

盛欣, 徐宇辉, 郑文骞, 等. 四个牛族物种染色体水平基因组完全型微卫星分布特征分析[J]. 江苏农业学报, 2026, 42(1): 110-119.
doi: 10.3969/j.issn.1000-4440.2026.01.012

四个牛族物种染色体水平基因组完全型微卫星分布特征分析

盛欣^{1,2,3}, 徐宇辉^{1,2,3}, 郑文骞^{1,2,3}, 陈生梅^{1,2,3}, 李瑞哲^{1,2,3}, 王贵元⁴, 马志杰^{1,2,3}
(1. 青海大学畜牧兽医科学院, 青海 西宁 810016; 2. 农业农村部青藏高原畜禽遗传育种重点实验室, 青海 西宁 810016;
3. 青海省高原家畜遗传资源保护与创新利用重点实验室, 青海 西宁 810016; 4. 天峻县畜牧兽医站, 青海 海西 817200)

摘要: 牛族物种作为农业生产与生态系统中的关键类群, 其遗传资源的保护与可持续利用对畜牧业发展及维持生物多样性具有重要意义。然而, 当前针对牛族物种基因组中完全型微卫星(Perfect microsatellites, P-SSR)的系统比较研究相对匮乏。本研究基于染色体水平的高质量基因组序列, 采用Krait v1.4.0软件对普通牛、牦牛(含家牦牛和野牦牛)、瘤牛、水牛基因组中的P-SSR进行系统检测与综合分析。分别从普通牛、家牦牛、野牦牛、瘤牛及水牛全基因组中鉴定出817 477个、775 897个、740 044个、739 975个、777 401个P-SSR; 其中, 普通牛基因组中P-SSR相对频率最高, 其次是水牛、野牦牛和瘤牛, 家牦牛基因组中P-SSR相对频率最低。在水牛基因组中, P-SSR序列总长度所占比例最高, 普通牛、家牦牛及野牦牛次之, 瘤牛最低。在4个牛族物种的基因组中, 单核苷酸重复类型P-SSR的出现频率最高, 其次为二核苷酸重复、三核苷酸重复、五核苷酸重复、四核苷酸重复、六核苷酸重复。4个牛族物种的单核苷酸重复至五核苷酸重复类型的P-SSR优势重复基序相同, 分别为A、AC、AGC、AAAT和AACTG。在各牛族物种中, 相同重复类型的P-SSR重复基序数量分布区间高度一致。单核苷酸重复至六核苷酸重复类型的重复基序数量分布范围分别是12~28次、7~24次、5~13次、4~8次、4~7次和4~10次。相关性分析结果表明, 4个牛族物种基因组中各条染色体长度与其所含P-SSR数量呈极显著正相关($P < 0.001$), 而各条染色体P-SSR相对频率和相对密度均与其染色体G+C含量呈显著或极显著负相关($P < 0.05$ 、 $P < 0.01$ 、 $P < 0.001$)。长度大于100 bp和1 000 bp的P-SSR数量在家牦牛基因组中最多。综上, 本研究结果系统揭示了各牛族物种基因组中P-SSR的丰度与分布规律, 为深入探究牛族物种基因组中重复序列的组成差异、进化机制及其生物学功能提供了重要依据。

关键词: 牛族; 基因组; 完全型微卫星(P-SSR); 丰度; 分布规律

中图分类号: S823 **文献标识码:** A **文章编号:** 1000-4440(2026)01-0110-10

Microsatellite distribution characteristics in chromosome-level genomes of four bovini species

SHENG Xin^{1,2,3}, XU Yuhui^{1,2,3}, ZHENG Wenqian^{1,2,3}, CHEN Shengmei^{1,2,3}, LI Ruizhe^{1,2,3},
WANG Guiyuan⁴, MA Zhijie^{1,2,3}

(1. Academy of Animal Science and Veterinary, Qinghai University, Xining 810016, China; 2. Key Laboratory of Livestock and Poultry Genetics and Breeding on the Qinghai-Tibet Plateau, Ministry of Agriculture and Rural Affairs, Xining 810016, China; 3. Plateau Livestock Genetic Resources Protection and Innovative Utilization Key Laboratory of Qinghai Province, Xining 810016, China; 4. Tianjun Animal Husbandry and Veterinary Station, Haixi 817200, China)

收稿日期: 2025-08-18

基金项目: 青海省“昆仑英才·高端创新创业人才”计划项目(领军人才); 第三次全国畜禽遗传资源普查项目

作者简介: 盛欣(2000-), 女, 广东汕头人, 硕士研究生, 主要从事动物遗传资源研究。(E-mail) 2603043164@qq.com

通讯作者: 马志杰, (E-mail) zhijiema@126.com

Abstract: Bovini species play a crucial role in agricultural production and ecosystem functioning, and the conservation and sustainable utilization of these genetic resources are of great significance for the development of animal husbandry and the maintenance of biodiversity. However, systematic comparative studies on perfect mi-

crossatellites (P-SSRs) in the genomes of bovine species are still scarce. To address this gap, this study conducted systematic detection and comprehensive analysis of P-SSRs in cattle (*Bos taurus*), yak (*Bos grunniens* and *Bos mutus*), zebu (*Bos indicus*), and buffalo (*Bubalus bubalis*) using Krait v1.4.0 software based on high-quality chromosome-level genome sequences. The results showed that a total of 817 477, 775 897, 740 044, 739 975, and 777 401 P-SSRs were identified in the genomic sequences of cattle, domestic yak, wild yak, zebu and buffalo, respectively. Among them, the P-SSRs showed the highest relative frequency in cattle, followed by buffalo, wild yak, and zebu, with the lowest in domestic yak. Moreover, the proportion of the total length of P-SSRs was the highest in buffalo, followed by cattle, domestic yak, and wild yak, with zebu showing the lowest proportion. In the genomic sequences of the four bovine species, the relative frequency of mononucleotide repeats of SSRs was the highest, followed by dinucleotide, trinucleotide, pentanucleotide, tetranucleotide, and hexanucleotide repeats. In all four bovine species, the dominant repeat motifs of P-SSRs in mononucleotide, dinucleotide, trinucleotide, tetranucleotide, and pentanucleotide repeats were identical, specifically A, AC, AGC, AAAT, and AACTG. Among different bovine species, the distribution ranges of repeat motif counts for each corresponding P-SSR type were highly conserved. The distribution ranges for the number of repeat motifs were 12–28 times for mononucleotide, 7–24 times for dinucleotide, 5–13 times for trinucleotide, 4–8 times for tetranucleotide, 4–7 times for pentanucleotide, and 4–10 times for hexanucleotide repeats. Correlation analysis revealed that across the four bovine species, chromosome length showed a highly significant positive correlation with the number of P-SSRs ($P < 0.001$). In contrast, both the relative frequency and relative density of P-SSRs on each chromosome were significantly or highly significantly negatively correlated with G+C content ($P < 0.05$, $P < 0.01$, $P < 0.001$). The number of P-SSRs longer than 100 bp and 1 000 bp was greatest in the domestic yak genome. In conclusion, the findings of this study systematically reveal the abundance and distribution patterns of P-SSRs in bovine genomes, providing a valuable foundation for further exploration of the compositional divergence, evolutionary mechanisms, and biological functions of repetitive sequences in these species.

Key words: bovine; genome; perfect microsatellites (P-SSRs); abundance; distribution pattern

微卫星又称简单序列重复 (SSR) 或短串联重复 (STR), 是由 1~6 个核苷酸为单元组成的串联重复 DNA 序列, 广泛存在于真核生物和原核生物的基因组中^[1-2]。根据核心序列基序的结构特征, 微卫星可以分为 3 类: 完全型微卫星 (P-SSR)、不完全型微卫星 (I-SSR) 和复合型微卫星 (C-SSR)^[3], 其中, 完全型微卫星又称单纯型微卫星^[4]。作为一种重要的分子标记, 微卫星已被广泛应用于群体遗传学的多个研究领域, 包括基因连锁分析^[5]、品种溯源鉴定^[6]、遗传资源保护^[7]、物种行为习性探索^[8]、系统进化分析^[9]等。

牛族 (Bovini) 隶属于牛科 (Bovidae) 牛亚科 (Bovinae), 主要是由野牛属 (*Bison*)、牛属 (*Bos*) 和水牛属 (*Bubalus*) 等类群的物种构成^[10-11], 涵盖对人类具有重要价值的家牛 (普通牛和瘤牛)、水牛和牦牛。牛族物种是哺乳动物中的重要类群, 具有广泛的经济、生态和文化价值^[12-14]。然而, 当前针对牛族物种基因组中 P-SSR 的研究仍存在明显不足: 一方面, 研究缺乏系统性, 现有研究多聚焦于单一物种 (如普通牛、家养牦牛) 或特定染色体基因组 (如 Y

染色体、X 染色体), 仅揭示了部分牛种或局部基因组的微卫星特征^[15-20], 尚未对牛族核心物种 (如普通牛、牦牛、瘤牛、水牛) 的 P-SSR 分布规律进行跨物种系统比较分析, 故难以明确牛种间的基因组遗传差异与进化关联; 另一方面, 前人研究多基于重叠群 (Contig) 或支架序列 (Scaffold) 组装水平的基因组数据进行 P-SSR 分析^[21-25], 其受限的基因组组装连续性对全基因组范围内微卫星分布特征的精准解析会产生影响。

随着高通量测序技术的迅速发展, 普通牛^[26]、瘤牛^[27]、水牛^[28]及牦牛^[29-30]染色体级别的高质量基因组序列均已完成测序, 这为系统比较各牛族物种基因组微卫星的分布特征与差异提供了重要基础。本研究基于普通牛、瘤牛、牦牛 (含家牦牛与野牦牛) 及水牛 4 个牛族物种染色体组装水平的基因组序列 (包括常染色体和 X 染色体), 系统分析各牛种的基因组微卫星分布规律及其差异, 旨在揭示其基因组重复序列的丰度特征, 进而为各牛种遗传资源的保护与可持续利用提供理论基础与数据支撑。

1 材料与amp;方法

1.1 4个牛族物种基因组数据集来源

本研究中4个牛族物种(普通牛、瘤牛、牦牛、水牛)的基因组序列从NCBI数据库(<https://www.ncbi.nlm.nih.gov/>)下载获得(表1),各基因组序列均已组装到染色体水平,将下载的基因组常染色体和X染色体基因组序列数据(表2)以FASTA格式保存以备后续分析。

1.2 4个牛族物种基因组P-SSR搜索及统计分析

使用Krait v1.4.0软件(<https://github.com/lm-du/krait>)对4个牛族物种基因组序列的P-SSR进行

综合分析。参考已有检测标准^[21],本研究将单核苷酸至六核苷酸重复类型的SSR基序最小重复次数分别设定为:12次(单核苷酸重复)、7次(二核苷酸重复)、5次(三核苷酸重复)、4次(四核苷酸重复)、4次(五核苷酸重复)及4次(六核苷酸重复)。此外,将可循环的微卫星序列及其反补序列归为同一类别的微卫星。4个牛族物种各自每条染色体上P-SSR的丰度分析数据保存为tsv格式,利用Excel 2021提取tsv文件中不同重复类型的P-SSR信息,并进行统计分析生成相应的表格数据文件,然后用OriginPro 2024b软件(<https://www.originlab.com/>)对统计结果进行可视化处理。

表1 4个牛族物种基因组序列数据集信息

Table 1 Information of the genome sequences of four bovini species

研究对象	普通牛	牦牛		瘤牛	水牛
		家牦牛	野牦牛		
基因组版本	ARS-UCD 2.0	BosGru 3.1	NWIPB_WYAK_1.0	ASM293397 v1	NDDB_SH_1
基因组序列号	GCF_002263795.3	GCA_005887515.3	GCA_027580195.1	GCA_002933975.1	GCF_019923935.1
基因组大小(Gb)	2.77	2.83	2.63	2.74	2.60
Scaffold N_{50} (Mb)	103.310	114.390	103.900	104.300	117.000
Contig N_{50} (Mb)	26.400	0.045	38.280	0.065	9.600
基因组发布时间(年-月-日)	2023-09-28	2019-06-10	2023-01-09	2018-02-22	2021-09-10

Contig N_{50} :重叠群(Contig)按长度降序累加至基因组总长50%时对应的Contig长度,反映短片段拼接连续性;Scaffold N_{50} :支架序列(Scaffold)按长度降序累加至基因组总长50%时对应的Scaffold长度,反映组装框架连续性。

2 结果与分析

2.1 4个牛族物种基因组P-SSR总数、相对频率和相对密度

从表2可知,本研究中,4个牛族物种常染色体和X染色体基因组大小介于2 571.98~2 796.83 Mb,其中家牦牛的基因组最长;其次是瘤牛、普通牛和水牛,野牦牛的常染色体和X染色体基因组最

小;普通牛基因组中P-SSR的相对频率最高,为311.02 loci/Mb,其次是水牛、野牦牛、瘤牛,家牦牛最低,为277.42 loci/Mb;水牛基因组中P-SSR相对密度最高,为5 267.71 bp/Mb,其次是家牦牛、普通牛、野牦牛,瘤牛基因组中P-SSR相对密度最低,为4 901.34 bp/Mb;水牛基因组中P-SSR占比最高,为0.53%,其次是家牦牛、普通牛、野牦牛,瘤牛基因组中P-SSR占比最低,为0.48%。

表2 4个牛族物种基因组特征及单纯型微卫星(P-SSR)的分布与丰度数据

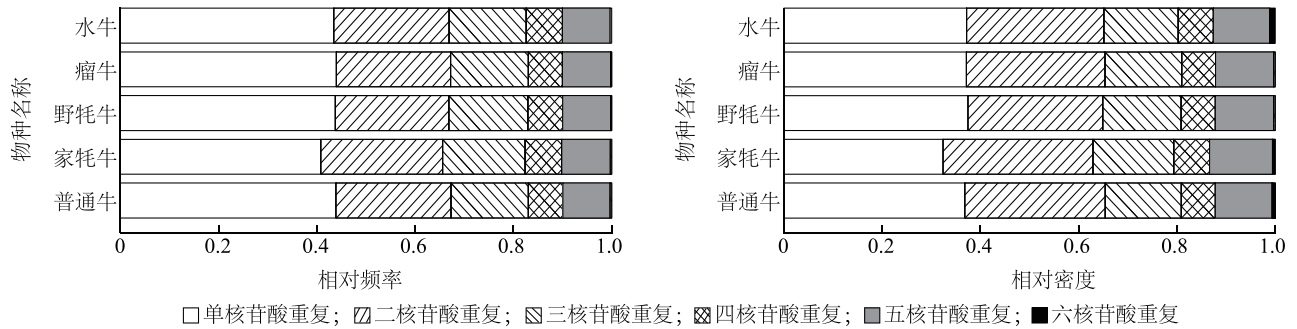
Table 2 Genome characteristics and distribution and abundance data of perfect microsatellites (P-SSRs) in four bovini species

检测指标	普通牛	牦牛		瘤牛	水牛
		家牦牛	野牦牛		
常染色体和X染色体基因组大小(Mb)	2 628.40	2 796.83	2 571.98	2 660.66	2 622.44
P-SSR总数(个)	817 477	775 897	740 044	739 975	777 401
P-SSR总长度(bp)	13 404 458	14 284 921	12 913 750	12 883 966	13 814 273
P-SSR相对频率(loci/Mb)	311.02	277.42	287.73	281.87	296.44
P-SSR相对密度(bp/Mb)	5 099.86	5 107.54	5 020.95	4 901.34	5 267.71
P-SSR占比(%)	0.51	0.51	0.50	0.48	0.53

2.2 4个牛族物种基因组中不同核苷酸重复类型的P-SSR丰度

从图1和表3可知,4个牛族物种基因组中6种重复类型的P-SSR相对频率与相对密度呈现相似的规律性分布,即单核苷酸重复>二核苷酸重复>三核苷酸重复>五核苷酸重复>四核苷酸重复>六核苷酸重复。单核苷酸重复的P-SSR相对频率和相对密度水

牛最高,家牦牛最低;二核苷酸重复的P-SSR相对频率水牛最高,瘤牛最低,相对密度家牦牛最高,瘤牛最低;三核苷酸重复的P-SSR相对频率和相对密度家牦牛最高,瘤牛最低;四核苷酸重复和六核苷酸重复的P-SSR相对频率和相对密度水牛最高,瘤牛最低;五核苷酸重复的P-SSR相对频率水牛最高,瘤牛最低;相对密度家牦牛最高,瘤牛最低。



A:不同核苷酸重复类型的完全型微卫星(P-SSR)相对频率;B:不同核苷酸重复类型的完全型微卫星(P-SSR)相对密度。

图1 4个牛族物种基因组中不同核苷酸重复类型的完全型微卫星(P-SSR)相对占比

Fig.1 Relative proportions of perfect microsatellites (P-SSRs) with different nucleotide repeat types in the genomes of four bovine species

表3 4个牛族物种基因组中不同核苷酸重复类型的完全型微卫星(P-SSR)相对频率和相对密度

Table 3 Relative frequency and density of perfect microsatellites (P-SSRs) with different nucleotide repeat types in the genomes of four bovine species

核苷酸重复类型	统计指标	普通牛	牦牛		瘤牛	水牛
			家牦牛	野牦牛		
单核苷酸重复	相对频率(loci/Mb)	127.49	113.18	125.90	121.10	128.77
	相对密度(bp/Mb)	1 878.78	1 654.58	1 879.90	1 780.56	1 958.60
二核苷酸重复	相对频率(loci/Mb)	68.00	68.76	66.49	64.10	69.60
	相对密度(bp/Mb)	1 455.43	1 558.76	1 379.23	1 354.72	1 472.20
三核苷酸重复	相对频率(loci/Mb)	45.73	46.64	46.29	43.51	46.49
	相对密度(bp/Mb)	791.97	838.85	799.72	749.55	797.32
四核苷酸重复	相对频率(loci/Mb)	20.47	20.66	20.39	19.19	21.87
	相对密度(bp/Mb)	352.13	373.06	350.53	328.84	377.11
五核苷酸重复	相对频率(loci/Mb)	28.03	27.39	28.11	27.03	28.77
	相对密度(bp/Mb)	590.12	655.19	595.99	567.80	605.20
六核苷酸重复	相对频率(loci/Mb)	0.84	0.79	0.56	0.55	0.93
	相对密度(bp/Mb)	31.43	27.09	15.57	15.27	57.28

2.3 4个牛族物种基因组中不同核苷酸重复基序类别的P-SSR频率

将相对频率大于1.00 loci/Mb的P-SSR类别定义为高频核苷酸重复基序类别进行统计分析,结果(表4)发现,有14种高频核苷酸重复基序类别的P-

SSR是4个牛族物种共有的,分别为A、AC、AG、AT、AAC、AAT、ACC、AGC、AAAC、AAAG、AAAT、AATG、AACTG、AGATC。单核苷酸重复类型的P-SSR有A和C2种重复基序类别,其中A的相对频率在牦牛、普通牛、瘤牛及水牛中均占据主导,在单

核苷酸中的占比均超过 96%。在二核苷酸重复类型的 P-SSR 中,AC 为优势重复基序类别,相对频率介于 37.86~41.91 loci/Mb,其次为 AT、AG。三核苷酸重复类型的 P-SSR 中以 AGC 为优势重复基序类别(相对频率为 32.88~34.37 loci/Mb),其次为 AAC、AAT、ACC。四核苷酸重复类型的 P-SSR 中相对频率排前 3 位的重复基序类别分别是 AAAT、AAAC、AAAG;五核苷酸重复类型的 P-SSR 中重复基序类别相对频率由高到低分别是 AACTG、AG-ATC、AAAAC。六核苷酸重复类型的 P-SSR 总体占比较低,其重复基序类别相对频率均小于 1.00 loci/Mb,其中普通牛、瘤牛、家牦牛和水牛的优势六核苷酸重复单元为 AACCT,野牦牛为 AAAAAC。

表 4 4 个牛族物种全基因组中不同核苷酸重复基序类别的相对频率

Table 4 Relative frequencies of different nucleotide repeat motif categories in genomes of four bovini species

重复基序类别	相对频率(loci/Mb)				
	普通牛	家牦牛	野牦牛	瘤牛	水牛
A	124.18	112.41	124.32	117.48	124.63
C	3.31	-	1.58	2.60	4.15
AC	40.80	41.91	40.39	37.86	41.38
AG	6.49	6.36	6.26	6.19	6.62
AT	20.52	20.38	19.70	19.36	21.40
AAC	3.30	3.62	3.61	3.17	3.84
AAT	2.69	2.75	2.69	2.59	2.88
ACC	1.76	1.89	1.83	1.47	1.86
AGC	34.25	34.03	34.37	32.88	34.21
AGG	-	1.47	1.02	-	-
CCG	-	1.04	-	-	-
AAAC	3.37	3.39	3.36	3.27	3.85
AAAG	1.77	1.76	1.72	1.71	1.90
AAAT	5.60	5.67	5.58	5.34	6.25
AATG	1.52	1.57	1.54	1.47	1.55
ATCC	1.31	1.30	1.29	-	1.33
AAAAC	-	1.04	1.02	-	1.32
AACTG	19.81	19.33	19.90	18.89	20.13
AGATC	4.44	4.25	4.42	4.22	4.35

-表示该核苷酸重复基序类别的相对频率低于 1.00 loci/Mb。

2.4 P-SSR 在 4 个牛族物种染色体上的分布频率和密度

统计结果(图 2)显示,在 4 个牛族物种中,1 号染色体含有的 P-SSR 数量最多。其中,水牛 P-SSR 数

量位列前 3 的染色体为 1 号染色体、2 号染色体和 3 号染色体。除水牛外,其他牛种 P-SSR 数量位列前 3 的染色体分别是 1 号染色体、X 染色体和 2 号染色体。普通牛和瘤牛的 25 号染色体含有的 P-SSR 数量最少,家牦牛和野牦牛的 29 号染色体 P-SSR 数量最少,水牛 24 号染色体 P-SSR 数量为最少。

相关性分析结果表明,牦牛各染色体长度与其所含 P-SSR 数量之间存在极显著的正相关关系(Pearson 相关系数 0.992~0.995, $P < 0.001$; Spearman 相关系数 0.983~0.994, $P < 0.001$),说明染色体越长,其所含 P-SSR 的数量相应越多。同时,4 个牛种中,各条染色体 G+C 含量与其所拥有的 P-SSR 相对频率及相对密度之间的关联性分析结果显示,染色体 G+C 含量与其 P-SSR 的相对频率(Pearson 相关系数 -0.620~-0.531, $P < 0.01$; Spearman 相关系数 -0.650~-0.648, $P < 0.05$)和相对密度(Pearson 相关系数 -0.652~-0.548, $P < 0.001$; Spearman 相关系数 -0.729~-0.630, $P < 0.01$)存在显著或极显著负相关。

2.5 4 个牛族物种基因组中不同重复类型 P-SSR 的重复基序数量

从图 3 可见,相同重复类型的 P-SSR 在 4 个牛族物种中均表现出相似的重复基序数量分布区间,并呈现一致的变化规律,即随着重复基序数量的增加,P-SSR 数量呈递减趋势。具体而言,单核苷酸重复与二核苷酸重复类型的 P-SSR 重复基序数量分布范围较宽,单核苷酸重复类型的重复基序数量主要集中于 12~28 次,二核苷酸重复类型的重复基序数量主要集中于 7~24 次;三核苷酸重复类型的重复基序数量多分布于 5~13 次;四核苷酸、五核苷酸及六核苷酸重复类型的重复基序数量分布范围相对较窄,分别集中于 4~8 次、4~7 次和 4~10 次。

对基因组长度大于 100 bp 的 P-SSR 分析结果(表 5、图 4)显示,家牦牛 P-SSR 数量最多,共有 1 171 个,总长度 896 830 bp,而瘤牛 P-SSR 数量最少,仅有 35 个,总长度为 5 665 bp。从表 5 可见,4 个牛族物种基因组中长度 >100 bp 的 P-SSR 数量排序依次为:家牦牛>水牛>野牦牛>普通牛>瘤牛,P-SSR 总长度排序则依次为:家牦牛>水牛>普通牛>野牦牛>瘤牛。从图 4 可见,在单核苷酸重复 P-SSR 至五核苷酸重复 P-SSR 之间,家牦牛长度大于 100 bp 的 P-SSR 相对频率和相对密度明显高于其他牛种。在六核苷酸重复 P-SSR 中,水牛中长度大于

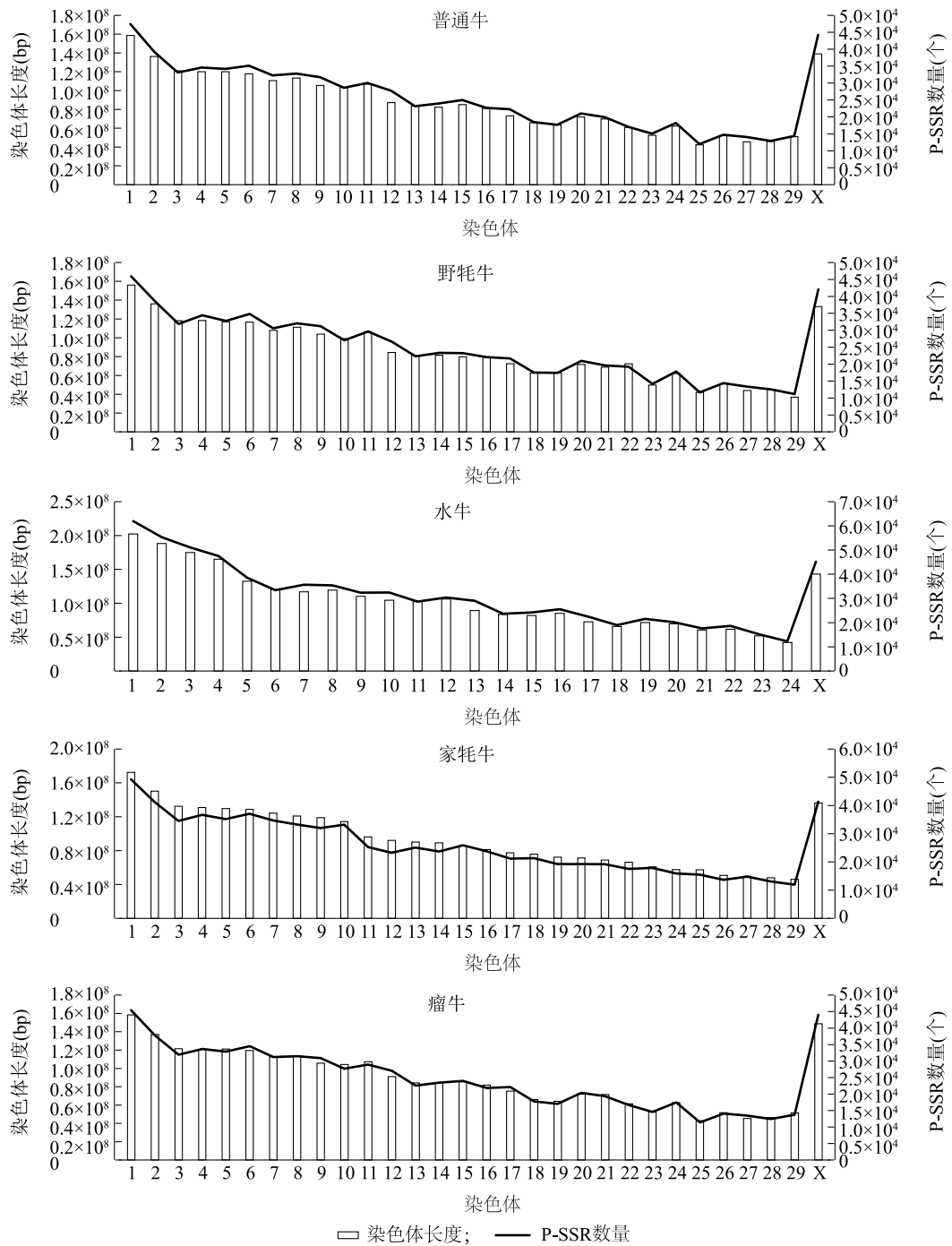


图2 4个牛族物种染色体长度与其所含 P-SSR 数量

Fig.2 Chromosome length and the number of perfect microsatellites (P-SSRs) in four bovini species

100 bp 的 P-SSR 相对频率和相对密度最高。野牦牛中长度大于 100 bp 的 P-SSR 相对频率和相对密度的优势核苷酸重复类型分别是二核苷酸重复和五核苷酸重复,普通牛和水牛中优势核苷酸重复类型均为六核苷酸重复,而家牦牛和瘤牛的优势核苷酸

重复类型均为二核苷酸重复。

对基因组长度大于 1 000 bp 的 P-SSR 分析结果(表 5)显示,家牦牛 P-SSR 数量(353 个)和总长度(621 540 bp)最高,而瘤牛 P-SSR 数量和总长度均为 0。从表 5 可知,P-SSR 数量的排序依次为:家

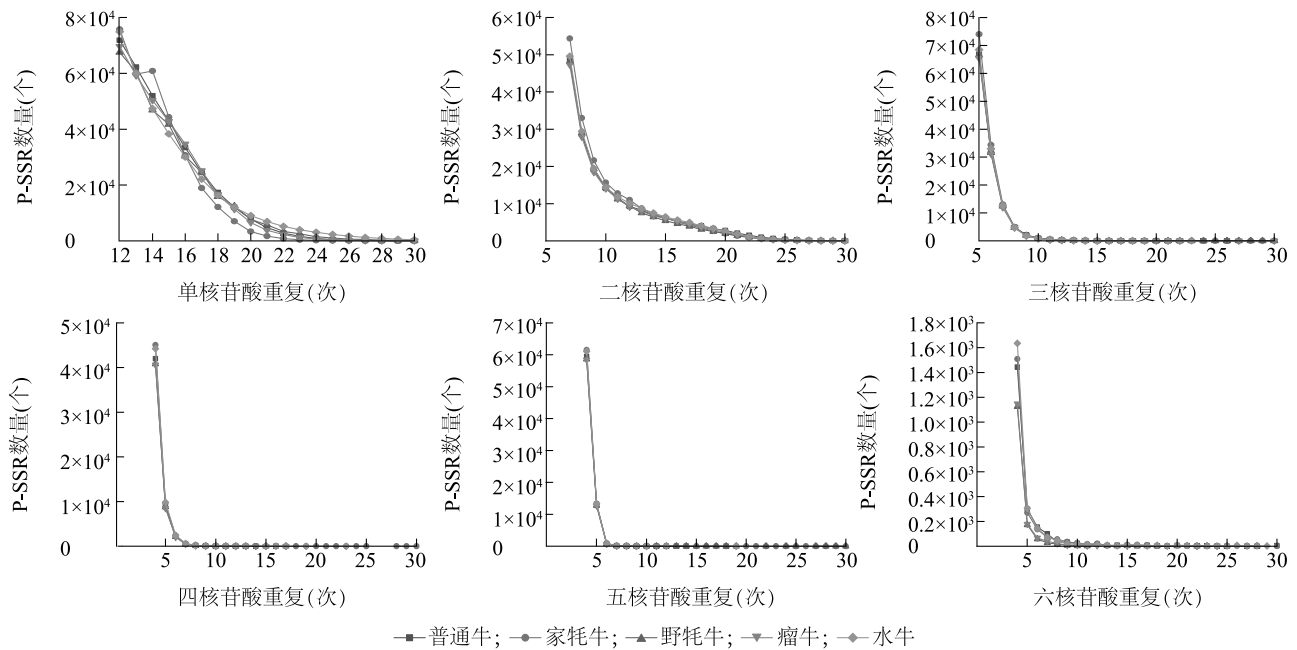


图3 4个牛族物种基因组中不同重复类型不同重复次数的完全型微卫星(P-SSR)数量

Fig.3 Number of perfect microsatellites (P-SSRs) with different repeat types and repeat times in the genomes of four bovine species

牦牛>水牛>野牦牛>普通牛>瘤牛, P-SSR 总长度排序依次为:家牦牛>水牛>普通牛>野牦牛>瘤牛。进一步分析发现,普通牛和水牛基因组长度大于1 000 bp的P-SSR类型均为六核苷酸重复,野牦牛中基因组长度大于1 000 bp的P-SSR类型为五核苷酸重复。由于家牦牛P-SSR基因组长度大于1 000 bp的重复类型达353个,本研究对其重复类型进行了深入分析,结果(图5)显示,在所有重复类型中二核苷酸重复的P-SSR数量(182个)和长度(307 734 bp)均居首位,六核苷酸重复的P-SSR数量(3个)和长度(4 728 bp)均最低。

表5 4个牛族物种基因组中不同长度完全型微卫星(P-SSR)的数量及总长度

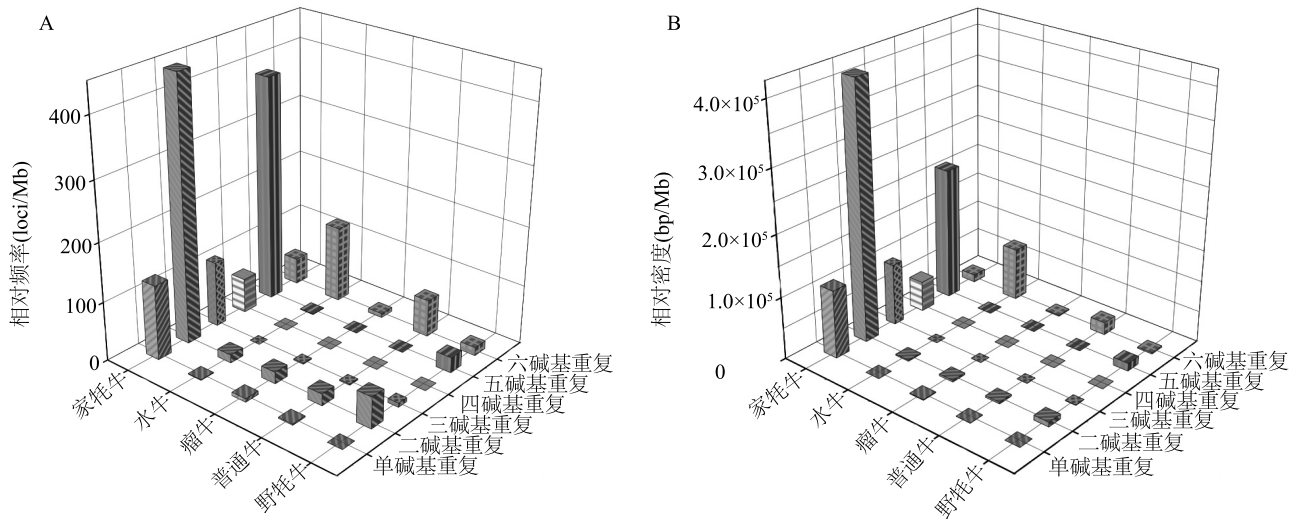
Table 5 Number and total length of perfect microsatellites (P-SSRs) by length in genomes of four bovine species

物种	基因组长度>100 bp P-SSR		基因组长度>1 000 bp P-SSR	
	数量(个)	总长度(bp)	数量(个)	总长度(bp)
普通牛	87	23 634	1	7 626
家牦牛	1 171	896 830	353	621 540
野牦牛	100	23 429	3	7 500
瘤牛	35	5 665	0	0
水牛	146	87 126	24	53 316

3 讨论

本研究首次基于染色体水平基因组系统揭示了4个牛族物种基因组中完全型微卫星(P-SSR)的丰度特征,发现水牛P-SSR占比最高(0.53%),家牦牛(0.51%)、普通牛(0.51%)、野牦牛(0.50%)次之,瘤牛最低(0.48%),这一结果与Qi等^[21]报道的部分牛科动物基因组P-SSR的占比(0.45%~0.57%)结果一致,且与骆驼(0.52%)^[31]、藏羚羊(0.54%)^[32]等物种的基因组P-SSR丰度接近,但显著低于猪(0.85%)^[33]、大鼠(1.41%)^[34]等物种。导致微卫星丰度在不同物种间存在显著差异的原因,不仅源于检测工具、统计标准和基因组大小,更与物种固有的生物学特性密切相关。不同物种在基因组大小、核苷酸组成及序列结构等方面均表现出多样性差异与特异性^[35-36],这些本质差异可能是导致微卫星分布丰度在牛种间存在显著差异的重要原因。

本研究中4个牛族物种基因组P-SSR的相对频率排序均为单核苷酸重复>二核苷酸重复>三核苷酸重复>五核苷酸重复>四核苷酸重复>六核苷酸重复,这一结果与山羊和藏羚羊^[32]等牛科近缘物种的研究结果一致,体现了牛科物种在P-SSR进化上的保守性。而在猪^[33]和骆驼^[31]基因组中,6种重复类型的P-SSR相对



A: 完全型微卫星 (P-SSR) 分布相对频率; B: 完全型微卫星 (P-SSR) 分布相对密度。

图 4 4个牛族物种基因组中长度大于 100 bp 的完全型微卫星 (P-SSR) 分布相对频率和相对密度

Fig.4 Relative frequency and density of perfect microsatellites (P-SSRs) longer than 100 bp in the genomes of four bovidae species

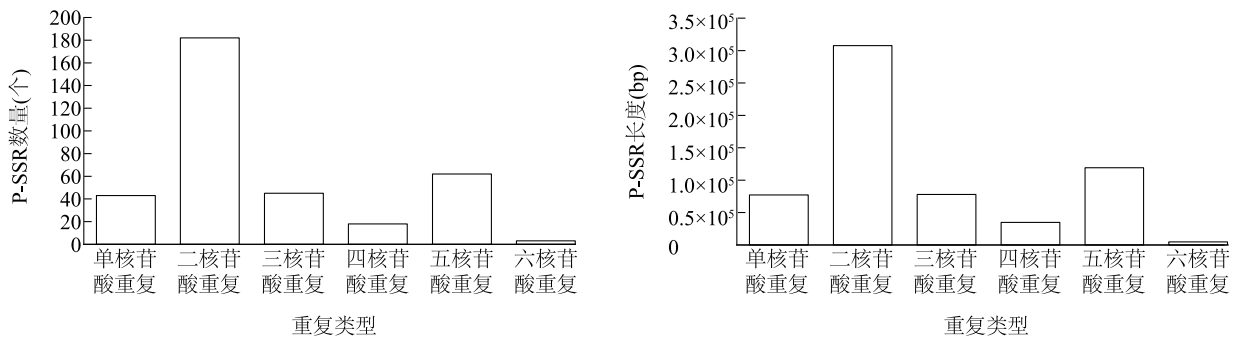


图 5 家牦牛基因组中长度大于1 000 bp 的 P-SSR 数量和长度

Fig.5 Number and length of perfect microsatellites (P-SSRs) longer than 1 000 bp in the genome of domestic yak

频率分布是单核苷酸重复>二核苷酸重复>四核苷酸重复>三核苷酸重复>五核苷酸重复>六核苷酸重复;在大鼠^[34]基因组中,6种重复类型的 P-SSR 相对频率的分布规律则是二核苷酸重复>单核苷酸重复>四核苷酸重复>三核苷酸重复>五核苷酸重复>六核苷酸重复;在鸟类基因组中^[37],四核苷酸重复类型的 P-SSR 出现的频率仅次于单核苷酸重复类型的 P-SSR。上述结果表明,物种间亲缘关系越近,其基因组中不同重复类型 P-SSR 的相对频率排序特征越相似。

本研究结果显示,在4个牛族物种基因组中,单核苷酸至五核苷酸重复类型的 P-SSR 优势重复单元依次为 A、AC、AGC、AAAT、AACTG;普通牛、瘤牛、家牦牛和水牛的优势六核苷酸重复基序均为 AACCT,而野牦牛的优势六核苷酸重复基序则为 AAAAAC。4个牛族物种基因组的 P-SSR 均表现出显著的 A 核苷酸重复基

序偏好性,其中单核苷酸中 A 重复基序的占比均高于 96%。此外,骆驼(96%)^[31]、鸟类(87.7%~98.7%)^[37]、猪(97%)^[33]等物种基因组的 P-SSR 也表现出单核苷酸 A 重复基序的偏好特征,这表明 A 重复基序是单核苷酸重复类型 P-SSR 的核心优势重复基序类别。本研究中4个牛族物种基因组中 P-SSR 优势二核苷酸重复基序是 AC,前人研究结果显示,骆驼^[31]、猪^[33]、猕猴^[38]、山羊及藏羚羊^[32]的优势二核苷酸重复类型的重复基序也均为 AC,说明动物中 AC 重复基序类别通常比 AG 或 AT 重复基序更占优势^[39]。三核苷酸重复类型的 P-SSR 多分布于基因编码区^[23],可通过编码特定氨基酸序列进而调控蛋白质结构与功能;这类重复序列的重复次数变异会直接导致编码产物的氨基酸重复单元发生改变,进而导致多种表型变异,如狗的骨骼形态差异及人类受体蛋白表达水平变化均被证实与基因

组三核苷酸重复序列的多态性有关^[40-41]。本研究4个牛族物种基因组中优势三核苷酸重复基序类别均为AGC也体现了物种间的保守性。随着重复单元长度的增加,优势重复基序的相对丰度呈现显著递减趋势,有些六核苷酸重复基序只出现1次或者不出现。Srivastava等^[39]研究结果表明,进化中脊椎动物一般不出现ACGCGT和AGCGCT重复基序。本研究中,ACGCGT和AGCGCT在牛族物种中也呈现出完全缺失状态。

本研究结果表明,在4个牛族物种中,各染色体长度与其所含P-SSR数量间呈现极显著的正相关关系($P<0.001$),这一结果与Hancock^[42]提出的假说一致。同时,染色体上P-SSR的相对频率和相对密度与其染色体G+C含量存在显著或极显著负相关关系($P<0.05$ 、 $P<0.01$ 、 $P<0.001$),这与Qi等^[21]的研究结果一致,也进一步证实了基因组简单重复序列和其G+C含量具有一定相关性的假设^[43]。

前人研究结果表明,多细胞生物更容易将SSR积累为长重复序列,哺乳动物尤为明显^[39]。本研究发现在4个牛族物种基因组中,不同重复类型P-SSR的重复基序数量分布范围存在较大差异,但各牛种间相同重复类型的P-SSR重复基序数量分布范围和变化趋势总体一致。具体而言,单核苷酸重复类型至六核苷酸重复类型的重复基序数量分布范围分别是12~28次、7~24次、5~13次、4~8次、4~7次和4~10次。所有6类P-SSR的数量均随重复基序数量增加而呈递减趋势,该现象同样也在骆驼^[31]、猕猴^[38]等多个物种的基因组研究中被发现,推测这一现象的成因在于,P-SSR序列越长,其结构稳定性越差,越容易因核苷酸突变发生序列改变,进而导致可检测到的P-SSR数量减少。

4 结论

本研究基于染色体水平的高质量基因组序列,对普通牛、牦牛(含家牦牛与野牦牛)、瘤牛及水牛的基因组完全型微卫星(P-SSR)开展系统性检测与综合性分析。结果显示,水牛基因组的P-SSR占比最高,家牦牛、普通牛、野牦牛次之,瘤牛最低。4个牛族物种基因组中,6种重复类型P-SSR的相对频率呈现一致的分布模式,具体排序为:单核苷酸重复>二核苷酸重复>三核苷酸重复>五核苷酸重复>四核苷酸重复>六核苷酸重复。在单核苷酸重复至五核苷酸重复类型的P-SSR中,优势重复基序依次为A、AC、AGC、AAAT、AACTG;

而不同牛种间的优势六核苷酸重复基序存在差异,其中普通牛、瘤牛、家牦牛和水牛的优势六核苷酸重复单元为AACCCCT,野牦牛则为AAAAAC。此外,4个牛族物种中同一重复类型P-SSR的重复基序数量分布范围高度一致,6类P-SSR的丰度均随重复基序数量的增加呈递减趋势。相关性分析结果表明,4个牛族物种各染色体长度与其所含P-SSR数量间呈极显著正相关($P<0.001$);染色体上P-SSR的相对频率及相对密度均与其染色体G+C含量呈显著或极显著负相关($P<0.05$ 、 $P<0.01$ 、 $P<0.001$)。本研究结果不仅为阐明牛族物种基因组微卫星的组成特征与分布规律提供了重要参考,也为后续基于微卫星DNA标记的遗传图谱构建、遗传资源评估体系建立奠定了理论基础。

参考文献:

- [1] SHARMA P C, GROVER A, KAHL G. Mining microsatellites in eukaryotic genomes[J]. Trends in Biotechnology, 2007, 25(11): 490-498.
- [2] TAUTZ D, RENZ M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes[J]. Nucleic Acids Research, 1984, 12(10): 4127-4138.
- [3] CHEN M, ZENG G M, TAN Z Y, et al. Compound microsatellites in complete *Escherichia coli* genomes[J]. FEBS Letters, 2011, 585(7): 1072-1076.
- [4] WEBER J L. Informativeness of human $(dC-dA)_n \cdot (dG-dT)_n$ polymorphisms[J]. Genomics, 1990, 7(4): 524-530.
- [5] SELKOE K A, TOONEN R J. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers[J]. Ecology Letters, 2006, 9(5): 615-629.
- [6] KANNUR B H, FAIROZE M N, GIRISH P S, et al. Breed traceability of buffalo meat using microsatellite genotyping technique[J]. Journal of Food Science and Technology, 2017, 54(2): 558-563.
- [7] ZHOU C, LI F J, WEN Q C, et al. Characterization of microsatellites in the endangered snow leopard based on the chromosome-level genome[J]. Mammal Research, 2021, 66(2): 385-398.
- [8] CHEN Y, GUO C, ZHOU S L, et al. The mating system of Himalayan marmots as inferred by microsatellite markers[J]. Current Zoology, 2023, 69(6): 654-657.
- [9] BOGA KURU B, KRMZBAYRAK T, ÖZŞENSOY Y. Zavot cattle genetic characterization using microsatellites[J]. Tropical Animal Health and Production, 2022, 54(6): 363.
- [10] WALL D A, DAVIS S K, READ B M. Phylogenetic relationships in the subfamily bovinæ (Mammalia: Artiodactyla) based on ribosomal DNA[J]. Journal of Mammalogy, 1992, 73(2): 262-275.
- [11] HASSANIN A, AN J, ROPIQUET A, et al. Combining multiple autosomal introns for studying shallow phylogeny and taxonomy of

- Laurasiatherian mammals; application to the tribe Bovini (Cetartiodactyla, Bovidae) [J]. *Molecular Phylogenetics and Evolution*, 2013, 66(3):766-775.
- [12] 丁子旭,徐祎雪,冯 彤,等. 水牛泌乳期与干乳期差异血清生化指标及代谢物研究[J]. *基因组学与应用生物学*, 2024, 43(6):1054-1065.
- [13] 郑秉武,智 宇,李彦军,等. 一例大林羚肠毒血症的诊断[J]. *野生动物学报*, 2022, 43(3):788-792.
- [14] WANG X D, PEI J, GUO S K, et al. Characterization of N(6)-methyladenosine in domesticated yak testes before and after sexual maturity[J]. *Frontiers in Cell and Developmental Biology*, 2021, 9:755670.
- [15] 王月月,刘雪雪,董坤哲,等. 7种家养动物全基因组微卫星分布的差异研究[J]. *中国畜牧兽医*, 2015, 42(9):2418-2426.
- [16] 徐宇辉,马志杰,郭卫兴,等. 牦牛全基因组单纯型微卫星分布特征分析[J]. *中国农业大学学报*, 2024, 29(3):134-142.
- [17] 汪鹏飞. 家牛基因组短串联重复变异研究[D]. 杨凌:西北农林科技大学, 2024.
- [18] MA Z J. Abundance and characterization of perfect microsatellites on the cattle Y chromosome [J]. *Animal Biotechnology*, 2017, 28(3):157-162.
- [19] 颜芳楠,马志杰,徐宇辉,等. 牦牛 Y 染色体基因组单纯型微卫星丰度分析[J]. *中国畜牧杂志*, 2023, 59(6):172-178.
- [20] 徐宇辉,徐东辉,曹 萍,等. 5个牛种 X 染色体单纯型微卫星分布规律及其差异研究[J]. *中国畜牧杂志*, 2024, 60(4):176-183.
- [21] QI W H, JIANG X M, DU L M, et al. Genome-wide survey and analysis of microsatellite sequences in bovid species [J]. *PLoS One*, 2015, 10(7):133667.
- [22] 戚文华,蒋雪梅,甘丽萍,等. 牛科动物基因组 GC 含量与微卫星含量的相关性分析[J]. *基因组学与应用生物学*, 2015, 34(11):2381-2386.
- [23] QI W H, JIANG X M, YAN C C, et al. Distribution patterns and variation analysis of simple sequence repeats in different genomic regions of bovid genomes[J]. *Scientific Reports*, 2018, 8:14407.
- [24] 戚文华,蒋雪梅,肖国生,等. 牛和绵羊全基因组微卫星序列的搜索及其生物信息学分析[J]. *畜牧兽医学报*, 2013, 44(11):1724-1733.
- [25] 戚文华,蒋雪梅,杜联明,等. 牦牛和水牛全基因组微卫星分布规律及其比较分析[J]. *基因组学与应用生物学*, 2015, 34(7):1406-1412.
- [26] ROSEN B D, BICKHART D M, SCHNABEL R D, et al. *De novo* assembly of the cattle reference genome with single-molecule sequencing[J]. *GigaScience*, 2020, 9(3):21.
- [27] PORRELLI S, GERBAULT-SEUREAU M, ROZZI R, et al. Draft genome of the lowland anoa (*Bubalus depressicornis*) and comparison with buffalo genome assemblies (Bovidae, Bubalina) [J]. *G3*, 2022, 12(11):234.
- [28] NAJI M M, UTSUNOMIYA Y T, SÖLKNER J, et al. Investigation of ancestral alleles in the Bovinae subfamily [J]. *BMC Genomics*, 2021, 22(1):108.
- [29] ZHANG S Z, LIU W Y, LIU X F, et al. Structural variants selected during yak domestication inferred from long-read whole-genome sequencing [J]. *Molecular Biology and Evolution*, 2021, 38(9):3676-3680.
- [30] GAO X, WANG S, WANG Y F, et al. Long read genome assemblies complemented by single cell RNA-sequencing reveal genetic and cellular mechanisms underlying the adaptive evolution of yak [J]. *Nature Communications*, 2022, 13:4887.
- [31] MANEE M M, ALGARNI A T, ALHARBI S N, et al. Genome-wide characterization and analysis of microsatellite sequences in camelid species [J]. *Mammal Research*, 2020, 65(2):359-373.
- [32] 戚文华,严超超,肖国生,等. 山羊和藏羚羊全基因组微卫星分布规律及其生物信息学分析[J]. *四川大学学报(自然科学版)*, 2016, 53(4):937-944.
- [33] 戚文华,蒋雪梅,肖国生,等. 猪全基因组中微卫星分布规律 [J]. *畜牧与兽医*, 2014, 46(8):9-13.
- [34] 涂飞云,刘晓华,杜联明,等. 大鼠全基因组微卫星分布特征研究[J]. *江西农业大学学报*, 2015, 37(4):708-711.
- [35] 杨宣叶,高明阳,胡欣妍,等. BVDV-2d 毒株的分离鉴定及同义密码子使用模式分析[J]. *江苏农业学报*, 2024, 40(11):2111-2121.
- [36] 周 迪,赵忠海,王 府,等. 基于转录组测序的关岭牛肌肉生长发育关键基因筛选与鉴定[J]. *南方农业学报*, 2024, 55(3):611-622.
- [37] HUANG J, LI W J, JIAN Z Y, et al. Genome-wide distribution and organization of microsatellites in six species of birds [J]. *Biochemical Systematics and Ecology*, 2016, 67:95-102.
- [38] LIU S X, HOU W, SUN T L, et al. Genome-wide mining and comparative analysis of microsatellites in three macaque species [J]. *Molecular Genetics and Genomics*, 2017, 292(3):537-550.
- [39] SRIVASTAVA S, AVVARU A K, SOWPATI D T, et al. Patterns of microsatellite distribution across eukaryotic genomes [J]. *BMC Genomics*, 2019, 20(1):153.
- [40] BROCKSCHMIDT F F, NÖTHEN M M, HILLMER A M. The two most common alleles of the coding GGN repeat in the androgen receptor gene cause differences in protein function [J]. *Journal of Molecular Endocrinology*, 2007, 39(1):1-8.
- [41] FONDON J W 3RD, GARNER H R. Molecular origins of rapid and continuous morphological evolution [J]. *PNAS*, 2004, 101(52):18058-18063.
- [42] HANCOCK J M. Simple sequences and the expanding genome [J]. *BioEssays*, 1996, 18(5):421-425.
- [43] ADELSON D L, RAISON J M, EDGAR R C. Characterization and distribution of retrotransposons and simple sequence repeats in the bovine genome [J]. *PNAS*, 2009, 106(31):12855-12860.

(责任编辑:黄克玲)