

Communication

A Comprehensive Strategy Combining Feature Selection and Local Optimization Algorithm to Optimize the Design of Low-Density Chip for Genomic Selection

Ruihan Mao ¹, Lei Zhou ¹ , Zhaojun Wang ², Jianliang Wu ² and Jianfeng Liu ^{1,*}¹ College of Animal Science and Technology, China Agricultural University, Beijing 100193, China² Beijing Zhongyu Pig Breeding Co., Ltd., Beijing 100194, China

* Correspondence: liujf@cau.edu.cn

Abstract: Design of low-density SNP chips provides an opportunity for wide application of genomic selection at lower cost. A novel strategy referred to as the “block-free” method is proposed in this study to select a subset of SNPs from a high-density chip to form a low-density panel. In this method, Feature Selection using a Feature Similarity (FSFS) algorithm was first performed to remove highly correlated SNPs, and then a Multiple-Objective, Local-Optimization (MOLO) algorithm was used to pick SNPs for the low-density panel. Two other commonly used methods called the “uniform” method and the “block-based” method were also implemented for comparison purposes. A real pig dataset with 7967 individuals from three breeds containing 43,832 SNPs was used for comparison of the methods. In terms of genotype imputation accuracy and genomic prediction accuracy, our strategy was superior in most cases when the densities were lower than 1K. The genotype imputation accuracy from the low-density chip compared to the original high-density chip was higher than 90% in all pig breeds as the density increased to 1K. In addition, the accuracies of predicted genomic breeding values (GEBV) calculated using the imputed panel were nearly 90% of estimates from the original chip for all traits and breeds. Our strategy is effective to design low-density chips by making full use of information of close relationships for genomic selection in animals and plants.

Keywords: genomic selection; low-density SNP chip; feature selection; local optimization; imputation



Citation: Mao, R.; Zhou, L.; Wang, Z.; Wu, J.; Liu, J. A Comprehensive Strategy Combining Feature Selection and Local Optimization Algorithm to Optimize the Design of Low-Density Chip for Genomic Selection.

Agriculture **2023**, *13*, 614.<https://doi.org/10.3390/agriculture13030614>

Academic Editors: Manuel García-Herreros and Pedro Manuel Aponte

Received: 7 February 2023

Revised: 28 February 2023

Accepted: 1 March 2023

Published: 3 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Genomic selection (GS) was first proposed by Meuwissen et al. in 2001 [1]. It is also referred to as genomic prediction following its wide implementation in animal and plant breeding. Given the advantages of reducing breeding costs and shortening the generation interval, GS has become a routine genetic evaluation strategy for many livestock species [2]. Currently, it plays an important role in the livestock breeding industry, especially for animals with high economic value, such as dairy cattle, pigs, and chickens [3], and for traits with low heritability or that are hard to measure, such as reproductive traits and meat quality traits. Although the costs of SNP (single nucleotide polymorphism) chips genotyping have significantly decreased in recent years, it is still too expensive to genotype all candidate individuals for most highly prolific species, such as pigs and chickens. An appealingly suitable solution is to apply lower-density SNP chips, and then impute them to medium- or high-density genotypes for practical GS application.

Many studies have investigated the effectiveness of low-density chips for different livestock species, including studies on dairy and beef cattle [4,5], pigs [6–8], sheep [9–11], and broiler and laying chickens [12,13], and all of them have proved that low-density chips were beneficial in GS with a proper genotype imputation processing procedure. Commercial low-density SNP chips, such as Illumina Bovine3K BeadChip and BovineLD BeadChip (6K), were designed to support imputation to higher-density genotypes in dairy

and beef cattle and were verified to perform well in terms of both imputation accuracies and genomic prediction accuracies [14,15].

Therefore, developing a low-density SNP panel and then imputing to higher density panels is a cost-effective solution for practical application of GS in many livestock species. Many strategies have been proposed for designing low-density SNP panels. For example, a simple strategy was to maximize the information entropy of the low-density SNP panels, in which evenly spaced SNPs were selected with the highest minor allele frequencies (MAF) [14]. However, this approach could not ensure minimization of the linkages between the selected SNPs. Another proposed strategy was to choose a set of SNPs that captures the largest proportion of genetic variance based on the results of previous studies [16], but this was not generalized due to trait specificity. In general, as far as we know, there has been no widely acceptable or standard strategy to design a low-density panel from a previous high-density panel.

The objective of this study was to propose a new strategy for designing low-density panels based on former high-density panels. To demonstrate the effectiveness of the proposed method, a real pig breeding dataset was used for method comparisons, and we compared our new method with other two state-of-the-art approaches. Genotype imputation accuracies and prediction of genomic estimated breeding values (GEBV) accuracies were compared for different low-density panel design methods.

2. Materials and Methods

2.1. Low-Density Chip Selection Methods

“Uniform” Method: This was considered the baseline method. SNPs in the low-density panel were selected uniformly across the map of high-density panel. This was achieved through the K-means algorithm because SNPs were not absolutely evenly distributed across the genome. First, each chromosome was divided into a fixed number of blocks according to its length and total number of designed SNPs on the chromosome. Then, the SNP marker in the center of each block was selected as the representative marker on the low-density panel.

“Block-Based” Method: This is the most commonly used method in recent studies. In information theory, Shannon Entropy (H) is defined as the average level of “information”, “surprise”, or “uncertainty” inherent in the variable’s possible outcomes [17]. Shannon Entropy of each locus was calculated with the following formula:

$$H = \sum_{i=1}^2 p_i \left[\log_2 \frac{1}{p_i} \right] = -[p_i \log_2 p_i + (1 - p_i) \log_2 (1 - p_i)]$$

where p_i is the MAF of the i th locus (only available for bi-allelic loci). The maximum Shannon Entropy value for a locus is one when MAF of this locus equals 0.5. Each chromosome was divided into multiple blocks as described above in the “uniform” method. The averaged MAF of each SNP considering all breeds was calculated. In each block, the SNP with the highest averaged MAF was chosen as an SNP for the low-density panel. This method could achieve the highest Shannon Entropy for the low-density panel.

“Block-free” Method: This is the key method proposed in this study. The “block-free” method is based on two approaches: the first one is called the “Feature Selection using Feature Similarity” (FSFS) algorithm [18–20], in which a machine learning strategy is utilized to solve the “tagging SNP problem” [21]. This algorithm selects features by first grouping them into homogeneous subsets and then choosing a representative feature from each subset. The second approach is the “Multiple-Objective, Local Optimization” (MOLO) [5,22] algorithm, in which both Shannon Entropy and uniformity are considered simultaneously, and a local optimal solution is obtained by a formula.

The method included the following three steps. Step 1: The FSFS algorithm was first performed to select SNPs. The extent of linkage between SNPs was expressed as $1 - r^2$, in which r^2 was the measure of the linkage disequilibrium (LD) between SNPs. If SNPs with strong LD were in the same cluster, then the SNP in the middle of each cluster was

chosen to represent the cluster. The threshold of r^2 can be adjusted to control the size of clusters. If there is more than one breed in the data, the breed with the largest number of genotyped animals would be chosen as the representative breed in this step. Step 2: The SNPs with MAF lower than a given threshold in any of the breeds were excluded. Step 3: The MOLO algorithm was used to finally select SNPs for the low-density panel in this step. The selection index of each SNP was computed as follows:

$$f = w_1 E^{t_1} + w_2 U^{t_2}$$

where E score was defined as the averaged Shannon Entropy of each SNP and U score was calculated to reflect the uniformity of the SNPs. Calculation details are available in the references [5,22]. w_1 and w_2 were the corresponding weights for E and U , respectively, under the restriction of $0 \leq w_1 \leq 1$, $0 \leq w_2 \leq 1$, and $w_1 + w_2 = 1$; t_1 and t_2 were shrinkage parameters. These parameters can be set up based on the data structure used in the research.

Each chromosome was divided into blocks. The SNPs were ranked by their respective scores within each block, and the SNP with the highest score was selected to represent the block. The MOLO algorithm in step 3 was realized through R package selectSNPs [22].

2.2. Data for Method Comparison

A dataset with three pure-breed breeds from a pig breeding farm in Inner Mongolia of China were used to validate the low-density SNP panel selection methods. There were 1518 Duroc (DD), 1702 Landrace (LL) and 4789 Yorkshire (YY) pigs with SNP genotypes in the dataset, and all pigs were born between 2014 and 2020. Two growth traits adjusted for age at 100 kg weights (AGE) and adjusted back-fat thickness at 100 kg weights (BF) were analyzed for genomic prediction. Details of the phenotype data collection and processing procedures are explained in Appendix A. All pigs were genotyped using a KPSISUS50-V1 chip consisting of 43,832 SNPs across the genome (detailed information presented in Appendix B).

Principal component analysis (PCA) was performed based on the raw genotype data to identify clusters using GCTA [23] as Supplementary Figure S1 showed. All individuals were clustered using the K-means method ($k = 3$). In total, 7967 individuals were grouped to the correct cluster, while 42 pigs were excluded as they obviously deviated from their breed clusters. Thus, total numbers of 1507, 1691, and 4769 individuals in DD, LL, and YY breeds were left, respectively, in the subsequent analysis.

Genotyping quality control (QC) for SNPs includes the following criteria: (1) genotype call rate > 95%, (2) MAF > 0.01, and (3) were not located on sex chromosomes. Filtration of SNPs was performed using the Plink v1.90 software [24], and it was performed for each breed separately. After QC, the number of remaining SNPs was 28,411, 34,111 and 34,768 for DD, LL and YY, respectively. There were 15,412 SNPs removed after QC in DD population, which leads to a substantial reduction of SNPs in the original panel. Finally, a total number of 24,161 SNPs were retained as common SNPs for these three pig breeds, and they were used as the original high-density panel to design the low-density panels (Figure 1a).

2.3. Method Comparison Strategies

We used the above pig breeding dataset to compare our newly proposed SNP panel selection method (“block-free method”) with the counterpart methods (“uniform method” and “block-based method”). For each SNP selection method, seven levels of low-density panels were generated using the three methods mentioned previously. To fully investigate the effectiveness of low-density panels, the total number of SNPs for the seven levels of densities were set as 300, 500, 800, 1000 (1K), 1500 (1.5K), 2000 (2K), and 3000 (3K), respectively. Each of the three methods was implemented to create these seven levels of low-density panels. For the “block-free” method, degree of linkage disequilibrium was calculated by Plink v1.90 [24]. In step one, the threshold of $1 - r^2$ was set to be 0.8. In step two, the MAF threshold was chosen as 0.1. In step three, E score was calculated by

the average MAF of all three breeds, and the parameters for the selection index of the MOLO algorithm were set as: $w_1 = 0.1$, $w_2 = 0.9$, $t_1 = 1$, $t_2 = 1$. Panels with the same density contained the same amount of SNPs on each chromosome for these three breeds. After the low-density panels were selected by each method, the usefulness of different low-density panels was compared both in terms of imputation and genomic selection.

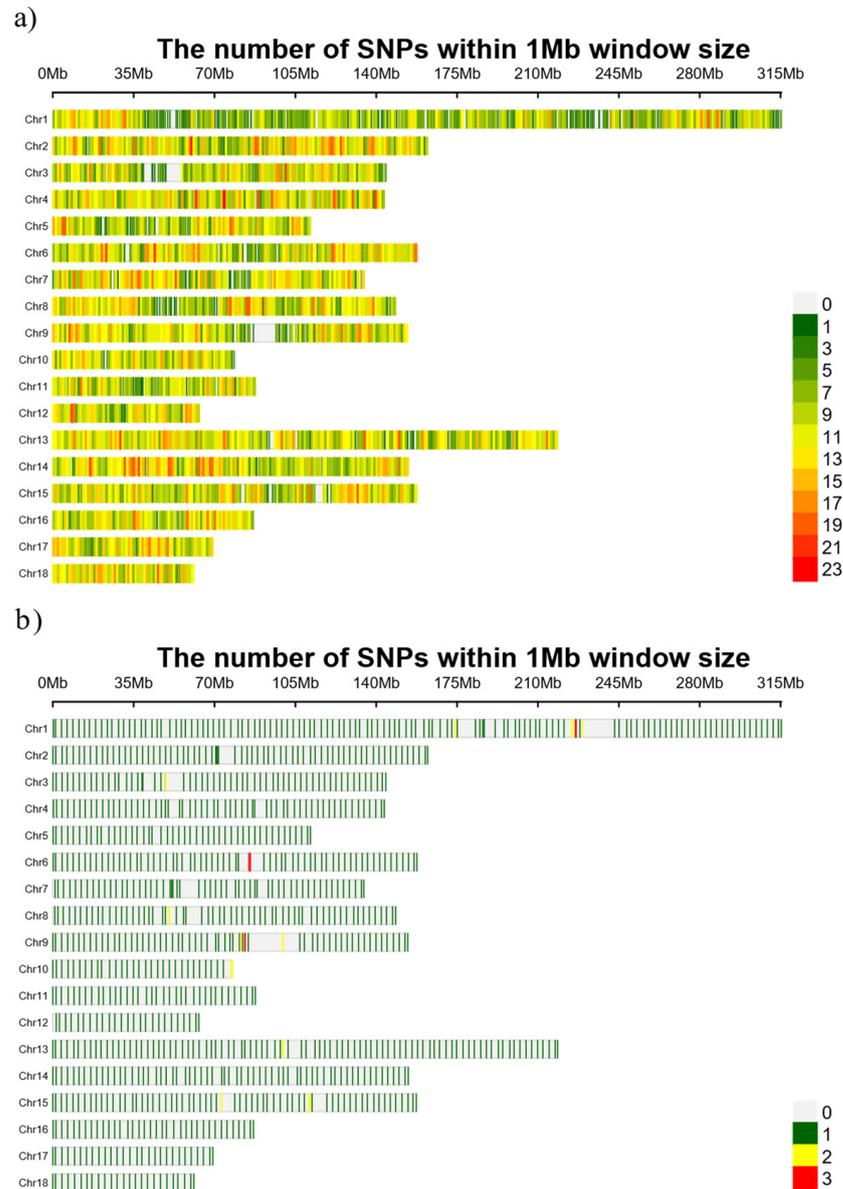


Figure 1. Distribution of SNPs on the genome. (a) Distribution of SNPs in the raw chip, (b) Distribution of SNPs in the 1K chip designed with “block-free” method. One vertical line corresponds to the number of SNPs in 1 Mb length and the numbers represented by different colors are as shown to the right of the figure.

The reference and validation populations in the process of genotype imputation and genomic selection were defined based on the year of birth, while only individuals with genome information were used in genomic selection. Individuals in the reference population were born up to 2019, while others born in 2020 were included in the validation populations. The details of the reference population sizes are showed in Table 1.

Table 1. Number of animals in the training and validation populations.

		Pig Breeds		
Populations		Duroc	Landrace	Yorkshire
Genotype imputations	Training population	1117	1194	4024
	Validation population	390	497	745
	Total	1507	1691	4769
Genomic predictions	Training population	1050	1132	3621
	Validation population	174	256	322
	Total	1224	1388	3943

The accuracy of imputation was assessed as the mean of the percentage of correctly imputed genotypes for all imputed SNPs. Correlation coefficients of the genotype matrices of the imputed and the real genotypes (coded 0, 1, 2) were also calculated as another measurement of genotype imputation accuracy. Details of the structure of validation population in imputation are shown in Supplementary Table S1.

A bivariate genomic best linear unbiased model (GBLUP) for AGE and BF was used to calculate genomic estimated breeding values (GEBVs) by DMU software [25], and the model was as follows:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1\mu_1 \\ 1\mu_2 \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \quad (1)$$

where y_1 and y_2 were the vector of observations for AGE and BF trait; μ_1 and μ_2 were the trait means for AGE and BF; g_1 and g_2 were the additive genomic values; and e_1 and e_2 were model residuals.

The genomic prediction accuracy was accessed by the correlation of GEBVs and corrected phenotypes (y_c). A bivariate BLUP model was used for the full dataset, and y_c were calculated as the sum of estimated breeding values (EBVs) and estimated residuals [26,27].

3. Results

In this study, we examined three strategies for the design of low-density chips: the “uniform”, “block-based”, and “block-free” methods, wherein the “block-free” method was a novel strategy we proposed. Under each strategy, we designed seven panels consisting of 300, 500, 800, 1K, 1.5K, 2K, and 3K SNP markers. To verify the advantage of our novel strategy, a pig breeding dataset of 1507 Duroc, 1691 Landrace, and 4769 Yorkshire pigs with high-density SNP chip was used to evaluate different strategies. The pigs of each breed were divided into the reference and validation sets. Then, we masked all markers except those that belonged to the low-density panels in the validation population and imputed them into the original density. Accuracies of genotype imputation and genomic prediction were then calculated.

3.1. Accuracy of Genotype Imputation

The Shannon Entropy, accuracy of imputation, and correlation coefficients of the genotype matrices were strongly determined by the number of SNPs in the low-density panels. As is shown in Figure 2, the “block-based method” always achieved the highest Shannon Entropy, as expected, while the tendency of imputation accuracy and correlation coefficients of the matrices were not exactly consistent with the rank of Shannon Entropy. As the density increased up to 1K for the low-density panels, the genotype imputation accuracies of “block-free method” were best (Figure 3a,b) in DD and YY breeds. Meanwhile, as the density increased to above 1K, the “block-based method” outperformed the other two methods in all breeds; that is, Shannon Entropy began to play a critical role. As the densities of the panel were increased from 300 to 1K, the imputation accuracies were improved rapidly, and a density of at least 1K was required for reaching genotype imputation accuracy around 90% for each breed.

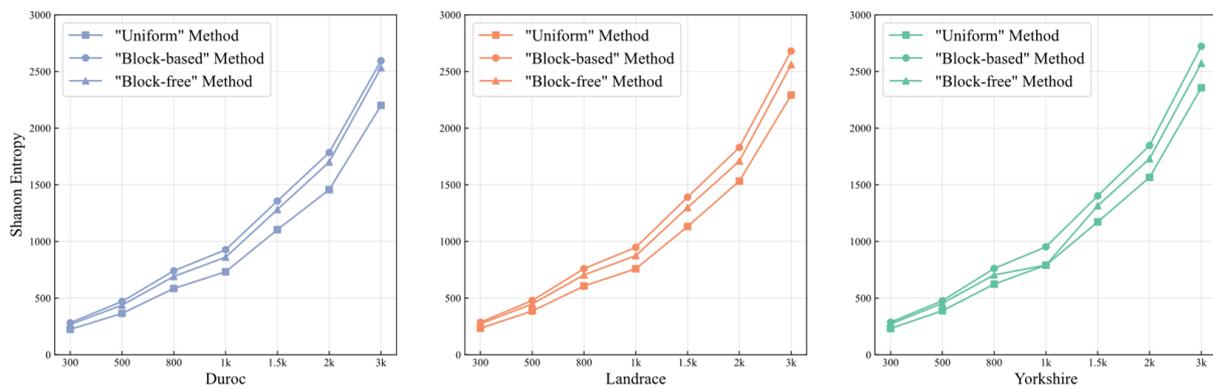


Figure 2. Shannon Entropy of different low-density panels. The horizontal coordinate represented the density of low-density panels, and the vertical coordinate indicated the Shannon Entropy. The shapes of square, circle, and triangle represented the “uniform” method, the “block-based” method, and the “block-free” method, respectively. The colors blue, orange, and green represented Duroc, Landrace, and Yorkshire, respectively.

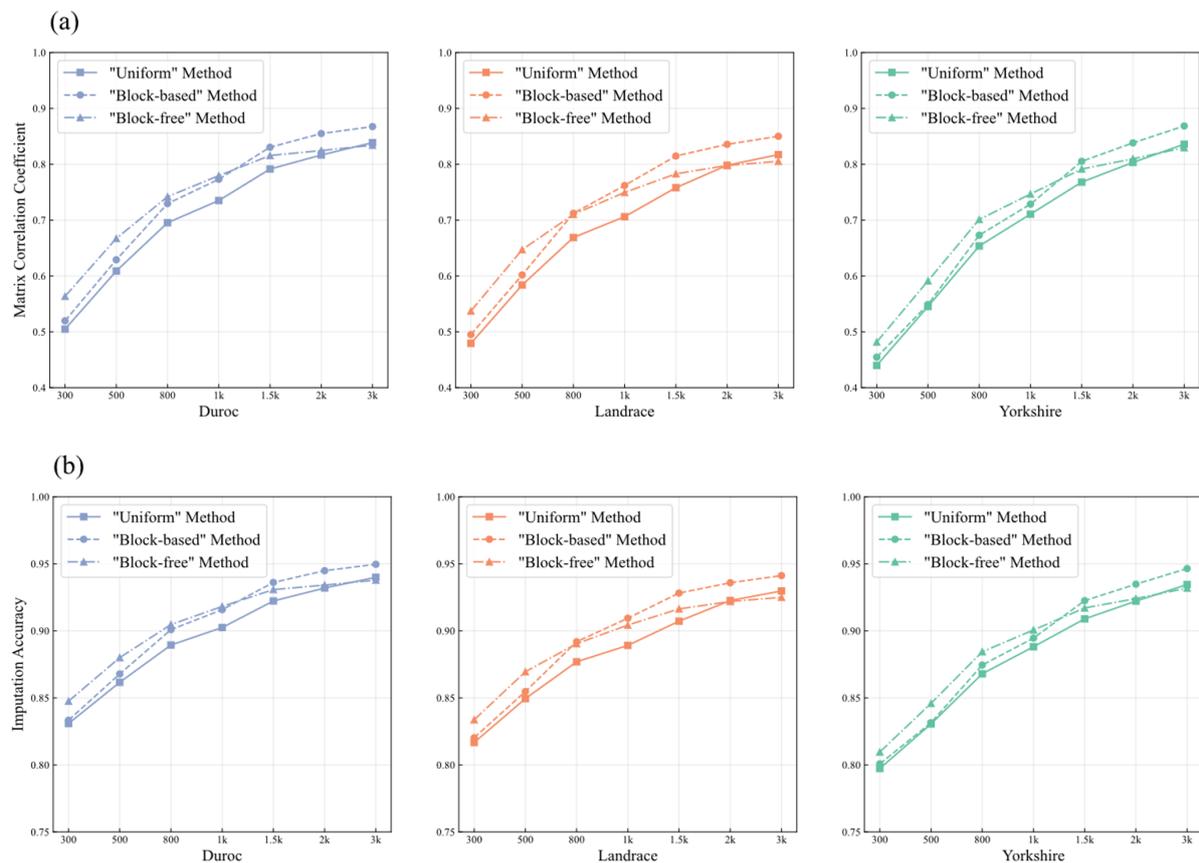


Figure 3. Genotype imputation performance of different low-density panels. The horizontal coordinate represented the density of low-density panels, and the vertical coordinate indicated (a) genotype imputation accuracies, (b) correlation coefficients of the matrices of the imputed and raw genotypes. The shapes of square, circle, and triangle represented the “uniform” method, the “block-based” method, and the “block-free” method, respectively. The colors blue, orange, and green represented Duroc, Landrace, and Yorkshire, respectively.

3.2. Accuracy of Genomic Estimated Breeding Values Prediction

The accuracies of GEBV for DD, LL, and YY when using the raw chip were 0.1314, 0.1165, and 0.1896, respectively, for AGE; and 0.2188, 0.1900, and 0.2408 for BF. The heri-

tabilities of AGE were 0.1502, 0.2323, and 0.2543, and BF were 0.3530, 0.4131, and 0.4139 for DD, LL, and YY, respectively as Supplementary Table S2 showed. The accuracies of predictions were similar across breeds, and accuracies for BF changed more obviously than AGE, which may be due to the higher heritability.

For both AGE and BF, in all breeds, the “block-free” method was superior to the other two strategies when the densities were below 1K, especially at a very low density. As the density increased, the accuracies of the three methods became closer and closer. However, this tendency was not obvious in the DD breed.

The accuracies of GEBVs for AGE and BF calculated from imputed SNPs were close to 90% in efficiency compared with the raw chip when the density reached 1K. This was consistent with the results of the imputation accuracy. The GEBV accuracies were quite close (>95% efficiency) for different low-density panel design strategies when the density increased to 3K. The accuracies of GEBVs for BF were slightly higher than AGE on the whole, due to its higher heritability (Figure 4a,b).

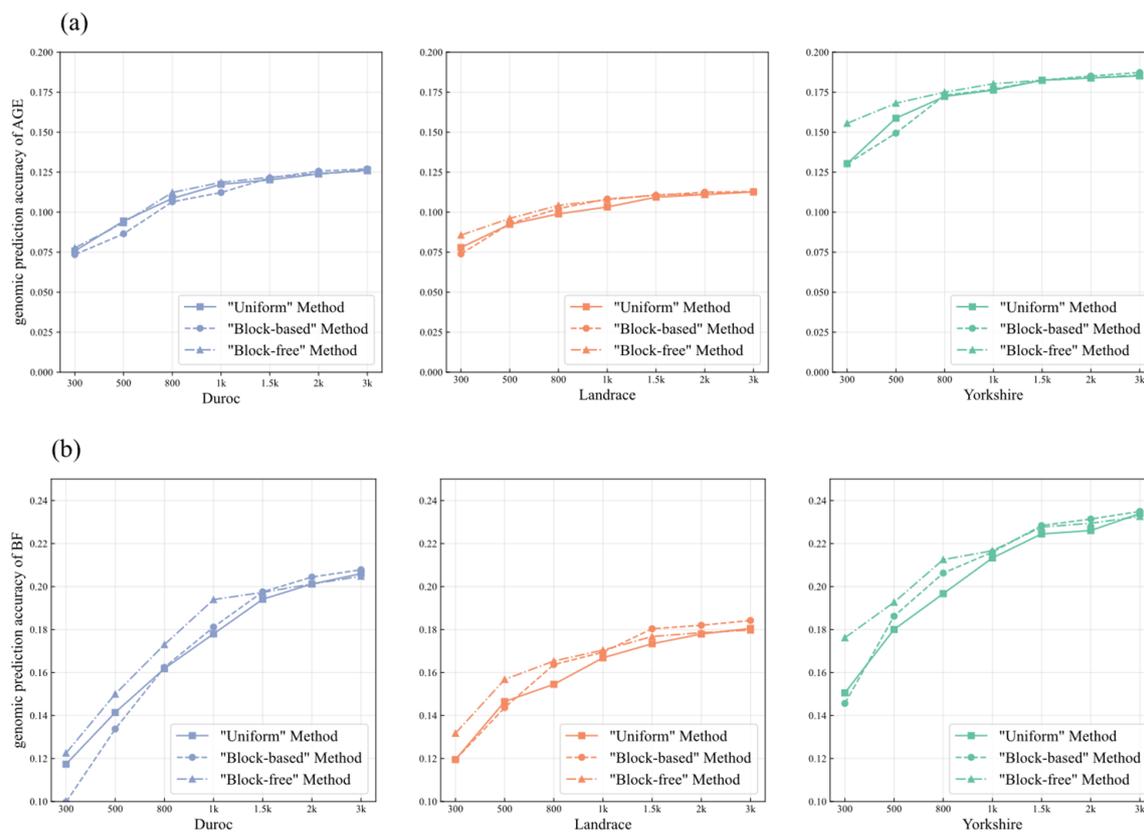


Figure 4. GEBV prediction accuracies of different low-density panels. The horizontal coordinate represents the density of low-density panels, and the vertical coordinate indicates (a) accuracies of GEBVs prediction of AGE calculated by imputed panels, and (b) accuracies of GEBVs prediction of BF calculated by imputed panels, respectively. The shapes of square, circle, and triangle represent the “uniform” method, the “block-based” method, and the “block-free” method, respectively. The colors blue, orange, and green represent Duroc, Landrace, and Yorkshire, respectively.

As the result of genotype imputation showed, a density of 1K is essential in meaningful practical application. The performance of the panel with 1K SNPs selected with our novel “block-free” method (Figure 1b) was superior, taking the balance of information loss and cost-saving into consideration. Accuracies of GEBV calculated by this imputed 1K chip and the raw chip for each breed are very close (Table 2).

Table 2. Prediction accuracies of GEBVs calculated by imputed 1K panel and raw chip.

	Populations	Breeds		
		Duroc	Landrace	Yorkshire
AGE	1K Panel	0.1186	0.1077	0.1803
	Raw Chip	0.1314	0.1165	0.1896
BF	1K Panel	0.1939	0.1704	0.2167
	Raw Chip	0.2188	0.1900	0.2408

4. Discussion

In this study, we proposed a new method for selection of SNPs in low-density panels. Using our new method and other competing methods, a series of different low-densities SNP chip panels were designed, and their application in genomic prediction was evaluated with a real pig breeding dataset. The results can be summarized that when marker density is lower than 1K, the performance of our “block-free” method was superior. With increase in marker density, the performance of the three methods tended to be similar, and the advantages of our method became less obvious. It is generally agreed that the low-density panel can be put into meaningful practical use only when the genotype imputation accuracy is greater than 90%. According to our results, to meet this condition, the density of the chip should be greater than 1K. The results of GEBV prediction also confirmed this view.

In general, 1K is an economical viable density, given that the accuracies of imputation and GEBV prediction all consistently reached 90% of that obtained from the raw chip. The imputation and GEBV accuracies at higher-density chips increased at a rather slow rate, which may offset cost considerations. These results were in line with previous studies on low-density chips for pig genomic selection [6,7]. Different data may produce different results; therefore, the low-density SNP selection strategy may depend on the data structure.

In a study of a low-density chip in a pig sire line [6], 1152 SNPs were necessary to obtain imputation accuracy of 90% when their sires were genotyped. A similar study used the same three pig breeds examined in this study with 2609 individuals for all breeds. [8]. Their study was based on a block method [4] and showed that at least 600 SNPs were needed for LL and YY breeds to obtain >90% imputation accuracy when both parents had dense genotypes, while more SNPs were required in the DD breed to reach an equivalent result. Their study showed that acceptable imputation quality could be achieved only when both parents were genotyped using a high-density chip. In some cases, the DD breed was suggested to be considered separately.

These results indicated that all animals can be genotyped using a low-density chip at an early age at low cost. Then, imputation and GEBV calculation can be performed to obtain GEBVs that reflect the genetic potential of each individual for economically important traits. The individuals with low GEBV can be eliminated to enable early effective selection.

Our results showed that imputation accuracies were highly related to the Shannon Entropy, and it was also necessary to consider LD between SNPs to select subsets through feature selection. This is so because in the imputation from low-density to medium- or high-density chips, the necessary information required is mainly from two sources: first, the relationship between reference population and validation population, and second, the linkage disequilibrium (LD) between current SNPs and those to be imputed.

The high imputation accuracy from this study resulted in little loss of information in GEBVs prediction. However, the reference population and the validation population in this study were closely related, but the characteristics (MAF and LD) used in the design may be population specificity. Therefore, the performance of the low-density chip designed for other pig populations using our method or under other relationship strategies needs to be verified further.

Most strategies for low-density chip design take into account the physical distance between SNPs and the information entropy carried by a single SNP, while the correlation and linkage of information between SNPs is ignored, which results in information redun-

dancy or loss. Our results showed that it is meaningful to take the LD between SNPs into account at the genomic or chromosome level. To further optimize the SNPs simplification strategy, more advanced machine learning methods should be attempted to achieve more efficiency rather than using only feature selection method. There is no doubt that it is feasible to design alternative methods to overcome the challenge of reducing genotyping costs by reducing chip density using smart strategies. A thorough study with complex genetic architecture and different traits would be needed to shed more light on this issue.

5. Conclusions

In conclusion, our novel strategy for creating a low-density SNP panel showed superiority in real data validation. It accounted for the correlation between SNPs, the uniformity of SNPs, and the Shannon Entropy of the panel, thereby combining the advantages of other existing methods. Thus, this strategy has great potential for designing low-density chips, especially when marker density is about 1K.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agriculture13030614/s1>, Table S1: Structure of validation population in Imputation, Table S2: Heritability of each trait for each breed, Figure S1: Principal coordinate analysis (PCA) of raw genotypes of pigs from three breeds.

Author Contributions: Conceptualization, J.L.; methodology, R.M.; validation, R.M.; writing—original draft preparation, R.M.; resources, Z.W. and J.W.; writing—review and editing, L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Earmarked Fund for National Natural Science Foundation of China (No.31972563), Beijing Municipal Commission of Science and Technology (NO. Z211100004621005), National key research and development program (NO. 2021YFD1300800), China Agriculture Research System (No. CARS-pig-35), and Science and Technology Program of Shenzhen (JSGG20210802153807024).

Institutional Review Board Statement: This study was conducted in strict accordance with the protocol approved by the Institutional Animal Care and Use Committee (IACUC) at the China Agricultural University (Beijing, People's Republic of China; permit no. DK996).

Data Availability Statement: The code of FSFS we implemented in this article has been uploaded to GitHub (<https://github.com/maoruihan/MaoChip> accessed on 6 February 2023). The pig grow data, pedigree, and genotype data used in this study are available at figshare: <https://doi.org/10.6084/m9.figshare.21971234>; <https://doi.org/10.6084/m9.figshare.21971219>; <https://doi.org/10.6084/m9.figshare.21971204> (accessed on 6 February 2023).

Acknowledgments: We are grateful to Mrode Raphael at the International Livestock Research Institute (ILRI) for providing the polish work, and the Best Genetics breeding farm for providing genetic and phenotypic data.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Three purebred pig breeds, Duroc (DD), Landrace (LL), and Yorkshire (YY), were used in this study. Phenotypes were recorded from 2017 to 2019 to keep the environmental conditions consistent as much as possible. All individuals in this study remained in good health during the testing period. Backfat thickness was measured between the 10th and 11th ribs of the pigs using real-time B-mode ultrasound at the end of testing.

Age at 100 kg (*AGE*) was adjusted using the following formula:

$$AGE = AGE_{test} + (100 - BW) \left(\frac{AGE_{test} - A}{BW} \right)$$

where AGE_{test} represents the age at the end of testing; BW is the body weight at the end of testing; and A is the constant correction coefficient for boars and dams, with A shown as follows:

Table A1. Constant correction coefficients for AGE of pig breeds.

Breed	Male	Female
Duroc	55.289	49.361
Landrace	48.441	51.014
Yorkshire	50.775	46.415

Backfat thickness at 100 kg (BF) was adjusted using the following formula:

$$BF = BF_{test} + (100 - BW) \left(\frac{BF_{test}}{BW - B} \right)$$

where BF_{test} represents backfat thickness at the end of testing; BW is the body weight at the end of testing; and B is the constant correction coefficient for boars and dams, with B shown as follows:

Table A2. Constant correction coefficients for BF of pig breeds.

Breed	Male	Female
Duroc	−6.240	−4.481
Landrace	−5.623	−3.315
Yorkshire	−7.277	−9.440

Appendix B

The ear tissues of 1314 samples were collected, preserved with 75% alcohol, and stored in -20°C freezers. Genomic DNA was extracted from the collected frozen ear tissue samples using a Qiagen DNeasy Tissue kit (Qiagen, Germany), and they were then analyzed using spectrophotometry and agarose gel electrophoresis to ensure that they were of high quality. All DNA samples were suitable for genotyping with a ratio of light absorption (A_{260}/A_{280}) between 1.8 and 2.0, a concentration $>50\text{ ng}/\mu\text{L}$, and a total volume $< 50\ \mu\text{L}$.

The SNP chip array named KPSISUS50-V1 used in this study was designed using the Illumina platform based on pig genome version 10.2. SNP markers were chosen from multiple sources, including SNPs related to major economic traits from the published literature (5706 SNPs), SNPs with favorable polymorphisms detected in multiple pig breeds at home and abroad (34,262 SNPs), and SNPs from the NCBI pig SNP database (8265 SNPs).

References

1. Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **2001**, *157*, 1819–1829. [[CrossRef](#)]
2. Van Eenennaam, A.L.; Weigel, K.A.; Young, A.E.; Cleveland, M.A.; Dekkers, J.C. Applied Animal Genomics: Results from the Field. *Annu. Rev. Anim. Biosci.* **2014**, *2*, 105–139. [[CrossRef](#)]
3. García-Ruiz, A.; Cole, J.B.; VanRaden, P.M.; Wiggans, G.R.; Ruiz-López, F.J.; Van Tassell, C.P. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E3995–E4004. [[CrossRef](#)] [[PubMed](#)]
4. Judge, M.M.; Kearney, J.F.; McClure, M.C.; Sleator, R.D.; Berry, D.P. Evaluation of developed low-density genotype panels for imputation to higher density in independent dairy and beef cattle populations1. *J. Anim. Sci.* **2016**, *94*, 949–962. [[CrossRef](#)]
5. Wu, X.-L.; Xu, J.; Feng, G.; Wiggans, G.R.; Taylor, J.F.; He, J.; Qian, C.; Qiu, J.; Simpson, B.; Walker, J.; et al. Optimal Design of Low-Density SNP Arrays for Genomic Prediction: Algorithm and Applications. *PLoS ONE* **2016**, *11*, e0161719. [[CrossRef](#)] [[PubMed](#)]
6. Wellmann, R.; Preuß, S.; Tholen, E.; Heinkel, J.; Wimmers, K.; Bennewitz, J. Genomic selection using low density marker panels with application to a sire line in pigs. *Genet. Sel. Evol.* **2013**, *45*, 28. [[CrossRef](#)]
7. Grossi, D.A.; Brito, L.F.; Jafarikia, M.; Schenkel, F.S.; Feng, Z. Genotype imputation from various low-density SNP panels and its impact on accuracy of genomic breeding values in pigs. *Animal* **2018**, *12*, 2235–2245. [[CrossRef](#)]

8. Shashkova, T.I.; Martynova, E.U.; Ayupova, A.F.; Shumskiy, A.A.; Ogurtsova, P.A.; Kostyunina, O.V.; Khaitovich, P.E.; Mazin, P.V.; Zinovieva, N.A. Development of a low-density panel for genomic selection of pigs in Russia1. *Transl. Anim. Sci.* **2020**, *4*, 264–274. [[CrossRef](#)] [[PubMed](#)]
9. Bolormaa, S.; Gore, K.; van der Werf, J.H.J.; Hayes, B.J.; Daetwyler, H.D. Design of a low-density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy. *Anim. Genet.* **2015**, *46*, 544–556. [[CrossRef](#)]
10. Raoul, J.; Swan, A.A.; Elsen, J.-M. Using a very low-density SNP panel for genomic selection in a breeding program for sheep. *Genet. Sel. Evol.* **2017**, *49*, 76. [[CrossRef](#)]
11. O'Brien, A.C.; Judge, M.M.; Fair, S.; Berry, D.P. High imputation accuracy from informative low-to-medium density single nucleotide polymorphism genotypes is achievable in sheep1. *J. Anim. Sci.* **2019**, *97*, 1550–1567. [[CrossRef](#)]
12. Wang, C.; Habier, D.; Peiris, B.L.; Wolc, A.; Kranis, A.; Watson, K.A.; Avendano, S.; Garrick, D.J.; Fernando, R.L.; Lamont, S.J.; et al. Accuracy of genomic prediction using an evenly spaced, low-density single nucleotide polymorphism panel in broiler chickens. *Poult. Sci.* **2013**, *92*, 1712–1723. [[CrossRef](#)]
13. Herry, F.; Hérault, F.; Druet, D.P.; Varenne, A.; Burlot, T.; Le Roy, P.; Allais, S. Design of low density SNP chips for genotype imputation in layer chicken. *BMC Genet.* **2018**, *19*, 108. [[CrossRef](#)]
14. Boichard, D.; Chung, H.; Dassonneville, R.; David, X.; Eggen, A.; Fritz, S.; Gietzen, K.J.; Hayes, B.J.; Lawley, C.T.; Sonstegard, T.S.; et al. Design of a Bovine Low-Density SNP Array Optimized for Imputation. *PLoS ONE* **2012**, *7*, e34130. [[CrossRef](#)]
15. Segelke, D.; Chen, J.; Liu, Z.; Reinhardt, F.; Thaller, G.; Reents, R. Reliability of genomic prediction for German Holsteins using imputed genotypes from low-density chips. *J. Dairy Sci.* **2012**, *95*, 5403–5411. [[CrossRef](#)] [[PubMed](#)]
16. Weigel, K.; Campos, G.D.L.; González-Recio, O.; Naya, H.; Wu, X.; Long, N.; Rosa, G.; Gianola, D. Predictive ability of direct genomic values for lifetime net merit of Holstein sires using selected subsets of single nucleotide polymorphism markers. *J. Dairy Sci.* **2009**, *92*, 5248–5257. [[CrossRef](#)] [[PubMed](#)]
17. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
18. Mitra, P.; Pal, S. Erratum: Correction to “unsupervised feature selection using feature similarity”. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 721. [[CrossRef](#)]
19. Phuong, T.M.; Lin, Z.; Altman, R.B. Choosing SNPs Using Feature Selection. *J. Bioinform. Comput. Biol.* **2006**, *4*, 241–257. [[CrossRef](#)]
20. Hill, W.G.; Robertson, A. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **1968**, *38*, 226–231. [[CrossRef](#)]
21. Wall, J.D.; Pritchard, J.K. Haplotype blocks and linkage disequilibrium in the human genome. *Nat. Rev. Genet.* **2003**, *4*, 587–597. [[CrossRef](#)] [[PubMed](#)]
22. Wu, X.; Li, H.; Ferretti, R.; Simpson, B.; Walker, J.; Parham, J.; Mastro, L.; Qiu, J.; Schultz, T.; Tait, R.G.; et al. A unified local objective function for optimally selecting SNPs on arrays for agricultural genomics applications. *Anim. Genet.* **2020**, *51*, 306–310. [[CrossRef](#)]
23. Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.* **2011**, *88*, 76–82. [[CrossRef](#)] [[PubMed](#)]
24. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [[CrossRef](#)] [[PubMed](#)]
25. Madsen, P.; Jensen, J. *A User's Guide to DMU*, Version 6, Release 5.1; Aarhus University: Aarhus, Denmark, 2012.
26. Christensen, O.; Madsen, P.; Nielsen, B.; Ostensen, T.; Su, G. Single-step methods for genomic evaluation in pigs. *Animal* **2012**, *6*, 1565–1571. [[CrossRef](#)]
27. Guo, X.; Christensen, O.F.; Ostensen, T.; Wang, Y.; Lund, M.S.; Su, G. Improving genetic evaluation of litter size and piglet mortality for both genotyped and nongenotyped individuals using a single-step method1. *J. Anim. Sci.* **2015**, *93*, 503–512. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.