



王坤,张元庆,陈卓,刘彦杰,王泽昭,李博,程景,牛晓艳.“太行云牛”群体遗传结构与基因组选择信号分析[J]. 中国农业大学学报,2025,30(08):96-106.
WANG Kun, ZHANG Yuanqing, CHEN Zhuo, LIU Yanjie, WANG Zezhao, LI Bo, CHENG Jing, NIU Xiaoyan. Population genetic structure and genomic selection signals analysis in “Taihang Yun” cattle[J]. *Journal of China Agricultural University*, 2025,30(08):96-106.
DOI: 10.11841/j.issn.1007-4333.2025.08.09

“太行云牛”群体遗传结构与基因组选择信号分析

王坤¹ 张元庆¹ 陈卓¹ 刘彦杰¹ 王泽昭² 李博¹ 程景¹ 牛晓艳^{1*}

(1. 山西农业大学 动物科学学院, 山西 太谷 030801;
2. 中国农业科学院 北京畜牧兽医研究所, 北京 100193)

摘要 为研究“太行云牛”群体遗传结构,筛选与“太行云牛”重要经济性状相关的基因组选择信号区域及候选基因,本研究利用“育芯1号”牛200 K液相芯片,分别对113头“太行云牛”和26头晋南牛进行了SNP检测、群体遗传结构分析、连锁不平衡衰减分析;应用 F_{ST} 和 θ_{π} 方法筛选“太行云牛”中受选择的基因组信号区域,并通过GO和KEGG注释挖掘潜在的候选基因。结果显示:1)对芯片数据进行质控后共获得187 988个SNP位点,主成分分析和邻接树结果均显示“太行云牛”和晋南牛可清楚的分为2支,且2个群体均表现出连锁不平衡随遗传距离增加逐渐衰减的规律,“太行云牛”的连锁不平衡程度高于晋南牛。2)采用ADMIXTURE进行祖先成分分析,当 $K=3$ 时,结果最接近真实的品种祖先成分,交叉验证误差值最小。3)2种分析方法共筛选到618个在“太行云牛”中受选择的位点,涉及257个基因,其中包括2个与毛色相关基因(*KIT*和*MITF*),2个与脂肪沉积和肉嫩度相关基因(*SOAT2*和*CYP2R1*),4个与繁殖过程相关基因(*EIF4B*和*MAP3K1*和*IFNAR2*和*TGF β 2*),4个与抗病及应激相关的基因(*PDGFRA*和*FGF6*和*KDR*和*RPS6KA6*)。综上,本研究采用 F_{ST} 和 θ_{π} 基因组选择信号分析在两群体基因组中筛选到618个受选择区域和257个受选择基因,这些选择信号涉及到毛色、脂肪代谢、繁殖、抗病、生长发育等生物学过程。本研究为“太行云牛”优良性状的分子机理提供了重要遗传信息,鉴定到的受选择位点和基因也为后续该品种分子育种和选种提供数据依据。

关键词 “太行云牛”; 基因芯片; 遗传结构; 选择信号; 候选基因

中图分类号 S823 文章编号 1007-4333(2025)08-0096-11 文献标志码 A

Population genetic structure and genomic selection signals analysis in “Taihang Yun” cattle

WANG Kun¹, ZHANG Yuanqing¹, CHEN Zhuo¹, LIU Yanjie¹, WANG Zezhao², LI Bo¹,
CHENG Jing¹, NIU Xiaoyan^{1*}

(1. College of Animal Sciences, Shanxi Agricultural University, Taigu 030801, China;
2. Institute of Beijing Husbandry and Veterinary Medicine, Chinese Academy of Agricultural Sciences, Beijing 100193, China)

Abstract The purpose of this paper is to explore the genetic structure of “Taihang Yun” cattle, and to select important genomic selection signals and candidate genes related with economic traits in “Taihang Yun” cattle. This study utilized the “YuXin-1” bovine 200 K liquid phase SNP chip to conduct SNP genotyping, population genetic structure analysis, and linkage disequilibrium decay analysis in 113 “Taihang Yun” cattle and 26 Jinnan cattle; F_{ST} and θ_{π} methods were applied combinedly to select the genomic signal areas, also

收稿日期: 2025-01-04

基金项目: 山西省关键核心技术攻关子项目(NYGG22-02); 和顺县现代肉牛产业发展研究院项目(XDHZHSY2024-03); 山西省科技重大专项计划“揭榜挂帅”项目(202201140601026)

第一作者: 王坤(ORCID:0009-0005-2448-8308), 硕士研究生, E-mail: wk19834340420@163.com

通讯作者: 牛晓艳(ORCID:0000-0002-5036-0157), 研究员, 主要从事动物遗传育种研究, E-mail: nxysxau2022@163vip.com

the potential candidate genes were explored through GO and KEGG annotations. The results showed that : 1) After quality control of the chip data, a total of 187 988 SNPs loci were obtained. Principal component analysis and NJ tree analysis both clearly divided “Taihang Yun” cattle and Jinnan cattle into two clades, which both exhibited a gradual decay of linkage disequilibrium with increasing distance. The degree of linkage disequilibrium in “Taihang Yun” cattle was higher than that in Jinnan cattle. 2) ADMIXTURE was used for ancestral component analysis, and when $K=3$, the result was closest to the real ancestral content, with the lowest cross validation value. 3) Two analysis methods detected 618 selected sites in “Taihang Yun” cattle, including 257 genes, which contained 2 genes related with coat color (*KIT*, *MITF*), 2 genes related with fat deposition and meat tenderness (*SOAT2*, *CYP2R1*), 4 genes related with reproduction (*EIF4B*, *MAP3K1*, *IFNAR2*, *TGF β 2*), 4 genes related with disease resistance and stress (*PDGFRA*, *FGF6*, *KDR*, *RPS6KA6*). In conclusion, F_{ST} and θ_{π} method were applied in both populations and detected 618 selected regions and 257 genes, which involved biological processes as coat color, fat metabolism, reproduction, disease resistance and growth and development. The results provided essential genetic information to molecular mechanisms of excellent traits in “Taihang Yun” cattle, the sites and candidate genes identified also provided data bases for molecular breeding and selection.

Keywords “Taihang Yun” cattle; DNA chip; genetic structure; selection signal; candidate genes

肉牛作为一种节粮型家畜,在保障畜产品有效供给,丰富居民膳食结构方面发挥着重要作用^[1]。近年来,牛肉在我国居民肉类消费中的比重不断增加,并且仍呈现不断上升趋势^[2]。肉牛养殖在山西省历史悠久、分布广泛、数量众多,牛肉产量从2013年5.20万t增长到2020年的7.37万t^[3]。

“太行云牛”的核心产区在和顺县,是在中国西门塔尔牛太行类群的基础上,经黄牛改良、级进杂交和横交选育等多个阶段,历时40余年而形成的专门化肉牛品种。相较晋南牛,“太行云牛”具有体型高大、适应性强、耐粗饲、饲料转化率高、产肉率高、适合山区放牧等特点,目前已成为山西省供港牛肉的首选,在和顺地区推广面积较大,农户饲养积极性很高^[4]。该品种不仅适应舍饲、放牧、半舍饲半放牧等多种饲养管理模式,同时能适应不同地域的风土气候和资源条件,于2011年入选成为国家地理标志保护产品^[5]。2024年,山西农业大学动物科学学院在北京联合大学旅游学院首都餐饮文化研究发展基地举办了“太行云牛”牛肉品鉴会,专家组一致认为该牛肉整体达到国家标准规定的特级胴体S级分割肉质量水平,与美国肉牛胴体等级标准最优质量水平相当,具备生产开发中高档牛肉产品的潜力^[6]。“十四五”以来,国家高度重视并大力发展民族种业,2021年农业农村部发布《全国畜禽遗传改良计划2021—2035》,提出要支持培育群体规模大,特色优势明显的地方品种。目前,山西省农业厅已开展育种联合攻关项目对“太行云牛”进行深入研究

和开发,希望挖掘与相关的基因区段和功能位点,并从基因组层面解析其优势性状的发生机制。

基因组选择信号(Selection signature)包括“选择性清除”和“搭便车”效应^[7],它们都是选择作用在动物基因组上留下的明显特征,不仅可探索物种进化的分子机制,而且可寻找与生物学功能或表型之间存在相关性的基因组区域。目前,在动物基因组水平上研究选择作用和寻找受选择的区段(基因)已成为遗传学上的一个热点。Zhang等^[8]基于30头皮南牛的全基因组测序数据和世界范围内已发表的169头牛的基因组数据,利用核苷酸多态性比值(θ_{π})、固定指数(F_{ST})以及CLR和XP-EHH分析方法,筛选到与免疫(*BOLA-DQA2*、*BOLA-DQB*、*LSM14A*、*SEC13*和*NAALADL2*)、生长性状(*CYP4A11*、*RPL26*和*MYH10*)、胚胎发育(*REV3L*、*NT5E*、*CDX2*、*KDM6B*和*ADAMTS9*)、无角性状(*CH21orf62*)和气候适应(*ANTXR2*)相关基因。Song等^[9]利用30头夏南牛的全基因组测序数据和178头已发表的牛基因组数据,利用 θ_{π} 和CLR相结合以及 F_{ST} 和XP-EHH相结合的分析方法,筛选出可能影响生长发育(*WSCD1*、*TMEM68*、*MFN1*和*NCKAP5*)、肉质性状(*MCCC1*)、和免疫力(*IL11RA*、*CNTFR*、*CCL27*、*SLAMF1*、*SLAMF7*、*NAA35*和*GOLM1*)的基因。鲍麒麟等^[10]基于野牦牛(*Bos mutus*)和天祝白牦牛(*Bos grunniens*)的基因组重测序数据,利用GCTA软件获得SNP变异位点,然后利用 F_{ST} 和Tajima'D法分析2个牦牛群体的选择信号分布情况。Zhang等^[11]采用 F_{ST} 和 θ_{π} 方法

对南阳牛和欧洲商业牛品种进行全基因组选择信号分析,揭示了2个品种中与生殖性能和免疫相关的基因中存在显著的选择信号,并筛选出可能与繁殖(*MAP2K2*、*PGR*和*GSE1*)、免疫反应(*NCOA2*、*HSF1*和*PAX5*)和嗅觉(*TAS1R3*)相关的基因。

作为一个地方培育品种,“太行云牛”经历了长期的人工选择,为更详细和深入了解该类群形成过程中品种的遗传多样性、基因组特征、改良驯化过程等,本研究采用晋南牛作为对照品种,利用“育芯1号”200 K牛液相芯片,通过全基因组选择信号分析,从分子层面揭示了“太行云牛”和晋南牛的遗传差异,为新品种的选育、开发利用和未来分子育种的实施提供理论依据和数据参考。

1 材料与方法

1.1 样品采集

26头晋南牛颈静脉血样采集自运城市国家级晋南牛遗传资源基因保护中心,113头“太行云牛”颈静脉血样分别采集自山西银河湾农牧科技有限公司和和顺县绿和生态农牧业开发有限公司。尽量选择3~5岁彼此间无亲缘关系的成年母牛,保定后采用4#针头采集颈静脉血样5 mL,EDTA抗凝处理后低温下运送至实验室,−20℃冷冻保存备用。

1.2 DNA提取和基因芯片分型

基因组DNA采用MinibestUnibest®动物血液基因组提取试剂盒提取(OMEGA公司,广州),提取完毕后,采用1.0%琼脂糖凝胶电泳检测基因组片段的完整性,采用Nanodrop 2000检测DNA样本的纯度($OD_{260}/OD_{280}=1.8\sim 2.0$),采用Quibit核酸蛋白定量仪(Thermo Fisher, Invitrogen公司)精确测定DNA浓度。

基因组DNA打断成长度为200 bp左右的片段,经过末端修复后,连接测序接头,探针杂交后获得目标区域捕获文库进行测序,采用北京康普森农业科技有限公司开发的“育芯1号”牛育种液相芯片对采集的139头个体进行片段杂交、捕获及重测序。原始数据经过去接头、过滤低质量序列等质控步骤后得到Clean data,将Clean data与参考基因组*Bos taurus* ARS-UCD1.2进行比对后通过GCTA软件^[12]的Haplotype caller模块进行变异检测,并从中获取目标位点的分型结果。最后采用ANNOVAR软件^[13]进行变异位点注释。

采用PLINK(Version 1.9)^[14]对芯片数据进行质控。质控标准为:①只使用常染色体上的位点;②SNP检出率(Call rate) $\geq 90\%$;③个体检出率 $\geq 90\%$;④最小等位基因频率(Minimum allele frequency, MAF) ≥ 0.01 ;⑤哈代-温伯格(Hardy weinberg) $P \leq 1.0 \times 10^{-6}$ 。质控后剩余的SNPs用于后续分析。

1.3 遗传结构分析

为了解晋南牛和“太行云牛”2个群体间的遗传关系和群体结构,使用GCTA软件分别计算2个群体的亲缘关系矩阵(Genetic relationship matrix, GRM),之后采用PLINK(Version 1.9)进行主成分分析,采用R语言ggplot2程序包实现可视化。

采用邻接法构建常染色体的系统发育树,使用Perl语言自编脚本将数据转化后输入MEGA软件^[15](Version 7.0),然后使用FigTree(Version 1.4.3)对发育树进行可视化。

使用ADMIXTURE软件^[16](Version 1.3.0)进行晋南牛和“太行云牛”祖先成分的分析,该软件从多位点SNP基因型数据集中对祖先成分进行最大似然分析,使用更加快速的数值优化算法计算祖先成分。

1.4 连锁不平衡分析

利用PopLDdecay(Version 3.4)软件^[17]计算两群体SNP位点间的连锁不平衡,计算随着SNP位点间距离的增加,连锁不平衡 r^2 值的衰减程度,采用R(Version 4.0.2)绘制连锁不平衡衰减图。

1.5 基因组选择信号分析

1.5.1 F_{ST} 检验

根据Weir等^[18]提出的统计学方法,以晋南牛群体为对照,“太行云牛”群体为试验群体,群体间遗传分化指数(Fixation index, F_{ST})、核苷酸多样性(Nucleotide diversity, π)均由VCFtools(Version 0.1.16)^[19]基于芯片数据集计算,窗口大小为100 kb,滑动步长为10 kb。具体参数:--fst-window-size 100 000 --fst-window-step 10 000。采用Origin软件(Version 2021)绘制2个群体不同区域的 F_{ST} 值分布图。遗传分化指数 F_{ST} 的计算公式如下:

$$F_{ST} = \frac{H_T - H_S}{H_T} \quad (1)$$

式中:ST为群体间变异数; H_S 是各亚群自由交配情

况下群体内的平均杂合度; S 为群体内变异数; H_T 是种群中自由交配情况下的杂合度; T 为总变异数。

1.5.2 θ_π Ratio 检验

同样利用 VCFtools 计算滑动窗口内的 θ_π 值,窗口大小同样设置为 100 kb,滑动步长设置为 10 kb,具体参数为 --window-pi 100 kb --window-pi-step 10 000。其中两群体 θ_π 比率 (θ_π Ratio) 的计算公式为:

$$\theta_\pi \text{ Ratio} = \frac{\theta_{\pi \text{ 太行云牛}}}{\theta_{\pi \text{ 晋南牛}}} \quad (2)$$

选择同时具有显著低 θ_π Ratio 和显著高 θ_π Ratio (分别为 0.5% 左尾和 0.5% 右尾) 以及显著高 F_{ST} 值 (1% 右尾) 的窗口作为沿基因组具有强选择性扫描信号的区域。采用 R 语言中的 merge 函数合并并筛选 F_{ST} 和 θ_π Ratio 前 1% 的重叠区域,并作为后续功能基因注释和富集分析的基础。2 种方法获得的受选择的信号区域均分别采用 R 语言中 CMplot 包绘制曼哈顿图。

1.6 基因注释及功能的富集分析

基于参考基因组 *Bos taurus* ARS-UCD1.2, 使用 snpEFF 软件^[20] 对前 1% 重叠区域的受选择基因进行功能注释。使用 DAVID (Version 2023) (<https://david.ncifcrf.gov/home.jsp>) 进行基因本体 (Gene Ontology, GO) 功能富集和京都百科全书 (Kyoto encyclopedia of genes and genomes) 通路富集分析,以 $P < 0.05$ 作为显著性基因和通路的筛选标准。

2 结果与分析

2.1 数据的变异检测

139 头个体基因组 DNA 无明显降解与蛋白质污染,浓度和纯度均符合建库要求。Raw data 中,每个样本获得数据 6.95~15.03 G,经过滤后的 Clean data 每个样本获得数据 6.9~14.97 G,比对有效率达到 99% 以上,GC 含量在 43.18%~44.93%。通过 BWA 软件将 Clean data 比对到参考基因组上,比对率在 99% 以上。在“太行云牛”和晋南牛中共鉴定到 189 067 个 SNPs,经 PLINK 质控后获得 187 988 个 SNPs。对于两群体 SNP 位点分布的分析显示,大多数 SNPs (40.1%) 位于基因间区域,其中少部分的 SNPs 分别位于开放阅读框的上游 (5.4%)、下游 (5.5%) 以及基因的非翻译区 (0.7%),其余 SNPs 位于外显子 (11%) 和内含子区 (37.2%) (图 1)。

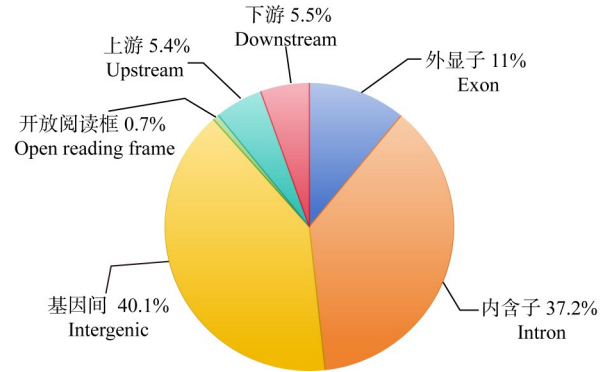


图 1 2 个群体共 139 头个体基因组 SNP 注释

Fig. 1 SNP annotation of 139 individuals of two cattle populations

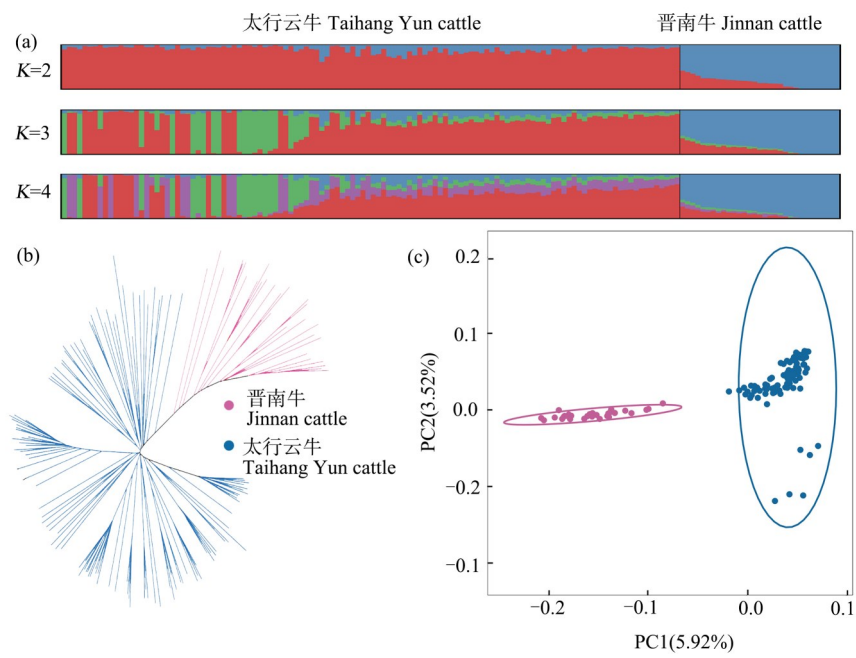
2.2 群体遗传结构分析

“太行云牛”为西门塔尔牛与本地黄牛多代级进杂交后选育形成的群体,而晋南牛为地方品种,两者在体型、毛色、角型和产肉性能方面存在较大差异。基于 2 个群体的 SNP 信息,采用 PCA 主成分分析解析“太行云牛”和晋南牛群体之间的遗传结构。结果如图 2(c) 所示,其中第一主成分和第二主成分分别解释了 5.92% 和 3.52% 的总变异,可直观看出 2 个群体各自聚为一类,基本无杂交现象,这也与两品种 (群体) 培育历程和所处的地理位置相一致,并为后续的选择信号分析奠定了基础。采用基因组中非连锁位点构建邻接树,进一步阐释 2 个品种的遗传结构。NJ 树的结果 (图 2(b)) 与 PCA 结果一致,两者互为印证。从结果可知,晋南牛与“太行云牛”较为明显的分为两支。

遗传结构分析中,最后采用祖先成分分析对晋南牛和“太行云牛”进行了研究。结果如图 2(a) 所示。结果显示,当 $K=2$ 时,显示晋南牛和“太行云牛”具有 2 种祖先类型。当 $K=3$ 时,结果最接近真实的品种祖先成分,交叉验证误差值最小,这也与“太行云牛”培育过程中使用了晋南牛、西门塔尔牛和红白花荷斯坦牛一致。当 $K=4$ 时,结果与 $K=3$ 相似,可能与“太行云牛”培育过程中采用了多系别西门塔尔牛有关。

2.3 连锁不平衡衰减分析

使用 PopLDdecay 软件分别计算了晋南牛和“太行云牛”的连锁不平衡程度,2 个群体均表现出随着位点之间距离的增加连锁不平衡程度 r^2 降低的规律 (图 3)。在 0~150 kb,“太行云牛”的连锁不平衡程度高于晋南牛,在 150 kb 之后,晋南牛的连锁不平衡程度高于“太行云牛”。



(a)“太行云牛”和晋南牛品种祖先成分分析(ADMIXTURE);(b)139头牛的邻接系统发育树,不同颜色表示不同品种;(c)使用芯片数据对2个品种进行PCA分析。
(a)The ancestor mixture analysis of “Taihang Yun” cattle and Jinnan cattle ; (b) Adjacent phylogenetic tree of 139 individuals, with different breeds displayed in different colors; (c) PCA analysis of two cattle breeds.

图2 “太行云牛”和晋南牛2群体的遗传结构分析

Fig. 2 Genetic structure of “Taihang Yun” cattle and Jinnan cattle

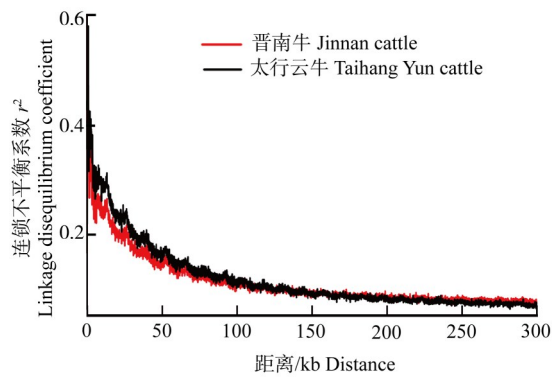


图3 “太行云牛”和晋南牛的连锁不平衡衰减分析

Fig. 3 Linkage disequilibrium of “Taihang Yun” cattle and Jinnan cattle

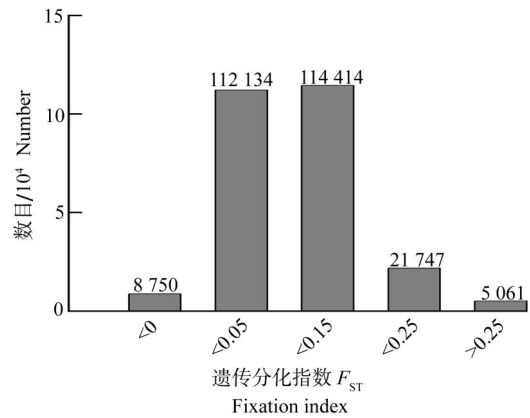


图4 “太行云牛”和晋南牛两群体间 F_{ST} 值经验分析

Fig. 4 The empirical distribution of F_{ST} between “Taihang Yun” cattle and Jinnan cattle

2.4 基于群体分化的 F_{ST} 检验分析

采用VCFtools对质控后获得的187 988个SNP位点进行计算,得到2个群体之间每个窗口区域的 F_{ST} 值,采用Origin软件绘制了晋南牛和“太行云牛”2个群体间每个区域的 F_{ST} 值(图4)。结果显示, $F_{ST}>0.05$ 的区域141 222个, $F_{ST}>0.25$ 的区域5 061个,表明两群体基因组之间存在一定程度的遗传分化。

全基因组水平上 F_{ST} 基因组选择信号的曼哈顿图(Manhattan plot)如图5(a)所示。全基因组水平上取 F_{ST} 值的前1%为阈值,共鉴定出受选择区域2 622个,经功能注释获得895个受选择基因。对受选择基因进一步进行GO功能注释,发现生物学过程(Biological process, BP)显著富集到磷酸化、RNA聚

合酶 II、细胞内信号转导、细胞迁移、干细胞增殖等条目。分子功能(Molecular function, MF)显著富集到 ATP 结合、钙依赖性磷脂酶 A2 活性、降钙素受体结合、转录抑制因子活性等条目。细胞组分(Cellular component, CC)显著富集到细胞质、胞质囊泡、质膜等条目。KEGG 通路分析富集到 10 条显著通路,主要集中在 Ras 信号通路、cAMP 信号通路,PI3K-Akt 信号通路上。

2.5 基于核苷酸位点差异的 θ_π Ratio 检验分析

计算芯片上每一个 SNP 位点的 θ_π Ratio 值后,选择前 1% 的区域作为受选择窗口,共筛选到

2 618 个受选择区域(图 5(b)),经注释获得 1 029 个受选择基因,GO 功能富集结果显示,其中分子功能(BP)显著富集到蛋白结合、钙粘着蛋白结合、转录共调节因子活性等条目;细胞组分(CC)显著富集到连环蛋白复合体、细胞质、肌动蛋白细胞骨架等条目;生物过程(MF)显著富集到化学突触传递、多细胞生物发育、信号转导等条目。KEGG 通路富集分析共获得 31 个显著富集信号通路,主要包括神经活性配体-受体相互作用、孕激素介导的卵母细胞成熟、PI3K-Akt 信号通路、黑素瘤等信号通路。

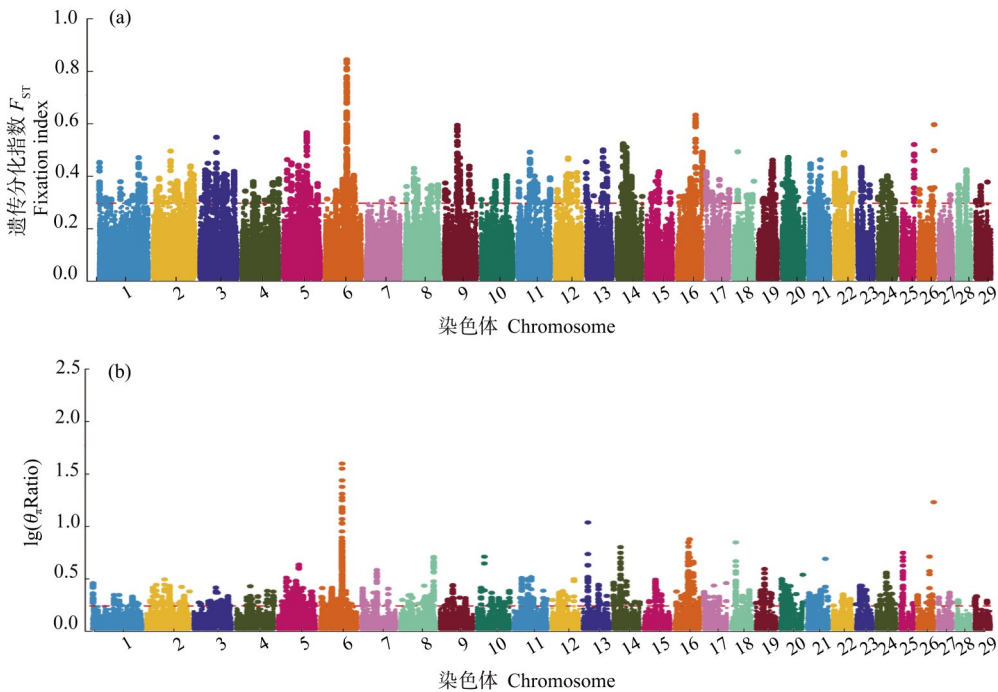


图 5 F_{ST} (a)和 θ_π Ratio(b)2 种方法获得的基因组选择信号曼哈顿图
Fig. 5 Manhattan plot of genomic selection signature of F_{ST} (a) and θ_π Ratio (b)

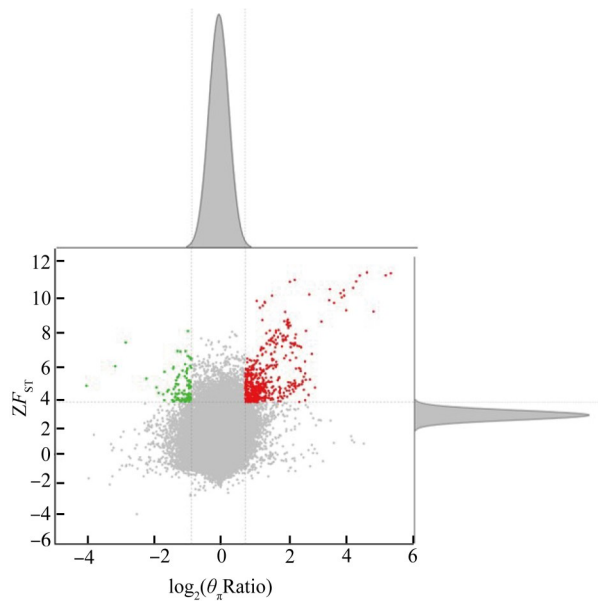
2.6 2 种选择信号方法重叠位点

本研究中,采用 F_{ST} 法选择前 1% 为阈值($F_{ST} \geq 0.2910$), θ_π Ratio 选择头部 0.5% 和尾部 0.5% 为阈值($\theta_\pi \text{ Ratio} \geq 1.6701$),同时达到上述 2 个阈值的窗口被认为是受选择区域,窗口内的位点为受选择位点(图 6),结果共获得 618 个基因组区段,包含 257 个基因。GO 和 KEGG 富集分析发现,选择信号较强烈的区域包含一些与牛毛色相关基因(如: KIT ($F_{ST}=0.32$, $\theta_\pi \text{ Ratio}=4.31$))、胚胎发育相关基因(如: $GATA3$ ($F_{ST}=0.26$, $\theta_\pi \text{ Ratio}=4.01$))、免疫调节和细胞炎症反应相关基因(如: $TNFSF12$ ($F_{ST}=$

0.31 , $\theta_\pi \text{ Ratio}=3.68$))。

2.7 功能基因的注释

使用 DAVID 软件对 F_{ST} 和 θ_π Ratio 2 种方法所获得的共 257 个重叠的受选择基因进行 GO 富集分析和 KEGG 功能注释。共富集到 21 个显著的 GO 条目($P < 0.05$), 7 条显著的 KEGG 通路($P < 0.05$)。其中,富集最显著的 GO 条目为 GO: 0005737 即 Cytoplasm(细胞质),该生物学过程涉及到的基因共 20 个(图 7, 表 1); 最显著的 KEGG 通路是 bta04145: PI3K-Akt 信号通路,该通路涉及到的基因为 11 个(表 2)。



灰色区域代表未受到选择性清除的区域;红色和绿色代表受到选择性清除的区域。 F_{ST} 已标准化并转换为 ZF_{ST} 。
Grey area represents the area that is not selectively swept, and the red and green represent the area that is selectively swept. F_{ST} is standardized and converted to ZF_{ST} .

图6 F_{ST} 和 θ_π Ratio联合分析
Fig. 6 Combination analysis of F_{ST} and θ_π Ratio

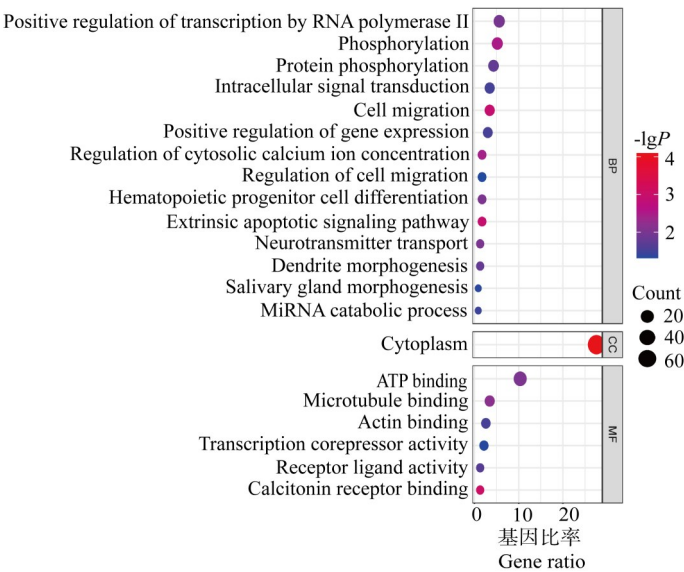


图7 选择信号显著富集的GO条目
Fig. 7 Significantly enriched GO term of selection signature

3 讨论

“太行云牛”属山西省地方培育品种,具有一系列独特的优良特性,利用选择信号的方法对其进行研究有助于更深入和精准的开发利用这一地方培育品种。本研究采用全基因组SNP芯片数据,利用

基于群体分化的 F_{ST} 方法和基于核苷酸多样性的 θ_π Ratio方法分析了“太行云牛”群体基因组上的受选择信号,共发现618个受到选择的区域和257个受选择的基因。进一步的GO和KEGG注释发现,部分候选基因与“太行云牛”的毛色、免疫、繁殖、泌乳以及肉质相关。

表 1 显著富集的 GO 条目及相关基因

Table 1 Significantly enriched GO term and related genes

编号 ID	功能注释 Function annotation	P 值 P value	基因 Gene
0005737	细胞质	0.000 080	<i>AK2</i> 、 <i>MICAL2</i> 、 <i>RNF19B</i> 、 <i>ECRG4</i> 、 <i>CYP2R1</i> 、 <i>POLR2A</i> 、 <i>SH3KBP1</i> 、 <i>HMGB3</i> 、 <i>CDK10</i> 、 <i>FGF23</i> 、 <i>GAS8</i> 、 <i>ZC3H12A</i> 、 <i>UBE2Q2</i> 、 <i>MSRB3</i> 、 <i>KALRN</i> 、 <i>RGS20</i> 、 <i>MC1R</i> 、 <i>PIK3C3</i> 、 <i>PDCL2</i> 、 <i>TNNI3K</i>
0097191	外源性凋亡信号通路	0.001 120	<i>TGFB2</i> 、 <i>TNFSF12</i> 、 <i>KRT18</i>
0051480	胞质钙离子浓度调节	0.003 825	<i>CALCA</i> 、 <i>OC407218</i> 、 <i>TP2B2</i>
0002244	造血干细胞分化	0.009 280	<i>DNAI4</i> 、 <i>PDGFRA</i> 、 <i>KIT</i> 、 <i>DOCK1</i>
0031716	降钙素受体结合	0.000 870	<i>CALCA</i> 、 <i>OC407218</i>
0003779	肌动蛋白结合	0.029 410	<i>SPIRE2</i> 、 <i>ICAL2</i> 、 <i>SPN</i> 、 <i>NNT2</i>
0006468	蛋白质磷酸化	0.017 193	<i>KIT</i> 、 <i>PDGFRA</i> 、 <i>RPS6KA6</i>

表 2 显著富集的 KEGG 通路及相关基因

Table 2 Significantly enriched KEGG pathway and related genes

编号 ID	功能注释 Function annotation	P 值 P value	基因 Gene
bta04151	PI3K-Akt 信号通路	0.006 291	<i>KIT</i> 、 <i>PDGFRA</i> 、 <i>FGF6</i> 、 <i>PIK3CD</i> 、 <i>IKBKB</i> 、 <i>G6PC3</i> 、 <i>KDR</i> 、 <i>ITGB7</i> 、 <i>EIF4B</i> 、 <i>IFNAR2</i> 、 <i>FGF23</i>
bta05218	黑素瘤	0.006 670	<i>PDGFRA</i> 、 <i>MITF</i> 、 <i>FGF6</i> 、 <i>PIK3CD</i> 、 <i>FGF23</i>
bta04973	碳水化合物消化吸收	0.009 932	<i>PIK3CD</i> 、 <i>SLC2A5</i> 、 <i>G6PC3</i> 、 <i>ATP1B3</i>
bta04010	MAPK 信号通路	0.011 083	<i>KIT</i> 、 <i>PDGFRA</i> 、 <i>RPS6KA6</i> 、 <i>TGFB2</i> 、 <i>FGF6</i> 、 <i>IKBKB</i> 、 <i>MAP3K1</i> 、 <i>KDR</i> 、 <i>FGF23</i>
bta04014	Ras 信号通路	0.012 661	<i>KIT</i> 、 <i>PDGFRA</i> 、 <i>RASAL2</i> 、 <i>FGF6</i> 、 <i>PIK3CD</i> 、 <i>IKBKB</i> 、 <i>KDR</i> 、 <i>FGF23</i>
bta05202	癌症中的转录失调	0.015 602	<i>SS18</i> 、 <i>MITF</i> 、 <i>DDX5</i> 、 <i>RUNX2</i> 、 <i>SUPT3H</i> 、 <i>PLAT</i> 、 <i>ITGB7</i>
bta00100	类固醇生物合成	0.017 193	<i>SOAT2</i> 、 <i>CYP2R1</i> 、 <i>MSMO1</i>

本研究通过质控,共获得 187 988 个高置信的 SNP 位点,通过 PCA 聚类分析,发现“太行云牛”和晋南牛明显分为 2 个类群,这说明尽管“太行云牛”是在晋南牛基础上培育出的,但经过 40 余年的选育,两者在遗传基础上已经出现了一定程度的差异。连锁不平衡是指位于某一座位的特定等位基因与另一座位的等位基因同时出现的概率大于群体中因随机分布的 2 个等位基因同时出现的概率,通常采用相关系数(r^2)来衡量品种/群体的连锁不平衡程度^[21]。通过对两群体的连锁不平衡分析可看出,在 0~150 kb, r^2 值衰减速度“太行云牛”高于晋南牛,在 150~300 kb,两者 r^2 值几乎重合,晋南牛略高于“太行云牛”。这说明在本地黄牛中引入欧洲牛血缘后,加速了其衰减速度。连锁不平衡程度也可以用于评估畜禽品种(群体)的人工干预和驯化程度,衰减速度越慢,说明该群体驯化程度越高,

选择强度越大^[22]。本研究中,晋南牛 r^2 值低于“太行云牛”,说明“太行云牛”群体的选育程度低于本地品种,这也与两群体的培育历史相吻合。杨伟杰^[23]对 14 头海南黄牛进行全基因组重测序,通过与现代普通牛、印度瘤牛、中国瘤牛 3 个类群的对比分析,发现海南黄牛的 r^2 值低于其他 3 个类群。常铤晋^[24]采集了四川地区 27 头川驴样本进行全基因组重测序,并结合云南驴和西藏驴、关中驴、滚沙驴以及埃塞俄比亚驴、西班牙驴和澳大利亚驴 9 个品种的重测序数据进行分析,发现川驴和西藏驴的 r^2 值最低,澳大利亚驴和滚沙驴的 r^2 值最高。

相比晋南牛,“太行云牛”的毛色发生了极大的改变,由枣红色演变为头部,下腹部和尾帚为白色,躯体为红白花或黄白花^[4]。毛色是动物品种的显著标识之一,受到皮肤色素沉积、饲料和环境的影响,其中,黑色素生成激素受体(Melanocyte-stimulating hormone,

MSH)在调控黑素细胞中真黑素和棕黑素生成的过程中发挥重要作用^[25]。在本研究中,筛选到与毛色相关的2个基因 *KIT*和 *MITF*,它们分别与动物的黑白斑点^[26]以及白色毛色^[27]相关,并已在日本褐牛^[28]、意大利肉牛和奶牛^[29]以及舟山牛^[30]中进行研究。相比晋南牛,“太行云牛”的肉质更加鲜嫩多汁,属于专门化肉用品种。在“太行云牛”受选择区域中也筛选到2个与脂肪沉积及肉质性状密切相关的基因,如:*SOAT2*基因在牛的肝脏细胞中表达,对于维持胆固醇的正常水平,保证肝脏的正常代谢功能,调控牛体内脂肪的合成、储存和分解过程具有重要作用^[31]。*CYP2R1*基因则能够催化维生素D₃产生骨化二醇。犍牛 *CYP2R1*表达水平的降低可能导致维生素D₃的积累,有助于改善牛肉的嫩度^[32]。除此之外,还筛选到4个与牛繁殖相关的基因。如:*EIF4B*和 *MAP3K1*与产奶量和乳脂率有关^[33-34]; *IFNAR2*与胎儿与母体的信号识别及妊娠维持有关^[35]; *TGFB2*则与哺乳动物睾丸发育和精子发生有关,该基因通过PI3K-Akt信号通路调控睾丸支持细胞、精原干细胞的增殖和存活来间接促进精子的发生^[36]。其余还筛选到与哺乳动物免疫、疾病、应激响应相关的基因,如 *PDGFRA*^[37]、*FGF6*^[38]、*KDR*^[39]和 *RPS6KA6*^[40]。上述基因在未来“太行云牛”的遗传改良和群体选育中都应被重点关注。

4 结 论

本研究以晋南牛作为对照群体,利用“育芯1号”200 K牛液相SNP芯片对“太行云牛”进行了群体遗传结构和选择信号分析,发现两群体分离明显,并且“太行云牛”的连锁不平衡强度高于晋南牛。采用 F_{ST} 和 θ_{π} 基因组选择信号分析在两群体基因组中筛选到618个受选择区域和257个受选择基因,这些选择信号涉及到毛色、脂肪代谢、繁殖、抗病、生长发育等生物学过程。这些基因有助于深入认识“太行云牛”品种形成的遗传基础,为后续开发利用和标记辅助选择提供科学依据。

参考文献 References

- [1] 张翔飞,王之盛,孟庆翔,万发春,张海波,邹华国. 我国肉牛养殖在发展阶段粮型畜牧业中的贡献和发展潜力[J]. 中国畜牧杂志, 2013, 49(18): 12-16
Zhang X F, Wang Z S, Meng Q X, Wan F C, Zhang H B, Zou H W. Contribution and development potential of beef cattle breeding in developing

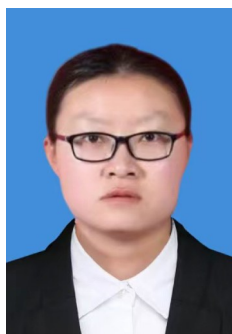
- grain-saving animal husbandry in China[J]. *Chinese Journal of Animal Science*, 2013, 49(18): 12-16 (in Chinese)
- [2] 韩磊. 中国肉类供需形势及稳产保供对策研究[J]. 价格理论与实践, 2020(7): 57-61
Han L. Study on situation of meat supply and demand and countermeasures of ensuring stable production and supply in China[J]. *Price: Theory & Practice*, 2020(7): 57-61 (in Chinese)
- [3] 赵晓强,张元庆. 山西省肉牛产业链发展状况研究[J]. 中国牛业科学, 2022, 48(3): 69-72, 76
Zhao X Q, Zhang Y Q. Study on the development of beef cattle industry chain in Shanxi Province[J]. *China Cattle Science*, 2022, 48(3): 69-72, 76 (in Chinese)
- [4] 山西省农业科学院畜牧兽医研究所,和顺县人民政府. 太行云牛[M]. 北京: 中国农业科学技术出版社, 2019
Animal Husbandry and Veterinary Research Institute of Shanxi Academy of Agricultural Sciences, Heshun County People's Government. *Taihang Yun Cattle*[M]. Beijing: China Agricultural Science and Technology Press, 2019 (in Chinese)
- [5] 王坤,杨效民,刘彦杰,李博,张元庆,牛晓艳. 西门塔尔牛太行类群生长发育性状的遗传参数估计[J]. 中国畜禽种业, 2024, 20(10): 7-16
Wang K, Yang X M, Liu Y J, Li B, Zhang Y Q, Niu X Y. Genetic parameter estimation of growth and development traits of Simmental cattle Taihang Group[J]. *The Chinese Livestock and Poultry Breeding*, 2024, 20(10): 7-16 (in Chinese)
- [6] 王薇,彭宗璐. 和顺肉牛产业走上新质发展道路[N]. 中国食品报, 2024-11-22(2)
Wang W, Peng Z L. Heshun Beef industry embark on a new path of high-quality development [N]. *Chinese Food Newspaper*, 2024-11-22(2) (in Chinese)
- [7] 潘章源,贺小云,刘秋月,郭晓飞,曹晓涵,胡文萍,狄冉,王翔宇,储明星. 全基因组测序(WGS)在畜禽群体进化和功能基因挖掘中的应用[J]. 农业生物技术学报, 2016, 24(12): 1945-1954
Pan Z Y, He X Y, Liu Q Y, Guo X F, Cao X H, Hu W P, Di R, Wang X Y, Chu M X. Application of whole genome sequencing(WGS) in population evolution and gene discovery of domestic animals [J]. *Journal of Agricultural Biotechnology*, 2016, 24(12): 1945-1954 (in Chinese)
- [8] Zhang S J, Yao Z, Li X M, Zhang Z J, Liu X, Yang P, Chen N B, Xia X T, Lyu S J, Shi Q T, Wang E Y, Ru B R, Jiang Y, Lei C Z, Chen H, Huang Y Z. Assessing genomic diversity and signatures of selection in Pinan cattle using whole-genome sequencing data[J]. *BMC Genomics*, 2022, 23(1): 460
- [9] Song X Y, Yao Z, Zhang Z J, Lyu S J, Chen N B, Qi X S, Liu X, Ma W D, Wang W S, Lei C Z, Jiang Y, Wang E Y, Huang Y Z. Whole-genomic sequencing reveals genomic diversity and selection signatures in Xia'nan cattle [J]. *BMC Genomics*, 2024, 25(1): 559
- [10] 鲍麒,梁春年,郭宪,李吉叶,褚敏,阎萍. 全基因组选择信号解析野牦牛和天祝白牦牛的遗传差异[J]. 基因组学与应用生物学, 2022, 41(4): 731-741
Bao Q, Liang C N, Guo X, Li J Y, Chu M, Yan P. Analysis of genetic differences between wild yak and Tianzhu white yak by genome-wide selection signals [J]. *Genomics and Applied Biology*, 2022, 41(4): 731-741 (in

- Chinese)
- [11] Zhang Y, Wei Z T, Zhang M, Wang S W, Gao T Y, Huang H T, Zhang T L, Cai H F, Liu X, Fu T, Liang D. Population structure and selection signal analysis of Nanyang cattle based on whole-genome sequencing data[J]. *Genes (Basel)*, 2024, 15(3): 351
- [12] Yang J, Lee S H, Goddard M E, Visscher P M. GCTA: A tool for genome-wide complex trait analysis[J]. *The American Journal of Human Genetics*, 2011, 88(1): 76-82
- [13] Wang K, Li M Y, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data[J]. *Nucleic Acids Research*, 2010, 38(16): e164
- [14] 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, Sham P C. PLINK: A tool set for whole-genome association and population-based linkage analyses [J]. *American Journal of Human Genetics*, 2007, 81(3): 559-575
- [15] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary genetics analysis across computing platforms[J]. *Molecular Biology and Evolution*, 2018, 35(6): 1547-1549
- [16] Montana G, Hoggart C. Statistical software for gene mapping by admixture linkage disequilibrium[J]. *Briefings in Bioinformatics*, 2007, 8(6): 393-395
- [17] Zhang C, Dong S S, Xu J Y, He W M, Yang T L. PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files[J]. *Bioinformatics*, 2019, 35(10): 1786-1788
- [18] Weir B S, Cockerham C C. Estimating F-statistics for the analysis of population structure[J]. *Evolution*, 1984, 38(6): 1358-1370
- [19] Danecek P, Auton A, Abecasis G, Albers C A, Banks E, DePristo M A, Handsaker R E, Lunter G, Marth G T, Sherry S T, McVean G, Durbin R, 1000 genomes project analysis group. The variant call format and VCFtools [J]. *Bioinformatics*, 2011, 27(15): 2156-2158
- [20] Cingolani P. Variant annotation and functional prediction: SnpEff[J]. *Methods in Molecular Biology*, 2022, 2493: 289-314
- [21] 潘章源, 贺小云, 王翔宇, 郭晓飞, 曹晓涵, 胡文萍, 狄冉, 刘秋月, 储明星. 家养动物选择信号研究进展[J]. *遗传*, 2016, 38(12): 1068-1078
- Pan Z Y, He X Y, Wang X Y, Guo X F, Cao X H, Hu W P, Di R, Liu Q Y, Chu M X. Selection signatures in domesticated animals [J]. *Hereditas*, 2016, 38(12): 1068-1078 (in Chinese)
- [22] 梁艳, 顾以韧, 刘鹏亮, 郝晓东, 弥世荣, 陶璇, 杨雪梅, 杨跃奎, 王言, 陈晓晖, 龚建军, 吕学斌, 何志平. 川乡黑猪基因组选择信号分析[J]. *中国畜牧杂志*, 2024, 60(12): 164-168
- Liang Y, Gu Y R, Liu P L, Hao X D, Mi S R, Tao X, Yang X M, Yang Y K, Wang Y, Chen X H, Gong J J, Lv X B, He Z P. Analysis of genome selection signal of Chuanxiang black pig[J]. *Chinese Journal of Animal Science*, 2024, 60(12): 164-168 (in Chinese)
- [23] 杨伟杰. 海南黄牛群体遗传多样性及全基因组选择信号研究[D]. 海口: 海南大学, 2023
- Yang W J. Study on genetic diversity and genome-wide selection signal of Hainan yellow cattle population[D]. Haikou: Hainan University, 2023(in Chinese)
- [24] 常铤晋. 中国西南驴全基因组遗传多样性与选择信号分析[D]. 杨凌: 西北农林科技大学, 2023
- Chang T J. Genetic diversity and selection signal analysis of the whole genome of China southwest donkey [D]. Yangling: Northwest A&F University, 2023 (in Chinese)
- [25] Klungland H, Vage D I, Gomez-Raya L, Adalsteinsson S, Lien S. The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination[J]. *Mammalian Genome*, 1995, 6(9): 636-639
- [26] Charon K M, Lipka K R. The effect of a coat colour-associated genes polymorphism on animal health: A review [J]. *Annals of Animal Science*, 2015, 15(1): 3-17
- [27] Brenig B, Beck J, Floren C, Bornemann-Kolatzki K, Wiedemann I, Hennecke S, Swalve H, Schütz E. Molecular genetics of coat colour variations in White Galloway and White Park cattle[J]. *Animal Genetics*, 2013, 44(4): 450-453
- [28] Matsumoto H, Kojima M, Takamuku H, Kimura S, Kashimura A, Imai S, Yamauchi K, Ito S. MC1R c.310G>- and c.871G>A determine the coat color of Kumamoto sub-breed of Japanese Brown cattle[J]. *Animal Science Journal*, 2020, 91(1): e13367
- [29] Fontanesi L, Scotti E, Russo V. Analysis of SNPs in the *KIT* gene of cattle with different coat colour patterns and perspectives to use these markers for breed traceability and authentication of beef and dairy products[J]. *Italian Journal of Animal Science*, 2010, 9(2): 217-221
- [30] Jiang L H, Kon T, Chen C Y, Ichikawa R, Zheng Q Y, Pei L Y, Takemura I, Nsobi L H, Tabata H, Pan H, Omori Y, Ogura A. Whole-genome sequencing of endangered Zhoushan cattle suggests its origin and the association of MC1R with black coat colour[J]. *Scientific Reports*, 2021, 11(1): 17359
- [31] Marshall S M, Gromovsky A D, Kelley K L, Davis M A, Wilson M D, Lee R G, Crooke R M, Graham M J, Rudel L L, Mark Brown J, Temel R E. Acute sterol o-acyltransferase 2 (SOAT2) knockdown rapidly mobilizes hepatic cholesterol for fecal excretion[J]. *PLoS One*, 2014, 9(6): e98953
- [32] Stroia C M, Ghitea T C, Vranceanu M, Muresan M, Bimbo-Szuha E, Pallag C R, Pallag A. Relationship between vitamin D3 deficiency, metabolic syndrome and *VDR*, *GC*, and *CYP2R1* gene polymorphisms[J]. *Nutrients*, 2024, 16(9): 1272
- [33] Chen B, Chen Y H, Rai K R, Wang X F, Liu S S, Li Y Y, Xiao M, Ma Y, Wang G Q, Guo G J, Huang S L, Chen J L. Deficiency of eIF4B increases mouse mortality and impairs antiviral immunity [J]. *Frontiers in Immunology*, 2021, 12: 723885
- [34] Wang J J, Kimura E, Mongan M, Xia Y. Genetic control of MAP3K1 in eye development and sex differentiation[J]. *Cells*, 2022, 11(1): 34
- [35] López-Bielma M F, Falfán-Valencia R, Abarca-Rojano E, Pérez-Rubio G. Participation of single-nucleotide variants in *IFNAR1* and *IFNAR2* in the immune response against SARS-CoV-2 Infection: A systematic review[J]. *Pathogens*, 2023, 12(11): 1320
- [36] La Y F, Ma X M, Bao P J, Chu M, Guo X, Liang C N, Yan P. Identification and profiling of microRNAs during yak's testicular development[J]. *BMC Veterinary Research*, 2023, 19(1): 53
- [37] Guérit E, Arts F, Dachy G, Boulouadnine B, Demoulin J B. PDGF receptor

- mutations in human diseases[J]. *Cellular and Molecular Life Sciences*, 2021, 78(8): 3867-3881
- [38] Armand A S, Laziz I, Chanoine C. FGF6 in myogenesis[J]. *Biochimica et Biophysica Acta: Molecular Cell Research*, 2006, 1763(8): 773-778
- [39] Cui Y N, Zhang P P, Liang X, Xu J L, Liu X Y, Wu Y M, Zhang J L, Wang W, Zhang F, Guo R H. Association of KDR mutation with better clinical outcomes in pan-cancer for immune checkpoint inhibitors[J]. *American Journal of Cancer Research*, 2022, 12(4): 1766-1783
- [40] Sun Y, Cao S, Yang M, Wu S, Wang Z, Lin X, Song X, Liao D J. Basic anatomy and tumor biology of the RPS6KA6 gene that encodes the p90 ribosomal S6 kinase-4[J]. *Oncogene*, 2013, 32(14): 1794-1810
- 责任编辑: 秦梅



第一作者简介: 王坤,山西农业大学动物科学学院畜牧学专业硕士研究生。2023—2024学年获得山西农业大学硕士三等学业奖学金,2024—2025学年获得山西农业大学优秀研究生荣誉称号。



通讯作者简介: 牛晓艳,博士,山西农业大学动物科学学院研究员,硕士研究生导师,山西省畜禽遗传资源委员会特养分会委员。主要从事肉牛遗传育种与新品种培育,动物功能基因挖掘等相关研究。主持山西省农业关键核心技术攻关子项目、山西省自然科学基金面上项目,山西省留学回国人员科研项目,中央引导地方科技发展资金项目等。以第一作者发表SCI论文3篇,核心期刊论文15篇,参与制订地方标准3项,获山西省科学技术进步二等奖1项(第5完成人),获科技成果9项。