



徐宇辉, 马志杰, 郭卫兴, 莫延新, 陈生梅. 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(03): 134-142.

XU Yuhui, MA Zhijie, GUO Weixing, MO Yanxin, CHEN Shengmei. Analysis on the distribution features of perfect microsatellite in the whole genome of yak [J]. *Journal of China Agricultural University*, 2024, 29(03): 134-142.

DOI: 10.11841/j.issn.1007-4333.2024.03.12

牦牛全基因组单纯型微卫星分布特征分析

徐宇辉¹ 马志杰^{1*} 郭卫兴¹ 莫延新² 陈生梅¹

(1. 青海大学 畜牧兽医科学院/农业农村部青藏高原畜禽遗传育种重点实验室/青海省高原家畜遗传资源保护与创新利用重点实验室, 西宁 810016;

2. 青海省畜禽遗传资源保护利用中心, 西宁 810016)

摘要 为探究牦牛全基因组中重复序列的组成特征, 本研究以染色体水平的牦牛全基因组(BosGru3.1)为研究对象, 利用生物信息学方法对牦牛基因组中的单纯型SSRs序列进行了检测, 并对其分布特征进行综合分析。结果表明: 1) 在牦牛全基因组(2.83 Gb)中, 共筛选出778 413个单纯型SSRs, 总长度为14.42 Mb, 占全基因组序列总长的0.51%, 相对频率和相对密度分别为274.92 loci/Mb和5 094.43 bp/Mb。2) 牦牛31条染色体中, 1号染色体所含的SSRs数量最多(48 592个, 6.24%), 29号染色体所含的SSRs数量最少(11 758个, 1.51%), 染色体DNA序列长度与其所含SSRs数量呈显著正相关($P < 0.001$)。3) 6种重复类型的单纯型SSRs在牦牛基因组中分布不均匀, 其中单碱基重复类型的SSRs数量最多(320 592个, 41.19%), 其次是两碱基(197 824个, 25.41%)、三碱基(130 441个, 16.76%)、五碱基(75 056个, 9.64%)、四碱基(52 523个, 6.75%)和六碱基(1 977个, 0.25%)。4) 各重复类型分别以A、AC、AT、AGC、AAC、AAAT、AAAC、AACTG、AGATC和AACCCCT类别的SSRs分布较多, 表现出明显的A(T)碱基组成优势。5) 各重复类型的SSRs其重复拷贝数分别集中在12~25次(单碱基)、7~26次(两碱基)、5~13次(三碱基)、4~8次(四碱基)、4~7次(五碱基)和4~6次(六碱基)。综上, 本研究揭示了牦牛全基因组中单纯型SSRs的分布特征, 表明各重复类型的SSRs其频率、密度、优势碱基类别及重复拷贝数均存在差异, 呈现出一定的规律性。本研究为深入了解牦牛基因组中重复序列特别是单纯型SSRs的组成和分布特征提供参考, 也为后续发掘牦牛特异性SSRs标记、构建遗传图谱及开展标记与性状间的关联分析等研究奠定基础。

关键词 牦牛; 基因组; 单纯型微卫星; 丰度

中图分类号 S823.8+5

文章编号 1007-4333(2024)03-0134-09

文献标志码 A

Analysis on the distribution features of perfect microsatellite in the whole genome of yak

XU Yuhui¹, MA Zhijie^{1*}, GUO Weixing¹, MO Yanxin², CHEN Shengmei¹

(1. Plateau Livestock Genetic Resources Protection and Innovative Utilization key Laboratory of Qinghai Province/Key Laboratory of Livestock and Poultry Genetics and Breeding on the Qinghai-Tibet Plateau of Ministry of Agriculture and Rural Affairs/Academy of Animal Science and Veterinary Medicine, Qinghai University, Xining 810016, China;

2. Qinghai Provincial Livestock Genetic Resources Conservation and Utilization Center, Xining 810016, China)

Abstract To explore the distribution and feature of repetitive sequences in yak genome (BosGru3.1), the perfect SSRs were detected by bioinformatics methods in the yak genome, and its distribution characteristics were comprehensively analyzed. The results showed that: 1) A total of 778 413 perfect

收稿日期: 2023-07-28

基金项目: 国家自然科学基金项目(31960656); 青海省“昆仑英才·高端创新创业人才”计划项目(领军人才)

第一作者: 徐宇辉(ORCID:0009-0005-7150-9220), 硕士研究生, E-mail: xuyhui0911@126.com

通讯作者: 马志杰(ORCID:0000-0001-7544-3533), 副研究员, 主要从事动物遗传资源研究, E-mail: zhijiema@126.com

SSRs were screened out in yak genome (2.83 Gb), with a total length of 14.42 Mb, accounting for 0.51% of the total length of the genome sequence. The relative frequency and relative density were 274.92 loci/Mb and 5 094.43 bp/Mb, respectively. 2) Among the 31 chromosomes of yak, chromosome 1 contained the largest number of SSRs (48 592, 6.24%), and chromosome 29 contained the lowest number of SSRs (11 758, 1.51%). The length of the chromosomal DNA sequence was significantly positively correlated with the number of SSRs contained. 3) In yak genome, the distribution of six repeat types of perfect SSRs was uneven. Among them, the number of SSRs with mononucleotide repeat type was the highest (320 592, 41.19%), followed by dinucleotide (197 824, 25.41%), trinucleotide (130 441, 16.76%), pentanucleotide (75 056, 9.64%), tetranucleotide (52 523, 6.75%), and hexanucleotide (1 977, 0.25%). 4) Repeat copy category of A, AC, AT, AGC, AAC, AAAT, AAAC, AACTG, AGATC, and AACCT was predominate in six repeat types, showing a distinct A(T) base composition advantage. 5) The repeat copy numbers of different type of perfect SSRs were concentrated at 12-25 (mononucleotide), 7-26 (dinucleotide), 5-13 (trinucleotide), 4-8 (tetranucleotide), 4-7 (pentanucleotide), and 4-6 (hexanucleotide). In conclusion, there are differences in the frequency, density, dominant base categories, and copy numbers among different types of SSRs, showing certain regularity. The results of this study can provide a reference for a deeper understanding of the composition and distribution features of repetitive sequences, especially perfect SSRs, and lay a foundation for subsequent studies on mining yak-specific SSR markers, constructing genetic maps, and carrying out marker-trait association analyses.

Keywords yak; genome; perfect microsatellite; abundance

微卫星(Microsatellite),即短串联重复序列(Short tandem repeats, STRs),又称为简单重复序列(Simple sequence repeats, SSRs),是指以1~6个核苷酸为基本单元串联重复而成的DNA序列,广泛分布于真核生物和原核生物的基因组中^[1-2]。微卫星由高突变的核心序列和较为保守的侧翼序列组成。依据核心序列的构成,微卫星被分为3种类型:完全型(Perfect)、不完全型(Imperfect)和复合型(Compound),而完全型微卫星也被称之为单纯型微卫星(Perfect SSRs)^[3]。微卫星在基因组中具有数量多、分布广、杂合率高和共显性遗传等特点,已在种群遗传图谱构建、遗传多样性评估和亲缘关系鉴定等研究领域被广泛应用^[4-6]。

牦牛(*Bos grunniens*)生活在青藏高原及其毗邻的高山、亚高山地区,为当地牧民提供奶、肉、毛等生活必需品,是青藏高原牧民重要的生产、生活和经济来源,也是当地畜牧业经济可持续发展中不可缺少的重要畜种。在先前的研究中,微卫星标记已被广泛用于牦牛品种(群体)的分子遗传多样性、群体遗传结构和遗传渐渗状况等相关评估和分析中^[7-12]。2012年,牦牛基因组序列首次测定完成,初步绘制了Scaffold水平上的牦牛全基因组序列变异草图^[13]。为探究牦牛全基因组微卫星分布特征及与其他物种间的差异,Ma^[14]分析了6种完全型微卫星

在牦牛全基因组中的分布状况,共搜索出723 172个完全型微卫星,平均长度为17.34 bp,相对频率和密度分别为272.18 loci/Mb和4 719.25 bp/Mb;发现单碱基重复类型的微卫星数量最多(44.04%),之后依次为二碱基(24.11%)、三碱基(15.80%)、五碱基(9.50%)、四碱基(6.40%)和六碱基重复类型(0.15%),其中A、AC和AGC等15种重复拷贝类别的微卫星在牦牛基因组中数量较多。随后,戚文华等^[15]通过研究牦牛和水牛全基因组中的完全型微卫星分布特征和差异,发现牦牛和水牛全基因组中的完全型微卫星数量分别为968 134个和1 052 443个,两物种间完全型微卫星分布的相对频率(366.01 loci/Mb和371.07 loci/Mb)与相对密度(5 686.00 bp/Mb和5 799.34 bp/Mb)均特别接近;二者不同重复类型的微卫星丰度分布模式均为:单碱基>二碱基>三碱基>五碱基>四碱基>六碱基,且相同重复类型的微卫星分布特征也基本一致。

综上,研究者已初步探究了牦牛基因组中单纯型微卫星的分布状况及与其他牛种间的差异^[14-15]。然而,上述已有研究仍存在一定的缺陷与不足,如各研究均基于组装到Scaffold水平的牦牛基因组而开展分析研究^[14-15],故不能完全揭示牦牛基因组中各染色体所含微卫星的分布状况及其特征差异。

然而,随着高通量测序和组装技术的发展,2019年基于二、三代测序及Hi-C技术结合的牦牛高质量基因组(BosGru3.1,GCA_005887515.3)已测序完成且组装到了染色体水平^[16],这为深入探究牦牛全基因组染色体水平上的微卫星分布特征提供了契机。鉴于此,本研究以染色体水平的牦牛全基因组序列为研究对象,对其单纯型微卫星丰度及分布特征进行系统分析,以期深入探究牦牛全基因组重复序列的组成、分布特征及发掘特异性分子标记奠定基础。

1 材料与方法

1.1 牦牛全基因组序列来源

本研究分析所用的牦牛全基因组(BosGru3.1)序列从NCBI数据库(<https://www.ncbi.nlm.nih.gov/>)下载,发布于2019年6月,相关文献于2021年8月报道,登录号为GCA_005887515.3,序列总长度为2 831.46 Mb,Scaffold N50和Contig N50分别为114.39和44.72 Mb,其全基因组序列均已组装定位到31条染色体上^[16],序列下载后以FASTA格式保存备用。

1.2 微卫星统计术语定义

微卫星相关术语的定义参照Ma^[14]的研究报道,即:1)重复类型(Repeat type):指微卫星序列的核心序列由几个碱基(bp)组成。根据核心序列的碱基数目组成,将微卫星分为单碱基(Mononucleotide)、两碱基(Dinucleotide)、三碱基(Trinucleotide)、四碱基(Tetranucleotide)、五碱基(Pentanucleotide)和六碱基(Hexanucleotide)重复类型,本研究主要对上述6种重复类型的单纯型SSRs进行搜索分析;2)重复拷贝类别(Repeat copy category):指SSRs具体由哪几个碱基组成,如二碱基重复类型中的AG、AC、AT和GC属于4种不同的重复拷贝类别;3)重复拷贝数(Repeat copy number):指在一个特定的SSRs位点,该序列中核心序列的重复次数。例如(AC)₅,表示在这个SSRs位点,AC这个核心序列的重复次数为5;4)相对频率(Relative frequency, loci/Mb):指每Mb基因组序列中包含的SSRs数量(loci);5)相对密度(Relative density, bp/Mb):指每Mb基因组序列中包含的SSRs序列长度(bp)。

1.3 微卫星搜索方法及统计分析

本研究利用MSDB v2.4(<http://msdb.biosv.com>)软件^[17]搜索牦牛全基因组中的单纯型SSRs序列,SSRs搜索标准如下:1~6个碱基重复类型的SSRs其核心序列的最小重复次数分别设置为12、7、5、4、4和4次,微卫星座位两端的侧翼序列长度设置为200 bp。统计分析中,将所有可循环的微卫星序列及其反补序列归为同一类别,如AGG、GGA和GAG以及CCT、TCC和CTC属同一重复拷贝类别。基于SPSS v25软件进行牦牛各条染色体序列长度与其所含单纯型SSRs数量之间的相关分析。

2 结果与分析

2.1 牦牛全基因组中不同重复类型的单纯型微卫星丰度分析

本研究在牦牛全基因组中共搜索到778 413个单纯型SSRs,总长度为14 424 667 bp,占全基因组序列长度的0.51%,相对频率为274.92 loci/Mb,相对密度为5 094.43 bp/Mb(表1)。6种重复类型的单纯型SSRs中,单碱基重复类型的SSRs数目最多,共320 592个,占总数的41.19%,其次是两碱基(25.41%)、三碱基(16.76%)、五碱基(9.64%)、四碱基(6.75%)和六碱基重复类型的单纯型SSRs,六碱基重复类型的SSRs数目最少,仅有1 977个,占总数的0.25%(图1和表1)。

2.2 单纯型微卫星在牦牛染色体上的分布频率和密度

由图2可知,在牦牛的31条染色体中,1号染色体上单纯型SSRs数量最多(48 592个),占总数的6.24%,其次是2号染色体(40 712个,占5.23%)、X染色体(40 699个,占5.23%)、6号染色体(36 633个,占4.71%)和4号染色体(36 235个,占4.65%),而29号染色体上的单纯型SSRs分布数量最少(11 758个,占1.51%)。相关分析表明,牦牛各条染色体序列长度与其所含的SSRs数量间具有高度正相关($r=0.988$, $P<0.001$; Spearman's ρ , $r=0.984$, $P<0.001$),即染色体DNA序列越长,所含的单纯型SSRs数量越多。各条染色体所含单纯型SSRs的相对频率和相对密度之间也存在差异,其中Y染色体上单纯型SSRs的相对频率和相对密度最高,分别为481.75 loci/Mb和12 370.83 bp/Mb,而12号染色体上单纯型SSRs的相对频率(248.72 loci/Mb)和相对密度(4 558.09 bp/Mb)最低(图3)。相关分析表明,染色体上单纯型SSRs的相对频率($r=-0.271$, $P=$

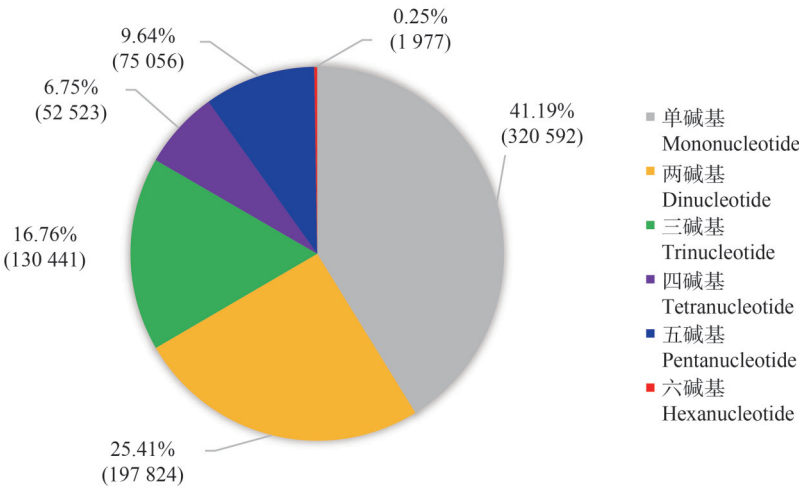


图 1 牦牛全基因组中不同重复类型的单纯型微卫星数目及所占比例

Fig. 1 Number and proportion of different type of perfect microsatellites in yak genome

表 1 牦牛全基因组序列中不同重复类型的单纯型 SSRs 数目、长度、频率和密度

Table 1 Number, length, frequency and density of different type of perfect microsatellites in yak genome

重复类型 Repeat type	数目 Number	总长度/bp Total length	平均长度/bp Average length	相对频率/(loci/Mb) Relative frequency	相对密度/(bp/Mb) Relative density
单碱基 Mononucleotide	320 592	4 708 137	14. 69	113. 22	1 662. 79
两碱基 Dinucleotide	197 824	4 516 744	22. 83	69. 87	1 595. 20
三碱基 Trinucleotide	130 441	2 353 536	18. 04	46. 07	831. 21
四碱基 Tetranucleotide	52 523	965 072	18. 37	18. 55	340. 84
五碱基 Pentanucleotide	75 056	1 811 080	24. 13	26. 51	639. 63
六碱基 Hexanucleotide	1 977	70 098	35. 46	0. 70	24. 76
合计 Total	778 413	14 424 667	18. 53	274. 92	5 094. 43

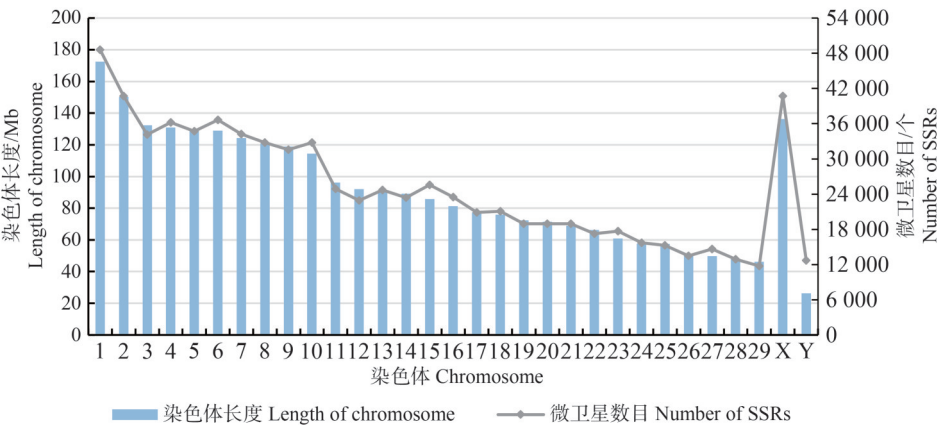


图 2 牦牛 31 条染色体上单纯型微卫星的分布数量

Fig. 2 Number of perfect microsatellites in 31 chromosomes of yak

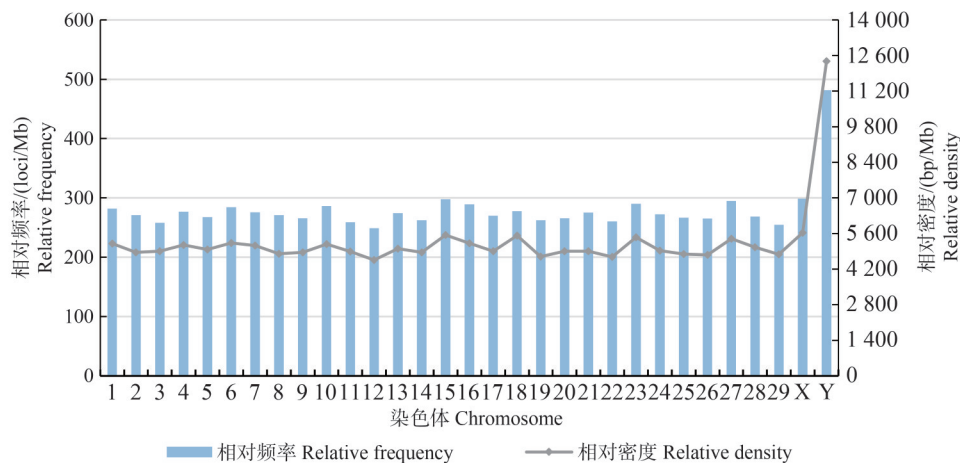


图3 牦牛31条染色体上单纯型微卫星的相对频率和相对密度

Fig. 3 Relative frequency and density of perfect microsatellites in 31 chromosomes of yak

0.141; Spearman’s rho, $r=0.066$, $P=0.724$)和相对密度($r=-0.297$, $P=0.105$; Spearman’s rho, $r=0.090$, $P=0.629$)与其染色体序列长度相关性不显著。

2.3 牦牛全基因组中不同重复拷贝类别的单纯型SSRs频率和密度大小

由表2可知,在单碱基重复类型中,A(T)碱基重复拷贝类别的SSRs数目最多,共318 425个,占单碱基重复类型SSRs总数的99.32%,总长度为4 679 981 bp,相对频率和相对密度分别为112.46 loci/Mb和1 652.85 bp/Mb。与A(T)碱基重复拷贝类别相比,C(G)碱基重复拷贝类别的SSRs出现频率较低,仅有2 167个,在单碱基重复类型中占比0.68%,总长度、相对频率和相对密度分别为28 156 bp、0.77 loci/Mb和9.94 bp/Mb。两碱基重复类型有AC、AT、AG和GC 4种重复拷贝类别,其中AC重复拷贝类别的SSRs数目最多,共121 982个,占两碱基重复类型SSRs总数的61.66%,相对频率为43.08 loci/Mb;其次是AT(29.12%)和AG(9.06%)重复拷贝类别的SSRs,相对频率分别为20.34 loci/Mb和6.33 loci/Mb(表2);GC重复拷贝类别的SSRs数量最少,仅324个,占比0.16%,频率为0.11 loci/Mb。牦牛基因组中三碱基重复拷贝类别有AGC、AAC、AAT、ACC、AGG、CCG、AAG、ACT、ATC和ACG 10种,其中AGC类别相对频率最高(33.89 loci/Mb),其次是AAC(3.44 loci/Mb)、AAT(2.61 loci/Mb)和ACC(1.85 loci/Mb)类别,ACG类别相对频率最低(0.02 loci/Mb)。在四碱基重复类型中,共确定

了32种重复拷贝类别,其中相对频率较高的是AAAT(4.68 loci/Mb)和AAAC(2.89 loci/Mb)类别,其次是AATG(1.52 loci/Mb)、AAAG(1.49 loci/Mb)和ATCC(1.43 loci/Mb)类别,其余27种重复拷贝类别的SSRs相对频率较低,均小于1.00 loci/Mb(表2)。统计分析发现,牦牛五碱基重复类型中共有88种重复拷贝类别,但仅有2种重复拷贝类别的SSRs相对频率大于1.00 loci/Mb,即AACTG重复拷贝类别的相对频率最高(19.09 loci/Mb),其次是AGATC重复拷贝类别SSRs(4.20 loci/Mb)。本研究共检测到156种六碱基重复拷贝类别的SSRs,但其相对频率均小于1.00 loci/Mb,其中AACCCT重复拷贝类别的SSRs相对频率最高,为0.17 loci/Mb。

2.4 牦牛全基因组中不同重复类型的单纯型SSRs重复拷贝数特征分析

分析发现,牦牛全基因组中各重复类型的单纯型SSRs其核心序列的重复拷贝数主要集中在4~30次,最小拷贝数为4次,最大拷贝数为3 080次。其中单碱基重复类型的单纯型SSRs重复拷贝数介于12~3 080次,但主要集中在12~25次,其SSRs总数为320 081个,占单碱基重复类型SSRs总数的99.84%(图4);两碱基重复类型的SSRs重复拷贝数分布在7~1 967次,重复拷贝数在7~26次的SSRs共196 761个,占两碱基重复类型SSRs总数的99.46%(图4);三碱基重复类型的SSRs重复拷贝数介于5~827次,集中在5~13次的SSRs共130 006个,

表 2 牦牛全基因组中不同重复拷贝类别的单纯型 SSRs(相对频率大于 1 loci/Mb)数目、长度、频率和密度

Table 2 Number, length, frequency and density of different category of perfect microsatellites
(frequency above 1 loci/Mb) in yak genome

基序 Motif	数目 Number	总长度/bp Total length	平均长度/bp Average length	相对频率/(loci/Mb) Relative frequency	相对密度/(bp/Mb) Relative density
A	318 425	4 679 981	14. 70	112. 46	1 652. 85
AC	121 982	2 776 730	22. 76	43. 08	980. 67
AT	57 597	1 390 796	24. 15	20. 34	491. 19
AG	17 921	344 144	19. 20	6. 33	121. 54
AGC	95 957	1 742 907	18. 16	33. 89	615. 55
AAC	9 750	167 379	17. 17	3. 44	59. 11
AAT	7 388	127 068	17. 20	2. 61	44. 88
ACC	5 249	95 424	18. 18	1. 85	33. 70
AGG	4 164	71 427	17. 15	1. 47	25. 23
CCG	2 898	61 842	21. 34	1. 02	21. 84
AAAT	13 256	238 264	17. 97	4. 68	84. 15
AAAC	8 181	166 344	20. 33	2. 89	58. 75
AATG	4 305	73 884	17. 16	1. 52	26. 09
AAAG	4 217	77 112	18. 29	1. 49	27. 23
ATCC	4 046	76 708	18. 96	1. 43	27. 09
AACTG	54 062	1 295 945	23. 97	19. 09	457. 69
AGATC	11 894	298 100	25. 06	4. 20	105. 28

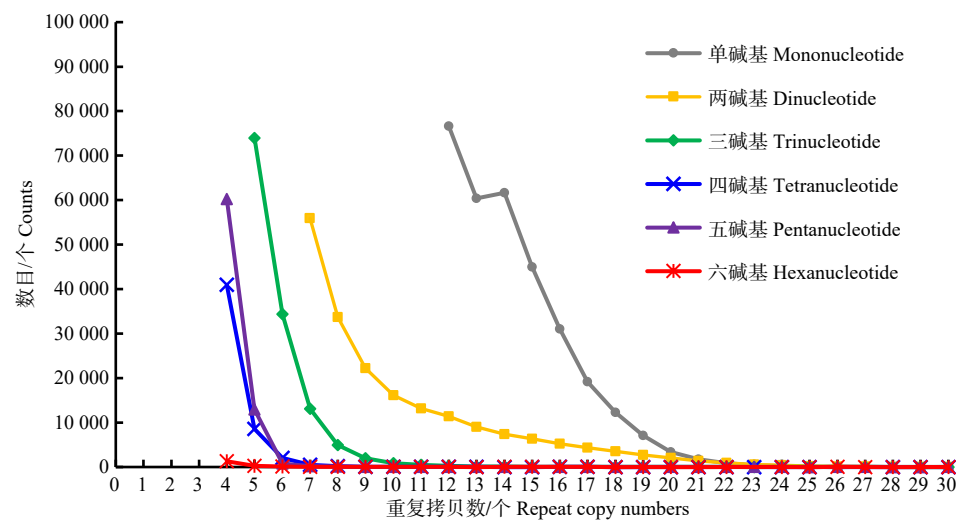


图 4 牦牛全基因组中不同重复类型的单纯型 SSRs 其重复拷贝数情况

Fig. 4 Repeat copy numbers of different type of perfect microsatellites in yak genome

占三碱基重复类型的单纯型 SSRs 总数的 99.67% (图 4); 四碱基重复类型的 SSRs 重复拷贝数介于 4~1 153 次, 其中重复拷贝数在 4~8 次的 SSRs 有 52 272 个, 占四碱基重复类型 SSRs 总数的 99.52%; 五碱基重复类型的 SSRs 重复拷贝数分布范围为 4~680 次, 但以 4~7 次为主 (74 210 个 SSRs), 占五碱基重复类型 SSRs 总数的 98.87% (图 4); 六碱基重复拷贝类型的 SSRs 重复拷贝数分布在 4~393 次, 重复拷贝数在 4~6 次的 SSRs 有 1 673 个, 占六碱基重复类型 SSRs 总数的 84.62% (图 4)。可以看出, 每种类型的单纯型微卫星中, 微卫星的数量均随着其核心序列拷贝数的增加而有减少的趋势。

3 讨论

本研究在牦牛的全基因组中共筛选到 778 413 个单纯型 SSRs, 其序列总长为 14.42 Mb, 占全基因组序列总长度的 0.51%。与先前研究中普通牛 (*Bos taurus*) (0.48%)、绵羊 (*Ovis aries*) (0.48%)、水牛 (*Bubalus bubalis*) (0.57%) 基因组中所含单纯型 SSRs 比例相比, 牦牛基因组中单纯型 SSRs 所占比例与上述物种均较为接近^[15,18]。而与人 (*Homo sapiens*) (3.00%)、小鼠 (*Mus musculus*) (2.85%) 相比, 单纯型 SSRs 在牦牛全基因组中的分布比例相对较低^[19-20]。究其原因, 可能不但与所使用的检测软件、统计标准不同有关, 而且还与各物种本身的特性也有关, 即不同物种其基因组大小及其核苷酸组成和排列方式均存在一定的差异, 进而导致其基因组所含微卫星重复序列存在较大差异。

在本研究牦牛 6 种重复类型的单纯型 SSRs 中, 单碱基重复类型的 SSRs 数量最多, 其次是两碱基、三碱基、五碱基、四碱基和六碱基重复类型的 SSRs, 这与普通牛、水牛、山羊 (*Capra hircus*)、绵羊和藏羚羊 (*Pantholops hodgsoni*) 等物种基因组中各重复类型的 SSRs 分布规律基本一致^[15,18,21]。在猪 (*Sus scrofa*)、犬 (*Canis lupus familiaris*) 和马 (*Equus caballus*) 基因组中, 6 种重复类型的单纯型 SSRs 分布规律则是单碱基重复类型的 SSRs 数量最多, 其次是两碱基、四碱基、三碱基、五碱基和六碱基重复类型的 SSRs^[22-23]。而在鸡 (*Gallus gallus*) 基因组中, 除单碱基重复类型的 SSRs 外, 数量最多的则是四碱基重复类型的 SSRs^[22]。上述研究结果表明不

同物种基因组中单纯型 SSRs 的分布虽存在差异, 但也存在一定的保守性, 即物种间亲缘关系越近, SSRs 的分布特征也更为相似。此外, 本研究分析结果表明, 牦牛各条染色体长度与其所含单纯型微卫星的数量呈显著正相关 ($r=0.988$, $P<0.001$; Spearman's rho, $r=0.984$, $P<0.001$), 即染色体 DNA 序列长度越长, 其所含的单纯型微卫星数量就越多, 这与先前 Hancock 提出的假说一致^[24]。

在牦牛单碱基重复类型的 SSRs 中, A 碱基重复类别的 SSRs 所占比例为 99.32%, 与 C 碱基重复类别 (0.68%) 的 SSRs 相比具有明显优势, 这种现象同样被发现于普通牛、绵羊和猪等物种基因组中^[15,18-23]。而在两碱基、三碱基、四碱基、五碱基和六碱基重复类型的 SSRs 中, 发现以 AC、AT、AGC、AAC、AAAT、AAAC、AACTG、AGATC 和 AACCT 重复拷贝类别的 SSRs 分布数量较多, 表现出明显的 A(T) 碱基优势, 相反富含 C(G) 碱基的 SSRs 出现频率较低, 这与人、虎皮鹦鹉 (*Melopsittacus undulatus*)、恒河猴 (*Macaca mulatta*) 等物种基因组中的 SSRs 分布规律基本一致^[19,25-26]。通常, 物种基因组中不同碱基排列编码的蛋白质拥有的功能也不同, 位于编码区的 SSRs 行使着不同的功能, 微卫星序列中 A(T)、C(G) 碱基分布规律的深入研究将有助于了解 SSRs 的生物学功能^[27-28]。先前的研究表明, 三碱基重复类型的 SSRs 不同于其他重复类型, 其与生物体的遗传疾病具有一定的相关性^[29], 外显子中的三碱基重复类型的 SSRs 也总是最丰富的^[30]。因此, SSRs 在牦牛基因组中的功能特征、调控机制以及与某些疾病或性状间的关联分析都有待进一步深入探索。

研究表明, 随着 SSRs 序列长度的增加, 其稳定性会逐渐下降, 即 SSRs 重复序列拷贝数越高, 越容易发生碱基突变, 使原来的单纯型 SSRs 变为复合型 SSRs 或非微卫星序列, 从而导致 SSRs 数目减少^[31]。在本研究中, 牦牛全基因组中各重复类型的 SSRs 其核心序列的重复拷贝数存在较大差异, 如单碱基、两碱基、三碱基、四碱基、五碱基和六碱基重复类型的单纯型 SSRs 其重复拷贝数分别主要集中在 12~25 次、7~26 次、5~13 次、4~8 次、4~7 次和 4~6 次。此外, 6 种重复类型的单纯型 SSRs 数量均随着其核心序列拷贝数的增加而减少, 该现象同样被发现在水牛、普通牛、绵羊、人和小鼠等多个物种

的基因组中^[14-15,18-26],提示各物种基因组中的不同类型的SSRs可能经历了不同的动态进化过程或处于基因组内进化的不同阶段。

4 结 论

牦牛全基因组中,6种重复类型的单纯型SSRs分布不均匀,单碱基重复类型的SSRs数量最多,其次依次是两碱基、三碱基、五碱基、四碱基和六碱基重复类型的SSRs。31条染色体中,1号染色体上的单纯型SSRs数量最多,29号染色体上单纯型SSRs分布数量最少,各条染色体序列长度与其所含的SSRs数量呈显著正相关。6种重复类型分别以A、AC、AT、AGC、AAC、AAAT、AAAC、AACTG、AGATC和AACCCT重复拷贝类别的SSRs分布相对较多。各重复类型的SSRs其核心重复序列的拷贝数存在较大差异,且6种重复类型的单纯型SSRs数量均随着其核心序列拷贝数的增加而减少。本研究明确了牦牛全基因组中单纯型SSRs的分布特征,为后续深入探究牦牛基因组中重复序列的组成、分布特征、生物学功能及发掘特异性分子标记提供了理论和数据支撑,也为进一步探讨牦牛基因组中重复序列的进化规律和生物学意义奠定基础。

参考文献 References

- [1] Tautz D, Renz M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes[J]. *Nucleic Acids Research*, 1984, 12(10): 4127-4138
- [2] Morgante M, Hanafey M, Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes[J]. *Nature Genetics*, 2002, 30(2): 194-200
- [3] Weber J L. Informativeness of human (dC-dA)_n·(dG-dT)_n polymorphisms[J]. *Genomics*, 1990, 7(4): 524-530
- [4] Wu X P, Feng Y W, Jiang H L, Liu X Q, Pan Y. A preliminary genetic linkage map of *Sinonovacula constricta* (Lamarck, 1818) based on microsatellites derived from RAD sequencing[J]. *Journal of Ocean University of China*, 2018, 17(4): 947-956
- [5] 胡江. 野牦牛、家牦牛及其杂交种的生物学特性和遗传多样性研究[D]. 兰州: 甘肃农业大学, 2001
Hu J. Study on biological characteristics and genetic diversity of wild yak, domestic yak and their hybrids[D]. Lanzhou: Gansu Agricultural University, 2001(in Chinese)
- [6] 王敏, 张新浩, 崔冉, 刘宇, 杨莉, 李海静, 赵春江. 利用微卫星标记鉴定德州驴亲子关系[J]. *中国畜牧兽医*, 2019, 46(7): 2003-2011
Wang M, Zhang X H, Cui R, Liu Y, Yang L, Li H J, Zhao C J. Parentage testing of Dezhou donkey with microsatellite marker[J]. *China Animal Husbandry & Veterinary Medicine*, 2019, 46(7): 2003-2011(in Chinese)
- [7] 马志杰, 钟金城, 韩建林, 徐惊涛, 刘仲娜, 白文林. 牦牛分子遗传多样性研究进展[J]. *遗传*, 2013, 35(2): 151-160
- [8] Ma Z J, Zhong J C, Han J L, Xu J T, Liu Z N, Bai W L. Research progress on molecular genetic diversity of the yak (*Bos grunniens*)[J]. *Hereditas*, 2013, 35(2): 151-160(in Chinese)
- [9] Nguyen T T, Genini S, Ménétrey F, Malek M, Vögeli P, Goe M R, Stranzinger G. Application of bovine microsatellite markers for genetic diversity analysis of Swiss yak (*Poephagus grunniens*)[J]. *Animal Genetics*, 2005, 36(6): 484-489
- [10] Qi X B, Jianlin H, Wang G, Rege J E O, Hanotte O. Assessment of cattle genetic introgression into domestic yak populations using mitochondrial and microsatellite DNA markers[J]. *Animal Genetics*, 2010, 41(3): 242-252
- [11] Dorji T, Goddard M, Perkins J, Robinson N, Roder W. Genetic diversity in Bhutanese yak (*Bos grunniens*) populations using microsatellite markers[C]. In: Han J L, eds. *Proceedings of the third International Congress on Yak*. Lhasa, Tibet, China, 2000: 197-201
- [12] Wang M Q, Weigend S, Barre-Dirie A, Carnwath J W, Zhonglin L, Niemann H. Analysis of two Chinese yak (*Bos grunniens*) populations using bovine microsatellite primers[J]. *Journal of Animal Breeding and Genetics*, 2003, 120(4): 237-244
- [13] Li Q F, Zhao X B, Yanhua L, Ning L, Wu C X, Zhang X. Construction of microsatellite-enriched library of yak and phylogenetic study of six Chinese yak populations using yak-specific microsatellites[C]. In: *Proceedings of the fourth International Congress on Yak*. Chengdu, Sichuan, China, 2004: 109-111
- [14] Qiu Q, Zhang G J, Ma T, Qian W B, Wang J Y, Ye Z Q, Cao C C, Hu Q J, Kim J, Larkin D M, Auvil L, Capitanu B, Ma J, Lewin H A, Qian X J, Lang Y S, Zhou R, Wang L Z, Wang K, Xia J Q, Liao S G, Pan S K, Lu X, Hou H L, Wang Y, Zang X T, Yin Y, Ma H, Zhang J, Wang Z F, Zhang Y M, Zhang D W, Yonezawa T, Hasegawa M, Zhong Y, Liu W B, Zhang Y, Huang Z Y, Zhang S X, Long R J, Yang H M, Wang J, Lenstra J A, Cooper D N, Wu Y, Wang J, Shi P, Wang J, Liu J Q. The yak genome and adaptation to life at high altitude[J]. *Nature Genetics*, 2012, 44(8): 946-949
- [15] Ma Z J. Genome-wide characterization of perfect microsatellites in yak (*Bos grunniens*)[J]. *Genetica*, 2015, 143(4): 515-520
- [16] 戚文华, 蒋雪梅, 杜联明, 肖国生. 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
Qi W H, Jiang X M, Du L M, Xiao G S. Distribution regularities and comparative analysis of microsatellite in the whole genomes of yak and water buffalo[J]. *Genomics and Applied Biology*, 2015, 34(7): 1406-1412(in Chinese)
- [17] Zhang S Z, Liu W Y, Liu X F, Du X, Zhang K, Zhang Y, Song Y W, Zi Y N, Qiu Q, Lenstra J A, Liu J Q. Structural variants selected during yak domestication inferred from long-read whole-genome sequencing[J]. *Molecular Biology and Evolution*, 2021, 38(9): 3676-3680
- [18] Du L M, Li Y Z, Zhang X Y, Yue B S. MSDB: A user-friendly program for reporting distribution and building databases of microsatellites from genome sequences[J]. *Journal of Heredity*, 2013, 104(1): 154-157
- [19] 戚文华, 蒋雪梅, 肖国生, 黄小云, 杜联明. 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
Qi W H, Jiang X M, Xiao G S, Huang X Y, Du L M. Seeking and bioinformatics analysis of microsatellite sequence in the genomes of cow and sheep[J]. *Chinese Journal of Animal and Veterinary Sciences*, 2013, 44(11): 1724-1733(in Chinese)
- [20] Subramanian S, Mishra R, Singh L. Genome-wide analysis of

- microsatellite repeats in humans: Their abundance and density in specific genomic regions[J]. *Genome Biology*, 2003, 4: R13-R13
- [20] 童晓玲, 代方银, 李斌, 余泉友, 夏庆友, 鲁成. 小鼠基因组中的微卫星重复序列的数量、分布和密度[J]. *动物学报*, 2006(1): 138-152
- Tong X L, Dai F Y, Li B, Yu Q Y, Xia Q Y, Lu C. Microsatellite repeats in mouse: Abundance distribution and density[J]. *Current Zoology*, 2006, 52(1): 138-152(in Chinese)
- [21] 都玉蓉, 谭春敏, 徐永涛, 周智红, 马建滨, 郭松长. 藏羚羊基因组微卫星分析与初步验证[J]. *四川动物*, 2016, 35(6): 845-851
- Du Y R, Tan C M, Xu Y T, Zhou Z H, Ma J B, Guo S C. Genome-wide characterization and identification of microsatellites in *Pantholops hodgsoni* [J]. *Sichuan Journal of Zoology*, 2016, 35(6): 845-851 (in Chinese)
- [22] 王月, 刘雪, 董坤哲, 陈潇飞, 叶绍辉, 马月辉. 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
- Wang Y Y, Liu X X, Dong K Z, Chen X F, Ye S H, Ma Y H. Distribution difference of microsatellite in 7 domestic animals genomes[J]. *China Animal Husbandry & Veterinary Medicine*, 2015, 42(9): 2418-2426 (in Chinese)
- [23] 戚文华, 蒋雪梅, 肖国生, 黄小云. 猪全基因组中微卫星分布规律[J]. *畜牧与兽医*, 2014, 46(8): 9-13
- Qi W H, Jiang X M, Xiao G S, Huang X Y. Distribution regularities of microsatellites in the pig genome [J]. *Animal Husbandry & Veterinary Medicine*, 2014, 46(8): 9-13(in Chinese)
- [24] Hancock J M. Simple sequences and the expanding genome [J]. *BioEssays*, 1996, 18(5): 421-425
- [25] 黄杰, 原宝东, 杨承忠. 虎皮鹦鹉全基因组中微卫星分布规律研究[J]. *野生物学报*, 2017, 38(3): 422-426
- Huang J, Yuan B D, Yang C Z. Distribution of microsatellites in genome of budgerigar (*Melopsittacus undulatus*) [J]. *Chinese Journal of Wildlife*, 2017, 38(3): 422-426(in Chinese)
- [26] Webster M T, Smith N G C, Ellegren H. Microsatellite evolution inferred from human - chimpanzee genomic sequence alignments [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(13): 8748-8753
- [27] Schlötterer C. Genome evolution: Are microsatellites really simple sequences [J]. *Current Biology*, 1998, 8(4): R132-R134
- [28] 魏朝明, 孔光耀, 廉振民, 刘慧, 范永文, 张慧. 蜜蜂全基因组中微卫星的丰度及其分布[J]. *昆虫知识*, 2007, 44(4): 501-504
- Wei Z M, Kong G Y, Lian Z M, Liu H, Fan Y W, Zhang H. Abundance and distribution of microsatellites in the entire *Apis mellifera* genome [J]. *Chinese Bulletin of Entomology*, 2007, 44(4): 501-504(in Chinese)
- [29] Pearson C E, Sinden R R. Trinucleotide repeat DNA structures: Dynamic mutations from dynamic DNA [J]. *Current Opinion in Structural Biology*, 1998, 8(3): 321-330
- [30] Tóth G, Gáspári Z, Jurka J. Microsatellites in different eukaryotic genomes: Survey and analysis [J]. *Genome Research*, 2000, 10(7): 967-981
- [31] Wierdl M, Dominska M, Petes T D. Microsatellite instability in yeast: Dependence on the length of the microsatellite [J]. *Genetics*, 1997, 146(3): 769-779

责任编辑:秦梅



第一作者简介:徐宇辉,青海大学畜牧专业在读硕士研究生,2022—2023年在西北农林科技大学动物科技学院学习。在校期间曾获国家励志奖学金,获优秀学生干部、先进个人等称号。现阶段参与青海省“昆仑英才·高端创新创业人才”计划(领军人才)、国家自然科学基金(31960656)等项目,参与发表论文1篇。



通讯作者简介:马志杰,博士,青海大学硕士生导师,副研究员。国家第三次畜禽遗传资源普查牛专业组技术专家组成员,英国卡迪夫大学公派访问学者,中国畜牧兽医学会畜禽遗传标记学会理事,青海省牛产业科技创新平台建设岗位专家,青海省重大项目谋划专家库成员,青海省种业联盟副秘书长,青海省专家人才联合会成员。《中国牛业科学》杂志编委, *Animal genetics*、*Heredity*、《畜牧兽医学报》等期刊审稿专家。主要从事牦牛、黄牛遗传资源评估、遗传育种和基因组学研究。2016年入选青海省“自然科学与工程技术学科”学术带头人,2017年入选青海省高端创新人才千人计划“拔尖人才”,2018年入选中国科学院“西部青年学者”,2022年入选青海省“昆仑英才·高端创新创业人才计划”领军人才。多次被青海大学评为“先进工作者”和“优秀共产党员”,荣获“青海大学2023届毕业研究生优秀学位论文指导教师”称号。主持完成国家自然科学基金、青海省科技厅、西部之光等项目10余项,参与申报获批5个国家牦牛、绵羊遗传资源,完成地方标准2项,获批专利4项、软件著作权11项。发表学术论文80余篇(第一作者发表SCI论文20余篇)。参编世界首部大型高原医学系列专业丛书——《高原适应的生物基础》专著1部。目前,正在主持国家自然科学基金、国家农业农村部及省科技厅项目3项。