

阿勒泰白头牛基因组结构变异分析

惠科铭¹, 李莎¹, 曹琴琴¹, 陈秋明¹, 郭杨²

(1. 新疆农业大学动物科学学院, 乌鲁木齐 830052; 2. 新疆维吾尔自治区畜牧总站, 乌鲁木齐 830004)

摘要:阿勒泰白头牛是新疆阿勒泰地区的乳肉役兼用地方品种, 具有白头、抗病力强和性情温顺特点。结构变异(structure variations, SVs)是影响牛表型性状的重要遗传变异形式。本研究旨在获得阿勒泰白头牛基因组 SVs 的特征和种质特征的候选基因。以 20 头阿勒泰白头牛测序深度为 17.32X 的基因组序列为基础, 通过 Delly、Lumpy 和 Manta 三个软件进行 SVs 的检测, 并通过 SURVIVOR 软件取交集进行整合, 总共鉴定到 37847 个 SVs。从类型看, 缺失、易位、倒位、复制和插入的比例分别为 55.5%、26.7%、8.7%、8.9% 和不足 0.1%。从长度看, 缺失、倒位和复制的分布呈现右偏态。以频率大于 0.9 为标准, 获得 636 个高频 SVs, 涉及 865 个蛋白编码基因。基因注释和富集分析发现, 阿勒泰白头牛的高频 SVs 与免疫抗病和神经功能相关, 契合了抗病力强和性情温顺特征, 可作为相关特征的候选基因(免疫抗病: *BoIA-DQB*、*TNIP3*、*IL1RAP* 和 *IL1R1*, 神经性情: *EPHA6*、*GRM7* 和 *HTR2A*)。值得注意的是, 我们发现 *ASIP* 基因 5' 非编码区存在一个缺失, 该基因可减少真黑素的合成, 在哺乳动物毛色调控中起着重要作用。研究的结果可以为阿勒泰白头牛种质特征的遗传机制解析奠定基础, 也可以为阿勒泰白头牛的资源利用提供基础资料。

关键词:阿勒泰白头牛; 基因组; 结构变异; 种质特征

中图分类号: S823 **文献标识码:** A **文章编号:** 1001-9111(2023)03-0001-04

阿勒泰白头牛是新疆阿勒泰地区的乳肉役兼用地方品种, 主要分布于布尔津、吉木乃和哈巴河等县, 其典型特征是头部呈现白色, 兼具耐粗饲、抗病力强等特点^[1]。目前该品种的研究非常少。以其为关键词在中国知网进行搜索, 只显示 16 篇文献。

在分子遗传学方面, 微卫星标记的方法发现阿勒泰白头牛属于普通牛亚种^[2], 事实上, 牛基因组问世后, 牛的遗传学研究进入后基因组时代, 包括基因芯片在内的各种高通量技术应运而生。Liu 等利用 100K SNP 芯片, 获得了阿勒泰白头牛观测杂合度和近交系数等遗传多样性指标^[3]。随着测序成本的降低, 目前群体水平的全基因组测序已经在很多品种上得以实施。截至 2023 年 6 月, NCBI SRA 数据库已经释放了超过 150 个品种 8200 个个体的

全基因组序列。最近, chen 等利用基因组数据, 发现阿勒泰白头牛主要以东亚普通牛和欧洲普通牛血统为主, 并发现 *EPB41L5*、*SCG5* 和 *KIT* 可能是环境适应性和白头特征的种质基因^[4]。

全基因组序列的另一个应用场景为结构变异(structure variations, SVs)检测。SVs 是遗传变异的重要形式之一, 为五十到几百万碱基序列的变异, 包含不平衡的拷贝数变异(Copy Number Variations, CNVs)(缺失、重复和插入)以及平衡的重排(倒位、染色体间和染色体内易位)。与容易检测的 SNP 相比, SVs 影响了更多核苷酸, 在遗传多样性和进化中具有重要作用。另一方面, SVs 可以通过改变基因剂量、改变基因结构、暴露隐性等位基因和改变基因调控等特定机制对表型产生影响^[5], 如 212 bp 的重

收稿日期: 2023-05-06 修回日期: 2023-05-22

基金项目: 新疆维吾尔自治区自然科学基金资助项目(2022D01A64); 新疆农业大学 2023 年大学生创新项目(dxscx2023180)

作者简介: 惠科铭(2002—), 男, 本科生, 研究方向为动物遗传育种。

* 通讯作者: 郭杨(1989—), 女, 高级畜牧师, 研究方向为动物遗传育种。

复导致安格斯牛无角^[6]。

本研究以阿勒泰白头牛的基因组序列为基础,通过 Delly、Manta 和 Lumpy 软件进行 SVs 的检测,获得 SVs 的特征,并通过筛选高频 SVs,进行相应的注释,获得阿勒泰白头牛种质特征的候选基因,为解析阿勒泰白头牛种质特征的遗传机制提供基础资料。

1 材料与方法

1.1 试验材料

从西北农林科技大学雷初朝教授实验室获得 20 头阿勒泰白头牛的基因组数据,该批数据样本通过新疆畜牧科学院董红研究员联系采自新疆阿勒泰地区布尔津县禾木乡阿勒泰白头牛核心群,并且已经释放可以开放获取^[4]。

1.2 基因组数据的质控与比对

使用 Trimmomatic 软件^[7]对测序下机的原始读段(raw reads)进行了质控处理获得清洗读段(clean reads)。使用 BWA MEM 算法^[8]将 clean reads 比对到参考基因组(ARS - UCD1.2);使用 PICARD 软件的 SortSam 和 MarkDuplicates 模块将比对结果进行排序和去除 PCR 重复序列。最后使用 Qualimap 软件^[9]进行比对率等参数的计算。

1.3 SVs 检测与注释

首先,分别使用 Delly 软件^[10]、Lumpy 软件^[11]和 Manta 软件^[12]进行 SVs 的检测和分型。其后,使用 SURVIVOR 软件^[13]对前述三种工具的结果进行整合。以家牛参考基因组的注释文件为基础,利用 ANNOVAR 软件^[14]进行 SVs 的注释。以注释的蛋白编码基因集合为基础,使用 DAVID 网站^[15]进行京都基因与基因组百科全书(KEGG)和基因本体(GO)数据库的富集分析。

2 结果与分析

2.1 基因组数据描述

20 头阿勒泰白头牛基因组序列质控产生了 6 423 956 161 条 clean reads 和 955.78 Gb 碱基。比对到参考基因组后的比率为 99.83%,深度为 17.32 X。

2.2 SVs 的鉴定

利用 Delly、Lumpy 和 Manta 三款软件检测到 37 847 个 SVs。从染色体分布来看,数量最高的是 1 号染色体(2192),最低的是 25 号染色体。从类型上看,缺失占比超过 50%,最少的为插入,不足 0.1%。从长度来看,缺失、倒位和复制的分布都呈现右偏态,即长度越大,数量越少(图 1)。

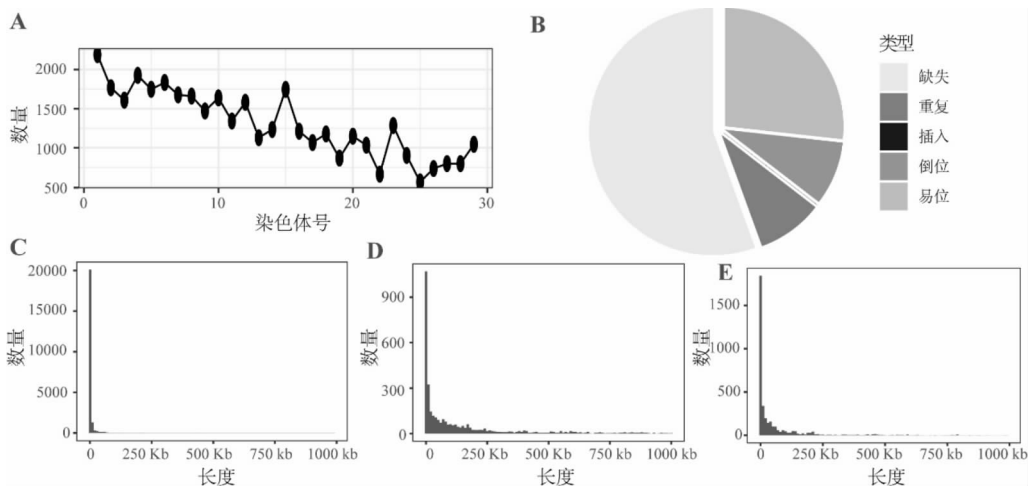


图 1 SVs 特征

A:SVs 的染色体分布;B:SVs 的类型;C:缺失长度的直方图;D:倒位长度的直方图;E:复制长度的直方图

2.3 高频 SVs 的功能注释

以频率大于 0.9 为标准,获得 636 个高频 SVs,涉及 865 个蛋白编码基因,如 *TNIP3* 基因的外显子缺失, *BOLA - DQB* 基因的外显子复制, *EPHA6*、*IL1RAP*、*IL1R1*、*HTR2A* 和 *GMR7* 等基因的内含子缺失, *ASIP* 基因 5'非编码区缺失。

通过对阿勒泰白头牛高频 SVs 区间的基因进行了 GO 富集分析,发现 10 个 GO 条目显著富集,这 10 个 GO 条目中,除通过质膜粘附分子实现的嗜同性细胞粘附和黏着连接与免疫疾病功能相关外,其余都与神经功能相关。同时,KEGG 富集分析发现 3 条显著的通路都与神经功能相关(表 1)。

表1 SVs 区域基因的富集

类别	名称	基因数	P	FDR
BP	通过质膜粘附分子实现的嗜同性细胞粘附	19	3.67×10^{-7}	8.04×10^{-4}
CC	谷氨酸能突触	23	3.59×10^{-6}	1.53×10^{-3}
CC	受体复合体	20	9.29×10^{-7}	1.98×10^{-3}
CC	突触	26	1.15×10^{-4}	1.63×10^{-2}
CC	黏着连接	13	3.58×10^{-4}	3.39×10^{-2}
CC	细胞质	157	4.09×10^{-4}	3.39×10^{-2}
CC	连环蛋白复合体	7	4.78×10^{-4}	3.39×10^{-2}
MF	钙粘蛋白结合	11	1.08×10^{-5}	7.41×10^{-3}
MF	钙离子结合	42	2.13×10^{-5}	7.41×10^{-3}
MF	PDZ 结构域结合	10	9.34×10^{-5}	2.16×10^{-2}
KEGG	钙信号通路	27	1.03×10^{-6}	3.07×10^{-4}
KEGG	肾素分泌	11	2.48×10^{-4}	3.69×10^{-2}
KEGG	神经激活配体-受体相互作用	29	3.83×10^{-4}	3.83×10^{-2}

3 讨论

3.1 SVs 的特征

由于影响的核苷酸较多,SVs 成为遗传变异检测的重要方向。基于高通量的基因组序列,目前已经开发出很多算法进行 SVs 的检测。Delly 集成了读段配对(Read - Pair, RP)、拆分读段(Split - read, SR)和读段深度(Read - depth, RD)算法进行 SVs 鉴定^[10]。Lumpy 整合 RP、SR、RD 和其他先验知识检测基因组的 SVs^[11]。Manta 软件以 RP 和 SR 为基础进行 SVs 的检测^[12]。本研究利用这三款软件,以取交集的方法检测阿勒泰白头牛的基因组 SVs,从而减少假阳性的发生。

在获得的 37 847 个 SVs 中,1 号染色体数量最多,随着染色体号的增加而出现降低的趋势,这与哺乳动物染色体的命名规则有关(染色体较大,编号较小)。事实上,单核苷酸多态性在染色体上的数量分布也有此趋势^[16]。从类型上看,缺失占比超过 50%,而插入不到 0.1%,这是由于二代测序技术产生的读长较短,难以检测大片段插入所致。从长度来看,缺失、复制和倒位的分布都呈现右偏态,这与基因组的保守性相关,大片段的变异容易引起表型的剧烈变化,在中性进化的理论中,适应性较低,不易存活。事实上,本研究得到的类型和长度分布趋势与此前的报道的结果是一致的^[17],提示分析结果的可靠程度高。

3.2 阿勒泰白头牛种质特征的候选 SVs

阿勒泰白头牛最显著的特征是头部白色。ASIP 基因编码刺鼠信号蛋白,可以通过阻断黑色素皮质激素受体的合成,进而减少真黑素和增加褐黑素的产生^[18]。在内洛尔牛上,发现 ASIP 基因的 SV 与毛色的深浅有关^[19]。我们的研究发现,阿勒泰白头牛 5' 端非编码区存在 8 399 bp 的缺失,这个缺失可以作

为该品种毛色特征的候选基因。

作为地方品种,阿勒泰白头牛具有抗病能力强的特点。本研究发现 *Bola - DQB* 基因的外显子存在复制,该基因编码主要组织相容性复合体,其多态性与中国西门塔尔牛和三河牛的乳房炎发生有关^[20]。在 *TNIP3* 基因的外显子上,我们发现一个缺失,该基因为 LPS 诱导的 NF - kB 激活负调节因子,其 mRNA 表达量在重度抑郁症患者 TNF - α 分泌细胞中显著高于健康对照组^[21]。我们在 *IL1RAP* 基因的内含子上,也发现一个缺失,该基因为 *IL1* 型受体复合物的亚基,在健康奶牛子宫组织中的相对表达量显著高于患子宫内膜炎奶牛子宫组织^[22]。此外,我们在 *IL1R1* 基因的内含子上也发现一个缺失,该基因编码白介素 1 受体 I 蛋白,在抗病性好的瘤牛与其杂交牛的外周血单核细胞中存在显著差异^[23]。在人类的研究中,发现该基因的遗传变异与先兆子痫有关^[24]。因而,本研究发现的这些 SVs 可以作为阿勒泰白头牛抗病能力强的候选基因。事实上,本研究的富集分析结果也检测到 2 个显著的与免疫疾病相关的 GO 条目,从侧面反映了阿勒泰白头牛的免疫抗病特性。

阿勒泰白头牛这类地方品种的役用特点对中华文明的诞生和发展起着重要作用,该特点提示地方品种在性情方面比较温和。我们的研究发现 *HTR2A* 基因的内含子区存在 87 bp 的缺失,该基因编码 5 - 羟色胺 2A 受体蛋白,其多态性与精神分裂症的治疗效果有关^[25]。在 *EPHA6* 基因的内含子上,我们也发现一个缺失,该基因编码酪氨酸蛋白激酶受体,在人类该基因的易位,导致智力障碍和神经发育紊乱^[26]。此外,我们也在 *GRM7* 基因的内含子上发现缺失,该基因编码谷氨酸受体蛋白,其拷贝数变异与人类的精神分裂症相关^[27]。因此,本研究发现的这些 SVs 可以作为阿勒泰白头牛役用、温顺特

性的候选基因。事实上,本研究在富集分析中发现的显著 GO 条目和 KEGG 通路大多与神经功能相关,也能从侧面反映阿勒泰白头牛的役用、温顺特性。

4 结论

本研究鉴定的 37 847 个基因组 SVs 丰富了阿勒泰白头牛的遗传变异资料,与免疫疾病、神经性情相关的高频 SVs 可以作为阿勒泰白头牛抗病能力强、役用性温的候选基因,ASIP 基因 5'非编码区的缺失可以作为阿勒泰白头牛白头典型特征的候选基因,这些候选基因的鉴定为解析相关种质特征的遗传机制奠定基础。

参考文献:

- [1] 吾热力哈孜,别克木汗,巴合提,阿达力,沙肯,哈提帕.阿勒泰白头牛简介[J].黑龙江畜牧兽医,2004,(05):20-21.
- [2] 罗永发,王志刚,李加琪等.采用微卫星标记分析 13 个中外牛品种的遗传变异和品种间的遗传关系[J].生物多样性,2006(06):498-507.
- [3] LIU B, TAO W, FENG D, et al. Revealing Genetic Diversity and Population Structure of Endangered Altay White-Headed Cattle Population Using 100 k SNP Markers[J]. Animals, 2022, 12(22): 3214.
- [4] CHEN J, WANG Y, QI X, et al. Genome-wide analysis emancipates genomic diversity and signature of selection in Altay white-headed cattle of Xinjiang, China[J]. Frontiers in Genetics, 2023, 14: 1144-1249.
- [5] WEISCHENFELDT J, SYMMONS O, SPITZ F, et al. Phenotypic impact of genomic structural variation; insights from and for human disease[J]. Nature Reviews Genetics, 2013, 14(2): 125-138.
- [6] CARLSON D F, LANCTO C A, ZANG B, et al. Production of hornless dairy cattle from genome-edited cell lines[J]. Nature biotechnology, 2016, 34(5): 479-481.
- [7] BOLGER A M, LOHSE M, USADEL B. Trimmomatic: a flexible trimmer for Illumina sequence data[J]. Bioinformatics, 2014, 30(15): 2114-2120.
- [8] LI H, DURBIN R. Fast and accurate short read alignment with Burrows - Wheeler transform [J]. bioinformatics, 2009, 25(14): 1754-1760.
- [9] GARCÍA-ALCALDE F, OKONECHNIKOV K, CARBONELL J, et al. Qualimap: evaluating next-generation sequencing alignment data[J]. Bioinformatics, 2012, 28(20): 2678-2679.
- [10] RAUSCH T, ZICHNER T, SCHLÄTTL A, et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis[J]. Bioinformatics, 2012, 28(18): i333-i339.
- [11] LAYER R M, CHIANG C, QUINLAN A R, et al. LUMPY: a probabilistic framework for structural variant discovery[J]. Genome biology, 2014, 15(6): 1-19.
- [12] CHEN X, SCHULZ-TRIEGLAFF O, SHAW R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications[J]. Bioinformatics, 2016, 32(8): 1220-1222.
- [13] JEFFARES D C, JOLLY C, HOTI M, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast[J]. Nature communications, 2017, 8(1): 1-11.
- [14] WANG K, LI M, HAKONARSON H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data[J]. Nucleic acids research, 2010, 38(16): e164-e164.
- [15] HUANG D W, SHERMAN B T, TAN Q, et al. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists[J]. Nucleic acids research, 2007, 35(suppl_2): W169-W175.
- [16] DE DONATO M, PETERS S O, MITCHELL S E, et al. Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing [J]. PloS one, 2013, 8(5): e62137.
- [17] UPADHYAY M, DERKS M F L, ANDERSSON G, et al. Introgression contributes to distribution of structural variations in cattle [J]. Genomics, 2021, 113(5): 3092-3102.
- [18] LETKO A, AMMANN B, JAGANNATHAN V, et al. A deletion spanning the promoter and first exon of the hair cycle - specific ASIP transcript isoform in black and tan rabbits[J]. Animal genetics, 2020, 51(1): 137-140.
- [19] TRIGO B B, UTSUNOMIYA A T H, FORTUNATO A A A D, et al. Variants at the ASIP locus contribute to coat color darkening in Nellore cattle [J]. Genetics Selection Evolution, 2021, 53(1): 1-12.
- [20] 高树新,许尚忠,李金泉等.2 种兼用型牛 BoLA-DQB · exon2 基因的多态性与乳房炎的相关性[J].西北农林科技大学学报(自然科学版),2008,210(03):8-12.
- [21] HUANG K W, WU M K, HUNG Y Y. Elevated TNIP3 mRNA Expression in TNF- α -Secreting Cells from Patients with Major Depressive Disorder[J]. Neuroimmunomodulation, 2019, 26(3): 153-158.
- [22] 陈秀竹,常迪,黄启震等. IL1RAP 在健康奶牛与患子宫内膜炎奶牛子宫内组织间表达差异的研究[J].北京农学院学报,2019,34(02):77-81.
- [23] KUMAR A, KUMAR GAUR G, PANIGRAHI M, et al. Comparative gene expression profile in circulating PBMCs of Bos indicus and crossbred cattle to understand disease tolerance mechanism [J]. Animal Biotechnology, 2022: 1-9.
- [24] SIVARAJ N, RACHEL K V, SUWARI T K, et al. Association of ILIR1 gene (SNP rs2071374) with the risk of preeclampsia [J]. Journal of Reproductive Immunology, 2022, 149: 103-103.
- [25] 董文涛,纪云鹤. 5-HTR1A, 5-HTR2A 基因多态性与 SSRIs 抗抑郁疗效的关系 [J]. 中国医师杂志, 2021, 23(12): 1857-1859.
- [26] SCHNEIDER A, PUECHBERTY J, NG B L, et al. Identification of disrupted AUTS2 and EPHA6 genes by array painting in a patient carrying a de novo balanced translocation t(3;7) with intellectual disability and neurodevelopment disorder[J]. American Journal of Medical Genetics Part A, 2015, 167(12): 3031-3037.
- [27] 赵亚丽,张克让,许琪等.代谢型谷氨酸受体 7 基因的拷贝数变异与精神分裂症的相关性[J].中国医学科学院学报, 2009,31(06):664-668.

- [6] 王景山,马万欣,孙雁征,李昊,何洋,何曙光,石顺利. 通辽市 2018 年—2020 年良种肉牛推广体系建设目标[J]. 中国牛业科学,2017,43(06):70-72.
- [7] 马云,谷帅锋,潘翠丽等. 宁夏回族自治区肉牛种业现状及问题及对策建议[J]. 中国畜禽种业,2023,19(02):9-13.
- [8] 王进宽. 宁夏原州区肉牛产业发展现状及建议[J]. 养殖与饲料,2021,20(07):129-130.
- [9] 蔡翠翠,丁晓婷,万鹏,等. 宁夏固原地区肉牛养殖业面临的挑战以及未来发展趋势[J]. 中国牛业科学,2020,46(05):73-75.
- [10] 王亚书. 肉牛疫病防控中存在的问题及建议[J]. 中国畜牧业,2022(13):93.
- [11] 高旭红,谢建亮,侯鹏霞,等. 宁夏肉牛产业现状及需求的调研报告[J]. 宁夏农林科技,2021,62(04):62-65.
- [12] 韩振,杨春. 美国肉牛产业发展及对我国的启示[J]. 中国畜牧杂志,2018,54(06):143-147.
- [13] 单文心,吴迪,舒坤良. 吉林省肉牛产业发展问题及对策研究[J]. 东北农业科学,2018,43(04):62-64.

The Current Situation, Problems and Suggestions of Ningxia Beef Cattle Industry Chain

FENG Lan¹, GU Shuai-feng¹, PAN Cui-li^{1,2}, WEI Da-wei¹, ZHANG Ling-kai¹, CAI Bei¹, MA Yun^{1*}

(1. Key Laboratory of Ruminant Molecular Cell Breeding in Ningxia Hui Autonomous Region, College of Animal Science and Technology, Ningxia University, Yinchuan 750021; 2. Ningxia Vocational and Technical College, College of Life Science and Technology, Yinchuan 750021)

Abstract: Beef cattle industry is an important part of China's agriculture, and also one of the "six special" industries determined by Ningxia Hui Autonomous Region. The development of beef cattle industry plays a vital role in the economic growth of Ningxia. From the perspective of industrial chain, this paper comprehensively analyzed the current situation and problems of beef cattle related industries such as genetic breeding, disease prevention and control, feed nutrition, breeding and fattening, production and processing, and marketing, and put forward suggestions for industrial development, providing a reference for the high-quality coordinated development of the multiple industries related to beef cattle in Ningxia Hui Autonomous Region.

Key words: Ningxia; beef cattle industry chain; varieties; seed production system

(上接第 4 页)

Identification of Genomic Structure Variation in Itay White-headed Cattle

HUI Ke-ming¹, LI Sha¹, CAO Qin-qin¹, CHEN Qiu-ming¹, GUO Yang^{2*}

(1. College of Animal Science, Xinjiang Agricultural University, Urumqi 830052, China;
2. Xinjiang Region Animal Husbandry Station, Urumqi 830004, China)

Abstract: Altay white-headed cattle, an indigenous breed for dairy, beef and draft purpose in the Altay region of Xinjiang, China, displays white head, strong disease resistance and gentle temperament. Structural variation (SV) is an important genetic variation that affect phenotypic traits in cattle. This study aims to characterize the SVs in the genome of Altay white cattle and identify candidate genes associated with their phenotypic features. Based on the whole genome sequences of 20 Altay white-headed cattle with a sequencing depth of 17.32X, SVs were detected using three software tools: Delly, Lumpy and Manta. Based on the intersection of two or three strategies, the SVs were further integrated using SURVIVOR software, resulting in a total of 37,847 SVs being identified. In terms of type, the proportions of deletion, translocation, inversion, duplication, and insertion were 55.5%, 26.7%, 8.7%, 8.9% and less than 0.1%, respectively. Regarding the length distribution, deletions, inversions, and duplications showed a right-skewed pattern. Using a frequency threshold of greater than 0.9, we obtained 636 high-frequency SVs, which involved 865 protein-coding genes. Gene annotation and enrichment analysis revealed that the high-frequency SVs in Altay white cattle were associated with immune response, disease resistance and neural functions, aligning with the breed's characteristics of strong disease resistance and gentle temperament. These genes, such as *Bola-DQB*, *TNIP3*, *IL1RAP* and *IL1R1* (immune response), *EPHA6*, *GRM7* and *HTR2A* (neural temperament), may serve as candidate genes related to these traits. It is worth noting that we identified a deletion in the 5' non-coding region of the *ASIP* gene, which plays an important role in regulating melanin synthesis and is involved in mammalian coat color determination. The findings of this study provide insights into the genetic mechanisms underlying the phenotypic traits of Altay white-headed cattle and serve as fundamental information for the utilization and conservation of this breed.

Key words: Altay white-headed cattle, genome, structure variations, germplasm characteristics