



# 大白猪基因组选择在不同应用策略下的比较研究

董林松<sup>1,2</sup>, 谈成<sup>1</sup>, 蔡更元<sup>1,2</sup>, 吴珍芳<sup>1,2\*</sup>

(1. 温氏食品集团股份有限公司, 新兴 527400;

2. 华南农业大学动物科学学院 国家生猪种业工程技术研究中心, 广州 510642)

**摘要:** 目前, 基因组选择(genomic selection, GS)技术已经在种猪育种中开展, 但为获得较高的收益, 还需研究一些应用策略, 如确定仔猪基因分型个体比例和早期仔猪留种比例。本试验选择温氏集团出生于 2011—2016 年的大白种猪作为研究对象, 共有超过 4.5 万条的生长测定记录, 超过 7 万条繁殖记录, 和 2 090 个个体的简化基因组测序(GBS)数据, 其中, 出生于 2016 年 7~12 月的 440 个体作为候选群体。研究性状包括两个生长性状(校正 100 kg 日龄和校正 100 kg 背膘厚)和一个繁殖性状(总产仔数)。为对比预测效果, 在候选群体进行育种值预测时, 按照是否利用其基因型或表型信息分为 4 种预测方案, 比较不同方案的预测可靠性和个体选择指数的排名情况。结果显示, 在预测候选群育种值时, 利用其表型或基因型信息均比不利用时的预测结果更加可靠。对生长性状终测前、后进行基因组选择指数计算, 发现, 终测后指数排名前 30% 的个体都位于终测前指数排名前 60% 内。若仔猪出生后仅选择常规 BLUP 预测指数排名前 60% 的个体, 会导致有接近 15% 的具有优秀潜力的个体被遗漏。本研究建议, 对所有新生健康仔猪都进行基因分型并计算基因组选择指数, 然后对指数排名靠前 60% 的个体进行性能测定。

**关键词:** 基因组选择; 预测可靠性; 育种策略; 种猪

中图分类号: S828.2

文献标志码: A

文章编号: 0366-6964(2020)09-2089-09

## Comparative Analysis of Different Implementation Strategies on Genomic Selection in Large White Pigs

DONG Linsong<sup>1,2</sup>, TAN Cheng<sup>1</sup>, CAI Gengyuan<sup>1,2</sup>, WU Zhenfang<sup>1,2\*</sup>

(1. WENS Foodstuff Group Co., Ltd., Xinxing 527400, China;

2. National Engineering Research Center for Breeding Swine Industry, College of Animal Science, South China Agricultural University, Guangzhou 510642, China)

**Abstract:** At present, genomic selection (GS) has been applied in pigs breeding, but some implementation strategies, such as the determination of genotyping ratios or early selection rates for piglets, are required to obtain a higher benefit using this technology. The Large White pigs born from 2011 to 2016 at WENS Foodstuff Group Co., Ltd were chose as the research objects, including more than 45 000 growth measurement records, more than 70 000 reproduction records and 2 090 individuals with genotyping-by-sequencing (GBS) data. The 440 individuals born from July to December in 2016 were used as the candidate individuals. The traits included two growth traits, age at 100 kg and backfat thickness at 100 kg, and one reproduction trait, number of total born. To compare the prediction effects, four prediction scenarios were designed according to

收稿日期: 2020-03-07

基金项目: 广东省自然科学基金(2018B030313011); 中国博士后科学基金资助项目(2019M653093)

作者简介: 董林松(1986-), 男, 山东威海人, 博士, 主要从事动物分子数量遗传学研究, E-mail: linsong325@163.com; 谈成(1988-), 男, 安徽桐城人, 博士, 主要从事生物信息及数量遗传学研究, E-mail: tanch@scau.edu.cn. 董林松与谈成为同等贡献作者

\* 通信作者: 吴珍芳, 主要从事动物遗传育种研究, E-mail: wzfemail@163.com

including or ignoring the phenotypic or genotypic information of candidate individuals when predicting their breeding values. The predictive reliability of different scenarios and rankings of selection indices of individuals would be compared. The results showed that the results using the phenotypic and genotypic information was more reliable than ignoring them to predict the breeding values of candidate individuals. When genomic selection indices were calculated before and after performances testing for the growth traits, the individuals ranking in the top 30% of indices after testing were all found in the individuals ranking in the top 60% of indices before testing. If the piglets with the top 60% of traditional BLUP indices were only selected, around 15% of individuals with good genetic potentials would be omitted. This study suggests that all healthy piglets after birth are genotyped and their genomic selection indices are calculated, and then the individuals ranking in the top 60% of indices are chose to perform growth measurement.

**Key words:** genomic selection; predictive reliability; breeding strategy; breeding pigs

\* **Corresponding author:** WU Zhenfang, E-mail: wzfemail@163.com

基因组选择(genomic selection, GS)是利用基因组信息对候选亲本进行基因组育种值(genomic estimated breeding values, GEBV)预测并进行选种的方法,最早由 Meuwissen 等<sup>[1]</sup>在 2001 年提出。该技术具有预测可靠、有效控制近交增量<sup>[2]</sup>和能够进行早期选择<sup>[3]</sup>等优点,但也需要较高的基因分型成本,前期仅在奶牛育种上有较广泛的应用<sup>[4]</sup>。但近年来,随着基因分型费用的降低,GS 技术已经开始在其他物种上展开研究与应用,其中,在种猪育种中的应用逐渐受到人们的重视<sup>[5-9]</sup>。在国内,温氏集团是率先开展此项工作的企业<sup>[10]</sup>,该企业于 2011 年已经对该技术进行研究与应用。截至目前,温氏集团已经在大白、长白和杜洛克等多个猪种中开展了 GS 应用。

GS 的原理和 GEBV 计算方法在大部分物种中都通用,但在实际应用策略上却有很大差别。例如,奶牛的主选性状与产奶有关,而选育优良公奶牛是其主要手段,因此,在奶牛 GS 育种上主要采用针对限性性状的选育策略<sup>[3]</sup>。再如,鱼类育种以获得较快生长速度和较好肉质为主,但由于早期成活率较低且不利于采样的原因,往往很难对其进行早期选择<sup>[11]</sup>。种猪的选育以长得快、产肉多、肉质好和产仔多等为主要目标,其中部分性状是限性性状,甚至是无法进行活体度量的性状。尽管 GS 已经应用于种猪育种并被证明可行,但考虑其物种特点和育种成本,以及不同方案下 GEBV 预测的可靠性不同,因此,有必要寻找适合其特点的 GS 育种方案。例如,早期选择是 GS 技术的一大优势,然而早期的留种比例如何确定?如若留种过多则会给栏舍造成压

力且增加生物安全风险,又因太多的公猪没有去势而影响肉质上的风味,在最终出栏时只能降价出售而增加育种成本;但若留种过少则可能错失掉一些具有优秀遗传潜力的种猪。再如,是否需要对所有新出生待选仔猪都进行基因分型?若都进行分型无疑会增加育种成本,可否先通过常规最佳线性无偏预测(best linear unbiased prediction, BLUP)方法<sup>[12]</sup>对所有仔猪进行育种值(estimated breeding values, EBV)预测,挑选 EBV 靠前的仔猪进行基因分型,以此来节省分型成本。

本研究选取温氏大白种猪群体,通过 4 种不同的策略对上述问题展开研究,探索既能节约成本又不丢失优秀遗传资源的基因组育种方案。

## 1 材料与方法

### 1.1 研究材料

选择来自于温氏集团且出生于 2011—2016 年底的大白种猪作为研究对象,一共有超过 52 万条系谱记录,超过 4.5 万条的生长表型测定记录,以及超过 7 万条繁殖记录,且有超过 2 000 个个体有简化基因组测序(genotyping-by-sequencing, GBS)数据。选取校正 100 kg 体重日龄(Age\_100 kg)、校正 100 kg 体重背膘厚(BF\_100 kg)和总产仔数(number of total born, NTB)3 个性状进行研究。Age\_100 kg 和 BF\_100 kg 为生长性状,NTB 为繁殖性状。

### 1.2 数据质控

剔除 Age\_100 kg 中大于 230 d 和小于 110 d 的表型数据;剔除 BF\_100 kg 中大于 35 mm 和小于 5 mm 的数据;在 NTB 中,只选择 1~8 胎的数据,超

过 8 胎的数据剔除,并剔除窝产总仔数低于 3 的窝次,且将窝产总仔数大于 22 的数据校正为 22。以上数据质控的目的是尽可能将错误数据剔除,且归

整出现频次较小的数据令总体服从正态分布。3 个性状质控后的表型统计结果见表 1。

表 1 大白猪 3 个数量性状的表型统计结果

Table 1 Phenotypic statistics for 3 quantitative traits in Large White pigs

性状 Trait	数量 Number	平均值 Mean	标准差 Standard deviation	最大值 Maximum value	最小值 Minimum value
100 kg 体重日龄/d Age_100 kg	45 554	156.62	12.06	221.2	117.07
100 kg 体重背膘厚/mm BF_100 kg	45 487	12.20	2.94	29.41	5.01
总产仔数/头 NTB	72 426	13.52	3.39	22	3

### 1.3 大白猪 GBS 文库构建及测序数据分析

GBS 是一种只对部分基因组序列进行测序的方法<sup>[13]</sup>,因此称为简化基因组测序。简而言之,使用两种限制性内切酶 *EcoR* I 和 *Msp* I 对基因组 DNA 进行双酶切,将酶切产物与特定设计的带有 barcode 的接头连接,通过两次磁珠筛选出长度在 300~500 bp 范围内的连接产物,再经过 PCR 扩增形成 GBS 文库,并在 Illumina HiSeq 2000 测序平台上进行双端测序。

使用 GBSX v1.3 软件<sup>[14]</sup>将原始数据切割成独立的个体 fastq 文件,并使用 FastQC v0.11.8 软件<sup>[15]</sup>对单个样品的测序数据进行质控,删除测序质量低于 1 M reads 或测序质量不合格的个体。将质控后的 reads 通过 BWA v0.7.15 软件<sup>[16]</sup>比对到猪参考基因组上(assembly sscrofa11.1, <https://www.ncbi.nlm.nih.gov/genome/?term=pig>),再使用 SNP 分型软件(the genome analysis toolkit 4.0, GATK 4.0)<sup>[17]</sup>进行基因组 SNP 挖掘。对所鉴定到的缺失率低于 10% 的 SNPs 位点使用 Beagle 4.1 软件<sup>[18]</sup>进行填充,使用 VCFtools v0.1.13 软件<sup>[19]</sup>将填充好的 VCF 文件按照基因型“AA Aa aa”

转换成可分析的“0 1 2”格式,并筛选最小等位基因频率(minor allele frequency, MAF)大于 0.01 且经过卡方检验符合哈代-温伯格平衡( $P > 1 \times 10^{-5}$ )的位点。剔除测序质量不合格或无表型数据的个体,最终获得 2 090 个有基因组信息的个体,且每个个体有 197 483 个高质量的 SNPs 位点用于后续分析。

### 1.4 传统和基因组育种值预测方案和模型

1.4.1 候选群体划分 本研究将 2 090 头有基因组信息的大白猪按照出生日期先后排列,排在最后的 440 头猪(出生日期范围为 2016 年 7~12 月)作为候选群体。所有 2 090 头猪都有生长终测数据,但只有其中的 1 078 个体有繁殖数据,这是因为公猪无法产仔,而且一些母猪还没有达到产仔的日龄。

1.4.2 传统和基因组育种值预测方案 采用 4 种方案来预测候选群的生长性状育种值,即根据是否用到候选群的基因组信息和表型来划分,表 2 中是每种方案的简要介绍。在繁殖性状中,候选群在生长性状终测时还没有产仔数据,因此只采用方案 2( $G_Y P_N$  组)和方案 4( $G_N P_N$  组)对育种值进行预测。

表 2 4 种预测候选群育种值的方案

Table 2 Four scenarios of breeding values prediction for candidate individuals

方案 Scenario	描述 <sup>a</sup> Description	英文缩写 Abbreviation	备注 Note
方案 1 Scenario 1	有基因型有表型	$G_Y P_Y$	相当于在终测结束后使用基因组预测 GEBV
方案 2 Scenario 2	有基因型无表型	$G_Y P_N$	相当于在仔猪出生后使用基因组预测 GEBV
方案 3 Scenario 3	无基因型有表型	$G_N P_Y$	相当于在终测结束后使用传统 BLUP 预测 EBV
方案 4 Scenario 4	无基因型无表型	$G_N P_N$	相当于在仔猪出生后使用传统 BLUP 预测 EBV

<sup>a</sup>. 指在预测候选群育种值时,是否利用了候选群的基因组或表型信息

<sup>a</sup>. Indicates whether the genomic or phenotypic information of candidate individual was used to predict their breeding values

### 1.4.3 传统和基因组育种值预测模型和方法

根据预测方案的不同采用不同的育种值预测方法,当选择方案 1 和 2 时,采用系谱结合基因组信息的一步基因组 BLUP 法(single-step genomic BLUP, ssGBLUP)<sup>[20-21]</sup>;当选择方案 3 和 4 时,只使用传统的 BLUP 方法<sup>[12]</sup>。两种方法的数学模型和求解方程组非常类似,只是在构建加性遗传相关矩阵上,前者使用系谱和基因组结合产生的 H 矩阵,而后者仅通过系谱计算生成 A 矩阵。 $H^{-1}$  的计算<sup>[22]</sup>:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & (\alpha G + \beta A_{22})^{-1} - A_{22}^{-1} \end{bmatrix}$$

其中, G 为基因组关系矩阵,由基因组 SNP 基因型计算而来<sup>[23]</sup>;  $A_{22}$  为有基因组信息的个体通过系谱计算的加性遗传相关矩阵;在本研究中,  $\alpha$  和  $\beta$  分别取值为 0.99 和 0.01,即在有基因组信息的情况下尽可能多地利用该信息计算个体间的亲缘关系。

使用双性状模型对 Age\_100 kg 和 BF\_100 kg 进行预测:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} K_1 & 0 \\ 0 & K_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

其中,  $y_i$ 、 $b_i$ 、 $c_i$ 、 $a_i$  和  $e_i$  分别表示第  $i$  个性状的表型值向量、固定效应向量、共同环境(窝组)效应向量、育种值向量和随机剩余值效应向量,其中,固定效应由性别和测定场-年-季构成,  $X_i$ 、 $K_i$  和  $Z_i$  则分别为  $b_i$ 、 $c_i$  和  $a_i$  的关联矩阵。本研究假设共同环境(窝组)效应和育种值在两性状间存在相关,而随机剩余值效应不存在,从而获得 3 个随机效应的方差-协方差矩阵:

$$\text{Var}(c) = \begin{bmatrix} I\sigma_{c_1}^2 & I\sigma_{c_1 c_2} \\ I\sigma_{c_1 c_2} & I\sigma_{c_2}^2 \end{bmatrix}$$

其中, I 为单位矩阵,  $\sigma_{c_1}^2$ 、 $\sigma_{c_2}^2$  和  $\sigma_{c_1 c_2}$  分别为第一个性状和第二个性状的窝组效应方差,以及二者的协方差:

$$\text{Var}(a) = \begin{bmatrix} M\sigma_{a_1}^2 & M\sigma_{a_1 a_2} \\ M\sigma_{a_1 a_2} & M\sigma_{a_2}^2 \end{bmatrix}$$

其中, M 在 ssGBLUP 方法中用 H 表示,在 BLUP 方法中用 A 表示,  $\sigma_{a_1}^2$ 、 $\sigma_{a_2}^2$  和  $\sigma_{a_1 a_2}$  分别为第一个性状和第二个性状的加性遗传方差,以及二者的协方差:

$$\text{Var}(e) = \begin{bmatrix} I\sigma_{e_1}^2 & 0 \\ 0 & I\sigma_{e_2}^2 \end{bmatrix}$$

其中,  $\sigma_{e_1}^2$  和  $\sigma_{e_2}^2$  分别为两个性状的剩余值方差。

对 NTB 性状使用包含永久环境效应的重复力模型进行预测:

$$y = Xb + Kp + Za + e$$

其中,  $p$  为永久环境效应向量,且  $p \sim N(0, I\sigma_p^2)$ ,  $\sigma_p^2$  为永久环境方差,  $K$  为其关联矩阵,其余变量的解释与生长预测模型类似,其中固定效应由分娩胎次和分娩场-年-季构成。

使用 BLUPF90 软件<sup>[24]</sup> 中 AIREMLF90 函数对上述性状进行遗传参数估计和育种值预测,程序的迭代收敛标准为相邻两次迭代的估计方差的差值小于  $1 \times 10^{-8}$ 。使用预测的可靠性(reliability)作为育种值预测准确与否的评价指标,其理论计算公式<sup>[25]</sup>:

$$\text{Reliability} = 1 - \frac{\text{PEV}}{\sigma_a^2}$$

其中, PEV 为育种值预测值的误差方差,可由 BLUPF90 计算而来,理论上可靠性的平方根即为真实育种值与估计育种值的相关系数<sup>[25-26]</sup>,即育种值预测准确性(accuracy)。在所有模型预测结束后,根据预测可靠性来比较各个预测方案的优劣。

在实际育种中,生长性状的  $G_Y P_Y$  组和繁殖性状的  $G_Y P_N$  组所预测的 GEBV 是终测时选种的主要依据之一,因此,其他组将通过计算与这两组结果的相关性,以及综合选择指数的排序情况来讨论基因组选择的优化策略。在进行选择指数排序前先对其进行计算,计算公式:

$$\begin{aligned} \text{选择指数} = & -0.3 \times \frac{\text{Age}_{100} \text{ kg 育种值}}{\text{Age}_{100} \text{ kg 育种值标准差}} - \\ & 0.3 \times \frac{\text{BF}_{100} \text{ kg 育种值}}{\text{BF}_{100} \text{ kg 育种值标准差}} + 0.4 \times \\ & \frac{\text{NTB 育种值}}{\text{NTB 育种值标准差}} \end{aligned}$$

其中,育种值标准差是通过 BLUPF90 估计的遗传参数值。将综合指数从高到低排序,并根据综合指数的秩序设定编号 1,2,3...,直到 440。编号值越小越符合选种标准,说明个体优秀的概率越大。

## 2 结果

### 2.1 大白猪 3 个性状的遗传参数估计结果

采用包含(ssGBLUP 模型)和不包含(BLUP 模

型)基因组信息两种模型来估计 3 个性状的遗传参数,估计结果见表 3。两个模型对各个组分的估计结果非常接近,且所有遗传参数的估计值均极显著大于 0( $P < 0.01$ ),表明利用其中任一种模型进行遗传参数估计都是可行的。总体上,NTB、Age\_100 kg 和 BF\_100 kg 的遗传力估计值分别位于 0.065、0.26 和 0.41 附近,NTB 最低而 BF\_100 kg 最高。

表 3 大白猪 3 个性状的遗传参数估计结果

Table 3 Genetic parameter estimates for 3 traits in Large White pigs

性状 Trait	模型 Model	$\sigma_a^2$	$\sigma_p^2$	$\sigma_e^2$	$\sigma_e^2$	狭义遗传力 $h^2$ Narrow sense heritability	重复力 $r_e$ Repeatability
Age_100 kg	BLUP 模型	22.327±1.348	—	13.967±0.479	50.960±0.804	0.256±0.014	—
	ssGBLUP 模型	22.927±1.364	—	13.998±0.478	50.747±0.804	0.262±0.014	—
BF_100 kg	BLUP 模型	1.654±0.070	—	0.335±0.017	2.056±0.039	0.409±0.014	—
	ssGBLUP 模型	1.665±0.069	—	0.338±0.017	2.060±0.039	0.410±0.014	—
NTB	BLUP 模型	0.693±0.056	0.755±0.054	—	8.999±0.057	0.066±0.006	0.139±0.004
	ssGBLUP 模型	0.667±0.055	0.780±0.053	—	8.999±0.057	0.064±0.006	0.139±0.004

## 2.2 传统和基因组育种值预测

2.2.1 预测可靠性 4 种方案的预测可靠性如表 4 所示。在同一性状中,当候选群有表型且使用基因组预测( $G_Y P_Y$  组)时,预测可靠性是最高的,极显著高于其他各组( $P < 0.01$ )。而当表型和基因型都没有( $G_N P_N$  组)时,相当于仔猪出生后利用常规 BLUP 对 EBV 进行预测,预测可靠性极显著低于其他各组( $P < 0.01$ )。而在  $G_Y P_N$  和  $G_N P_Y$  两组中,预测可靠性较为接近。当仔猪出生后,通过基因组预测 GEBV( $G_Y P_N$  组)比 BLUP 预测 EBV( $G_N P_N$  组)的可靠性在 Age\_100 kg、BF\_100 kg 和 NTB 3 个性

对 Age\_100 kg 和 BF\_100 kg 的双性状模型还可以估计出遗传相关和窝组效应相关。在遗传相关上, BLUP 模型估计结果是一 0.346±0.021, ssGBLUP 模型的结果是一 0.346±0.021;在窝组效应相关上, BLUP 模型结果是一 0.207±0.029, 而 ssGBLUP 模型结果是一 0.205±0.029, 可见各种相关的估计结果在两种模型中也几乎一致。

状中分别提高的比例为 53.2%、52.8% 和 56.7%, 说明通过基因组预测 GEBV 比常规 BLUP 准确得多,这符合绝大多数的研究结果<sup>[27-31]</sup>。当待选种猪结测后,通过基因组预测 GEBV( $G_Y P_Y$  组)比 BLUP 预测 EBV( $G_N P_Y$  组)的可靠性在 Age\_100 kg 和 BF\_100 kg 上分别提高的比例为 27.3% 和 20.1%, 效果也非常不错,但不如出生时明显。

在同一种预测方案中比较 3 个性状的预测可靠性, BF\_100 kg 为最高, Age\_100 kg 次之, 而 NTB 最差,与遗传力估计值的排序完全一样。这个规律不论是在  $G_Y P_N$  组还是  $G_N P_N$  组都是相同的。

表 4 大白猪 3 个性状在不同方案中的育种值预测可靠性(平均值±标准误)

Table 4 Predictive reliability of breeding values of different scenarios for 3 traits in Large White pigs(Mean±SE)

性状 Trait	$G_Y P_Y$	$G_Y P_N$	$G_N P_Y$	$G_N P_N$
Age_100 kg	0.573±0.001	0.458±0.002	0.450±0.001	0.299±0.002
BF_100 kg	0.663±0.001	0.527±0.002	0.552±0.001	0.345±0.002
NTB	—	0.293±0.002	—	0.187±0.002

### 2.2.2 不同组育种值相关性和指数分布情况

通过 4 种方案对 Age\_100 kg 和 BF\_100 kg 预测育种值的 Pearson 相关结果如表 5 所示, Spearman 秩

相关与 Pearson 相关的值十分接近,就不再展示。在 NTB 性状中只有  $G_Y P_N$  和  $G_N P_N$  两种方案的预测结果,其相关性为 0.544±0.026。在 Age\_100 kg

和 BF\_100 kg 中, 不论选取哪两组的结果进行比较, 都会发现 BF\_100 kg 中的结果值要高于 Age\_100 kg, 原因可能是 BF\_100 kg 的遗传力较高, 在  $G_N P_N$  组中育种值预测的已经较为准确, 因此, 即使再增加其他如表型或基因型信息, 所预测的育种值与  $G_N P_N$  组也会有相对较高相关性。而对于遗传力相对低的性状, 所增加的额外信息越多, 预测可靠性所提高的幅度越大, 不同组之间的育种值相关性

自然会低一些。而针对同一生产性状, 与  $G_Y P_Y$  组预测结果相关最高的是  $G_Y P_N$  组,  $G_N P_Y$  组则略低。但若对  $G_Y P_N$  和  $G_N P_Y$  两组结果进行比较会发现相关性并不高, 在 Age\_100 kg 性状中只有 49%, 而在 BF\_100 kg 中也仅有 66%。尽管这两组的预测可靠性比较接近, 各自与  $G_Y P_Y$  组预测育种值的相关系数也比较接近, 但根据这两组结果之间的相关性发现两组间预测结果的差异还是比较大的。

表 5 在 Age\_100 kg 和 BF\_100 kg 性状中 4 种方案预测育种值的 Pearson 相关结果

Table 5 Pearson correlations of breeding values predicted by 4 scenarios for Age\_100 kg and BF\_100 kg

	$G_Y P_Y$	$G_Y P_N$	$G_N P_Y$	$G_N P_N$
$G_Y P_Y$		0.885±0.016	0.859±0.018	0.669±0.027
$G_Y P_N$	0.838±0.019		0.660±0.028	0.700±0.026
$G_N P_Y$	0.779±0.022	0.490±0.034		0.810±0.021
$G_N P_N$	0.520±0.033	0.530±0.033	0.759±0.023	

下三角是 Age\_100 kg 结果, 上三角是 BF\_100 kg 结果(平均值±标准误差)

The results (Mean±SE) of Age\_100 kg are under the diagonal and results of BF\_100 kg are upper the diagonal

选择终测时  $G_Y P_Y$  组排名前 30% 的个体, 即最优秀的 132 个个体, 观察这些个体在其他各组排名区间的分布情况(表 6)。结果发现, 这些优秀的个体都位于  $G_Y P_N$  组所预测育种值的前 60%。而其余两组均有部分优秀个体遗漏在了前 60% 的区间之外。若在仔猪出生后使用常规 BLUP 预测(即  $G_N P_N$  组), 则  $G_Y P_Y$  组最优秀的 132 个个体中, 有 114 个个体位于

$G_N P_N$  组的前 60% 区间内, 所占比例超过 85%。

### 3 讨论

本研究所选择的 3 个性状, 是在大白猪育种中使用较为普遍的经济性状, 总产仔数、Age\_100 kg 和 BF\_100 kg 分别与猪的繁殖潜力、生长速度和瘦肉率有关, 且生长速度越快的猪其饲料利用效率往往更高。表 3 显示, 上述 3 个性状的遗传力分别位于低、中等和高遗传力范围内, 这符合之前学者的研究结果<sup>[32-34]</sup>, 同时也证明本研究所使用的数据是可靠的。但在彭潇等<sup>[35]</sup>的研究结果中, Age\_100 kg 的遗传力估计值高于 BF\_100 kg, 与本研究正好相反, 原因可能是在他们的评估模型中没有考虑窝组效应对表型的影响。在表 3 中可以看到, 窝组效应方差占总体方差的比重, 在 Age\_100 kg 中为 16% 左右, 但在 BF\_100 kg 中也占到了 8.3% 左右, 若不考虑此因素会导致遗传力估计值偏高, 尤其是对 Age\_100 kg 性状。根据 Daetwyler 等<sup>[36]</sup>的研究报道, 育种值预测的准确性与遗传力存在紧密关联, 遗传力越高育种值预测准确性也越高, 而本研究结果(表 4)也支持该结论。Age\_100 kg 与 BF\_100 kg 两性状之间存在极显著( $P < 0.01$ )的负遗传相关, 说明背膘厚的种猪其生长速度会相对较快, 因此需要通过综合选择指数来进行控制, 避免在选育其中

表 6 选择指数在  $G_Y P_Y$  组中排名前 30% 的个体在其他各组中的排名分布情况

Table 6 The rank distributions in other groups for the individuals ranking in the top 30% of selection indices in group  $G_Y P_Y$

其他组排名区间 Ranking intervals in other groups	$G_Y P_N$	$G_N P_Y$	$G_N P_N$
前 10%	44	30	29
10%~20%	42	31	33
20%~30%	27	19	20
30%~40%	13	22	14
40%~50%	3	12	12
50%~60%	3	3	6
60%~70%	0	5	4
70%~80%	0	5	9
80%~90%	0	4	3
90%~100%	0	1	2

一个性状时,另一个性状朝向不利的方向发展。

根据本研究结果,不管是高遗传力还是低遗传力性状,使用 GS 比常规 BLUP 选择所获得的额外收益都要高,其根本原因在于,基因组选择能够更准确地预测育种值中的孟德尔抽样误差项<sup>[2]</sup>。因为 GS 使用的是基因组信息来估计个体间的亲缘关系,能够准确观察到等位基因在世代间传递的具体情况,而不像系谱仅通过概率来推断个体间的亲缘关系。因此,从这个角度分析,任何性状的选育都可以采用 GS 技术,但在实际育种中,也需要同时考虑成本问题。种猪育种的特点是候选群的生长性状在终测时会有表型数据,但还没有繁殖数据。因此,对于生长性状,应该是在有表型的前提下,观察 GS 与常规 BLUP 两种方法的预测效果,即将  $G_Y P_Y$  组和  $G_N P_Y$  组结果进行比较;而对于繁殖性状,应该在没有表型的情况下,观察两种方法的预测效果,即将  $G_Y P_N$  组和  $G_N P_N$  组进行比较。通过对比可发现,GS 对繁殖性状的预测可靠性提高的幅度要明显高于生长性状,原因在于,生长性状遗传力高于繁殖性状而且生长性状在终测时可以获得表型,EBV 估计本来就较为准确,因此 GS 对其所产生的优势相对不会太明显。因此,建议在资金充足的情况下,父系猪和母系猪品种都可以选择 GS 育种方法,但若资金较为紧张,则应该优先考虑在母系猪品种中开展 GS 育种。

尽管 BLUP 也可以在仔猪出生时预测 EBV,但由于个体间亲缘关系是通过概率推断的,这会导致同一窝内的所有仔猪通过 BLUP 预测的 EBV 是相同的,而基因组选择可以令同一窝内仔猪预测出不同的 GEBV。因此,相对于常规 BLUP,GS 在早期选择上更加可靠,表 6 的结果也证实了这一推断。据本研究结果,若在仔猪出生后全部进行基因组预测,则只需将综合指数排名前 60% 的个体进行性能终测即可,因为终测后所选的前 30% 优秀个体全部都位于出生后基因组预测指数排名的前 60% 区间内,其余个体可转做他用,没必要为排名后 40% 的个体花费大量的人力、物力和财力。而若先进行 BLUP 预测,然后选择 EBV 排名靠前 60% 的个体进行基因分型,则会错失掉接近 15% 的优秀个体。因此,对于资金较为充足的企业,建议对所有健康仔猪全部进行 DNA 提取和基因分型,这样可以避免遗漏优秀的种猪。

## 4 结 论

本研究通过是否利用候选群表型或基因组信息等 4 种方案对自身育种值进行预测,结果显示,利用信息越多时预测的可靠性越高。对比不同预测方案下育种值相关性和选择指数的分布情况,建议对大白猪候选群体的基因组选择实施如下策略:仔猪出生时尽可能对所有健康仔猪进行基因分型,通过基因组预测其基因组育种值并计算选择指数,只选择指数值排名前 60% 的仔猪进行性能终测和下一步的选种工作,可保证在不丢失优秀种猪资源的前提下尽可能节省育种成本。

## 参考文献(References):

- [1] MEUWISSEN T H, HAYES B J, GODDARD M E. Prediction of total genetic value using genome-wide dense marker maps [J]. *Genetics*, 2001, 157 (4): 1819-1829.
- [2] DAETWYLER H D, VILLANUEVA B, BIJMA P, et al. Inbreeding in genome-wide selection[J]. *J Anim Breed Genet*, 2007, 124(6):369-376.
- [3] SCHAEFFER L R. Strategy for applying genome-wide selection in dairy cattle[J]. *J Anim Breed Genet*, 2006, 123(4):218-223.
- [4] HAYES B J, BOWMAN P J, CHAMBERLAIN A J, et al. *Invited review*: genomic selection in dairy cattle: progress and challenges[J]. *J Dairy Sci*, 2009, 92 (2):433-443.
- [5] OSTERSEN T, CHRISTENSEN O F, MADSEN P, et al. Sparse single-step method for genomic evaluation in pigs[J]. *Genet Sel Evol*, 2016, 48:48.
- [6] TUSELL L, GILBERT H, RIQUET J, et al. Pedigree and genomic evaluation of pigs using a terminal-cross model[J]. *Genet Sel Evol*, 2016, 48:32.
- [7] TRIBOUT T, LARZUL C, PHOCAS F. Efficiency of genomic selection in a purebred pig male line[J]. *J Anim Sci*, 2012, 90(12): 4164-4176.
- [8] 王 晨, 秦 珂, 薛 明, 等. 全基因组选择在猪育种中的应用[J]. *畜牧兽医学报*, 2016, 47(1):1-9.  
WANG C, QIN K, XUE M, et al. Application of genomic selection in swine breeding[J]. *Acta Veterinaria et Zootechnica Sinica*, 2016, 47(1):1-9. (in Chinese)
- [9] SONG H, ZHANG J, JIANG Y, et al. Genomic prediction for growth and reproduction traits in pig using an admixed reference population[J]. *J Anim Sci*, 2017, 95(8):3415-3424.

- [10] 王青来,李亚兰.全国首例全基因组选育特级种猪在粤诞生[J].农业知识(科学养殖),2014(2):25.  
WANG Q L,LI Y L. The first pig selected by genomic selection was appeared in Guangdong[J]. *Agricultural Knowledge*,2014(2):25. (in Chinese)
- [11] DONG L S,XIAO S J,WANG Q R, et al. Comparative analysis of the GBLUP, emBayesB, and GWAS algorithms to predict genetic values in large yellow croaker (*Larimichthys crocea*) [J]. *BMC Genomics*, 2016,17:460.
- [12] HENDERSON C R. Best linear unbiased estimation and prediction under a selection model[J]. *Biometrics*,1975,31(2):423-447.
- [13] ELSHIRE R J, GLAUBITZ J C, SUN Q, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species[J]. *PLoS One*, 2011,6(5):e19379.
- [14] HERTEN K, HESTAND M S, VERMEESCH J R, et al. GBSX: a toolkit for experimental design and demultiplexing genotyping by sequencing experiments [J]. *BMC Bioinformatics*,2015,16(1):73.
- [15] XI Y,DI L,FEI L, et al. HTQC:a fast quality control toolkit for Illumina sequencing data[J]. *BMC Bioinformatics*,2013,14(1):33.
- [16] LI H,DURBIN R. Fast and accurate long-read alignment with Burrows-Wheeler transform[J]. *Bioinformatics*, 2010, 26(5): 589-595.
- [17] MCKENNA A, HANNA M, BANKS E, et al. The genome analysis toolkit:a MapReduce framework for analyzing next-generation DNA sequencing data[J]. *Genome Res*,2010,20(9):1297-1303.
- [18] BROWNING B L,BROWNING S R. Genotype imputation with millions of reference samples[J]. *Am J Hum Genet*,2016, 98(1):116-126.
- [19] DANECHEK P, AUTON A, ABECASIS G, et al. The variant call format and VCFtools[J]. *Bioinformatics*, 2011,27(15):2156-2158.
- [20] LEGARRA A,AGUILAR I,MISZTAL I. A relationship matrix including full pedigree and genomic information[J]. *J Dairy Sci*, 2009, 92(9):4656-4663.
- [21] MISZTAL I,LEGARRA A,AGUILAR I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information[J]. *J Dairy Sci*,2009,92(9):4648-4655.
- [22] LIU Z, GODDARD M E, REINHARDT F, et al. A single-step genomic model with direct estimation of marker effects [J]. *J Dairy Sci*, 2014, 97 (9): 5833-5850.
- [23] VANRADEN P M. Efficient methods to compute genomic predictions[J]. *J Dairy Sci*,2008,91(11): 4414-4423.
- [24] MISZTAL I, TSURUTA S, STRABEL T, et al. BLUPF90 and related programs (BGF90)[C]//Proceedings of the 7th World Congress on Genetics Applied to Livestock Production. Montpellier,France, 2002.
- [25] GORJANC G,BIJMA P,HICKEY J M. Reliability of pedigree-based and genomic evaluations in selected populations[J]. *Genet Sel Evol*,2015,47(1):65.
- [26] AGUILAR I,FERNANDEZ E N,BLASCO A, et al. Effects of ignoring inbreeding in model-based accuracy for BLUP and SSGBLUP[J]. *J Anim Breed Genet*, 2020,137(4):356-364.
- [27] 李 勇,苗泽圃,蒙小云,等. RAD-seq 技术在大白猪繁殖性状基因组选择上的应用[J]. 农业生物技术学报,2017, 25(9): 1508-1515.  
LI Y,MIAO Z P,MENG X Y, et al. The Application of RAD-seq technology on genomic selection of fertility traits for Large White pigs (*Sus scrofa*) [J]. *Journal of Agricultural Biotechnology*, 2017, 25 (9): 1508-1515. (in Chinese)
- [28] FORNI S, AGUILAR I, MISZTAL I. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information [J]. *Genet Sel Evol*,2011,43(1):1.
- [29] TUSELL L, PÉREZ-RODRÍGUEZ P, FORNI S, et al. Genome-enabled methods for predicting litter size in pigs;a comparison[J]. *Animal*,2013,7(11): 1739-1749.
- [30] GUO X,CHRISTENSEN O F,OSTERSEN T, et al. Improving genetic evaluation of litter size and piglet mortality for both genotyped and nongenotyped individuals using a single-step method[J]. *J Anim Sci*, 2015,93(2):503-512.
- [31] OSTERSEN T, CHRISTENSEN O F, HENRYON M, et al. Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs [J]. *Genet Sel Evol*, 2011, 43(1):38.
- [32] BIDANEL J P,DUCOS A,GUÉBLEZ R, et al. Genetic parameters of backfat thickness, age at 100 kg and ultimate pH in on-farm tested French Landrace and Large White pigs[J]. *Livest Prod Sci*,1994,40(3): 291-301.



- [33] FERRAZ J B S, JOHNSON R K. Animal model estimation of genetic parameters and response to selection for litter size and weight, growth, and backfat in closed seedstock populations of large white and Landrace swine[J]. *J Anim Sci*, 1993, 71(4): 850-858.
- [34] YE J, TAN C, HU X X, et al. Genetic parameters for reproductive traits at different parities in Large White pigs[J]. *J Anim Sci*, 2018, 96(4): 1215-1220.
- [35] 彭 潇, 尹立林, 梅全顺, 等. 猪主要经济性状的基因组选择研究[J]. *畜牧兽医学报*, 2019, 50(2): 439-445.
- [36] PENG X, YIN L L, MEI Q S, et al. A study of genome selection based on the porcine major economic traits[J]. *Acta Veterinaria et Zootechnica Sinica*, 2019, 50(2): 439-445. (in Chinese)
- [36] DAETWYLER H D, VILLANUEVA B, WOOLLIAMS J A. Accuracy of predicting the genetic risk of disease using a genome-wide approach[J]. *PLoS One*, 2008, 3(10): e3395.

(编辑 郭云雁)