5 million SNPs and a GBLUP model, Ober et al. (2012) obtained a predictive correlation of 0.24 for starvation resistance.

Figure 3 Correlation between predicted genomic values and corrected phenotypes (COR) and Likelihood Ratio statistic (LRT) obtained in the entire data set using different number of SNPs and genomic best linear unbiased prediction (GBLUP) for body weight (BW), ultra-sound of breast muscle (BM) and hen house egg production (HHP). LRT was calculated as twice the difference in log-likelihoods between the full ($\sigma^2_g \neq 0$) and reduced ($\sigma^2_g = 0$) models.
in *Drosophila melanogaster*. Higher correlations have been reported in livestock (e.g. Pryce & Daetwyler 2012). Factors affecting predictive ability are discussed in de los Campos et al. (2013). Often, predictive correlations are divided by the square root of the trait heritability, producing a relative value that conveys a false impression of predictive ability. In general, larger samples for training are expected to provide more accurate out-of-sample predictions perhaps due to a better inference of individual SNP effects. Also, predictive correlations are larger when response variable is based on an average, such as EBV.

Table 2 presents the predictive ability obtained using different MAF bins. Similar to goodness of fit statistics, the predictive ability delivered by SNPs with low MAF was consistently better than that from SNPs with high MAF, especially for HHP. The highest correlation for BW was obtained with MAF bin 0.01–0.09, but for BM and HHP, this was obtained with MAF bin 0.09–0.20. For HHP, the decline in predictive ability with increasing MAF was more clear than for BW and BM.

The recently published 1000 Genome Project paper (Abecasis et al. 2012) reported an excess of rare variants (MAF <0.5%) at functional sites. Under the assumption that most mutations are deleterious, there would be a relatively higher selection pressure on genetic variants at functional sites than at other sites. This may explain why better predictions are obtained with low MAF markers, but confirming that such markers in our study are located in functional parts requires further research. There was a strong correlation ($r \approx 0.88$) between estimated genomic heritability in the training set and the predictive correlation. Previous studies on genomic selection have also indicated that the accuracy of selection is greater for traits with higher heritability (e.g. Hayes et al. 2010).

We examined the impact of MAF on characteristics of $G$: as MAF increased from 0.01 to 0.09 to 0.4–0.5, the variance of diagonal elements of $G$ decreased from 0.179 to 0.003, but the average and variance of off diagonals were almost the same across MAF bins (Table 3). Simeone et al. (2011) also found that marker MAF was negatively associated with the variance of diagonal elements of $G$. These authors claimed that ‘aberrant’ diagonal elements may be attributed to an admixed population, unreliable SNP sample data or low call rate. However, an increase of rare (<0.10) or low-frequency SNPs (0.01–0.05) and including monomorphic SNPs into $G$ can also lead to extreme diagonal elements of $G$. We observed that the diagonal element for a specific bird was 13.55 when low-frequency variants were used, but when including all SNPs, the diagonal element for the same bird decreased to 4.47. Thus, if rare variants are included in the data to increase the number of markers, this may lead to ‘outlier’ diagonal elements in $G$. Chen et al. (2011) pointed out that EBVs of genotyped animals and genetic parameter estimates could be biased depending on allele frequency edits. However, in our study, removing this bird from the data set did not change predictive ability.

The averages of diagonal and off diagonal elements of $G$ when using all markers in our study were similar to those in Forni et al. (2011) and Chen et al. (2011), but our variances were larger. Accuracies of EBV with genomic BLUP are also a function of the difference between averages of diagonal and off-diagonal elements of $G$ (Forni et al. 2011). Hence, differences between the performance of various MAF bins in prediction may be related to properties of the respective $G$ matrices. Chen et al. (2011) found that the performance of different sets of markers with MAF

### Table 2: Predictive ability estimated by genomic best linear unbiased prediction (GBLUP) using single nucleotide polymorphisms (SNPs) binned based on minor allele frequency (MAF) for body weight (BW), ultra-sound of breast muscle (BM), and hen house egg production (HHP) traits in broiler chickens

<table>
<thead>
<tr>
<th>MAF bin</th>
<th>COR1 ± SE</th>
<th>$r^2_{inv,3}$</th>
<th>COR ± SE</th>
<th>$h^2_{inv}$</th>
<th>COR ± SE</th>
<th>$h^2_{inv}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01–0.09</td>
<td>0.28$^{a,2}$ ± 0.002</td>
<td>0.29</td>
<td>0.30$^{a}$ ± 0.002</td>
<td>0.28</td>
<td>0.22$^{a}$ ± 0.002</td>
<td>0.18</td>
</tr>
<tr>
<td>0.09–0.20</td>
<td>0.26$^{a}$ ± 0.002</td>
<td>0.26</td>
<td>0.33$^{a}$ ± 0.001</td>
<td>0.31</td>
<td>0.23$^{a}$ ± 0.001</td>
<td>0.19</td>
</tr>
<tr>
<td>0.20–0.29</td>
<td>0.23$^{a}$ ± 0.002</td>
<td>0.23</td>
<td>0.29$^{a}$ ± 0.001</td>
<td>0.28</td>
<td>0.20$^{a}$ ± 0.002</td>
<td>0.17</td>
</tr>
<tr>
<td>0.29–0.40</td>
<td>0.21$^{a,2}$ ± 0.003</td>
<td>0.20</td>
<td>0.29$^{a}$ ± 0.001</td>
<td>0.27</td>
<td>0.18$^{a}$ ± 0.003</td>
<td>0.13</td>
</tr>
<tr>
<td>0.40–0.50</td>
<td>0.24$^{a,2}$ ± 0.004</td>
<td>0.19</td>
<td>0.29$^{a}$ ± 0.001</td>
<td>0.24</td>
<td>0.16$^{a,2}$ ± 0.003</td>
<td>0.14</td>
</tr>
<tr>
<td>All markers</td>
<td>0.27$^{a,2}$ ± 0.002</td>
<td>0.30</td>
<td>0.33$^{a}$ ± 0.001</td>
<td>0.33</td>
<td>0.21$^{a}$ ± 0.002</td>
<td>0.19</td>
</tr>
</tbody>
</table>

1 Correlation between genomic predicted breeding values and corrected phenotypes in testing set.
2 Standard error.
3 Genomic heritability estimated in the data used to train the model.
4 Different superscript letters indicate significant differences (p < 0.05).
thresholds ranging from none to 0.4 was very similar, in contrast with our results. Chen et al. (2011) combined high- and low-frequency SNPs together, so effects of low-frequency markers on predictive ability may have been cancelled out by those of high-frequency markers. Also, in our study, the number of markers for MAF bins was about equal, whereas in Chen et al. (2011), when the threshold for removing low-frequency SNPs increased, the number of SNPs decreased. As the number of SNP decreases, $G$ may be more prone to creating ‘fake’ relationships. Edriss et al. (2013) used GBLUP and BayesA and edited markers for MAF with frequency thresholds of no limit, 0.001, 0.01, 0.02, 0.05 and 0.10, and prediction accuracies were nearly the same.

In the MAF bin 0.01–0.09, many markers were nearly monomorphic, while in quintiles of higher frequency (in particular 4 and 5), the two alleles contributed to variance. Because of this, we expected that MAF bins 4 and 5 would deliver a better predictive ability, but it was not the case here.

The predictive performance of bins of SNPs constructed using effect sizes is presented in Table 4. Correlations increased when SNPs of stronger effect sizes were used for constructing $G$, and the best predictive performance was attained when all SNPs were included. Similar to binning SNPs by MAF, there was a strong association between heritability in the training sets and predictive correlation in the testing sets for BW and BM ($r$ ~0.86). Quintiles 4 and 5 led to a much larger estimate of genomic heritability in the training set of BW and BM with estimates close to 0 for quintiles 1 and 2. It must be pointed out that the predictive ability delivered by all markers together was equal to the predictive ability of markers in quintiles 4 and 5 binned by effect sizes for BW and BM. Although it seems that using a low-density panel consisted of SNPs with largest effect size is cost effective in breeding programmes, having multiple low-density trait-specific panels with a multitrait breeding objective is not practical. Hence, using a low-density panel consisted of largest effect size SNPs that are common among the important traits along with imputation may be a solution for enhancing the predictive ability in multitrait breeding objective.

Although unlikely, if a causal variant is tagged by many SNPs, perhaps their genetic signal is captured more than once, which could lead to an overestimation of $h^2$. Speed et al. (2012) found that genomic heritability estimates where highly sensitive to uneven LD between SNPs, such that contributions to $h^2$ were overstated from causal variants in regions of high LD and underestimated in regions of low LD. The impact of this on predictive ability is unknown because LD relationships between markers and QTLs are unknown. When the number of markers is large, perhaps many SNPs do not tag any causal variant, so prediction might be improved by the removal of redundant SNPs from $G$. While redundant SNPs may harm predictive ability, dense markers around a

Table 3 Features of genomic relationship matrices with single nucleotide polymorphisms (SNPs) by class of minor allele frequency (MAF)

<table>
<thead>
<tr>
<th>MAF bin</th>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagonal elements</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01–0.09</td>
<td>0.999</td>
<td>0.411</td>
<td>13.554</td>
<td>0.179</td>
</tr>
<tr>
<td>0.09–0.20</td>
<td>0.993</td>
<td>0.798</td>
<td>3.555</td>
<td>0.012</td>
</tr>
<tr>
<td>0.20–0.29</td>
<td>0.990</td>
<td>0.843</td>
<td>2.121</td>
<td>0.004</td>
</tr>
<tr>
<td>0.29–0.40</td>
<td>0.984</td>
<td>0.842</td>
<td>1.664</td>
<td>0.003</td>
</tr>
<tr>
<td>0.40–0.50</td>
<td>0.984</td>
<td>0.816</td>
<td>1.447</td>
<td>0.003</td>
</tr>
<tr>
<td>All markers</td>
<td>0.990</td>
<td>0.791</td>
<td>4.465</td>
<td>0.015</td>
</tr>
<tr>
<td>Off-diagonal elements</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01–0.09</td>
<td>–0.001</td>
<td>–0.362</td>
<td>1.275</td>
<td>0.004</td>
</tr>
<tr>
<td>0.09–0.20</td>
<td>–0.001</td>
<td>–0.197</td>
<td>0.929</td>
<td>0.005</td>
</tr>
<tr>
<td>0.20–0.29</td>
<td>–0.001</td>
<td>–0.221</td>
<td>0.964</td>
<td>0.005</td>
</tr>
<tr>
<td>0.29–0.40</td>
<td>–0.001</td>
<td>–0.249</td>
<td>0.982</td>
<td>0.005</td>
</tr>
<tr>
<td>0.40–0.50</td>
<td>–0.001</td>
<td>–0.266</td>
<td>0.990</td>
<td>0.005</td>
</tr>
<tr>
<td>All markers</td>
<td>–0.001</td>
<td>–0.154</td>
<td>0.931</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Table 4 Predictive ability delivered by genomic best linear unbiased prediction (GBLUP) using single nucleotide polymorphisms (SNPs) partitioned into quintiles of absolute marker effect sizes for body weight (BW), ultra-sound of breast muscle (BM) and hen house egg production (HHP) 1

<table>
<thead>
<tr>
<th>Marker effects</th>
<th>BW</th>
<th>BM</th>
<th>HHP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>COR ± SE</td>
<td>$h^2_{	ext{num}}$</td>
<td>COR ± SE</td>
</tr>
<tr>
<td>Q1</td>
<td>0.09$^a$ ± 0.003</td>
<td>0.01</td>
<td>0.16$^a$ ± 0.002</td>
</tr>
<tr>
<td>Q2</td>
<td>0.16$^b$ ± 0.002</td>
<td>0.01</td>
<td>0.18$^a$ ± 0.002</td>
</tr>
<tr>
<td>Q3</td>
<td>0.20$^b$ ± 0.002</td>
<td>0.03</td>
<td>0.25$^b$ ± 0.001</td>
</tr>
<tr>
<td>Q4</td>
<td>0.23$^b$ ± 0.001</td>
<td>0.34</td>
<td>0.30$^b$ ± 0.001</td>
</tr>
<tr>
<td>Q5</td>
<td>0.29$^b$ ± 0.001</td>
<td>0.58</td>
<td>0.32$^b$ ± 0.001</td>
</tr>
<tr>
<td>All markers</td>
<td>0.27$^b$ ± 0.002</td>
<td>0.30</td>
<td>0.33$^b$ ± 0.001</td>
</tr>
</tbody>
</table>

$^1$Symbols are as in Table 2.
causal variant may lead to overestimation of the effect of causal variants. It is important to keep in mind that inference about markers is heavily influenced by priors in high-dimensional models (Gianola 2013). Weigel et al. (2009) and Vazquez et al. (2010) evaluated strategies for selecting subsets of SNP based on the absolute values of their estimated effects, or using evenly spaced SNP across the genome in dairy cattle. These researchers reported that a low-density panel of markers (≈2000 markers) with large effects outperformed predictions based on evenly spaced SNP, but the latter preserve more reliability provided that imputation is used. In Vazquez et al. (2010), high-density panels delivered the best predictive ability, setting an upper bound for the predictive ability of low-density panel sets. These two studies were carried out with a 50 K SNP chip and only low-density marker sets consisting of 30–2000 SNPs were investigated. In our study, we partitioned SNPs based on absolute values of their effect sizes using five quintiles and the predictive ability delivered by each subset of SNP was compared with that from all markers. In the above-mentioned studies, the method was Bayesian LASSO, but here it was GBLUP. Furthermore, these authors combined genomic evaluation with traditional parent averages, but here only markers were used as predictors.

Designing a chip with low density can be important for economic reasons. The performance of SNP sets of various densities is presented in Table 5. Contrary to results presented in Figure 3 with goodness of fit statistics, the predictive correlation was almost constant for all SNP groups, with a small increase from 5 K to 10 K for BM and HHP. Moser et al. (2010) looked at the accuracy of GEBV for several dairy traits by using subsets of SNPs selected by different methods. A subset of 5000 SNP allowed attaining >90% of the accuracy obtained with a complete set of SNPs. Our results demonstrate that in contrary to human studies in which increasing the density of markers until 400 K improved the genome-enabled prediction (Makowsky et al. 2011); in chicken, increase in the marker density could not improve prediction accuracy. The high extent of LD due to low effective population size and family structure in training sets are plausible explanations for current findings.

### Table 5 Predictive ability delivered by genomic best linear unbiased prediction (GBLUP) models using a varying number of single nucleotide polymorphisms (SNPs) for body weight (BW), ultra-sound of breast muscle (BM) and hen house egg production (HHP)

<table>
<thead>
<tr>
<th>Marker density</th>
<th>BW</th>
<th>BM</th>
<th>HHP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>COR ± SE</td>
<td>h²&lt;sub&gt;TRN&lt;/sub&gt;</td>
<td>COR ± SE</td>
</tr>
<tr>
<td>5 K</td>
<td>0.28 ± 0.003</td>
<td>0.27</td>
<td>0.31 ± 0.002</td>
</tr>
<tr>
<td>10 K</td>
<td>0.28 ± 0.001</td>
<td>0.29</td>
<td>0.32 ± 0.001</td>
</tr>
<tr>
<td>20 K</td>
<td>0.28 ± 0.001</td>
<td>0.30</td>
<td>0.33 ± 0.001</td>
</tr>
<tr>
<td>40 K</td>
<td>0.28 ± 0.001</td>
<td>0.30</td>
<td>0.33 ± 0.001</td>
</tr>
<tr>
<td>100 K</td>
<td>0.28 ± 0.001</td>
<td>0.31</td>
<td>0.33 ± 0.001</td>
</tr>
<tr>
<td>200 K</td>
<td>0.27 ± 0.002</td>
<td>0.30</td>
<td>0.33 ± 0.001</td>
</tr>
<tr>
<td>All markers</td>
<td>0.27 ± 0.002</td>
<td>0.30</td>
<td>0.33 ± 0.001</td>
</tr>
</tbody>
</table>

1Symbols are as in Table 2.

### Conclusion

We found that SNPs with low MAF delivered better predictive ability than SNPs with high MAF and those markers produced models with better goodness of fit statistics. Estimated genomic heritability in a training set seemed to be an indicator of predictive ability. Predictive power brought by SNPs in a quintile with the largest effect sizes was largest. When binning based on number of SNPs, increasing the number of markers did not improve predictive ability. Therefore, using a subset of markers containing 10 K SNPs with the largest effect sizes SNPs could be used to predict breeding values of chicken in this population. In conclusion, a low-density chip with low-frequency SNPs and large effect sizes should be considered from the perspective of enhancing predictive ability. Such low-density chips may be cost-effective in chicken breeding programs.

### Conflict of Interest

The authors do not have any conflict of interest.

### Acknowledgement

The first author wishes to acknowledge Aviagen (Newbridge, Scotland) for providing the data, and the Ministry of Science, Research and Technology of Iran for financially supporting his visit to the University of Wisconsin-Madison. This study was partially...
supported by the Wisconsin Agriculture Experiment Station.

References


