

生物启发算法在肉牛分子育种基因组筛选中的应用

叶治兵 马义诚 崔繁荣 袁理星 李红波 *
新疆畜牧科学院畜牧研究所，新疆 乌鲁木齐 830011

摘要：随着农业现代化和食品安全需求的提升，分子育种技术逐渐成为优化畜牧业生产力的重要手段。然而，肉牛基因组数据的庞大规模、基因间复杂的交互作用以及表型与基因型间关联的不明确性，极大地限制了传统分析方法的效率与精准性。为了高效筛选与经济性状相关的关键基因和遗传标记，构建精准的育种值预测模型，本文探索了遗传算法、蚁群算法及粒子群算法等生物启发算法在基因选择、基因调控网络构建和基因组关联分析中的应用。这些方法通过模拟生物行为和群体优化机制，有效提升了基因组筛选的效率和模型预测的准确性，为精准育种的理论与实践提供了新思路。

关键词：生物启发算法；肉牛分子育种；基因组筛选

Application of Bioheuristic Algorithm in Molecular Breeding Genome Screening of Beef Cattle

Ye,Zhibing Ma,Yicheng Cui,Fanrong Yuan,Lixing Li,Hongbo*
Animal Husbandry Institute, Xinjiang Academy of Animal Sciences, Urumqi, Xinjiang, 830011, China

Abstract: With the improvement of agricultural modernization and food safety demand, molecular breeding technology has gradually become an important means to optimize the productivity of animal husbandry. However, the huge scale of beef cattle genome data, the complex interaction between genes and the ambiguity of the association between phenotype and genotype greatly limit the efficiency and accuracy of traditional analysis methods. In order to efficiently screen key genes and genetic markers related to economic traits and build an accurate breeding value prediction model, this paper explores the application of bio-heuristic algorithms such as genetic algorithm, ant colony algorithm and particle swarm optimization algorithm in gene selection, gene regulatory network construction and genome association analysis. These methods effectively improve the efficiency of genome screening and the accuracy of model prediction by simulating biological behavior and population optimization mechanism, and provide new ideas for the theory and practice of precision breeding.

Keywords: Biological heuristic algorithm; Molecular breeding of beef cattle; Genome screening

DOI: 10.62639/sspis22.20250201

引言

在农业现代化和食品安全日益受到重视的背景下，分子育种技术被广泛认为是提升畜牧业生产效率和优化资源利用的重要工具。肉牛育种中，筛选与经济性状密切相关的基因组标记已成为提高育种效率的关键。然而，庞大的基因组数据、复杂的基因间交互以及表型与基因型之间模糊的关联，使传统分析方法在精准性和效率上面临诸多限制。近年来，生物启发算法凭借其卓越的优化能力，为基因组筛选提供了全新的研究路径。遗传算法、蚁群算法和粒子群算法的应用展现了挖掘关键遗传标记和解析基因间复杂关系的潜力，为肉牛精准育种奠定了理论基础，也推动了该领域技术的进一步发展。

一、肉牛分子育种基因组筛选的关键问题与目标

(一) 肉牛分子育种中基因组筛选的关键问题

肉牛分子育种是现代育种技术的重要领域，其核心任务是通过基因组筛选识别出与经济性状

(稿件编号：IS-25-1-1020)

作者简介：叶治兵（1995-），男，汉，新疆乌苏人，研究实习员，主要研究方向为肉牛遗传育种与屠宰加工。
通讯作者：李红波（1980-），男，新疆博乐人，正高级畜牧师，博士在读，研究方向为肉牛遗传育种与饲养管理。
基金项目：国家肉牛牦牛产业技术体系（CARS-37），新疆肉牛产业技术体系XJARS-10-03。

相关的关键基因或遗传标记。然而，由于肉牛基因组的复杂性和表型遗传的多样性，基因组筛选面临以下几个关键问题（表1）：

表1 肉牛分子育种中基因组筛选的关键问题

序号	问题	简要描述
1	基因数量庞大	基因组包含数十万个基因和变异，需处理海量高维数据，筛选具有生物学意义的特征基因是巨大挑战。
2	基因间相互作用复杂	经济性状受多基因交互及环境影响，传统模型难以捕获非线性和多维交互，限制了筛选结果的解释力。
3	表型与基因型关联难以确定	表型与基因型的关系不明确，环境干扰和表型遗传力的不确定性增加了基因组关联分析的复杂性。
4	高噪声和数据质量问题	测序错误、数据缺失和噪声影响筛选结果，需提高数据预处理质量以增强筛选可靠性和准确性。

(二) 基因组筛选的主要目标

尽管基因组筛选面临诸多挑战，但其主要目标十分明确，即从海量基因组数据中筛选出与肉牛重要经济性状密切相关的基因和遗传标记，为精准育种提供科学依据。具体目标如下表2所示：

表 2 基因组筛选的主要目标

序号	目标
1	筛选关键基因和遗传标记
2	构建育种值预测模型
3	提升遗传改良效率
4	解析复杂性状遗传基础
5	促进精准育种发展

二、生物启发算法在肉牛分子育种基因组筛选中的应用

（一）基因选择与特征提取

基因选择与特征提取是肉牛分子育种基因组筛选中的核心环节，该步骤可以从海量的基因组数据中筛选出与目标性状高度相关的特征基因，构建高效的预测模型，助力育种决策的科学化。GA 具有强大的全局搜索能力，在高维空间优化问题上表现优异，被广泛应用于基因特征选择中。遗传算法在模拟自然选择的进化机制后，以“适者生存”为准则，在种群内进行连续迭代优化，逐步筛选出具有生物学意义的特征基因。具体而言，GA 首先会将基因组中的潜在特征基因表示为个体编码，采用二进制编码以指示基因是否被选择；随后，以初始化种群并定义适应度函数来评估个体的表现优劣，该适应度函数可以基于所筛选基因构建的预测模型的准确性，如基于表型与基因型的相关性或机器学习算法的预测性能。在每一代中，种群会通过选择操作保留高适应度的个体，再以交叉和变异操作探索新的解空间，以此逐步逼近全局最优解。在肉牛分子育种中，GA 能够有效减少基因组数据中的冗余信息，筛选出生长速度、肉质性状等能够解释目标性状遗传变异的关键基因，同时显著提高预测模型的精度。与传统的单因素基因选择方法相比，GA 能够更高效地处理基因间复杂的非线性交互关系，特别是在基因间可能存在协同作用的情况下，GA 可利用整体优化策略去避免漏选潜在的关键基因。此外，GA 还可以动态组合优化基因特征，进一步提高数据处理效率，减轻计算负担，与现代育种中越来越庞大的基因组数据规模高度适配。

（二）基因调控网络构建

基因调控网络构建是揭示基因间复杂关系和理解肉牛复杂性状遗传基础的关键手段，而生物启发算法在该领域的应用为基因调控网络的构建提供了新颖且高效的方法。基因调控网络的目的是刻画基因之间的调控关系和功能交互，在识别网络拓扑结构后，解析基因的核心调控节点和调控模式。ACO 就是一种典型的生物启发算法，拥有强大的全局优化能力和动态适应性，特别适用于高维、复杂的网络优化问题。在构建基因调控网络的过程中，ACO 模拟蚂蚁觅食路径选择的机制，利用信息素的累积和更新，引导解的搜索方向，从而构建具有生物学意义的网络结构。ACO 首先可将基因对的潜在调控关系表示为问题空间中的路径，每条路径的权重可以基于基因表达数据的相关性或互信息计算。初始阶段，算法为所有路径分配均匀的信息素，随后虚拟蚂蚁利用随

机漫步和基于启发函数的选择构建候选网络，并依据目标函数评估每一代的网络质量。优质网络路径上的信息素被逐渐强化，吸引更多蚂蚁选择这些路径，达到逐步优化网络拓扑结构的目的。

在肉牛分子育种的应用中，ACO 的动态搜索机制可以挖掘基因间隐藏的调控关系，特别是在生长性状、繁殖能力或抗病性状这一类复杂性状的遗传研究中，ACO 能够经网络结构优化识别关键的调控基因节点以及与目标性状密切相关的调控模块。相比传统的基因调控网络构建方法，ACO 的优势体现在以下几个方面：（1）能够有效应对大规模基因组数据中的高维非线性特征，以过分布式信息素更新机制提升网络搜索的鲁棒性；（2）拥有自适应搜索策略，可在早期阶段广泛探索解空间，避免陷入局部最优；（3）引入信息素蒸发机制后，可以防止解空间的过度搜索，从而在有限计算资源下提高算法效率和解的质量。

（三）基因组关联分析

全基因组关联分析（Genome-Wide Association Studies, GWAS）是肉牛分子育种研究中的重要方法，可以揭示基因组遗传变异与重要经济性状之间的关联，为精准育种提供理论基础。然而，GWAS 面临着高维数据处理和复杂模型参数优化的双重困境，PSO 这一类生物启发算法，在此过程中展现出显著优势，PSO 可模拟鸟群觅食行为的群体搜索机制，以种群中粒子的位置和速度动态优化目标函数，在全局范围内寻找最优解。具体到 GWAS 的应用中，PSO 主要用于优化关联分析模型的参数配置，提高模型对目标性状相关遗传变异的检测能力。线性混合模型（Linear Mixed Model, LMM）就是 GWAS 中常用的方法之一，但很大程度上会受到参数选择的影响；PSO 可组合初始化种群的随机参数，并依据目标函数对种群进行迭代优化，逐步调整模型参数，最终提高模型的适配性和泛化能力。

PSO 应于肉牛经济性状研究中，可以通过全局搜索找到最优参数组合，增强了模型对表型复杂性和基因型非线性效应的解释力；经个体间的信息共享与协同优化，PSO 还可以同时处理大量单核苷酸多态性（Single Nucleotide Polymorphisms, SNPs），有效降低了因多重检验带来的假阳性率。与传统的参数优化方法相比，PSO 的显著优势在于它的高效性和鲁棒性：PSO 在初期的随机搜索覆盖广阔的解空间，而后在局部搜索阶段集中精力优化优质解的周围区域，避免了陷入局部最优问题。此外，PSO 的动态更新机制和全局收敛特性也使其在高维数据中表现尤为突出，为挖掘稀有变异和检测复杂性状的多基因遗传基础提供了有力支持。

三、应用案例分析

（一）案例背景

某肉牛分子育种中的核心需求为通过筛选和分析与经济性状相关的遗传变异，优化育种决策。然而，面对数十万甚至上百万个 SNP 位点和复杂的表型数据，传统方法在处理效率和结果准确

性上存在明显不足。本案例采用生物启发算法, 通过遗传算法选择关键基因, 利用粒子群算法优化关联分析模型, 期望提高基因组筛选效率和模型预测准确性。该项目包括 50 万肉牛基因型数据以及对应的生长速率、背膘厚、肌肉评分等多种表型测量数据, 这些数据经过严格质量控制, 确保了后续分析的可靠性。

(二) 算法实施过程

数据预处理是算法实施的第一步, 对数据质量进行了严格控制, 研究中剔除了缺失率超过 10% 的 SNP 位点, 以主成分分析 (PCA) 校正种群结构对关联分析的潜在干扰。

在算法核心部分, 遗传算法 (GA) 用于筛选与目标表型相关的关键 SNP 位点, 将每个基因位点编码为二进制个体, 并定义适应度函数, 该函数依据表型预测模型的准确性评估基因选择的优劣。种群初始化后, GA 通过模拟交叉和变异操作生成新个体, 在每一代中动态优化种群, 从而逐步逼近最优基因集合。

粒子群算法 (PSO) 用于优化 GWAS 中的线性混合模型参数配置, 在 PSO 的实现中, 每个粒子代表一个模型参数组合, 其适应度由关联分析结果的 $-\log_{10}(P\text{-value})$ 得分决定。PSO 通过迭代更

新粒子位置和速度, 动态调整模型参数组合, 使 GWAS 具有更强的统计效能和遗传变异检测能力。

(三) 主要成果

本研究通过生物启发算法获得了多项重要成果, 显著提升了基因组筛选效率和模型性能。在基因选择方面, 遗传算法筛选出 35 个显著关联 SNP 位点, 解释表型遗传变异的比例从传统分析的 12% 提升至 25%。筛选出的 SNP 与文献中已知的相关基因高度一致, 例如与肌肉发育相关的 MYOG 基因和脂肪代谢相关的 PPAR γ 基因, 进一步验证了 GA 的科学性和实用性。

在模型优化方面, 粒子群算法提升了线性混合模型的适配性, 使得显著位点数量增加 15%, 假阳性率从 8.2% 降至 6.4%。此外, PSO 优化后的模型均方误差 (MSE) 降低了 15%, 表型预测能力显著增强。这些成果说明, GA 和 PSO 在肉牛分子育种的基因组筛选和模型优化方面具有显著优势, 为后续育种决策提供了坚实依据。

表 3 PSO 优化前后线性混合模型的性能指标

性能指标	优化前	优化后
显著位点数量	45	52
假阳性率	8.2%	6.4%
MSE	0.120	0.102

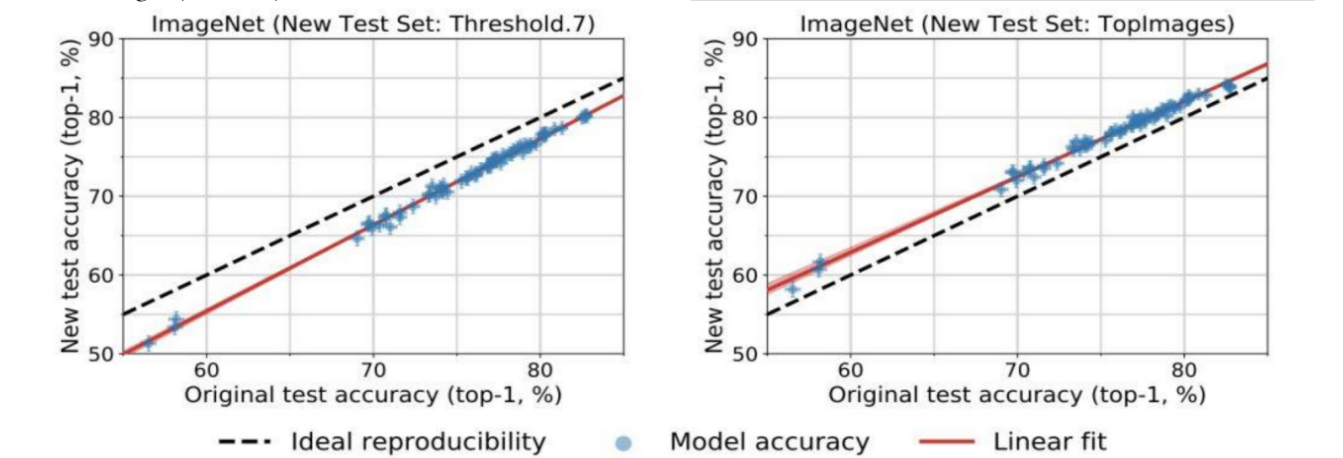


图 1 遗传算法优化前后预测模型的性能对比

上图 1 展示了优化前后模型的 ROC 曲线, 优化后的 AUC 值从 0.82 提升至 0.94, 表明模型预测能力显著提高。

四、结论

生物启发算法在肉牛分子育种基因组筛选中的应用, 不仅为解决传统方法在高维数据处理和复杂性状解析方面的瓶颈问题提供了新的思路, 也为精准育种的发展开辟了广阔的路径。从遗传算法在基因选择中的高效性, 到蚁群算法在基因调控网络构建中的适用性, 再到粒子群算法在基因组关联分析中的卓越表现, 这些方法展示了强大的优化能力和广泛的适用潜力。通过这些技术的应用, 研究者得以高效筛选出与经济性状相关的关键基因, 优化预测模型, 提升遗传改良效率, 为肉牛育种的科学决策提供了坚实的数据支撑。然而, 尽管生物启发算法展示了诸多优势, 其在

实际应用中仍面临一定的挑战。例如, 高维数据环境下的计算复杂度、模型过拟合的风险、算法参数调整的复杂性问题, 都需要进一步探索与优化。同时, 随着基因组数据规模的持续扩大, 如何在提升算法性能的同时降低计算成本, 也是未来研究的关键方向。值得一提的是, 结合深度学习、并行计算等前沿技术, 进一步提升算法的智能化和效率, 可能为基因组筛选的实践带来更大的突破。

参考文献:

[1] 赵学健, 赵可. 基于遗传算法的生物启发频繁项集挖掘策略 [J]. 计算机科学, 2023, 50 (S2): 636-643.
[2] 王殿伟, 刘旺, 房杰, 许志杰. 受生物视觉启发的无人机航拍低照度图像增强算法 [J]. 西北工业大学学报, 2023, 41 (01): 144-152.
[3] 姚红娟, 华翔, 王海, 李宝华. 一种生物启发的无线断网网络数据传输算法 [J]. 西安工业大学学报, 2021, 41 (02): 219-225.