



基于全基因组数据绘制全球绵羊 *LCORL* 基因 SNP 变异图谱

张新月¹, 张 硕¹, 王海燕², 霍永智², 刘小东², 姜 雨¹, 李 冉^{1*}

(1. 西北农林科技大学 动物科技学院, 陕西杨凌 712100;

2. 榆林市羊产业发展中心, 陕西榆林 719053)

摘 要: *LCORL* 基因是影响绵羊生长速度和体格大小的重要基因, 全面绘制该基因遗传变异图谱有助于系统评估该基因与生长性状的相关性, 并挖掘与生长相关的变异位点, 然而尚无关于该基因单核苷酸多态性(singlenucleotidepolymorphisms, 简称 SNPs)在绵羊不同品种中系统分布特征报道。本研究基于实验室收集的国内外 27 个品种 1 615 只羊的全基因组重测序数据, 共筛选出 4 777 个高可信度 SNPs, 其中 16 个为编码区突变, 包括 10 个错义突变和 6 个同义突变。此外, 本研究发现其中一个错义突变(g. 38075364G>A)在大体型的多浪羊中受到选择, 推测该位点可能与绵羊生长性状有关。本研究绘制了绵羊 *LCORL* 基因的遗传变异图谱, 为探索与绵羊生长性状相关的突变和遗传标记提供了宝贵的资源。

关键词: 绵羊; *LCORL* 基因; SNP; 全基因组重测序

[中图分类号] S813.3

[文献标志码] A

[文章编号] 1004-6704(2025)-03-0020-06

A SNPs Variation Map of *LCORL* Gene among Global Sheep Breeds Based on Whole-genome Resequencing

ZHANG Xinyue¹, ZHANG Shuo¹, WANG Haiyan², HUO Yongzhi²,

LIU Xiaodong², JIANG Yu¹, LI Ran^{1*}

(1. College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi 712100, China;

2. Yulin Sheep Industry Development Center, Yulin, Shaanxi 719053, China)

Abstract: The *LCORL* gene is an important gene that influences the growth rate and body size of sheep. Constructing a comprehensive genetic variation map of this gene will help systematically assess its correlation with sheep growth traits and identify growth-related variation sites. However, no study has systematically screened the single nucleotide polymorphisms(SNPs) of this gene in different sheep breeds. In this study, based on whole-genome resequencing data from 1 615 sheep across 27 domestic and international breeds, 4 777 high-confidence SNPs were identified, including 16 mutations in coding regions, with 10 missense mutations and 6 synonymous mutations. Additionally, this study found that one missense mutation(g. 38075364G>A) is under

selection in large-bodied Dorper sheep, suggesting that this site may be related to sheep growth traits. This study constructed a genetic variation map of the *LCORL* gene in sheep, providing valuable resources for exploring mutations and genetic markers related to sheep growth traits.

Key words: sheep; *LCORL* gene; SNP; whole-genome sequencing

[收稿日期] 2025-01-14

[基金项目] 农业生物育种国家科技重大专项(2022ZD04-01403);
陕西省双链融合项目(2022GD-TSLD-46-0401);
陕西省自然科学基金(2023-JC-YB-157); 陕西省
重点研发计划(2024NC2-GJHX-15)

[第一作者] 张新月(1998-), 女, 主要从事绵羊基因组学研究
工作。E-mail: 1273653596@qq.com

* [通信作者] 李 冉, E-mail: ran.li1986@hotmail.com

1 引 言

绵羊为人们的日常生活提供肉、毛、奶、皮等多种畜产品,是经济、文化和社会生活的重要组成部分^[1]。体重和体尺作为绵羊重要的生长性状^[2],在肉羊育种中得到极大关注,目前,通过全基因组关联分析和选择信号检测,已挖掘出影响绵羊重要生长发育性状(尤其是数量性状,如体重)的功能基因和遗传变异。研究发现,家畜的生长速度和体型大小由较少的基因发挥主导作用,如配体依赖性核受体辅阻遏物(ligand dependent nuclear receptor corepressor-like, *LCORL*)^[3],肌肉生成抑制素基因(Myostatin, *MSTN*)^[4]和多型性腺瘤基因 1(pleio-morphic adenoma gene 1, *PLAG1*)等^[3]。

LCORL 是一种可能在睾丸精子发生过程中发挥作用的转录因子,在不同物种间高度保守。Steven 等对 *LCORL* 基因进行了详细的功能分析,证实该基因通过调控生长激素(GH)和胰岛素样生长因子 1(IGF-1)信号通路,影响细胞增殖和分化,从而调节动物和人类的生长发育^[5]。*LCORL* 基因座的多样性与成人身高和骨骼尺寸高度相关^[6-10]。*LCORL* 基因在不同家畜物种中的体型和体重性状的遗传研究中表现出显著的影响。Rubin 等^[11]通过全基因组重测序和全基因组关联分析(GWAS)研究发现, *LCORL* 基因与家猪的体长显著相关,揭示了其在猪体型遗传中的关键作用。Graber 等^[12]则在山羊的研究中也发现 *LCORL* 基因与山羊的身高存在显著关联。在驴的研究中,Shen 等^[13]对 78 头驴进行选择扫描和差异表达分析,发现在不同体型、品种的驴中, *LCORL* 基因与体型性状密切相关。家养绵羊的相关研究同样证实了 *LCORL* 基因的作用,多个研究结果表明, *LCORL* 基因座受到了正选择,并与绵羊的体重和体型等性状显著相关^[14-16]。此外,Al-Mamun 等通过对澳大利亚美利奴肉羊进行全基因组关联分析,成功鉴定出与人类 *LCORL* 基因同源的 QTL 区域,该区域与绵羊体重显著相关,进一步证明了 *LCORL* 基因座是绵羊体重遗传调控中的一个重要候选基因。

本研究拟利用从中国和欧洲不同绵羊群体获得的共 27 个群体,1 615 只羊的 WGS 数据来研究 *LCORL* 基因内所有的 SNPs 位点,分析关键变异在不同绵羊品种中的频率分布,挖掘与生长性能潜在相关的候选变异。相关结果将为肉绵羊生长性状的分子育种提供重要参考。

2 材料和方法

2.1 样本采集

从 NCBI SRA 公共数据库(<https://www.ncbi.nlm.nih.gov/sra/>)收集了共 1 615 只绵羊的全基因组重测序数据。这些数据来自 27 个绵羊群体,包括来自欧洲地区 292 只和东亚地区 1 323 只(表 1)。

表 1 样本信息表
Table 1 Sample information

地区	品种	样本量	平均测序深度(×)
欧洲	特克赛尔羊	29	14.14
	东佛里生羊	55	13.73
	无角陶赛特羊	41	9.12
	萨福克羊	29	17.55
	德国美利奴	26	8.88
	有角美利奴	62	10.87
	无角美利奴	50	11.75
	阿勒泰羊	41	13.56
	巴什拜羊	33	12.48
	哈萨克羊	81	17.00
东亚	策勒黑绵羊	133	11.60
	罗布羊	10	15.48
	新疆卡拉库尔羊	196	5.56
	多浪羊	40	13.97
	和田羊	210	8.01
	湖羊	129	9.99
	巴音布鲁克羊	31	7.90
	大尾寒羊	12	20.03
	蒙古羊	24	11.99
	小尾寒羊	29	17.16
	滩羊	38	13.20
	乌珠穆沁羊	19	9.82
	泗水裘皮羊	11	18.42
	洼地羊	18	20.36
	迪庆绵羊	13	9.05
	藏羊	160	13.13
	云南乌骨羊	95	9.40

2.2 基因组比对与变异检测

首先,使用 fastp v0.20.0^[17]过滤原始 FASTQ 文件,使用赛乐公司基于 BWA 算法研发的 slaidx 软件先对绵羊参考基因组 ARS-UI_Ramb_v2.0(GCA_016772045.1)建立索引,再利用 SLA(SaiLe Aligner)软件进行全基因组序列比对,获得 bam 文件。利用 SLC(SaiLe Caller)软件(基于 GATK_v4.0 算法)生成 gvcf 文件,然后合并 gvcf 文件以调用原始 SNP。使用 BCFtools v1.13^[18]去除了测序和对齐错误,参数为“QD<2.0,QUAL<30.0,SOR>3.0,FS>60.0,MQ<40.0,MQRankSum<

-12.5 和 ReadPosRankSum<-8.0”,并使用参数“-m2-M2-iMAF≥0.05&F_MISSING<0.2”进一步过滤 SNP。最后,使用 Beagle v5.4^[19]对过滤后的 vcf 文件进行 Imputation(填充)和 Phase(定相),得到 phase 后的 vcf 文件获得全基因组 vcf 文件后,提取 *LCORL* 基因的 SNPs 用于下游分析。

2.3 功能注释

使用 ANNOVAR 软件(版本 2016-02-01)^[20]的 annotate_variation.pl 脚本对 1 667 只绵羊群体的 6 号染色体进行单点 SNP 注释,参数设置:perl annotate_variation.pl--outfile SV_GT-buildver reference prefix input. bed path。最后,使用 VCFtools v.0.1.16^[21]软件的--freq 参数计算不同品种位于 *LCORL* 基因外显子区 SNPs 的等位基因频率。

2.4 单倍型分析

利用 VCFtools v.0.1.16 软件的--keep、--chr 参数提取多浪羊与和田羊的 vcf 文件,使用编写的 Python 脚本绘制以上个体在 6 号染色体的 38.05~38.10 Mb 的单倍型图。

2.5 选择信号检测

在多浪羊中,使用 EHH 和 0π 方法筛选选择下的基因组区域,并与和田绵羊中的基因组区域进行比较。使用 VCFtools v.0.1.16 软件--IMPUTE 参数将目标群体的 vcf 文件转换为 hap 格式,再利用 Selscan v2.0.2 软件^[22]的--ehh 参数计算两个群体的 EHH 值。使用 VCFtools v.0.1.16 软件的--window-pi、--window-pi-step 计算两个群体的核苷酸多样性,窗口大小为 5 kb,窗口步长为 2 kb。

3 结果与分析

3.1 *LCORL* 基因 SNP 变异挖掘

利用中国 1 615 只绵羊的全基因组序列数据分

析了 *LCORL* 基因的所有 SNP 位点。共鉴定出 4 777 个 SNP 位点,其中大部分 SNPs 位于基因间隔区(92.46%) (图 1)。在编码区共鉴定到 16 个 SNPs,包括 10 个错义突变(g.38074033 C>T、g.38074702 A>T、g.38075043 G>A、g.38075364 G>A、g.38075381 T>A、g.38075862 A>G、g.38076530 A>C、g.38077008 T>C、g.38077301 A>G 和 g.38077886 A>G)(表 2)和 6 个同义突变(g.38073896 A>G、g.38074886 G>A、g.38075003 G>A、g.38075747 T>C、g.38076761 T>C 和 g.38078389 T>C)。

这 16 个编码区变异中,8 个 SNPs 在不同绵羊品种的频率分布具有一定的群体特异性,5 个 SNPs (g.38075043 G>A、g.38075747 T>C、g.38076761 T>C、g.38077008 T>C、g.38077301 A>G)在所有群体中的最小等位基因频率均在 0.3 以下(图 2)。值得注意的是,g.38074702A>T(p.L1521I)和 g.38075381T>A(p.E1294D)在不同的中国绵羊品种和欧洲肉用绵羊品种(萨福克羊、无角陶赛特羊和东佛里生羊)中的频率相对较高,接近 1。在特定绵羊品种中频率相对较高的 SNPs 有 8 个(g.38074033C>T、g.38074886G>A、g.38075003G>A、g.38075364G>A、g.38075862A>G、g.38076530A>C、g.38077886A>G 和 g.38078389T>C),大体型的绵羊品种(多浪羊、迪庆绵羊、策勒黑羊、小尾寒羊、阿勒泰羊、云南乌骨绵羊、特克赛尔羊、湖羊和乌珠穆沁羊)在这些 SNP 位点的突变频率均大于 0.5,多浪羊(突变频率=0.80)和迪庆绵羊(突变频率=0.77)的突变频率较高(图 2)。

3.2 *LCORL* 基因 g.38075364G>A 突变

本研究发现 *LCORL* 区间在多浪羊品种受到选择,形成了长连锁的单倍型(图 3A),而小体格的和田羊没有该单倍型。选取错义突变(g.38075364G

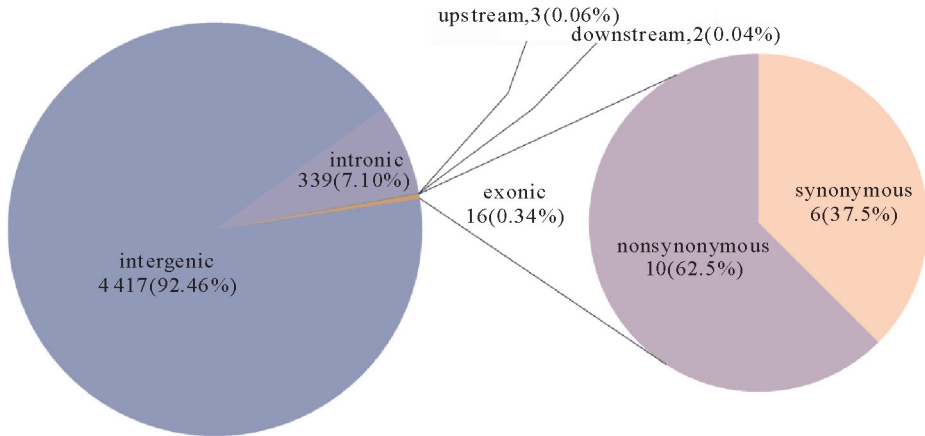


图 1 绵羊 *LCORL* 基因突变的分布

Fig. 1 Distribution of *LCORL* mutations in sheep

表 2 绵羊 *LCORL* 基因的编码区变异位点

Table 2 Variants found in the coding region of *LCORL* gene in sheep

变异位点	区域	类型	氨基酸变异	同义/错义
g. 38073896 A>G	exon8	c. T5367C	p. S1789S	同义
g. 38074033 C>T	exon8	c. G5230A	p. V1744I	错义
g. 38074702 A>T	exon8	c. T4561A	p. L1521I	错义
g. 38074886 G>A	exon8	c. C4377T	p. N1459N	同义
g. 38075003 G>A	exon8	c. C4260T	p. N1420N	同义
g. 38075043 G>A	exon8	c. C4220T	p. S1407L	错义
g. 38075364 G>A	exon8	c. C3899T	p. A1300V	错义
g. 38075381 T>A	exon8	c. A3882T	p. E1294D	错义
g. 38075747 T>C	exon8	c. A3516G	p. E1172E	同义
g. 38075862 A>G	exon8	c. T3401C	p. I1134T	错义
g. 38076530 A>C	exon8	c. T2733G	p. N911K	错义
g. 38076761 T>C	exon8	c. A2502G	p. E834E	同义
g. 38077008 T>C	exon7	c. A2257G	p. K753E	错义
g. 38077301 A>G	exon7	c. T1964C	p. F655S	错义
g. 38077886 A>G	exon7	c. T1379C	p. L460S	错义
g. 38078389 T>C	exon7	c. A876G	p. Q292Q	同义

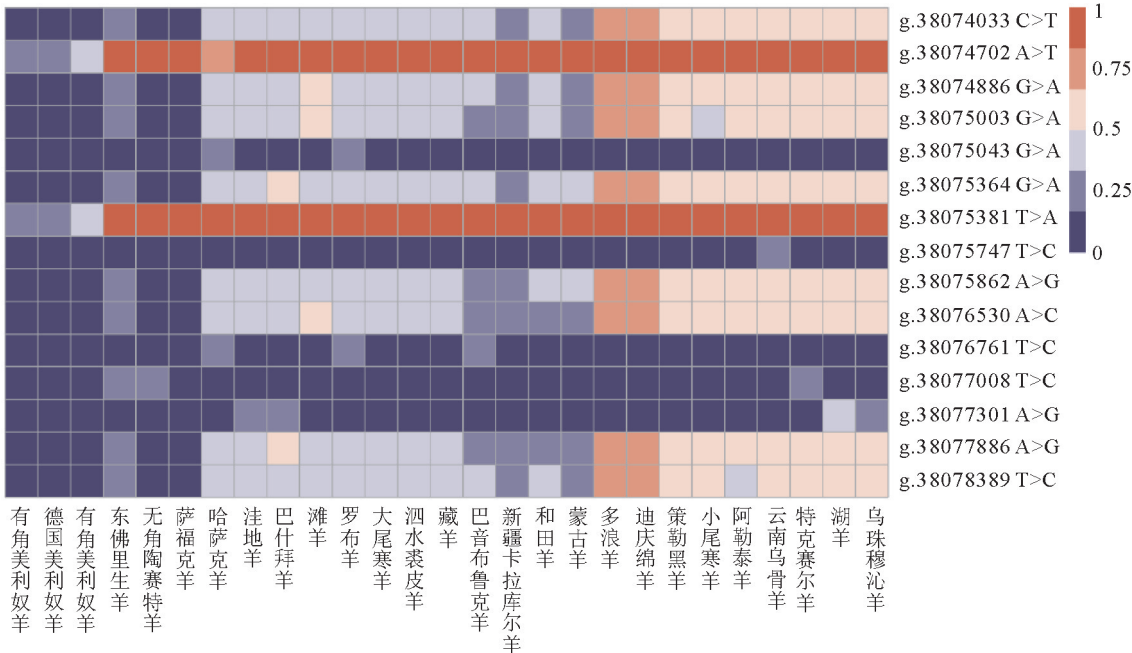


图 2 *LCORL* 基因 SNPs 在不同绵羊群体中的突变频率

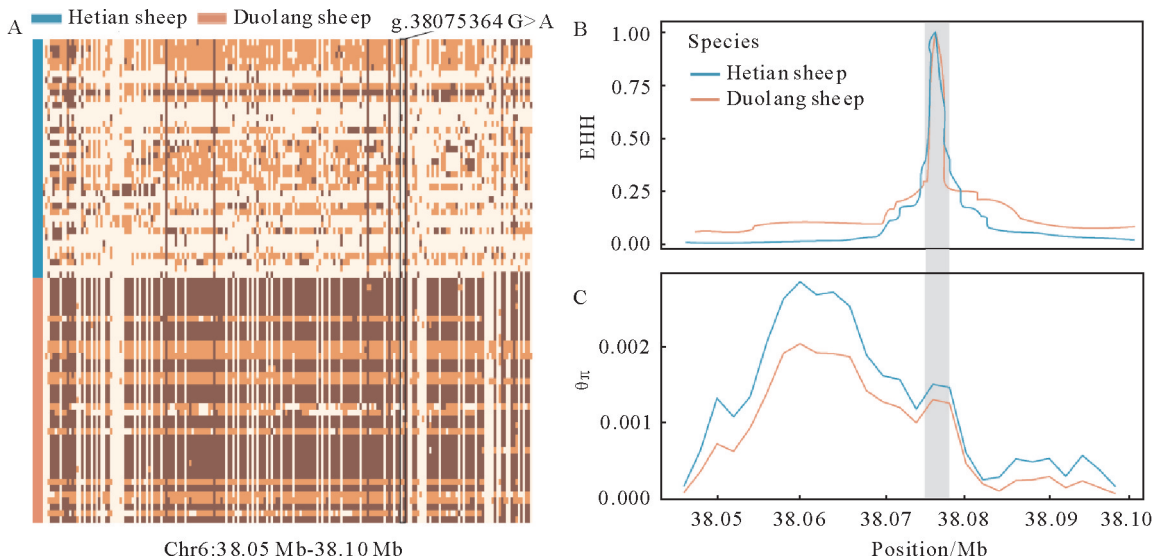
Fig. 2 Mutation frequency of SNPs in *LCORL* gene in different sheep populations

>A)代表该单倍型,其在多浪羊的频率为 0.80。EHH 的分析结果发现多浪羊在 g. 38075364G>A 周围区域的单倍型长于和田羊(图 3B)。多浪羊在该区域的低核苷酸多样性进一步证明该位点在多浪羊中处于选择状态,从而产生了长单倍型(图 3C)。此外,该突变不仅在多浪羊中出现的频率较高,而且在巴什拜羊(突变频率=0.52)、迪庆绵羊(突变频率=0.77)、策勒黑羊(突变频率=0.55)、小尾寒羊(突

变频率=0.59)、阿勒泰羊(突变频率=0.59)、云南乌骨绵羊(突变频率=0.57)、特克赛尔羊(突变频率=0.59)、湖羊(突变频率=0.64)和乌珠穆沁羊(突变频率=0.61)等其他中国群体中也有一定的频率,表明基于该位点可用于其他品种生长性状的选择。

4 讨论

LOCRL 基因是影响家畜体型大小等生长性状



A. g. 38075364G>A 突变在多浪羊具有的长单倍型;B. 多浪绵羊和和田绵羊中 EHH 选择信号的分析;C. 突变周围的核苷酸多样性(灰色区域)

图 3 g. 38075364G>A 突变在多浪羊中表现出强烈选择信号

Fig. 3 Regions surrounding the g. 38075364G>A mutation displays strong selective signatures in Duolang sheep

的主效基因,系统挖掘该基因在不同绵羊品种中的遗传变异情况,对绵羊分子标记挖掘的开展和标记辅助育种的应用具有重要指导意义,可为中国地方绵羊生长性状的遗传改良提供重要参考。本研究对中国 20 个和欧洲 7 个绵羊品种在 *LCORL* 基因区域的 SNPs 进行统计并计算了位于外显子的 SNPs 在所品种中的频率分布,通过系统扫描 *LCORL* 基因遗传变异图谱,共找到 4 777 个 SNPs,包括 16 个编码区突变。结果可见,7.10% 的 SNPs 位于内含子,只有 0.33% 位于外显子。一方面,这与内含子长度(161 882 bp)远大于外显子(6 760 bp)有关;更重要的是,相对于基因间区,外显子区的突变更可能对基因的生物学功能产生重要影响,会受到更强的负选择^[23-24],因此,在外显子区检测到的突变数目要远低于基因间区。

基于团队积累的国内外不同绵羊的基因组数据,我们进一步研究了这些突变在不同绵羊品种的分布规律,特别是发现了一系列在东、西方绵羊中频率存在显著差异的位点,如 g. 38074702A>T、g. 38075381T>A 等突变,表明这些位点可能在东、西方绵羊受到的人工选择强度不同。在西方肉羊品种中频率显著较高的 SNPs 可能与其大体格以及较快的生长速度有关,值得后续进行进一步的研究。另一方面,我们也关注了一些大体格的中国绵羊品种(如多浪羊等)频率较高的突变,特别是其中的错义突变,这些变异可能是影响中国地方绵羊品种体格大小的重要位点。

多浪羊是中国新疆优良肉脂兼用型绵羊品种,

具有体型大、生长快、繁殖率高等特点^[25],本研究在多浪羊中观察到错义突变 g. 38075364G>A (p. A1300V)的突变频率较高,并且在该基因座周围区域表现出低核苷酸多样性,形成了长单倍型,以上种种表明该突变在多浪羊中受到选择。迪庆绵羊属于藏系绵羊,是短毛型山地粗毛羊^[26],山地型藏羊的体型较大,暗示其体型可能受到 *LCORL* 基因的调控。此外,在中国地方绵羊品种中,g. 38075364G>A 的突变频率在大体型/中等偏大体型的绵羊品种(多浪羊、迪庆绵羊、乌珠穆沁羊、小尾寒羊、阿勒泰羊、云南乌骨绵羊、策勒黑绵羊和巴什拜羊)中普遍较高,而在小体型/中等体型绵羊(滩羊、哈萨克羊、泗水裘皮羊、藏羊、罗布羊、和田羊、巴音布鲁克羊、蒙古羊和新疆卡拉库尔羊)中普遍较低,进一步表明该位点与绵羊不同群体体型大小的变异有关。本研究丰富了绵羊生长发育遗传变异机制研究方面的基础资料,为肉羊分子育种提供了遗传学素材。

参考文献:

[1] MAZINANI M, RUDE B. Population, world production and quality of sheep and goat products[J]. American Journal of Animal and Veterinary Sciences, 2020, 15 (4): 291-299.

[2] THORNE J W, MURDOCH B M, FREKING B A, et al. Evolution of the sheep industry and genetic research in the United States: Opportunities for convergence in the twenty-first century [J]. Animal Genetics, 2021, 52 (4): 395-408.

[3] TAKASUGA A. PLAG1 and NCAPG-LCORL in live-

- stock[J]. *Animal Science Journal*, 2016, 87(2): 159-167.
- [4] CARNAC G, VERNUS B, BONNIEU A. Myostatin in the pathophysiology of skeletal muscle[J]. *Current Genomics*, 2007, 8(7): 415-422.
- [5] WYLER S C, GAHLOT S, BIDEYAN L, et al. *LCORL* regulates growth and metabolism[J]. *Endocrinology*, 2024, 165(12): 146.
- [6] GUDBJARTSSON D F, BRAGI WALTERS G, THORLEIFSSON G, et al. Many sequence variants affecting diversity of adult human height[J]. *Nature Genetics*, 2008, 40(5): 609-615.
- [7] SORANZO N, RIVADENEIRA F, CHINAPPEN-HORSELEY U, et al. Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size[J]. *PLoS Genetics*, 2009, 5(4): 1 000 445.
- [8] WEEDON M N, LANGO H, LINDGREN C M, et al. Genome-wide association analysis identifies 20 loci that influence adult height[J]. *Nature Genetics*, 2008, 40(5): 575-583.
- [9] New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism[EB/OL]. *Nature Genetics*, [2025-02-14]. <https://www.nature.com/articles/ng.2477>.
- [10] VAN DER VALK R J P, KREINER-MØLLER E, KOOLIJMAN M N, et al. A novel common variant in *DCST2* is associated with length in early life and height in adulthood[J]. *Human Molecular Genetics*, 2015, 24(4): 1 155-1 168.
- [11] RUBIN C J, MEGENS H J, MARTINEZ BARRIO A, et al. Strong signatures of selection in the domestic pig genome[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2012, 109(48): 19 529-19 536.
- [12] GRABER J K, SIGNER-HASLER H, BURREN A, et al. Evaluation of truncating variants in the *LCORL* gene in relation to body size of goats from Switzerland[J]. *Animal Genetics*, 2022, 53(2): 237-239.
- [13] SHEN J F, YU J, DAI X L, et al. Genomic analyses reveal distinct genetic architectures and selective pressures in Chinese donkeys[J]. *Journal of Genetics and Genomics*, 2021, 48(8): 737-745.
- [14] AL-MAMUN H A, KWAN P, CLARK S A, et al. Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight[J]. *Genetics, Selection, Evolution*, 2015, 47(1): 66.
- [15] POSBERGH C J, HUSON H J. All sheeps and sizes: A genetic investigation of mature body size across sheep breeds reveals a polygenic nature[J]. *Animal Genetics*, 2021, 52(1): 99-107.
- [16] MATIKA O, RIGGIO V, ANSELME-MOIZAN M, et al. Genome-wide association reveals QTL for growth, bone and in vivo carcass traits as assessed by computed tomography in Scottish Blackface lambs[J]. *Genetics, Selection, Evolution*, 2016, 48: 11.
- [17] CHEN S F, ZHOU Y Q, CHEN Y R, et al. Fastp: An ultra-fast all-in-one FASTQ preprocessor[J]. *Bioinformatics*, 2018, 34(17): 884-890.
- [18] LI H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data[J]. *Bioinformatics*, 2011, 27(21): 2 987-2 993.
- [19] BROWNING B L, BROWNING S R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals[J]. *American Journal of Human Genetics*, 2009, 84(2): 210-223.
- [20] 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): 164.
- [21] DANECEK P, AUTON A, ABECASIS G, et al. The variant call format and VCFtools[J]. *Bioinformatics*, 2011, 27(15): 2 156-2 158.
- [22] SZPIECH Z A. Selscan 2.0: Scanning for sweeps in unphased data[J]. *Bioinformatics*, 2024, 40(1): 6.
- [23] DAXINGER L, WHITELAW E. Understanding trans-generational epigenetic inheritance via the gametes in mammals[J]. *Nature Reviews Genetics*, 2012, 13(3): 153-162.
- [24] CHEN J, NI P Y, WU S W, et al. Prevalent use and evolution of exonic regulatory sequences in the human genome[J]. *Natural Sciences*, 2023, 3(2): 20 220 058.
- [25] WEI C H, WANG H H, LIU G, et al. Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds[J]. *BMC Genomics*, 2015, 16(1): 194.
- [26] 国家家养动物种质资源库[EB/OL]. [2025-02-14]. <https://cdad-is.org.cn/admin/Commdscribe/view?id=384>.