文章编号:1000-1638(2026)05-0449-12 DOI:10.13484/j.nmgdxxbzk.20260501

新型简化基因组测序技术的研发及其在蒙古牛生长性状中的应用*

裴栋超,海 超,宋丽爽,刘雪霏,白春玲,苏广华,杨 磊,李光鹏
(内蒙古大学生命科学学院省部共建草原家畜生殖调控与繁育国家重点实验室,呼和浩特 010070)

摘要:蒙古牛(*Bos taurus*)是我国北方地区的本土黄牛品种,具有耐寒、耐粗饲、抗逆性强等优良特性,是我国重要的畜种资源。目前,简化基因组测序(Reduced-representation genome sequencing,RRGS)已广泛应用于分子标记辅助育种,但该技术利用酶切法来简化基因组,操作复杂且对DNA质量要求高。本研究根据蒙古牛基因组中的高频碱基设计了简化基因组测序引物组,采用直接PCR策略实现基因组简化,并利用该引物组对384头蒙古牛进行简化测序,结合关联分析挖掘与蒙古牛生长性状相关的功能基因。结果共筛选出4个与体重显著相关的SNP位点和*METTL14*、*PRSS12*等8个候选基因;体高性状和体斜长性状的关联分析结果相似度较高,共筛选出2个SNP位点和*IRS1*、*RHBDD1*等5个候选基因。本研究构建了一种新型简化基因组测序技术,为蒙古牛基因组选择育种提供了技术与理论支持。

关键词:简化基因组测序技术;蒙古牛;单核苷酸多态性;关联分析
中图分类号:S813 **文献标志码**:A

蒙古牛起源于蒙古高原地区,目前主要分布于我国的内蒙古地区、东北地区和华北地区,是我国北方地区的主要黄牛品种之一[1]。蒙古牛具有极强的抗逆性和良好的遗传稳定性,但其生长速度缓慢,产肉性能较低,在肉乳生产性能方面难以满足现代化养殖的需求[2]。为改善蒙古牛的生产性能和经济性状,各地先后引进多种国外优秀肉牛品种进行杂交改良[3]。经过多年选育,杂交肉牛在生长性状和经济性状方面显著优于本土蒙古牛。但大规模的杂交导致本土纯种蒙古牛种群急剧萎缩,许多优良的抗逆性状也因外源基因渗入而逐渐退化。

二代测序技术(Next-generation sequencing,NGS)凭借其高通量和低成本的优势,为动植物育种领域带来了革命性突破[4]。简化基因组测序是一种基于NGS的基因组分析方法,包括SLAF-seq、GBS、RAPD-seq等[5-7]。这些方法通过选择性测序基因组的部分片段来降低测序成本和基因组复杂度,目前已广泛应用于群体遗传结构解析、选择信号检测和分子标记辅助育种等领域[8]。然而,该技术依赖于限制性内切酶,对DNA质量要求较高,且建库过程较为复杂,难以在大量动植物样本的育种检测中广泛使用。

本研究以蒙古牛为研究对象,创新性地开发了一种基于直接聚合酶链式反应(Polymerase chain reaction,PCR)的简化基因组测序技术。结合关联分析挖掘与蒙古牛生长性状相关的分子标记和功能基因,为基因组选择育种提供理论依据和技术支持。

* **收稿日期**:2025-09-23;**修回日期**:2025-12-28
基金项目:内蒙古自治区科技计划项目(2025KYPT0101)
作者简介:裴栋超(2000—),男,山西忻州人,2025级博士研究生。主要从事动物生殖生物技术与生物育种研究。E-mail:pdc102035@163.com
通信作者:杨 磊(1987—),男,内蒙古呼和浩特人,教授,博士。主要从事动物生殖生物技术与生物育种研究。E-mail:mrknowall@126.com

1 材料与方法

1.1 动物材料

本研究所用的蒙古牛饲养于内蒙古自治区呼和浩特市和林格尔县的内蒙古大学草原家畜种质创新与繁育基地。试验全程严格遵循国家畜牧行业标准和动物伦理规范实施数据采集。所有受试牛群在标准化饲养管理体系下进行统一饲喂,日粮配方、饲喂频次、环境温/湿度控制等参数一致。试验期间经兽医临床监测,确认群体健康状态稳定,未出现任何可能影响试验结果的异常生理指标或临床症状,从而有效控制环境变量对研究数据的干扰,保障数据可重复性与结论可信度。

1.2 试验试剂

试验使用的主要试剂见表1。

表1 试验所用化学试剂

Table 1 Reagents used in the experiment

试剂名称	货号	试剂品牌
血液基因组DNA提取试剂盒	DP348-03	Tiangen
普通琼脂糖凝胶DNA回收试剂盒	DP209-03	Tiangen
Phanta Max Super-Fidelity DNA Polymerase	P505-d3	Vazyme
VAHTS DNA Clean Beads	N411-02	Vazyme
Green Taq Mix	P131-03	Vazyme

1.3 表型采集

试验对象为384头蒙古牛,其中包括181头母牛和203头公牛。采集其生长性状相关表型数据,具体包括体重、体高和体斜长性状(表2)。

表2 试验群体表型数据统计

Table 2 Phenotypic data analysis in experimental populations

性状	最大值	最小值	平均值	标准差
体重/kg	399.61	270.20	335.20	37.50
体高/cm	122.90	104.10	113.50	5.46
体斜长/cm	137.40	121.10	130.31	5.82

1.4 基因组DNA提取

利用尾静脉采血法对384头蒙古牛进行血液采集,每头采集5 mL,放入含有2%乙二胺四乙酸(EDTA)的采血管中,并于−20 ℃冰箱保存,备用。使用血液基因组DNA提取试剂盒从血液样本中提取DNA,具体操作步骤按照说明书进行。基因组DNA提取完成后,利用紫外分光光度计检测其浓度和纯度,检测合格后于−20 ℃冰箱保存,备用。

1.5 简化基因组测序引物合成

按照引物设计原则,设计简化基因组测序技术体系引物。根据蒙古牛基因组中的高频碱基,设计5对用于简化蒙古牛基因组的测序引物(表3),引物由生工生物公司合成。将所有的正向引物混合,配制成引物混合物Mix-F,随后将所有的反向引物混合,配制成引物混合物Mix-R,由此得到两种引物混合物,以备后续试验使用。

1.6 蒙古牛基因组简化

以蒙古牛基因组DNA为模板,利用配置好的Mix-F和Mix-R,采用直接PCR的策略对384头蒙古牛基因组进行简化。PCR反应体系(50 μL)为:3 μL模板DNA,25 μL 2×Phanta Max Buffer,

1 μL dNTP Mix，2 μL 引物 Mix-F（10 μmol/L），2 μL 引物 Mix-R（10 μmol/L），1 μL Phanta Max Super-Fidelity DNA Polymerase，16 μL RNase-free ddH_2O。PCR反应条件为：预变性，95 ℃，5 min；结合扩增（简化基因组测序引物与蒙古牛高频碱基结合），首先 95 ℃，1 min，然后 35 ℃，1 min，最后 72 ℃，1.5 min，此步骤进行5个循环；富集扩增（与高频碱基结合的序列进行富集），首先 95 ℃，1 min，然后 60 ℃，1 min，接着 72 ℃，1.5 min，此步骤进行 35 个循环；彻底延伸，72 ℃，7 min，4 ℃保存。

表 3　简化基因组测序引物

Table 3　Simplified genome sequencing primers

引物名称	正向引物（5′-3′）	反向引物（5′-3′）
Simplified-A	F：TGAGTCCAAACCGGATA	R：GACTGCGTACGAATTGAC
Simplified-B	F：TGAGTCCAAATATAAGC	R：GACTGCGTACGGCGCTGC
Simplified-C	F：TGAGTCCAAACCAAACC	R：GACTGCGTACGGGTTAAT
Simplified-D	F：TGAGTCCAAATTTTAAT	R：GACTGCGTACGGGGGTGA
Simplified-E	F：TGAGTCCAAAAAGGAAG	R：GACTGCGTACGCCTTAAC

1.7　目的片段回收

PCR反应结束后，对所得扩增产物进行2%琼脂糖凝胶电泳并成像。根据琼脂糖凝胶电泳的分析结果，使用普通琼脂糖凝胶DNA回收试剂盒（货号：DP209-03），回收长度在300～400 bp的DNA片段，具体操作步骤按照说明书进行。

1.8　高通量测序

胶回收后，将所有待测样本送至北京安诺优达基因科技有限公司进行建库和质检，质检合格后利用MGI测序平台PE150测序模式进行高通量测序。

1.9　测序数据质控

高通量测序的初始数据结果通常为图像文件，利用WriteFQ软件将其转化为FASTQ文件格式的原始测序数据（Raw reads）。为了确保后续数据分析的可靠性，使用Fqtools_Plus软件中的filter模块对原始测序数据进行过滤，具体处理如下：去除接头污染的Reads（Reads中接头污染的碱基数大于5 bp，若一端受到接头污染，则去掉两端的Reads）；去除低质量的Reads（Reads中质量值 $Q\leqslant 19$ 的碱基占总碱基的50%以上，若一端为低质量Reads，则去掉两端Reads）；去除含N比例大于5%的Reads（若一端含N比例大于5%，则去掉两端Reads）。

1.10　数据比对分析

利用BWA、SAMtools和GATK生物信息学分析软件对拆分后的Clean data进行变异检测，具体操作步骤如下：

首先，利用BWA 0.7.17软件[9]将拆分后的个体样本测序数据比对到牛的参考基因组（*Bos_taurus*. ARS-UCD1.3）上，使用SAMtools 1.19.2软件[10]的sort命令将比对结果进行排序并生成bam文件，再利用Picard 3.2.0中的MarkDuplicates命令去除bam文件中的假阳性。

然后，通过GATK 4.6.0.0软件[11]对每个个体进行变异检测，使用GATK中的HaplotypeCaller模块生成每个样本的gvcf文件，利用CombineGVCFs模块将所有样本的gvcf文件予以合并，再使用GenotypeGVCFs模块对合并后的gvcf文件执行基因分型操作，最终借助SelectVariants模块检出SNP，从而得到原始的变异文件。为获取高置信度的SNP，通过VariantFiltration模块对检出的SNP进行过滤。

1.11　群体结构分析

群体分层会引起关联结果产生假阳性位点，从而影响关联分析的准确度。本研究采用主成分分析（Principal component analysis，PCA）来调查试验群体中是否存在群体分层现象，以评估试验群体

内是否存在遗传差异,并减少假阳性关联带来的影响。

利用 Plink 1.9.0 软件将 SNP 数据转换成 Plink 格式,运行 Plink 的 PCA 分析命令,生成 Eigenval 和 Eigenvec 两个文件,最后使用 R 软件中的 ggplot2 包绘制 PCA 图。

1.12 关联分析

采用混合线性模型(Mixed linear model,MLM)对蒙古牛的 3 种表型性状和检测出的 SNPs 位点进行关联分析,通过显著度(P 值)筛选 SNPs。将群体结构和性别作为协变量代入分析模型,公式为

$$Y = X\alpha + W\beta + Z\gamma + \boldsymbol{Q}\theta + \varphi$$

其中,Y 是表型特征的观测值,X 是 SNP 的基因型值,α 是 SNP 效应系数,W 是群体结构效应,β 是群体效应系数,Z 是性别效应,γ 是性别效应系数,$\boldsymbol{Q}$ 是随机效应矩阵,θ 是随机效应系数,φ 是残差。

1.13 SNPs 注释和候选基因筛选

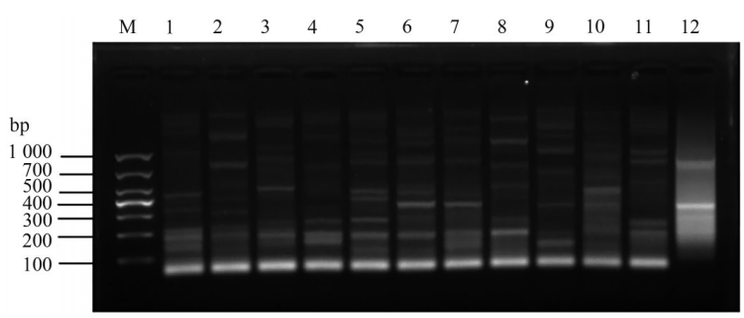
根据关联分析结果,将显著 SNPs 位点所在的物理位置上下游 300 kb 定义为候选基因区域。在此区域内,利用 SnpEff 软件将这些显著的 SNPs 注释到其对应的基因上。

2 结果与分析

2.1 蒙古牛基因组简化测序

本研究采用直接 PCR 的策略,使用针对蒙古牛设计的特异性引物进行基因组简化(图 1)。经凝胶电泳回收 300~400 bp 目标片段后,由测序公司进行建库。后续采用 Qubit 荧光定量仪与 LabChip Touch 核酸分析仪进行文库质量检测,所有样品浓度均显著高于测序平台要求的 3 ng/μL 阈值,符合高通量测序标准。

通过对 384 头蒙古牛进行简化基因组测序,共产生 2 577.7 Gb 原始测序数据,Q30 碱基百分比为 97.48%。利用 Fqtools_Plus 对原始数据进行过滤后,剔除接头污染 Reads 率和低质量 Reads 率较高的样本,最终保留 366 个有效样本,且样本的 Clean data 中 Q30 碱基比例在 96.98%~97.49% 之间,成功获得了高质量的测序数据,为后续变异检测提供了可靠的数据基础。



M:DL1000 DNA 分子标记;1~12:部分蒙古牛基因组样品。

图 1 部分蒙古牛样品的基因组简化

Fig. 1 Genome simplification in a subset of Mongolian cattle samples

2.2 SNP 检测结果

对 366 个有效样本进行变异检测,其中位于基因内含子区域的 SNPs 有 1 522 283 个,占比 59.99%,这类变异通常不影响蛋白质的氨基酸序列,但可能影响基因的表达;位于外显子区域的 SNPs 有 29 901 个,占比 1.18%,这类变异可能直接影响蛋白质的功能;变异位于基因间区域的 SNPs 有 635 872 个,占比 25.05%,这类变异可能影响基因的调控或产生新的基因调控元件;位于剪切位点的 SNPs 有 1 975 个,占比 0.08%,这类变异可能影响 RNA 剪接;位于基因编码区上游的 SNPs 有 173 091 个,占比 6.82%;位于基因编码区下游的 SNPs 有 153 954 个,占比 6.07%,这些区域的变异可能影响基因的表达调控;

位于mRNA 3′端的非翻译区的SNPs有14 170，占比0.56%；位于mRNA 5′端的非翻译区的SNPs有6 408个，占比0.25%，这些区域的变异可能影响mRNA的稳定性和翻译效率（图2，表4）。过滤后的SNPs在染色体上分布均匀且较为广泛（图3）。

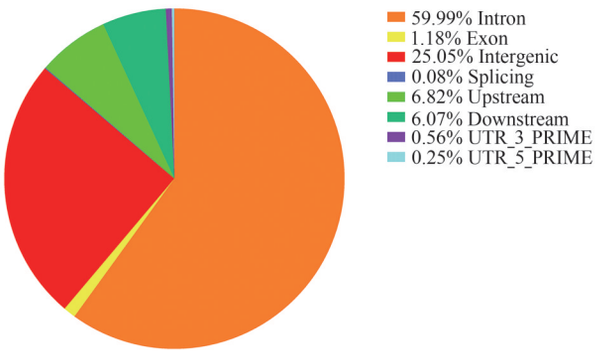


图2　SNP在不同区域的占比情况

Fig. 2　Proportions of SNPs in different regions

表4　SNP统计信息

Table 4　SNP statistical information

类别	SNP数量
内含子（Intron）	1 522 283
外显子（Exon）	29 901
基因间区域（Intergenic）	635 872
剪切位点（Splicing）	1 975
上游区域（Upstream）	173 091
下游区域（Downstream）	153 954
3′端非翻译区（UTR_3_PRIME）	14 170
5′端非翻译区（UTR_5_PRIME）	6 408

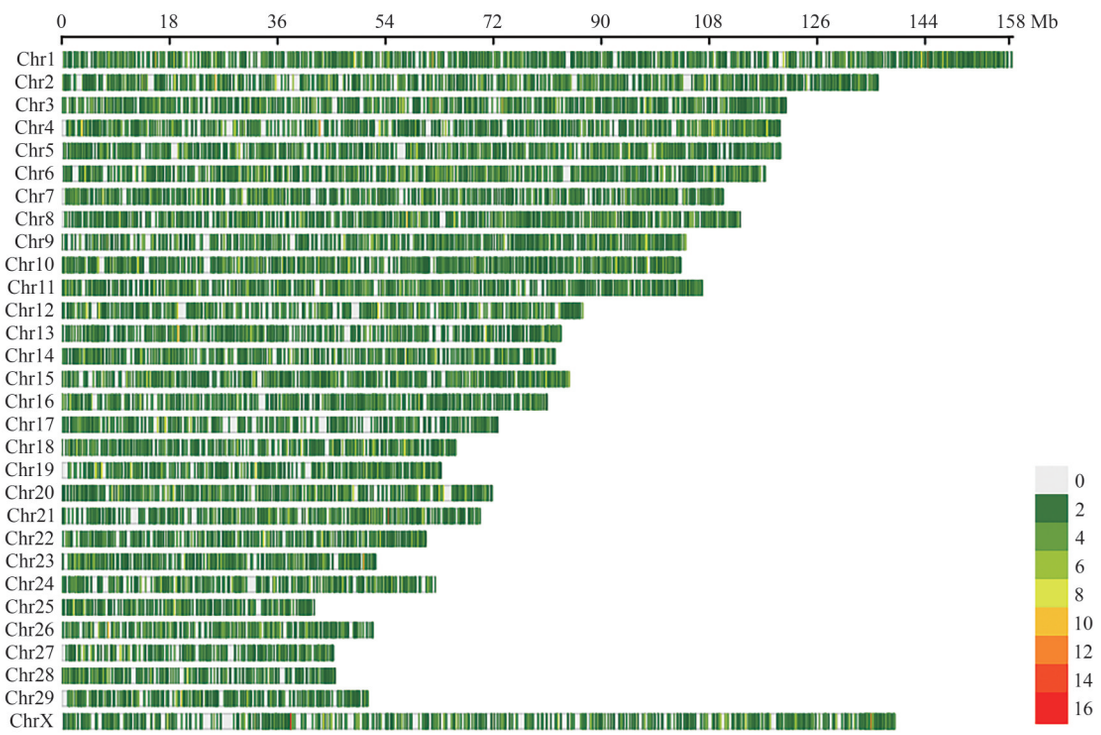


图3　群体的SNPs在染色体上的分布图

Fig. 3　Distribution of SNPs in the population on chromosomes

2.3　群体分层分析

群体分层是遗传背景不同而导致的群体内部和群体之间的系统性差异。在关联分析中，这种现象的存在可能会被错误地认为遗传标记位点与目标性状之间存在相关性。

利用试验群体的第一主成分值（PC1）和第二主成分值（PC2）绘制试验群体的结构聚类图。PC1和PC2是通过PCA得到的两个重要综合变量，它们分别代表了原始数据中能够解释一定比例变异的方向。其中，PC1解释了14.67%的变异，PC2解释了27.9%的变异，这两个主成分共同反映了该试验样本间的遗传变异特征。

试验群体出现了明显的分层现象，样本被划分为3个不同区域，说明样本之间存在遗传差异。大

多数样本聚集在红色区域,这表明该区域内的样本在遗传信息上具有较高的相似度,其他两个区域则包含了较少数量的样本,这可能意味着这些区域内的样本在遗传上与红色区域的样本存在较大差异(图4)。为了避免遗传标记与性状之间产生错误的关联信号,在分析过程中需要校正或控制群体分层带来的影响,以确保研究结果的准确性和可靠性。

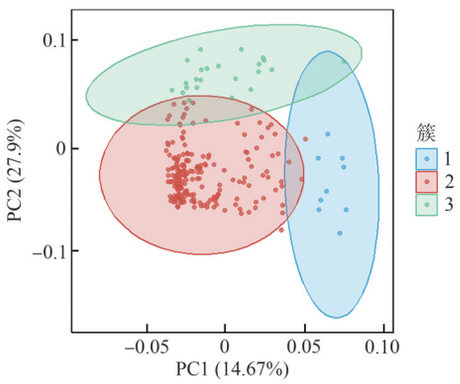


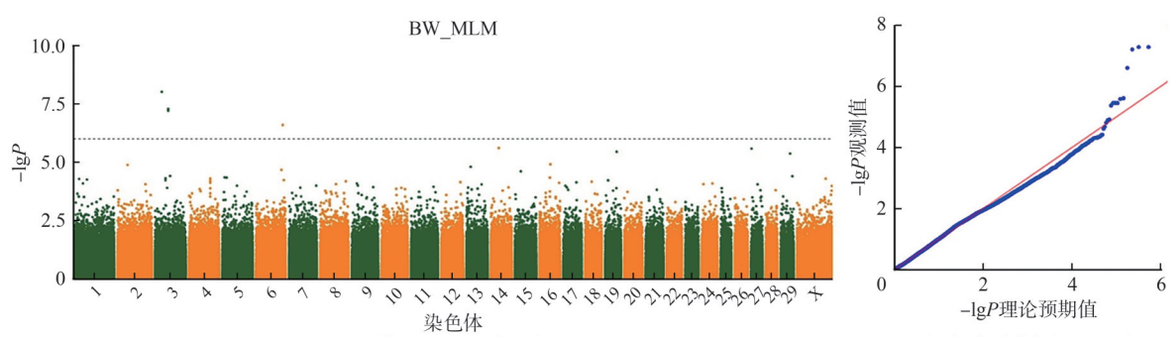
图4 蒙古牛群体主成分分析

Fig. 4 Principal component analysis of Mongolian cattle population

2.4 蒙古牛体重性状的关联分析

利用检测到的SNP位点对蒙古牛的体重性状进行关联分析,共检测出4个与体重性状显著相关的SNPs,如图5(a)所示。其中,有3个SNP位点位于3号染色体上,1个SNP位点位于6号染色体上。在蒙古牛体重性状关联分析的Q-Q图中,如图5(b)所示,大部分数据点集中在蓝色曲线(实际观测线)上,且蓝色曲线与红色直线(理论预期线)在图的中下部较为接近。这表明在多数情况下,实际观测到的$-\lg P$与理论预期值相符,即大部分位点的关联分析P值分布与理论分布相近,说明这些位点与蒙古牛体重性状的关联程度在理论预期范围内。然而,右上角区域存在部分数据点明显偏离理论预期线,且位于蓝色曲线上方,这些点对应的$-\lg P$观测值显著大于预期值,表明其与蒙古牛体重性状存在更强的关联性,极有可能是与蒙古牛体重性状真正相关的重要位点。

在这些显著SNPs位点所在的物理位置上下游300 kb区域内进行注释,共发现8个与体重性状相关的潜在候选基因(表5)。其中,在3:82397310和3:82397314这两个显著SNP位点附近注释到的候选基因均为*FOXD3*、*ALG6*和*ITGB3BP*;在3:86785756显著SNP位点附近注释到的候选基因只有*JUN*;在6:6774623显著SNP位点附近注释到的候选基因包括*SEC24D*、*METTL14*、*PRSS12*和*NDST3*。



(a) 体重性状的关联分析 (b) 关联分析的Q-Q图

图5 蒙古牛体重性状的关联分析结果

Fig. 5 Results of the association analysis of the body weight trait in Mongolian cattle

2.5 蒙古牛体高性状的关联分析

利用检测到的SNP位点对蒙古牛的体高性状进行关联分析,共检测出2个与体高性状显著相关的SNPs,均位于2号染色体上,如图6(a)所示。在蒙古牛体高性状关联分析的Q-Q图中,如图6(b)所示,大部分数据点分布在蓝色曲线上,且前期蓝色曲线与红色直线较为接近,说明多数位点的实际观测到的$-\lg P$值与理论预期值相符。但在图6(b)的右侧末端,部分数据点明显偏离理论预期值,位于蓝色曲线上方,表明这些位点的观测值大于预期值,意味着这些位点与蒙古牛体高性状的关联性比理论预期更强,可能是与体高性状相关的重要位点。

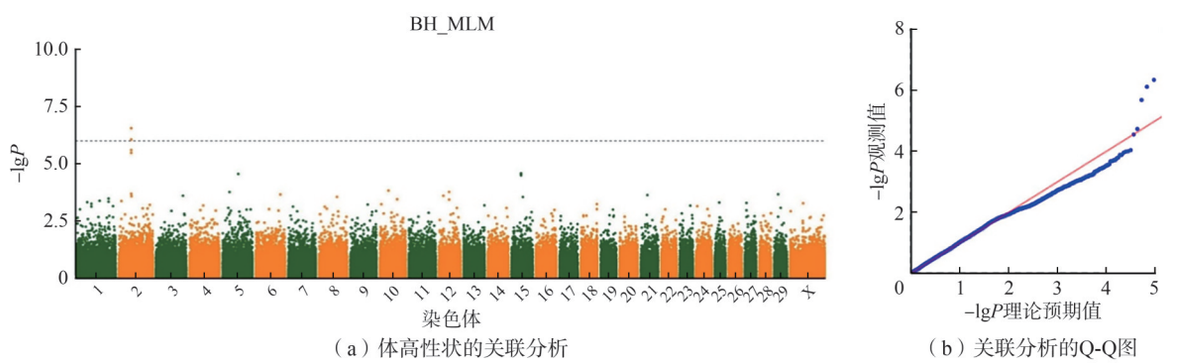
在这些显著SNPs位点所在的物理位置上下游300 kb区域内进行注释,共发现5个潜在候选基

因(表 5)。其中,在 2:115141339 显著 SNP 位点附近注释到的候选基因为 *IRS1*、*RHBDD1*、*COL4A4* 和 *COL4A3*;在 2:124003861 显著 SNP 位点附近注释到的候选基因为 *PTPRU*。

表 5　蒙古牛生长性状的显著关联位点

Table 5　Significantly associated loci related to the growth traits of Mongolian cattle

性状	染色体	物理位置	*P*	基因名称
体重	3	82397310	5.14×10^{-4}	*FOXD3*、*ALG6*、*ITGB3BP*
		82397314	5.14×10^{-4}	*FOXD3*、*ALG6*、*ITGB3BP*
		86785756	4.45×10^{-4}	*JUN*
	6	6774623	5.99×10^{-4}	*SEC24D*、*METTL14*、*PRSS12*、*NDST3*
体高	2	115141339	8.88×10^{-4}	*IRS1*、*RHBDD1*、*COL4A4*、*COL4A3*
		124003861	6.18×10^{-4}	*PTPRU*
体斜长	2	115141339	3.26×10^{-4}	*IRS1*、*RHBDD1*、*COL4A4*、*COL4A3*
		124003861	2.18×10^{-4}	*PTPRU*



(a) 体高性状的关联分析　　(b) 关联分析的Q-Q图

图 6　蒙古牛体高性状的关联分析结果

Fig. 6　Results of the association analysis of the body height trait in Mongolian cattle

2.6　蒙古牛体斜长性状的关联分析

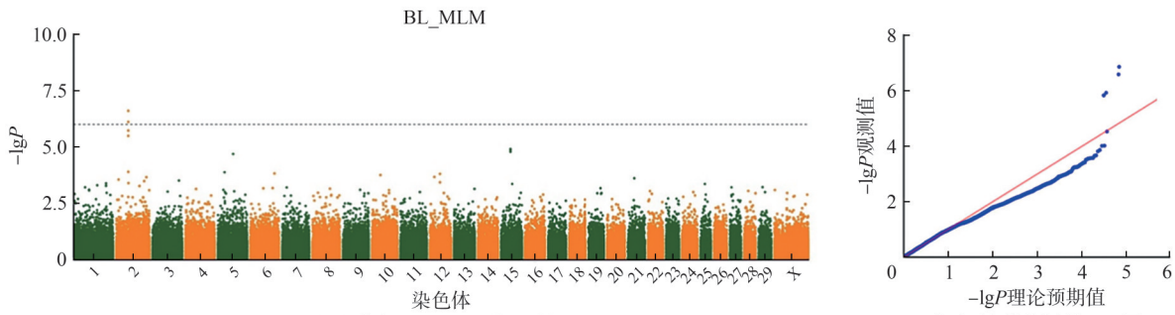
利用检测到的 SNP 位点对蒙古牛的体斜长性状进行关联分析,共检测出 2 个与体斜长性状显著相关的 SNPs,均位于 2 号染色体上,如图 7(a)所示。在蒙古牛体斜长性状关联分析的 Q-Q 图中,如图 7(b)所示,前期蓝色曲线(观测值)与红色曲线(预期值)较为接近,大部分位点分布在蓝色曲线上,表明多数位点的实际 $-\lg P$ 值与理论预期值相近。但在图 7(b)的右侧末端,部分位点明显偏离红色直线,位于蓝色曲线上方,意味着这些位点的观测值大于预期值,说明这些位点与蒙古牛体斜长性状的关联性超出理论预期,很可能是影响体斜长性状的关键位点。

在这些显著 SNPs 位点所在的物理位置上下游 300 kb 区域内进行注释,共发现 5 个潜在候选基因(表 5)。其中,在 2:115141339 显著 SNP 位点附近注释到的候选基因为 *IRS1*、*RHBDD1*、*COL4A4* 和 *COL4A3*;在 2:124003861 显著 SNP 位点附近注释到的候选基因为 *PTPRU*,该结果与体高性状的筛选结果一致。

3　讨论

3.1　蒙古牛基因组测序简化策略

蒙古牛是一种抗逆性极强的地方品种,具有耐粗饲、耐寒、抗风沙等特性。在半荒漠草原,无论是放牧、半舍饲还是全舍饲均表现出良好的适应性,能在极端恶劣的自然环境中生存。但蒙古牛生长速度缓慢、生产性能低,一定程度上限制了其在现代畜牧业中的发展。



（a）体斜长性状的关联分析 （b）关联分析的Q-Q图

图7 蒙古牛体斜长性状的关联分析结果

Fig. 7 Results of the association analysis of the body oblique length trait in Mongolian cattle

与常规限制性内切酶酶切法简化基因组的策略不同，本研究利用简化基因组测序技术对蒙古牛生长性状相关基因进行挖掘，是根据蒙古牛基因组序列中的高频碱基进行简化。在基因组序列中，某些碱基或碱基组合可能在特定区域频繁出现，这些高频碱基可能与基因的功能、调控或结构稳定性有关。在某些基因的启动子区域或编码区，特定的碱基模式可能影响基因的表达或蛋白质的结构[12]。通过选择这些高频碱基区域作为简化基因组引物的结合位点，可以提高引物与目标序列的结合效率和特异性。

3.2 简化基因组测序技术体系的构建

NGS技术因其能够并行测序大量DNA分子，在测序速度、通量和成本方面具有显著优势。目前，该技术已在生物育种领域得到广泛使用，并成为遗传学研究和分子育种工作中的重要工具[13-15]。在分子育种领域，常利用高密度SNP芯片、全基因组重测序以及简化基因组测序来识别遗传标记，以提高辅助标记选择（MAS）的效率和精确度，并加速育种进程。高密度SNP芯片是利用已知的SNP位点侧翼的序列设计特异性探针，利用杂交反应产生的荧光信号来鉴定探针位点的基因型[16-19]。全基因组重测序则是通过将测序得到的序列与参考基因组进行比对，以此来发现大量的SNPs。而简化基因组测序技术通过选择性地测序基因组的部分区域，实现了对基因组的简化处理，从而获得大量分布均匀的SNPs标记[20]。

本研究基于高通量测序技术，根据蒙古牛基因组中的高频碱基设计的简化基因组特异性引物，构建了一套操作简便、成本低廉且能够获得大量基因组变异信息的简化基因组测序技术体系。与高密度SNP芯片和全基因组重测序相比，本研究构建的技术体系不仅进一步降低了测序成本，而且还可以检测到未知的SNP位点。此外，相较于RAD、GBS、2b-RAD以及SLAF-Seq等其他高通量简化基因组测序技术[21]，本研究在多个方面进行了显著改进。其他高通量简化基因组测序技术在降低目标物种基因复杂度和文库构建时，通常需要依赖限制性内切酶酶切和超声波打断等操作，而本研究的技术无须依赖限制性内切酶进行酶切，仅通过简便的PCR即可完成蒙古牛基因组的简化和测序样本文库的构建，操作流程相对直观，可有效降低试验操作复杂度和技术门槛。本研究提出的方法对试验条件和参数依赖性较低，适用于样本规模较大或试验条件受限的研究场景。

3.3 体重性状相关候选基因

通过对蒙古牛的体重性状进行关联分析，共挖掘出8个与蒙古牛体重性状相关的潜在候选基因：*FOXD3*、*ALG6*、*ITGB3BP*、*JUN*、*SEC24D*、*METTL14*、*PRSS12*和*NDST3*。

*FOXD3*基因属于叉头框转录因子家族，在人类、小鼠等其他物种中，该基因在胚胎发育、细胞分化以及维持细胞多功能性等方面发挥着重要作用[22]。*ALG6*基因编码α-1,3-葡萄糖基转移酶，主要参与N-连接糖基化过程。这一过程对于蛋白质的正确折叠、功能运输至关重要，一旦发生异常，便会影响细胞的生理功能，进而引发多种疾病[23]。*ITGB3BP*基因编码整合素β3结合蛋白，该蛋白与细胞的黏附、迁移、增殖和凋亡等过程密切相关[24]。*JUN*基因编码转录因子c-Jun，是AP-1转录因子复合体的重要组成成分，参与调控细胞增殖、分化、生长、凋亡和应激反应等多种生物学过程。*SEC24D*

基因编码的蛋白质是COPⅡ coat complex家族的一部分，该复合物在细胞内主要负责蛋白质从内质网到高尔基体的运输任务，若该基因发生突变，可能导致蛋白质运输紊乱，进而影响细胞功能。此外，该基因还与小鼠的胚胎发育有关[25]。*METTL14*基因编码甲基转移酶，主要参与RNA的N6-甲基腺苷（m6A）修饰过程[26]。*PRSS12*基因编码一种丝氨酸蛋白酶，其编码的蛋白含有清道夫受体富含半胱氨酸域，可识别和结合特定分子，在细胞间传递信息或参与某些生物过程[27]。*NDST3*基因编码的蛋白质属于硫酸转移酶家族，主要功能是参与糖胺聚糖（GAGs）的合成过程[28]。

在这些潜在候选基因中，马兰花等[29]研究发现，*METTL14*基因在牦牛的7个不同组织中均有表达，且在牦牛生长发育的不同阶段，*METTL14*基因在前体脂肪细胞中的表达量也存在显著差异。这表明*METTL14*基因的表达与牦牛的生长阶段和组织类型有关，暗示其可能参与了牛的生长调控。此外，杨昕冉等[30]研究发现，在秦川牛不同组织以及前体脂肪细胞增殖和分化过程中，*METTL14*基因均有表达。在秦川牛前体脂肪细胞增殖期，*METTL14*基因表达量前期下降，之后有所上升但仍维持在较低水平；在分化过程中，其表达量在分化前期下降，中期上升，随后又下降。这表明*METTL14*基因参与的m6A修饰对秦川牛脂肪发育和沉积具有潜在影响，它可能调控牛前体脂肪细胞的脂肪生成，而脂肪发育和沉积也是影响牛体重性状的重要因素之一。综上所述，*METTL14*基因通过参与RNA的m6A修饰，在牛的肌肉发育和脂肪生成等过程中发挥作用，进而可能与牛的体重性状密切相关。熊金珂等[31]通过对绵羊的胸围、胸宽、体重和体长进行关联分析，筛选出2个位于*PRSS12*基因上的SNP位点，并将*PRSS12*作为潜在候选基因。由于绵羊和牛同属反刍动物，在遗传和生理特性上有一定相似性，未来或许可以从绵羊的研究中获得启示，为探究蒙古牛的*PRSS12*基因与生长性状的关系提供参考。除*METTL14*基因和*PRSS12*基因外，*FOXD3*、*ALG6*、*ITGB3BP*、*JUN*、*SEC24D*和*NDST3*基因目前尚无直接研究证明其与牛的体重或其他生长性状相关，有待进一步研究。

3.4　体高和体斜长性状相关候选基因

对蒙古牛的体重和体斜长性状进行关联分析，发现筛选出的显著SNP位点分布一致，均位于2号染色体上的115141339位置和124003861位置，共挖掘出5个与蒙古牛体高性状和体斜长性状相关的潜在候选基因：*IRS1*、*RHBDD1*、*COL4A4*、*COL4A3*和*PTPRU*。

*IRS1*基因主要编码胰岛素受体底物1，该蛋白在细胞信号传导中具有重要作用，能够将胰岛素信号传递到细胞内的各个部位。该基因介导的信号通路能够调节细胞的代谢、生长、增殖和分化。此外，*IRS1*基因还参与其他激素和生长因子的信号传导过程，与多种细胞生理过程关联，对维持机体的正常生理功能和内环境稳定具有重要意义[32]。*RHBDD1*基因编码的是一种跨膜蛋白，属于rhomboid家族的丝氨酸蛋白酶。该基因参与细胞信号通路调控，可以切割表皮生长因子受体（EGFR）家族的配体，如转化生长因子-α（TGF-α）等，使其从细胞膜上释放并激活EGFR信号通路，进而影响细胞的增殖、分化和存活等过程[33]。*COL4A4*基因编码胶原蛋白Ⅳ型α4链，是构成基底膜的主要成分之一，该基因能够调节组织架构并参与器官功能的维持，如肾脏过滤屏障、血管壁组成和眼部结构等。该基因突变可导致多种疾病，如慢性肾衰竭等[34]。*COL4A3*基因编码胶原蛋白Ⅳ型α3链，与*COL4A4*基因同属于胶原蛋白基因家族，二者结构相似，功能也较为接近[35]。*PTPRU*基因编码蛋白酪氨酸磷酸酶（PTP）家族蛋白，参与早期神经发育、组织修复和再生等过程。此外，该基因在免疫系统中也发挥重要作用，参与免疫细胞的信号转导和免疫反应的调节[36]。

在这些潜在候选基因中，已有研究表明*IRS1*基因在牛的生长发育过程中发挥着重要作用。Wu等[37]通过对100头日本黑牛研究发现，*IRS1*基因上存在一个SNP位点g.120947716T＞C，该SNP位点与牛的胴体重量和肋骨厚度显著相关。从进化角度看，*IRS1*基因在牛等多种生物中具有高度保守性，其编码的蛋白质在胰岛素信号通路中起着关键作用，参与细胞对葡萄糖的摄取、利用和储存，为牛的生长提供能量，进而可能影响牛的生长性状。马向辉等[38]对秦川牛体尺和肉质性状进行关联分析，发现*IRS1*基因外显子上的2 346 bp发生G-A突变，该位点不同基因型个体间在体斜长、腰角宽、胸围

和眼肌面积方面存在显著差异,表明*IRS1*基因与秦川牛的体斜长等体尺性状存在关联。虽然该研究未提及体高这一性状,但已证明*IRS1*基因与部分体尺性状密切相关,从侧面反映了*IRS1*基因对牛体尺性状的影响,为进一步研究*IRS1*基因与牛体高的关系提供了一定的理论基础。除*IRS1*基因外,*RHBDD1*、*COL4A4*、*COL4A3*和*PTPRU*基因目前尚无直接研究证明其与牛的体高、体斜长或其他生长性状相关,有待进一步研究。

基于体高性状和体斜长性状共享的SNP位点可以推测,体高和体斜长可能受到部分相同基因或遗传通路的影响。这些SNP位点可能位于与骨骼生长、发育调控相关的基因区域内,或者处于影响身体整体生长发育的信号通路相关基因上。此外,体高和体斜长可能是同一生长机制的不同表现形式,受到相似的遗传因素和环境因素调控。

在个体生长阶段,整体的生长激素水平、营养物质的分配和利用等因素可能对体高和体斜长的增长具有同步的影响。合适的体高和体斜长比例对于牛的运动、采食、繁殖等生理功能至关重要。共享的SNP位点可能暗示存在某种遗传机制,在塑造蒙古牛身体结构时,对体高和体斜长进行协同调控,以确保牛具有适应其生活方式和生产性能的理想体型。此外,本试验蒙古牛群体可能具有特定的遗传结构,在其进化历史中,可能经历过某些特定的选择压力,导致与体高和体斜长相关的基因区域发生了相似的遗传变化,或者该群体存在一定程度的近亲繁殖或遗传漂变,使得这些性状的遗传变异呈现出特定的模式,进而导致体高和体斜长相关的SNP位点表现出一致性。后续还需要进一步研究这些SNP位点对相关基因表达和功能的影响,结合连锁不平衡分析、基因注释等分析方法,深入了解这些位点所在的基因及其参与的生物学过程,以更准确地揭示体高和体斜长之间的遗传关系以及这些SNP位点的生物学意义。

3.5 候选基因功能验证的展望

尽管本研究通过群体水平的关联分析和功能注释筛选了若干与蒙古牛生长性状相关的候选基因,但其具体分子机制仍有待进一步验证。未来研究可从多个层面开展功能验证工作,以评估其与性状形成的相关性,从而直接验证其在性状调控中的作用。这些研究将有助于从功能层面深化对候选基因生物学意义的理解。

4 结论

本研究利用新型简化基因组测序技术挖掘与蒙古牛生长性状相关的分子标记和功能基因,该技术相比传统简化测序方法而言,操作简便快捷,能够快速完成基因组的简化。通过对检测出的SNP进行关联分析,挖掘出*METTL14*、*PRSS12*等8个与体重相关的候选基因和*IRS1*、*RHBDD1*等5个与体高、体斜长相关的候选基因,为蒙古牛基因组育种提供理论基础,并为蒙古牛分子标记辅助选择提供候选基因参考。

参考文献:

[1] Chen Q M, Shen J F, Hanif Q, et al. Whole genome analyses revealed genomic difference between European taurine and east Asian taurine[J]. Journal of Animal Breeding and Genetics, 2021, 138(1): 56-68.

[2] 齐昱,邢燕平,潘静,等. 基于转录组数据的蒙古牛皮肤组织抗寒相关信号通路及候选基因的筛选[J]. 畜牧兽医学报,2017,48(12):2301-2313.

[3] 李光鹏,白春玲,魏著英,等. 黄牛*Myostatin*基因编辑研究[J]. 内蒙古大学学报(自然科学版),2020,51(1):12-32.

[4] Metzker M L. Sequencing technologies-the next generation[J]. Nature Reviews Genetics, 2010, 11(1): 31-46.

[5] Davey J W, Hohenlohe P A, Etter P D, et al. Genome-wide genetic marker discovery and genotyping using next-generation sequencing[J]. Nature Reviews Genetics, 2011, 12(7): 499-510.

[6] 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.

[7] Robledo D, Palaiokostas C, Bargelloni L, et al. Applications of genotyping by sequencing in aquaculture breeding and genetics[J]. Reviews in Aquaculture, 2018, 10(3): 670-682.
[8] Harvey M G, Smith B T, Glenn T C, et al. Sequence capture versus restriction site associated DNA sequencing for shallow systematics[J]. Systematic Biology, 2016, 65(5): 910-924.
[9] Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform[J]. Bioinformatics, 2009, 25(14): 1754-1760.
[10] Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools[J]. Bioinformatics, 2009, 25(16): 2078-2079.
[11] Mckenna A, Hanna M, Banks E, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data[J]. Genome Research, 2010, 20(9): 1297-1303.
[12] 田蕾, 王峰, 班旭空, 等. 骨形态发生蛋白基因 4(*BMP4*)mRNA 在蒙古牛不同组织的表达[J]. 华北农学报, 2011, 26(4): 77-81.
[13] Frazer K A, Eskin E, Kang H M, et al. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains[J]. Nature, 2007, 448(7157): 1050-1053.
[14] 赵长英, 廖媛, 杨洁珂, 等. db/db小鼠基因型鉴定方法的比较研究[J]. 重庆医学, 2020, 49(20): 3451-3455.
[15] 谷明娟, 罗荣松, 高丽, 等. 全基因组重测序技术揭示蒙古牛无角性状变异[J]. 农业生物技术学报, 2020, 28(11): 1960-1969.
[16] 常晨城, 白音巴图, 周乐, 等. 内蒙古地方绵羊遗传多样性及尾椎数性状相关选择信号研究[J]. 中国畜牧杂志, 2023, 59(12): 109-115.
[17] Sheriff O, Ahbara A M, Haile A, et al. Whole-genome resequencing reveals genomic variation and dynamics in Ethiopian indigenous goats[J]. Frontiers in Genetics, 2024, 15: 1353026.
[18] 周东辉, 张昌政, 张任豹, 等. 基于 70K SNP 芯片评估牙山黑绒山羊保种群体的遗传结构[J]. 中国畜牧杂志, 2024, 60(4): 116-122.
[19] Meyermans R, Gorssen W, Wijnrocx K, et al. Unraveling the genetic diversity of Belgian milk sheep using medium-density SNP genotypes[J]. Animal Genetics, 2020, 51(2): 258-265.
[20] Liu J X, Shi L Y, Li Y, et al. Estimates of genomic inbreeding and identification of candidate regions that differ between Chinese indigenous sheep breeds[J]. Journal of Animal Science and Biotechnology, 2021, 12(1): 95-113.
[21] Baird N A, Etter P D, Atwood T S, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers[J]. PLoS One, 2008, 3(10): e3376.
[22] Li D, Mei H, Qi M, et al. FOXD3 is a novel tumor suppressor that affects growth, invasion, metastasis and angiogenesis of neuroblastoma[J]. Oncotarget, 2013, 4(11): 2021-2044.
[23] Bloch J S, Pesciullesi G, Boilevin J, et al. Structure and mechanism of the ER-based glucosyltransferase ALG6[J]. Nature, 2020, 579(7799): 443-447.
[24] Long L J, Zhang C, He Z Q, et al. LncRNA NR_045147 modulates osteogenic differentiation and migration in PDLSCs via ITGB3BP degradation and mitochondrial dysfunction[J]. Stem Cells Translational Medicine, 2025, 14(2): szae088.
[25] Adams E J, Khoriaty R, Kiseleva A, et al. Murine SEC24D can substitute functionally for SEC24C during embryonic development[J]. Scientific Reports, 2021, 11(1): 21100.
[26] Meng L P, Lin H, Huang X X, et al. METTL14 suppresses pyroptosis and diabetic cardiomyopathy by downregulating TINCR lncRNA[J]. Cell Death & Disease, 2022, 13(1): 31-45.
[27] Pratt J, Whitton L, Ryan A, et al. Genes encoding agrin (*AGRN*) and neurotrypsin (*PRSS12*) are associated with muscle mass, strength and plasma C-terminal agrin fragment concentration[J]. GeroScience, 2023, 45(3): 1289-1302.
[28] Tang Q, Liu M M, Liu Y, et al. NDST3 deacetylates α-tubulin and suppresses V-ATPase assembly and lysosomal acidification[J]. The EMBO Journal, 2021, 40(19): e107204.
[29] 马兰花, 顾亚荣, 张永峰, 等. 牦牛 METTL14 CDS 区克隆及表达谱分析[J]. 中国兽医学报, 2022, 42(5): 962-972, 980.
[30] 杨昕冉, 王建芳, 马鑫浩, 等. m6A 甲基化修饰相关酶基因在牛脂肪生成中的表达分析[J]. 生物技术通报, 2022, 38(7): 70-79.
[31] 熊金珂, 鲍晶晶, 毕亚珍, 等. 绵羊 *PRSS12* 基因多态性及其与生长性状的关联分析[J]. 农业生物技术学报,

2024,32(7):1587-1595.
[32] Sánchez-Sarasúa S,Meseguer-Beltrán M,García-Díaz C,et al. IRS1 expression in hippocampus is age-dependent and is required for mature spine maintenance and neuritogenesis[J]. Molecular and Cellular Neurosciences,2022,118:103693.
[33] Miao F,Zhang M M,Zhao Y C,et al. RHBDD1 upregulates EGFR via the AP-1 pathway in colorectal cancer[J]. Oncotarget,2017,8(15):25251-25260.
[34] Savige J. Heterozygous pathogenic *COL4A3* and *COL4A4* variants (autosomal dominant alport syndrome) are common,and not typically associated with end-stage kidney failure,hearing loss,or ocular abnormalities[J]. Kidney International Reports,2022,7(9):1933-1938.
[35] Yuan X H,Su Q,Wang H,et al. Genetic variants of the *COL4A3*,*COL4A4*,and *COL4A5* genes contribute to thinned glomerular basement membrane lesions in sporadic IgA nephropathy patients[J]. Journal of the American Society of Nephrology,2023,34(1):132-144.
[36] Zhu Z C,Liu Y J,Li K,et al. Protein tyrosine phosphatase receptor U (PTPRU) is required for glioma growth and motility[J]. Carcinogenesis,2014,35(8):1901-1910.
[37] Wu Y,Tong B,Wu K F,et al. Association of single nucleotide polymorphism in *IRS1* gene with growth-related trait in Japanese black beef cattle[J]. Journal of Animal and Veterinary Advances,2014,13(1):37-40.
[38] 马向辉,眷林森,高建斌,等. 秦川牛*IRS*-1基因多态性与体尺和肉质性状的相关性[J]. 畜牧兽医学报,2012,43(8):1200-1209.

(责任编委 那顺布和)

Development of Novel Simplified Genome Sequencing Technology and Its Application in Growth Traits of Mongolian Cattle

PEI Dongchao,HAI Chao,SONG Lishuang,LIU Xuefei,BAI Chunling,
SU Guanghua,YANG Lei,LI Guangpeng
(*State Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock*,
School of Life Sciences,*Inner Mongolia University*,*Hohhot* 010070,*China*)

Abstract: Mongolian cattle (*Bos taurus*) are an indigenous yellow cattle breed in northern China and represent an important livestock genetic resource, characterized by cold tolerance, adaptability to roughage, and strong resistance to environmental stress. Reduced-representation genome sequencing (RRGS) has been widely applied in molecular marker-assisted breeding; however, conventional RRGS relies on restriction enzyme digestion to reduce genome complexity, resulting in relatively cumbersome procedures and high requirements for DNA quality. To address these limitations, based on high-frequency nucleotide features of the Mongolian cattle genome, a novel RRGS primer set was designed and a direct PCR strategy was adopted to achieve genome simplification. Using the primer set, reduced-representation sequencing was performed on 384 Mongolian cattle, followed by association analyses to identify functional genes related to growth traits. The results identify 4 SNP loci significantly associated with body weight and 8 candidate genes, including *METTL14* and *PRSS12*. Association analyses of body height and body length traits show a high degree of consistency, revealing 2 SNP loci and 5 candidate genes, including *IRS1* and *RHBDD1*. In summary, the study establishes a novel reduced-representation genome sequencing strategy and provides both technical and theoretical support for genomic selection breeding in Mongolian cattle.

Key words: reduced-representation genome sequencing; Mongolian cattle; single nucleotide polymorphism; association analysis