

深度学习模型提高猪繁殖性状基因组选择准确性的研究

韦佳霖¹, 刘炎², 印遇龙¹, 孙敬春^{1*}, 徐康^{1*}

(1. 中国科学院亚热带农业生态研究所, 湖南长沙 410000; 2. 唐人神集团股份有限公司, 湖南株洲 412000)

摘要: 为了探究深度学习模型在猪繁殖性状中的基因组选择效果, 本研究以唐人神集团大白猪种猪群体的繁殖表型和芯片分型数据为研究对象, 采用基因型填充的策略增加遗传数据标记密度, 并评估填充前后数据集, 使用GBLUP、RF、LightGBM和深度学习模型DNNGP进行基因组选择的准确性。结果显示, 深度学习模型在芯片数据基因组选择中准确性比其他3款模型高0.032~0.155, 并且在基因型填充数据集中准确性比其他3款模型最高提升了0.248。结果表明深度学习模型在低遗传力性状和基因型填充数据集中均能够保持较好的基因组选择准确性, 并在猪智能化精准育种中具有较好的应用前景。

关键词: 大白猪; 深度学习; 基因组选择; 繁殖性状

中图分类号: S828.2

文献标识码: A

DOI 编号: 10.19556/j.0258-7033.20250802-04

在现代畜禽育种中, 基因组选择 (Genomic Selection, GS) 已成为加快遗传进展、提升繁殖效率的重要工具^[1-2]。相较于传统育种方法, 基因组选择能够利用分子标记预测个体的育种值, 从而在个体尚未表现出表型之前就进行精准筛选, 大幅缩短育种周期, 降低成本^[3]。在猪育种中, 繁殖性状普遍属于低遗传力性状, 遗传改良速度缓慢, 通过基因组选择可有效提高早期选择的准确性, 提升核心种猪群体的整体性能, 具有显著的经济与产业价值^[4]。近年来, 随着测序技术的发展与成本降低, 基因型填充 (Genotype Imputation) 已逐渐成为提高 SNP 覆盖密度、增强模型信息量和预测能力的重要手段^[5], 为后续的基因组选择模型优化提供了更丰富的基因组数据支持。

随着计算机硬件设施的迭代升级和人工智能的兴

起, 研究者开始探索机器学习和深度学习模型在基因组选择中的应用。随机森林 (Random Forest, RF)、轻量梯度提升机 (Light Gradient Boosting Machine, LightGBM) 都是基于决策树算法的机器学习模型, 通过将不同样本和特征输入多棵决策树中, 获得具有鲁棒性的准确结果。在以往研究中, 这类集成学习方法凭借对非线性关系的良好建模能力, 在多个性状预测中取得了优异结果^[6-7]。近年来, 卷积神经网络 (Convolutional Neural Network, CNN) 等深度学习方法被引入基因组选择, 这是一种能够快速选择重要特征的网络结构, 通过卷积核连续扫描全基因组变异信息并对其进行线性转换, 能够挖掘大规模基因型数据中的潜在结构与特征, 提升模型对复杂性状的预测能力^[8-9]。深度学习展现出强大的鲁棒性和泛化性能, 使其成为基因组选择研究的重要方向。

因此, 本研究基于唐人神集团大白种猪群体的繁殖表型与芯片基因型数据, 通过基因型填充扩展变异位点密度, 并结合主成分降维, 构建多种基因组选择模型, 系统比较基因组最佳线性无偏估计 (Genomic Best Linear Unbiased Prediction, GBLUP)、RF、LightGBM 与深度学习模型深度神经网络基因组预测 (Deep Neural Network Genomic Prediction, DNNGP) 在低遗传力繁殖性状上的预测性能, 旨在评估深度学习

收稿日期: 2025-08-02; 修回日期: 2025-09-28

资助项目: 农业生物育种国家科技重大专项 (2023ZD04046); 中国博士后科学基金面上资助 (2024M753485)

作者简介: 韦佳霖 (1999-), 男, 广西河池人, 硕士研究生, 主要从事动物遗传育种与繁殖方向研究, E-mail: wl18178889075@163.com

*通信作者: 徐康 (1988-), 男, 湖北荆门人, 博士, 副研究员, 主要从事动物遗传育种与繁殖方向研究, E-mail: xukang@isa.ac.cn; 孙敬春 (1991-), 男, 博士, 江苏南通人, 助理研究员, 主要从事动物遗传育种与繁殖方向研究, E-mail: sunjingchun@isa.ac.cn

模型在处理高维基因型信息中的准确性与稳定性,为猪的繁殖性状育种提供更高效率的基因组选择策略,推动智能化育种技术的发展与应用。

1 材料与方法

1.1 研究材料 本研究收集唐人神集团繁育场 2021—2024 年的大白母猪繁殖记录共 10 428 条,包含总产仔数 (Total Number Born, TNB)、产活仔数 (Number Born Alive, NBA)、健仔数 (Number Born Healthy, NBH)、出生窝重 (Litter Birth Weight, LBW),并收集系谱记录 17 783 条。随机选择 957 头大白母猪,采集耳样,使用康普森中芯一号进行基因分型,并使用 PLINK (v1.90)^[10] 软件中的 “--geno 0.1 --maf 0.01 --hwe 1e-6” 参数进行质控,保留 47 067 个常染色体 SNP 位点用于后续的分析。

1.2 遗传参数估计与繁殖表型校正 首先剔除了大白母猪繁殖记录中的极端值,保留了 2 头 ≤ 总产仔数 ≤ 36 头、产活仔数 ≥ 2 头、健仔数 ≥ 2 头、1.50 kg ≤ 出生窝重 < 35.00 kg 的繁殖记录。对于大白母猪繁殖性状,本研究考虑到永久环境效应,使用 ASReml (v4.2) 软件的重复力模型估计方差组分和遗传力,模型描述如下:

$$y = Xb + Za + Wpe + e$$

其中:繁殖性状的表型向量用 y 表示;固定效应 b 包含了分娩场-年-季和胎次效应; a 为母猪个体加性效应矩阵; pe 为永久环境效应向量; e 为随机残差效应向量; X 、 Z 、 W 分别为对应的结构矩阵。由于母猪繁殖性状存在一猪多胎的记录,本研究使用估计育种值 (Estimated Breeding Value, EBV) 与多个胎次的残差均值相加作为校正表型用于后续研究。

1.3 基因型数据预处理 为了增加芯片数据的变异标记密度从而提高基因组选择准确性,本研究使用 PHARP (<https://alphaindex.zju.edu.cn/PHARP/index.php/>)^[11] 在线填充面板对芯片数据进行基因型填充,选择 PHARPV4 面板,并采用 SHAPEIT2 + MINIMAC4 的策略对芯片数据进行填充^[12-13],随后剔除填充准确性 $R^2 < 0.8$ 的低质量填充位点,最终保留了 9 678 076 个变异位点用于后续的分析。对于基因组选择使用的机器学习模型和深度学习模型,本研究采用主成分分析 (Principal Components Analysis, PCA) 降维的策略对基因型数据进行预处理,从而加快模型训练速度,防止训练过程中

出现梯度爆炸等问题。具体来说,对于芯片数据和基因型填充数据,本研究统一保留了前 360 个主成分作为机器学习模型和深度学习模型的输入特征。

1.4 基因组选择模型比较 本研究采用 4 种模型进行基因组选择准确性,包括 GBLUP、RF、LightGBM 以及 DNNGP。其中,GBLUP 是应用最为广泛的基因组选择线性模型,其表达式如下:

$$y_c = \mu + Za + e$$

其中: y_c 是校正表型向量; μ 为总体均值; a 为加性效应向量,假设服从正态分布 $N(0, G\sigma_g^2)$; e 是残差向量。本研究采用十乘交叉验证的策略评估线性模型基因组选择准确性,即将群体随机分为数量基本一致的 10 组,每次测试依次将每组作为验证群,其他 9 组作为训练群体。计算每次测试时验证群的观测值与预测值之间的皮尔逊相关系数,该过程重复 20 次,取皮尔逊相关系数的均值作为基因组选择最终的准确性。 G 矩阵构建和育种值估计由 HIBLUP^[14] (v1.5.0) 实现,交叉验证由内部脚本实现。

本研究使用 RF 和 LightGBM 2 款机器学习模型比较基因组选择准确性。其中,RF 是一款经典的决策树集成学习模型,其通过将输入数据随机抽样,独立训练每颗树,最后将多颗树的预测结果平均得到最终的预测结果。其表达式如下^[6]:

$$y = \frac{1}{M} \sum_{m=1}^M t_m (\psi_m (y:X))$$

其中: y 是随机森林预测值; $t_m (\psi_m (y:X))$ 表示每棵树的预测值; M 表示决策树的数量。LightGBM 同样是一种基于梯度提升决策树的高效算法,并结合特征划分实现高效的并行计算与回归预测。决策树的数量和树的深度能够影响最终的预测效果,因此本研究使用十乘交叉验证评估每个超参数下模型的效果,并使用网格搜索策略寻找最佳超参数。RF、超参数网格搜索和交叉验证通过 Scikit-learn (v1.5.0) 库中的 GridSearchCV 模块实现。LightGBM 模型由 Python 包 lightgbm (v4.5.0) 实现。

近年来兴起的人工智能浪潮极大推进了深度学习模型在基因组选择领域中的应用。本研究选择了一款深度学习模型 DNNGP 用于基因组选择的比较。DNNGP 基于卷积神经网络的深度学习基因组选择模型,其包括 1 个输入层、3 个卷积层、1 个归一层、2 个 Dropout 层、

1 个扁平层、1 个全连接层和 1 个输出层组成，通过多个卷积层提取基因型 PCA 中的关键特征，并利用全连接网络对表型进行准确预测^[9]，其中卷积神经网络是 DNNGP 的核心网络模块，卷积输出的结果可表示为：

$$s(t) = (f * k)(t)$$

其中： $s(t)$ 为卷积输出； f 表示卷积核，卷积核在输入特征上移动 k 次，以获得卷积变换的结果^[9]；归一层能够将特征向量进行归一化处理，防止训练过程中梯度爆炸；Dropout 层能够按指定比例丢弃神经元从而减轻模型的过拟合程度。DNNGP 采用十乘交叉验证的策略评估模型训练准确性，并取 10 次准确性的均值作为模型准确性结果输出。本研究对 DNNGP 的输入批次大小、学习率、Dropout 比例和迭代期等超参数进行了网格搜索，保留了最佳超参数下的模型准确性。

2 结果与分析

2.1 表型数据统计 本研究对收集的大白猪总产仔数、产活仔数、健仔数和出生窝重 4 种繁殖表型数据进行描述性统计，其中 4 个繁殖性状的变异系数在 20.90%~25.17%（表 1），变异系数较高，该大白猪群体繁殖性状有较好的育种改良空间。

2.2 遗传参数估计 为了评估该大白母猪群体的遗传背景与繁殖性状遗传特征，本研究使用重复力模型估计其遗传参数。如表 2 所示，该大白母猪群体的总产仔数、产活仔数、健仔数和出生窝重 4 个繁殖性状遗传力范围在 0.04~0.10，为低遗传力性状^[15]。

2.3 基因型填充 本研究通过基因型填充，将 50K 芯片数据扩展到接近重测序覆盖的标记密度。为了保证后续分析位点的质量，根据基因型填充结果文件中 INFO 信息的 R^2 值进行低质量位点的剔除，随后统计染色体剩余位点的填充准确性以及平均准确性（图 1）。在剔除低质量填充位点后，本研究保留了平均填充准确性 R^2 为 0.950 的高质量填充位点用于后续分析（表 3）。

2.4 基因组选择准确性比较 为了评估机器学习模型和深度学习模型的基因组选择预测准确性，本研究选用了

表 2 大白猪繁殖性状遗传参数估计

性状	加性方差	永久环境方差	残差方差	遗传力（标准误）
总产仔数	0.48	0.86	10.71	0.04（0.01）
产活仔数	0.45	0.95	10.41	0.04（0.01）
健仔数	0.87	0.50	9.90	0.08（0.01）
出生窝重	1.55	1.60	11.92	0.10（0.02）

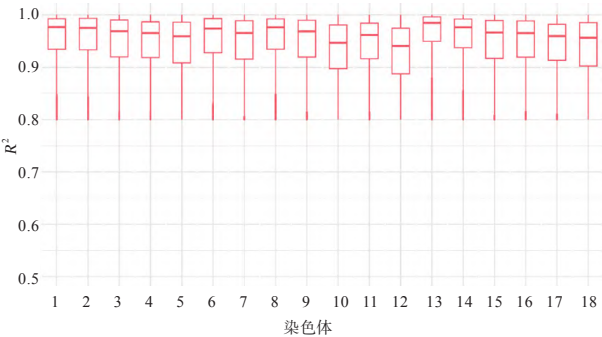


图 1 大白猪芯片数据基因型填充准确性 R^2

表 3 大白猪芯片数据基因型填充染色体平均准确性 R^2

染色体	位点数	平均准确性
1	831668	0.956
2	635152	0.955
3	622878	0.948
4	565763	0.946
5	411076	0.942
6	693923	0.953
7	602274	0.946
8	711429	0.955
9	587682	0.948
10	440565	0.934
11	435168	0.944
12	259114	0.927
13	784853	0.963
14	637485	0.957
15	489175	0.947
16	407978	0.947
17	339095	0.942
18	222798	0.939

2 款机器学习模型 RF 和 LightGBM，以及一款深度学习模型 DNNGP，与常规的线性模型 GBLUP 进行比较。结果发现，GBLUP 的预测准确性为 0.166~0.195，RF

表 1 大白母猪繁殖性状描述性统计

性状	平均值	最小值	最大值	标准差	变异系数，%	记录数
总产仔数，头	16.66	2	36	3.48	20.90	10428
产活仔数，头	15.05	2	27	3.44	22.85	10428
健仔数，头	13.34	2	27	3.36	25.17	10428
出生窝重，kg	17.85	1.50	32.54	3.82	21.40	4810

的预测准确性为 0.123~0.197, LightGBM 的预测准确性为 0.090~0.201, DNNGP 的预测准确性为 0.229~0.257 (图 2)。尽管 2 款机器学习在一部分性状的预测准确性比 GBLUP 更高, 而深度学习模型 DNNGP 比其他 3 款模型观察到了更高的预测准确性, 并且相比于 GBLUP 的预测准确性提升了 0.056~0.079。

除了对模型预测的准确性进行比较, 本研究还评估了模型预测值和观测值之间的误差, 包括均方误差 (Mean Square Error, *MSE*) 和平均绝对误差 (Mean Absolute Error, *MAE*), 以此更全面地比较 4 款模型的性能。如表 4 所示, GBLUP 模型在总产仔数、产活仔数和健仔数中的 *MSE* 比其他 3 个模型高 30.544~70.345, *MAE* 比其他 3 个模型高 2.842~6.337。该结果表明线性模型在低遗传力性状基因组选择中存在局限性, 机器学习和深度学习模型在低遗传力性状基因组选择任务中预测误差较低, 拥有良好的应用前景。

随着测序技术的成熟, 如今可供育种工作者使用的遗传数据越来越多样化, 除了 SNP 芯片数据, 基因型填充数据也在不断推广应用。本研究为了探究深度学习模型在多种类型的遗传数据中的应用, 对芯片数据进行了基因型填充, 并用于深度学习基因组选择的评估。如图 3 所示, GBLUP 的准确性在 0.147~0.202, RF 的预测准确性为 0.011~0.165, LightGBM 模型的预测准确性为 0.044~0.077, DNNGP 的预测准确性在 0.174~0.259。对于机器学习模型 RF 和 LightGBM, 基因型填充使得其预测准确性下降了 0.013~0.186; GBLUP 在产活仔数和总产仔数性状上准确性分别提升

表 4 基于大白猪芯片数据基因组选择均方误差和平均绝对误差

性状	GBLUP	RF	LightGBM	DNNGP
出生窝重	8.448	10.249	7.322	16.549
产活仔数	66.293	7.797	6.866	19.391
健仔数	51.714	4.936	8.257	21.170
总产仔数	76.475	6.130	9.217	20.395
出生窝重	1.973	2.252	2.011	3.301
产活仔数	7.719	2.085	2.000	3.890
健仔数	6.782	1.784	2.155	3.940
总产仔数	8.302	1.965	2.315	3.715

了 0.004 和 0.007, 但在出生窝重和健仔数上分别下降了 0.001 和 0.019; 对于深度学习模型 DNNGP, 基因型填充同样使得模型在出生窝重、产活仔数和总产仔数中准确性分别降低了 0.036、0.055 和 0.033, 但在健仔数性状上提升了 0.014, 并且 DNNGP 在出生窝重、健仔数和总产仔数性状中依然表现最佳。机器学习模型在处理基因型填充数据时性能明显下降, 深度学习模型则能够抵抗基因型填充引入的噪音位点。

比较 4 种模型的预测误差, 由于 GBLUP 只考虑了加性效应, 导致遗传力极低 (小于 0.1) 的产活仔数、健仔数和总产仔数 3 个性状的 *MSE* 和 *MAE* 远高于其他 3 种模型。对于基因型填充后基因组选择的 *MSE* 和 *MAE*, GBLUP 相比于填充前的变化并不明显 (<0.1); 对于 DNNGP, 出生窝重和健仔数的 *MSE* 在基因型填充后分别降低了 5.547 和 5.202, 在出生窝重、产活仔数和健仔数的 *MAE* 分别降低了 1.166、0.517 和 0.917 (表 5)。这些结果表明深度学习模型在面对基因型填充中

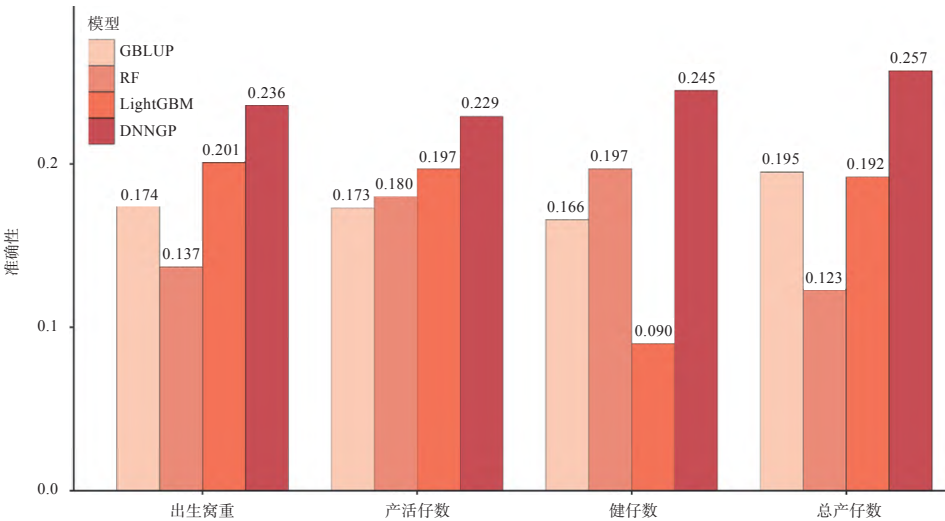


图 2 基于芯片数据的繁殖性状基因组选择准确性比较

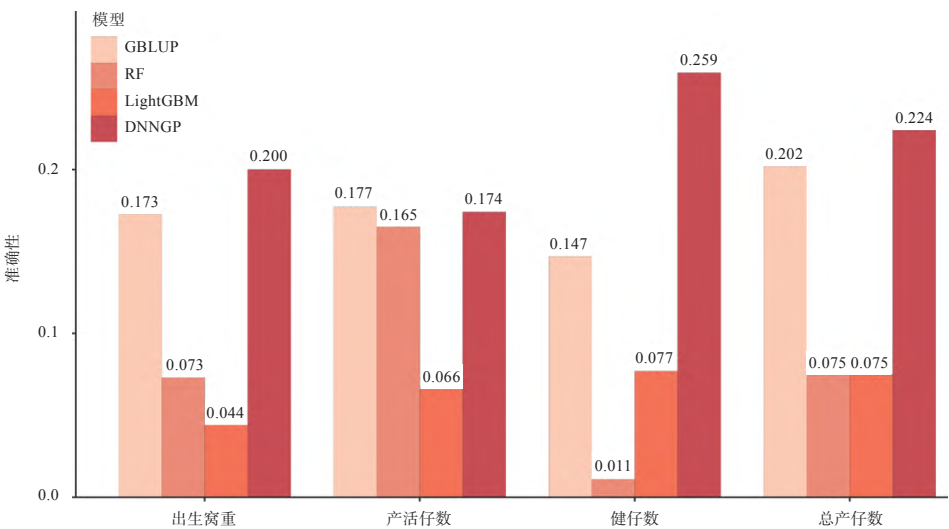


图 3 基于基因型填充数据的大白猪繁殖性状基因组选择准确性比较

表 5 基于大白猪基因型填充数据基因组选择均方误差和平均绝对误差

性状	GBLUP	RF	LightGBM	DNNGP
出生窝重	8.454	12.218	6.031	11.002
产活仔数	66.251	9.804	6.851	19.779
健仔数	51.732	6.237	6.087	15.968
总产仔数	76.413	7.818	8.077	21.250
出生窝重	1.970	2.433	1.718	2.135
产活仔数	7.716	2.274	2.023	3.373
健仔数	6.781	1.806	1.753	3.023
总产仔数	8.300	2.173	2.175	3.825

的复杂遗传信息时能够保持较好的鲁棒性与预测性能。

3 讨论

猪繁殖性状是重要的经济性状，猪群体的繁殖性能对养猪企业的经济效益有重要影响，因此，提升猪繁殖性状基因组选择的准确性尤为重要^[16]。近年来人工智能领域的飞速发展带动了各个领域对深度学习模型的应用，基因组选择领域同样展开了多样的深度学习模型尝试研究，利用诸如卷积神经网络^[17-18]、局部连接网络^[19]、长短期记忆网络^[20]、Transformer^[21]等神经网络架构进行基因组选择评估。然而先前报道的研究大多是基于高遗传力性状的基因组选择，对于猪繁殖性状这样的低遗传力性状研究甚少。因此，本研究选取了先前在作物基因组选择中研究的深度学习模型 DNNGP，与 GBLUP、RF 和 LightGBM 进行基因组选择比较，并分别在芯片和基因型填充数据集中分别评估 4 款模型在猪繁殖性状上的基因组选择准确性和误差。在本研究

几乎所有场景中，深度学习模型 DNNGP 的预测准确性比 GBLUP 更高或基本持平（提升了 -0.003~0.248），并且 DNNGP 在基因型填充数据中的抗噪能力比 RF 和 LightGBM 更强。

对于芯片数据的基因组选择，GBLUP 在总产仔数、产活仔数、健仔数和出生窝重 4 个繁殖性状中预测准确性均低于 0.2，符合低遗传力性状基因组选择的趋势^[22]。机器学习模型 RF 和 LightGBM 的预测准确性并不是在所有性状中都优于 GBLUP，这 2 款机器学习都基于决策树算法，目前该算法依然需要优化训练和提高预测的稳定性^[23]。深度学习模型 DNNGP 基于卷积神经网络对输入的主成分进行重要特征提取，并在多个神经网络处理层中不断学习遗传信息与表型之间的潜在关系，从而获得更准确的预测结果。

在先前的报道中，基因型填充带来了大量的有效位点，但同时也会引入错误的噪音位点，并难以通过质控完全剔除^[24-25]，错误分型会限制基因组选择准确性的提高^[15]。在本研究中，机器学习模型 RF 和 LightGBM 在基因型填充后的数据集中的准确性出现了下降（0.013~0.186），并完全低于 GBLUP 的准确性。表明基于决策树算法的机器学习模型在面对基因型填充这样的噪音数据集时存在一定的局限性，稳定的预测需要更抗噪的算法^[26]。而基于卷积神经网络的深度学习模型 DNNGP 利用深度卷积神经网络和神经元激活函数，识别并提取重要特征，结合梯度下降的算法优化神经元权重，最终实现噪音信息的部分过滤和准确预测^[19, 27]。本研究缺少在验证群体中的预测，而深度学习

模型在真实育种数据中的交叉验证准确性比 GBLUP、RF 和 LightGBM 更具优势, 表明深度学习模型在育种实践中具有良好的应用潜力。

目前深度学习在猪基因组选择中的研究应用较少, 该领域尚处于较初级的发展阶段, 部分种猪育种人员对深度学习模型在育种工作中的使用方法和遗传增益效果评估并没有足够的知识储备和实践经验。随着人工智能领域的快速发展和物联网智能化体系的推广, 生猪产业体系精准育种技术必将朝着智能化方向发展。因此, 探究深度学习模型在猪育种中的应用方法和效果, 能够为提高猪遗传改良速度和加速育种进展提供新的解决思路 and 方案, 为生猪产业向智能化精准育种方向发展提供理论基础和数据支撑。

4 结 论

本研究选用唐人神集团的 957 头大白母猪的总产仔数、产活仔数、健仔数和出生窝重 4 个低遗传力繁殖性状, 结合 SNP 芯片和基因型填充数据进行基因组选择。其中深度学习模型 DNNP 在芯片数据中预测准确性最优, 比 GBLUP、RF 和 LightGBM 3 款模型的预测准确性高 0.032~0.155; 在填充数据集上, 深度学习模型 DNNP 同样在总产仔数、健仔数和出生窝重性状中的预测准确性最优。

参考文献:

- [1] Liu B G, Liu H C, Tu J H, *et al.* An investigation of machine learning methods applied to genomic prediction in yellow-feathered broilers[J]. *Poult Sci*, 2025, 104(1): 104489.
- [2] Wei R, Zhang Z Y, Han H, *et al.* Integrative genomic analysis reveals shared loci for reproduction and production traits in Yorkshire pigs[J]. *BMC Genomics*, 2025, 26(1): 310.
- [3] 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.
- [4] Cordoba S, Balcells I, Castello A, *et al.* Endometrial gene expression profile of pregnant sows with extreme phenotypes for reproductive efficiency[J]. *Sci Rep*, 2015, 5: 14416.
- [5] Tong X K, Chen D, Hu J C, *et al.* Accurate haplotype construction and detection of selection signatures enabled by high quality pig genome sequences[J]. *Nat Commun*, 2023, 14(1): 5126.
- [6] Wang X, Shi S L, Wang G J, *et al.* Using machine learning to improve the accuracy of genomic prediction of reproduction traits in pigs[J]. *J Anim Sci Biotechnol*, 2022, 13(1): 60.
- [7] Zhao L Y, Tang P, Luo J J, *et al.* Genomic prediction with NetGP based on gene network and multi-omics data in plants[J]. *Plant Biotechnol J*, 2025, 23(4): 1190-1201.
- [8] Ma W L, Qiu Z X, Song J, *et al.* A deep convolutional neural network approach for predicting phenotypes from genotypes[J]. *Planta*, 2018, 248(5): 1307-1318.
- [9] Wang K, Abid M A, Rasheed A, *et al.* DNNP, a deep neural network-based method for genomic prediction using multi-omics data in plants[J]. *Mol Plant*, 2023, 16(1): 279-293.
- [10] Purcell S, Neale B, Todd-Brown K, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses[J]. *Am J Hum Genet*, 2007, 81(3): 559-575.
- [11] Wang Z, Zhang Z Y, Chen Z T, *et al.* PHARP: a pig haplotype reference panel for genotype imputation[J]. *Sci Rep*, 2022, 12(1): 12645.
- [12] Delaneau O, Marchini J, Zagury J F. A linear complexity phasing method for thousands of genomes[J]. *Nat Methods*, 2011, 9(2): 179-181.
- [13] Das S, Forer L, Schonherr S, *et al.* Next-generation genotype imputation service and methods[J]. *Nat Genet*, 2016, 48(10): 1284-1287.
- [14] Yin L L, Zhang H H, Tang Z S, *et al.* HIBLUP: an integration of statistical models on the BLUP framework for efficient genetic evaluation using big genomic data[J]. *Nucleic Acids Res*, 2023, 51(8): 3501-3512.
- [15] Sun J C, Wei J L, Pan Y, *et al.* Improving genomic prediction accuracy of pig reproductive traits based on genotype imputation using preselected markers with different imputation platforms[J]. *Animal*, 2025, 19(1): 101387.
- [16] Garcia-Vazquez F A. Artificial intelligence and porcine breeding[J]. *Anim Reprod Sci*, 2024, 269: 107538.
- [17] Liu Y, Wang D L, He F, *et al.* Phenotype prediction and genome-wide association study using deep convolutional neural network of soybean[J]. *Front Genet*, 2019, 10: 1091.
- [18] Gao P F, Zhao H N, Luo Z, *et al.* SoyDNGP: a web-accessible deep learning framework for genomic prediction in soybean breeding[J]. *Brief Bioinform*, 2023, 24(6): bbad349.
- [19] Lee H J, Lee J H, Gondro C, *et al.* DeepGBLUP: joint deep learning networks and GBLUP framework for accurate genomic prediction of complex traits in Korean native cattle[J]. *Genet Sel Evol*, 2023, 55(1): 56.
- [20] Wang C Y, Zhang D, Ma Y X, *et al.* WheatGP, a genomic prediction method based on CNN and LSTM[J]. *Brief Bioinform*, 2025, 26(2): bbaf191.
- [21] Wu C L, Zhang Y Y, Ying Z W, *et al.* A transformer-based genomic prediction method fused with knowledge-guided module[J]. *Brief Bioinform*, 2023, 25(1): bbad438.
- [22] Zhao Z D, Niu Q H, Wu J Y, *et al.* Integrating multi-layered biological priors to improve genomic prediction accuracy in beef cattle[J]. *Biol Direct*, 2024, 19(1): 147.
- [23] Shiralizadeh Z. A review and analysis of a parallel approach for decision tree learning from large data streams[J]. [2025-05-17]. <https://doi.org/10.48550/arXiv.2505.11780>.
- [24] Young A I, Nehzati S M, Benonisdottir S, *et al.* Mendelian

罗盘山猪脂肪间充质干细胞体外培养及生物学特性研究

王仲发^{1,2}, 李旻娟^{1,2}, 刘晏辰¹, 何玉楠^{1,3}, 刘文忠^{2*}, 关伟军^{1*}

(1. 中国农业科学院北京畜牧兽医研究所, 北京 100193; 2. 山西农业大学动物科学学院, 山西太谷 030801;
3. 西北农林科技大学动物科技学院, 陕西杨凌 712100)

摘要: 试验旨在通过体外分离培养罗盘山猪脂肪间充质干细胞 (Adipose-derived Mesenchymal Stem Cells, ADSCs), 探究其生物学特性和诱导分化潜能。采用 I 型胶原酶消化法体外获得罗盘山猪 ADSCs, 通过 RT-PCR 和免疫荧光技术鉴定 ADSCs 表面标志物的表达; 通过生长曲线绘制、划痕试验和克隆团形成试验, 探究其体外增殖趋势和生长发育规律; 通过成骨细胞、成脂细胞和成软骨细胞诱导, 探究其体外分化能力; 通过染色体核型分析, 探究其在体外环境中遗传物质的稳定性。RT-PCR 和免疫荧光结果表明, ADSCs 能够表达 CD29、CD44、CD73、CD90, 不表达与造血干细胞相关的 CD34; 通过胰蛋白酶消化法分离出的 ADSCs, 随着细胞代次的增加, 细胞活力逐渐降低; 细胞划痕和克隆结果表明, 随着代次的升高, 细胞体外增殖能力不断下降, 生长曲线呈“S”型; 体外诱导罗盘山猪 ADSCs, 能够诱导出成骨细胞、成脂肪细胞和成软骨细胞; 染色体核型分析结果表明, 在体外增殖过程中 ADSCs 能够保持稳定的染色体形态, 染色体基因组未发生改变 ($2n=38$)。本研究在体外成功分离出罗盘山猪 ADSCs, 且 ADSCs 具有较强的体外增殖和分化潜能, 研究结果可为罗盘山猪种质资源的保护和保存提供方法、为组织工程学研究提供参考依据。

关键词: 罗盘山猪; 脂肪; 间充质干细胞; 生物学特性

中图分类号: S828.2

文献标识码: A

DOI 编号: 10.19556/j.0258-7033.20250217-06

罗盘山猪因产自重庆罗盘山一带而得名, 具有适应性强、母性好、遗传稳定、抗逆性强、耐粗饲等特点^[1]。

收稿日期: 2025-02-17; 修回日期: 2025-04-22

资助项目: 国家重点研发计划 (2021YFD1200303)

作者简介: 王仲发 (1997-), 男, 甘肃定西人, 硕士研究生, 主要从事干细胞研究, E-mail: 2812539972@qq.com; 李旻娟 (2000-), 女, 山东临沂人, 硕士研究生, 主要从事干细胞研究, E-mail: 2171264303@qq.com

* 通信作者: 刘文忠 (1968-), 男, 山西沁水人, 博士, 教授, 主要从事动物遗传资源的分子评价与种质创新, E-mail: tglwzyc@163.com; 关伟军 (1966-), 男, 北京人, 博士, 教授, 主要从事畜禽遗传资源保存与应用研究, E-mail: guanweijun@caas.cn

但因各种原因导致罗盘山猪数量急剧减少, 在 2021 年, 罗盘山猪被列入《国家畜禽遗传资源品种名录》^[2]。罗盘山猪作为地方优良品种, 种质资源的保护和保存尤为重要, 而干细胞保存作为一种资源保存方式, 在畜禽种质资源保护和保存、组织工程学方面发挥了重要作用^[3]。罗盘山猪的养殖方式主要以农户散养为主, 没有形成集约化、规模化养殖, 保种方式还是以划定保护区、降低近交系数为主, 此种方式年限较长, 需要花费较大人力物力^[4-5]。有研究通过罗盘山猪体细胞克隆技术成功复原出 2 头健康、状态良好的罗盘山猪仔猪, 体细胞保存和克隆技术不仅能够大大缩短育种年限, 还能提高育种效率^[6]。另外体细胞在冷冻环境中能够

imputation of parental genotypes improves estimates of direct genetic effects[J]. Nat Genet, 2022, 54(6): 897-905.

[25] Zheng W G, Ma W L, Chen Z L, et al. DPlmpute: a genotype imputation framework for ultra-low coverage whole-genome sequencing and its application in genomic selection[J]. Adv Sci, 2025, 12(16): e2412482.

[26] Adcock B, Hientzsch B, Narayan A, et al. Hybrid least squares

for learning functions from highly noisy data[J]. [2025-07-03]. <https://doi.org/10.48550/arXiv.2507.02215>.

[27] Xiang T, Li T, Li J L, et al. Using machine learning to realize genetic site screening and genomic prediction of productive traits in pigs[J]. FASEB J, 2023, 37(6): e22961.

(责任编辑: 石丽丹)