DOI: 10.3724/SP.J.1006.2020.94135

研究简报

甘蔗栽培种单倍体基因组 SSR 位点的发掘与应用

王恒波 祁舒婷 陈姝琦 郭晋隆 阙友雄*

福建农林大学农业部福建甘蔗生物学与遗传育种重点实验室 / 国家甘蔗工程技术研究中心, 福建福州 350002

摘要: 甘蔗是世界上最重要的糖料作物之一, 由于尚未完全破译栽培种基因组, 导致 SSR 标记匮乏, 难以覆盖全基因组, 限制了甘蔗遗传研究的进展。本研究以栽培种 R570 的 4660 个 BAC 文库片段序列(累计总长为 382 Mb, 预测到 25,316 个编码蛋白基因)组装成的一套甘蔗单倍体基因组的模板, 利用 MISA (Microsatellite identification tool) 软件, 发掘 SSR 位点; 并综合分析其与 4 种禾本科植物(高粱、玉米、水稻和二岁短柄草)SSR 位点的分布特征; 选取 50 对以 TG 和 AG 重复基序的 SSR 引物, 分别利用 4 个甘蔗属材料(R570、ROC1、LA purple 和 SES208)和 24 个重要甘蔗亲本, 对 SSR 引物进行扩增效率验证和多态性分析。共发掘到 27,241 个 SSR 位点, 平均每个 BAC 片段有 6.29 个 SSR 位点, 平均密度为 71.33 个 SSR Mb⁻¹, 远低于高粱的平均密度(350.00 个 SSR Mb⁻¹)。在重复基序中, 占比前 2 位的分别为单核苷酸基序(11,079 个)和三核苷酸重复基序(6447 个), 合计占总 SSR 位点数的 64.33%。与甘蔗不同的是, 4 种禾本科植物中的三核苷酸基序类型数量最多、占比最大。此外, 在单核苷酸重复基序中, A/T 所占比例最高, 为 84.8%, C/G 所占比例最低, 为 15.2%; 在三核苷酸重复基序中, TGT/ACA 所占比例最高, 为 16.04%。总之, 禾本科植物基因组富含 A/T 的基序。在 50 对 SSR 引物(TG 基序 41 对和 AG 基序 9 对)的多态性验证中, 共有 45 对(90%)能够扩增出清晰的条带, 其中 35 对(70%)在 4 个甘蔗材料上呈现多态性。进一步利用 20 对多态性较高的 SSR 引物对 24 个甘蔗重要亲本材料进行分析, 共扩增到 95 个等位基因, 平均每对引物扩增 4.75 个, 验证了这些引物应用于甘蔗遗传多样性研究的可行性。本研究鉴定的甘蔗栽培种单倍体基因组 SSR 标记, 有效增加了甘蔗遗传研究中可用的分子标记数量, 可直接用于甘蔗群体遗传多样性分析和重要性状遗传机制的解析, 为甘蔗分子育种的深入研究奠定了基础。

关键词: 甘蔗栽培种; BAC 文库; SSR; 标记开发; 多态性

Development and application of SSR loci in monoploid reference genome of sugarcane cultivar

WANG Heng-Bo, QI Shu-Ting, CHEN Shu-Qi, GUO Jin-Long, and QUE You-Xiong*

Key Laboratory of Sugarcane Biology and Genetic Breeding (Fujian), Ministry of Agriculture, Fujian Agriculture and Forestry University / Sugarcane Research & Development Center, China Agricultural Technology System, Fuzhou 350002, Fujian, China

Abstract: Sugarcane is one of the most important sugar crops in the world. However, it is difficult to develop SSR on a large scale since the genome of cultivar has not been sequenced, which limits the genetic improvement of sugarcane. In this study, a template of monoploid sugarcane genome was assembled using a set of 4660 BAC library sequences (with a cumulative length of 382 Mb, predicting 25,316 genes) from cultivar ‘R570’. SSR loci were identified by using MISA (Microsatellite identification tool) software. The distribution characteristics of the monoploid genome ‘R570’ was comprehensively analyzed by comparing with the SSR loci of four Gramineae plants (*Sorghum bicolor*, *Zea mays*, *Oryza sativa*, and *Brachypodium distachyon*). Fifty pairs of primers with TG and AG repeat motifs were designed to verify the amplification efficiency and polymorphism by PCR amplification in four *Saccharum* clones (R570, ROC1, LA purple, and SES208) and twenty four core parents of sugarcane. A total of 27,241 SSR loci were identified, with an average of 6.29 SSR loci per BAC clone and an average density of 71.33 SSR Mb⁻¹.

本研究由引进国际先进农业科学技术计划(948 计划)项目(2014-S18), 国家现代农业产业技术体系建设专项(CARS-17)和福建农林大学校科技发展专项(KFA18025A)资助。

This study was supported by the Program of Introducing International Super Agricultural Science and Technology (948 Program) (2014-S18), the Agricultural Research System (CARS-17), and Fujian Agriculture and Forestry University Science and Technology Development Special Fund (KFA18025A).

* 通信作者(Corresponding author): 阙友雄, E-mail: queyouxiong@126.com, Tel: 0591-83852547

第一作者联系方式: E-mail: wanghengbo_0354@126.com, Tel: 0591-83789177

Received (收稿日期): 2019-09-11; Accepted (接受日期): 2019-12-26; Published online (网络出版日期): 2020-01-15.

URL: <http://kns.cnki.net/kcms/detail/11.1809.S.20200115.1136.020.html>

which was much lower than that of sorghum ($350.00 \text{ SSR Mb}^{-1}$). The mono-nucleotide (11,079) and tri-nucleotide repeat motifs (6447) accounted for 64.33% of the total SSR loci. The number and proportion of tri-nucleotide repeat motifs were the largest in the four Gramineae plants. In addition, A/T (accounting for 84.8%) motif had the highest proportion and C/G (accounting for 15.2%) motif the lowest proportion in the mono-nucleotide repeat motifs and TGT/ACA (accounting for 16.04%) motif had the highest proportion in the trinucleotide repeat motifs. In general, the genomes in Gramineae plants are rich in A/T repeat motifs. In the polymorphism validation of 50 pairs of primers (41 pairs of TG motif and 9 pairs of AG motif), 45 pairs of primers (90%) were found to be able to amplify successfully, of which 35 (70%) were polymorphic in 4 sugarcane clones. Furthermore, 20 pairs of polymorphic SSR primers were used to detect 24 core parents of sugarcane, a total of 95 alleles were amplified with an average of 4.75 alleles per primer, verifying the application feasibility of these primers for the genetic diversity analysis in sugarcane. The development of SSR markers from the monoploid genome of cultivars ‘R570’ not only enriches the number of SSR markers available in sugarcane genetic analysis, but also facilitates the genetic diversity analysis of sugarcane population and the genetic mechanism dissection of important agronomic traits, which provides a foundation for the in-depth research of molecular breeding in sugarcane.

Keywords: sugarcane cultivars; bacterial artificial chromosome library; SSR; development; polymorphism

甘蔗(*Saccharum* spp. hybrid)是人类最早利用的 C_4 植物, 是世界上最重要的糖料作物, 其食糖占世界总产的 80%, 同时也是一种重要生物能源作物, 其生物乙醇产量占世界总产的 40%^[1]。甘蔗是一种具有适应性强、生物量高、光合效率高、可连续多年种植及 CO_2 补偿点低的糖料作物。同时, 甘蔗是世界上生物量最大的作物之一, 2017 年甘蔗产量已经超过玉米, 位居世界第三^[2]。

1887 年, 在爪哇和西印度巴巴多斯试验场发现甘蔗种子可以产生幼苗, 开启了甘蔗有性杂交的历史^[3]。现代甘蔗栽培种是由甘蔗祖先热带种(*S. officinarum* L., $2n = 80$, $x = 10$)和割手密种(*S. spontaneum* L., $2n = 40\sim 128$, $x = 8$)杂交产生的真正意义上的甘蔗杂种, 为了恢复高含糖量性状, 将杂种后代与热带种回交 1 次, 母本性状通过 $2n$ 染色体传递给后代, 由于种间杂交和非孟德尔的遗传方式($2n+n$), 造成杂种后代具有高度杂合、多倍体和非整倍体性, 染色体数目在 $2n = 100\sim 130$ 之间^[4]。鉴于甘蔗栽培种的多倍及非整倍体的遗传背景, 其复杂性超过了大多数作物, 使相关的遗传研究、育种及基因组测序都面临极大的困难^[5-6]。

简单重复序列(simple sequence repeats, SSR)具有高度的多态性、广泛分布于真核生物的基因组^[5-6], 且分布随机^[7], 但更偏向于低重复、富含基因的区域^[8]。SSR 位点产生于 DNA 复制和修复时, DNA 聚合酶滑动或不均等重组^[9], 因而可以根据在种内或种间产生大量的长度变异^[10], 开发和筛选出多态性高、重复性好的 SSR 分子标记, 进而广泛应用于各种动、植物的品种指纹图谱鉴定^[11]、遗传多样性分析^[12-14]、遗传图谱构建^[15-17]及重要性状(基因)的遗传定位或解析^[18]等领域。然而, 相对于其他禾本科植物等模式作物, 甘蔗 SSR 分子标记开发及遗传连锁图谱的构建都比较落后, 相关的国内外报道较少。Singh 等^[13]从 4085 个 EST 序列中鉴定出 351 个 EST-SSRs, 验证后发现 134 个有多态性。Shamshad 等^[19]从 NCBI 数据库中获得 10,000 个 EST 序列, 鉴定出 406 个 SSRs, 验证了 63 个后发现 42 个具有多态性。Oliveira 等^[14]从甘蔗 EST 数据库中鉴定出 2005 个 SSRs, 验证了 342 个, 其中 224 个

(65.5%)呈多态性。甘蔗为同源多倍体作物, 染色体数目多(100~130), 遗传背景复杂, 基因组庞大(约 10 Gb)^[20], 大规模开发甘蔗基因组 SSR 标记面临很大的困难, 严重制约了甘蔗分子遗传研究相关工作的进展^[16]。迄今, 与其他禾本科植物(如高粱 *Sorghum bicolor*、小麦 *Triticum aestivum*、大麦 *Hordeum vulgare*、水稻 *Oryza sativa*、藜麦 *Chenopodium quinoa*、玉米 *Zea mays*、二穗短柄草 *Brachypodium distachyon*)相比, 甘蔗已开发的 SSR 标记数量少、多态性低, 难以满足甘蔗分子标记辅助育种和遗传作图等工作的要求。

甘蔗和高粱同源染色体片段(BAC)的比较分析表明, 二者存在较高的基因共线性和序列保守性, 且甘蔗单倍体基因组大小约为 800~900 Mb, 接近高粱基因组大小(750 Mb)^[20]。普遍认为, 一个单倍型序列可为其他单倍型同源染色体提供较好的参考。目前, 甘蔗栽培种 R570 具有一个 103,296 个克隆的 BAC 文库, 它代表 $14\times$ 单倍体基因组覆盖率和 $1.3\times$ R570 全基因组覆盖度^[21], 已被广泛应用于比较基因组分析^[22-24]。Garsmeur 等^[25]用全基因组分析(WGP)技术将甘蔗栽培种 R570 的 BAC 与高粱基因组比对, 确定了一个由 4660 个甘蔗 BAC 文库片段组成的覆盖甘蔗单倍体基因组常染色质的 BAC 的最小标记路径(MTP), 并完成了甘蔗栽培种 R570 的单倍体基因组测定。目前, 甘蔗栽培种全基因组 SSR 鉴定尚未报道。本研究旨在分析和验证甘蔗栽培种 R570 单倍体基因组数据, 利用生物信息学方法发掘 SSR 位点的特征及其分布规律, 并设计和合成 SSR 引物, 验证其多态性, 为甘蔗栽培种的分子指纹图谱构建、品种间遗传多样性分析、重要农艺性状的遗传机制研究及分子育种的推进奠定一定的基础。

1 材料与方法

1.1 材料

甘蔗属包括 1 个割手密种、1 个热带种及世界各国培育的甘蔗栽培种, 都来自于广州甘蔗糖业研究所海南甘蔗育种场(表 1)。

表 1 甘蔗品种资源名称和来源
Table 1 Name and origin of sugarcane varieties

序号 No.	名称 Name	育成品种数 Number of released varieties	类型 Type	来源 Origin
1	CP49-50	38	<i>Saccharun</i> hybrid	美国 USA
2	Co 419	25	<i>Saccharun</i> hybrid	印度 India
3	CP72-1210	17	<i>Saccharun</i> hybrid	美国 USA
4	NCo 310	13	<i>Saccharun</i> hybrid	印度 India
5	F108	12	<i>Saccharun</i> hybrid	中国台湾 Taiwan, China
6	华南 56-12 Huanan 56-12	10	<i>Saccharun</i> hybrid	中国 China
7	崖城 71-374 Yacheng 71-374	9	<i>Saccharun</i> hybrid	中国 China
8	粤农 73-204 Yuenong 73-204	9	<i>Saccharun</i> hybrid	中国 China
9	CP28-11	8	<i>Saccharun</i> hybrid	美国 USA
10	Co 1001	6	<i>Saccharun</i> hybrid	印度 India
11	桂糖 11 号 Guitang 11	6	<i>Saccharun</i> hybrid	中国 China
12	云蔗 65-225 Yunzhe 65-225	6	<i>Saccharun</i> hybrid	中国 China
13	川 73-219 Chuan 73-219	-	<i>Saccharun</i> hybrid	中国 China
14	ROC 1	6	<i>Saccharun</i> hybrid	中国台湾 Taiwan, China
15	科 5 Ke 5	4	<i>Saccharun</i> hybrid	菲律宾 Philippines
16	CP67-412	3	<i>Saccharun</i> hybrid	美国 USA
17	POJ2878	3	<i>Saccharun</i> hybrid	印度尼西亚爪哇岛 Java, Indonesia
18	华南 56-21 Huanan 56-21	3	<i>Saccharun</i> hybrid	中国 China
19	R570	-	<i>Saccharun</i> hybrid	法国 France
20	LA purple	-	<i>Saccharun officinarum</i>	美国 USA
21	SES208	-	<i>Saccharun spontaneum</i>	美国 USA
22	LCP85-384	-	<i>Saccharun</i> hybrid	美国 USA
23	ROC16	-	<i>Saccharun</i> hybrid	中国台湾 Taiwan, China
24	ROC22	-	<i>Saccharun</i> hybrid	中国台湾 Taiwan, China

1.2 基因组序列的来源

通过 EMBL-欧洲生物信息学研究所的公共数据库获得甘蔗栽培种 R570 基因组数据(登录号为 ERZ654945), 或者也可以从法国农业研究所甘蔗基因组中心(<http://sugarcane-genome.cirad.fr/>)直接获得。其他 4 种禾本科植物 SSR 位点序列特征来自郑燕等^[26]分析结果。

1.3 SSR 位点的查找与 SSR 引物的开发

应用 MISA (Microsatellite identification tool)软件扫描甘蔗栽培种 R570 的基因组 BAC 序列^[27], 该软件下载自 <http://pgrc.ipk-gatersleben.de/misa/>, 在配置文件中设置核苷酸重复基序 (motif) 分别为单 (mononucleotide repeats MDRs)、二 (dinucleotide repeats DNRs)、三 (trinucleotide repeats TNRs)、四 (tetranucleotide repeats TtNRs)、五 (pentanucleotide repeats PNRs)、六 (hexanucleotide repeats HNRs), 序列长度分别为 10、12、15、16、15、18。对 SSR 位点两侧各截取 200 bp 序列设计引物, 借助 MISA 软件提供的与 Primer3 的接口工具, 把 MISA 识别出来的

SSR 序列转为 Primer3 需要的格式。用 Primer3 (<http://frodo.wi.mit.edu/primer3/>)在线设计引物, 引物设计参数为 primer length: 18~28 bp; annealing temperature: 55~65℃; amplicon size: 100~500 bp; GC content: 45%~65%^[13]。

1.4 PCR 扩增和电泳分析

PCR 体系为 25 μL, 包含 DNA (25 ng μL⁻¹) 2.0 μL、正反向引物(10 μmol μL⁻¹)各 0.5 μL、2×*Taq* Plus Master Mix (Dye) 12.5 μL (试剂购自北京康为世纪生物科技有限公司)、9.5 μL ddH₂O。鉴于引物较多, 统一使用降落 PCR 程序, 在 T100 Thermal Cycler (Bio-Rad Research, USA)扩增仪上进行。扩增程序为 94℃预变性 5 min; 94℃变性 30 s, 65℃退火 30 s, 72℃延伸 30 s, 共 10 个循环, 每个循环退火温度降低 0.7℃; 94℃变性 30 s, 55℃退火 30 s, 72℃延伸 30 s, 共 25 个循环; 最后 72℃延伸 7 min, 4℃保存。PCR 产物经 6%非变性聚丙烯酰胺凝胶电泳分离, 140 V 恒压下电泳 3.0 h, 染色、照相及保存, 其中染色采用电泳后泡染

法,用水和 0.1 mol L⁻¹ NaCl 稀释 GelStain 10,000X 储备试剂 3300 倍,即标准染色液。GelStain 染料购自北京全式金生物技术有限公司(货号: GS101-01)。

1.5 数据统计分析

SSR 频率是指基因组中出现一个 SSR 位点的距离,即每若干 kb 出现一个 SSR 位点; SSR 丰度是指基因组中所有 SSR 位点的数量之和; SSR 相对丰度是指每百万个碱基中所含的 SSR 位点数量。按电泳扩增结果,选择清晰的扩增条带人工读带,在相同水平迁移位置上,有条带的记为“1”,没有条带的记为“0”,缺失数据记为“-”,根据统计的结果建立 0~1 矩阵。利用 NTSYS-pc 2.1 软件中的子程序 SIMQUAL 计算样品间的相似性系数(SM),然后用子程序 SAHN 中的非加权平均法 UPGMA (unweighted pair-group method with arithmetic means)进行聚类分析,最后利用 Tree plot 绘制树状聚类图。参考 Smith^[28]的方法计算引物的多态性信息量(polymorphism information content, PIC)。PIC=1- $\sum P_{ij}^2$, 其中 P_{ij} 为某标记 i 扩增的第 j 个等位基因出现的频率, PIC 的范围为 0~1, 当 PIC > 0.5

时为高多态性引物; 0.25<PIC<0.5 时为中度多态性引物; PIC < 0.25 时为低多态性引物。

2 结果与分析

2.1 甘蔗栽培种 SSR 位点的数量、类型及分布频率

利用 MISA 软件扫描甘蔗栽培种 R570 的 4660 个 BAC 文库序列(总长为 382 Mb), 通过分析 1~6 核苷酸重复基序, 设定 SSR 位点的长度不低于 10 bp, 共找到 27,241 个 SSR 位点(表 1), 平均每 1.08 个基因或 14.01 kb 含有 1 个 SSR 位点。其中单核苷酸重复基序类型出现的频率最高, 为 11,079 个位点, 占总数的 40.67%; 三核苷酸重复基序类型次之, 为 6447 个位点, 占总数的 23.67%, 两者合计占总数的 63.33%。而二、四、五、六核苷酸类型及复合型所占的比例相对较低, 分别为 11.97%、4.92%、9.32%和 9.45%, 合计占总数的 36.67%。

对于不同 SSR 核苷酸重复基序类型而言, 重复次数越少, 出现的频率越高。在 1~6 核苷酸重复基序类型中, 优势重复基序类型数量占比最多的重复次数都接近其筛

表 2 甘蔗栽培种 R570 基因组上各类核苷酸重复基序分布特征信息
Table 2 Distribution of various nucleotide repeat motifs in the genome of sugarcane cultivar R570

重复次数 Repeat number	核苷酸重复基序 Nucleotide repeat motif						合计 Total
	Mono-	Di-	Tri-	Tetra-	Penta-	Hexa-	
3					2044	2104	4148
4				941	313	355	1609
5			3718	225	112	72	4127
6		1210	1342	78	26	18	2674
7		527	636	29	9	6	1207
8		305	281	10	7	5	608
9		199	168	9	4	3	383
10	6297	128	77	8	4	0	6514
11	1997	84	71	5	2	4	2163
12	891	57	38	8	3	2	999
13	496	45	17	6	3	2	569
14	262	49	11	2	3	0	327
15	182	39	17	2	0	0	240
>15	954	619	71	18	8	3	1673
合计 Total	11079	3262	6447	1341	2538	2574	27241
优势重复次数 Dominant repeat number (%)	56.84	37.09	57.67	70.17	80.54	81.74	78.08
比例 Proportion (%)	40.67	11.97	23.67	4.92	9.32	9.45	100
平均重复次数 Mean repeat number	11.68	11.23	6.1	4.78	3.38	3.3	
设计引物的位点数 Number of loci primer designed	11079	3262	6447	1122	1995	1815	
比例 Proportion (%)	100	100	100	83.67	78.61	70.51	

选标准, 平均重复次数分别是 11.68、11.23、6.10、4.78、3.38、3.30。在单核苷酸重复基序类型数量中, 优势重复次数最多的是 10 次重复, 有 6297 个, 占总位点数 23.12%, 与三核苷酸重复基序总数基本相当, 但比二、四、五、六核苷酸重复基序的总数还多。在三核苷酸重复基序类型中, 优势重复次数最多的是 5 次重复, 有 3718 个, 占总位点数 13.65%, 除了单核苷酸重复类型外, 也都超过了其他 4 种类型重复基序类型的总数。二、四、五、六核苷酸重复基序类型优势重复次数相对较少, 其中四核苷酸重复基序的总数最少, 其优势重复次数为 4 次重复, 仅有 941 个, 占总位点数 3.45%。总体而言, 重复次数在 3~10 之间的总 SSR 位点数为 21,270 个, 占总位点数的 78.08%。在 1~6 核苷酸重复基序类型中, 优势重复次数占比分别从 56.83%、37.09%、57.67%、70.17%、80.54%和 81.74%, 除了二核苷酸重复基序较低外, 其他基序类型呈现出逐步递升趋势。此外, 在设计 SSR 引物成功率上, 随着核苷酸重复基序类型的增多和优势重复次数的降低, 设计到符合标准的 SSR 引物的成功率在逐渐下降。

2.2 甘蔗与其他 4 种禾本科植物 SSR 位点的数量和频率特征比较

选取了 4 种具有代表性的禾本科植物(高粱、玉米、水稻、二穗短柄草)和甘蔗比较, 它们的基因组变化范围在 272 Mb (二穗短柄草)和 2061 Mb (玉米)之间(表 3)。SSR 数量与基因组大小存在极显著的正相关($r = 0.92$; $P < 0.01$), 但 SSR 的相对丰度和频率与基因组大小没有相关

性。在 1~6 核苷酸重复基序中, 水稻 SSR 位点相对丰度都是最高的, 其次是二穗短柄草(除二和六核苷酸重复基序外), 而甘蔗的 SSR 相对丰度基本都是最低的(除单核苷酸重复基序外)。总的相对丰度从高到低分别为水稻(566.45)、二穗短柄草(361.15)、高粱(350.00)、玉米(152.54)、甘蔗(71.33), SSR 位点分布在禾本科植物中存在丰富的多样性。同时, SSR 出现的频率也与 SSR 数量成正比, 频率从高到低呈现相同的变化趋势。

5 种禾本科植物的 6 种不同核苷酸重复基序的 SSR 的相对数量(丰度)比较(表 3)表明, 除甘蔗的单核苷酸外, 其他物种都表现出三核苷酸的相对数量(丰度)最多, 六核苷酸重复基序次之, 且都是玉米的 SSR 数量最高, 水稻和甘蔗最少, 而相对数量(丰度)上水稻最高, 玉米和甘蔗最低。

甘蔗和高粱同属于禾本科黍亚科, 两者大约在八至九百万年前由共同祖先分化而来, 且高粱进化相对较慢, 保持了其祖先相对完整的基因组组成^[22]。甘蔗和玉米是 5 个禾本科物种中基因组最大的 2 个, 因此, 也最有可能出现长 SSR 序列。从表 4 可以看出, 所有物种的 1~6 核苷酸重复基序中, 前 3 种最长的 SSR 基序类型基本都是 A/T 碱基组成的重复基序类型, 而不是 C/G 重复类型。在单核苷酸重复和二核苷酸重复基序类型中, 最长的核苷酸重复 A(88)和 AC(910)均出现在玉米的基因组内, 而甘蔗和玉米相对都是最低的。在三核苷酸重复类型中, 最长的核苷酸重复(TGT)369 和(ACT)366 分别出现在甘蔗和高粱中, 水稻中则最低。对于四核苷酸到六核苷酸重复基序, 长的

表 3 5 种禾本科植物中 1~6 核苷酸重复基序类型的 SSR 数量和相对丰度
Table 3 SSR number and relative abundance of 1~6 nucleotide repeat motif types in five Gramineae plants

物种 Species	项目 Item	甘蔗 <i>S. spp.</i>	高粱 <i>S. bicolor</i>	玉米 <i>Z. mays</i>	水稻 <i>O. sativa</i>	二穗短柄草 <i>B. distachyon</i>
单核苷酸	数量 Number	11079.00	14294.00	30700.00	15311.00	7991.00
Mono-nucleotide	相对丰度 Relative abundance	29.00	19.34	14.90	41.16	29.38
二核苷酸	数量 Number	3262.00	38090.00	64663.00	35315.00	9175.00
Di-nucleotide	相对丰度 Relative abundance	8.54	51.54	31.37	94.93	33.73
三核苷酸	数量 Number	6447.00	80299.00	185973.00	77566.00	37005.00
Tri-nucleotide	相对丰度 Relative abundance	16.88	108.66	90.23	208.51	136.05
四核苷酸	数量 Number	1341.00	47062.00	58806.00	26411.00	17428.00
Tetra-nucleotide	相对丰度 Relative abundance	3.51	63.68	28.53	71.00	64.07
五核苷酸	数量 Number	2538.00	16630.00	38408.00	17080.00	7972.00
Penta-nucleotide	相对丰度 Relative abundance	6.64	22.50	18.64	45.91	29.31
六核苷酸	数量 Number	2574.00	62227.00	119813.00	38940.00	18629.00
Hexa-nucleotide	相对丰度 Relative abundance	6.74	84.20	58.13	104.68	68.49
SSR 数量(丰度) SSR number (abundance)		27241.00	258602.00	498363.00	210623.00	98200.00
基因组大小 Genome size (Mb)		382.00	739.00	2061.00	372.00	272.00
总的相对丰度 Relative abundance		71.33	350.00	152.54	566.45	361.15
SSR 频率 SSR frequency (1 kb ⁻¹)		14.02	2.86	6.56	1.77	2.77

表 4 5 种禾本科植物中前 3 种最长 SSR 基序类型

Table 4 The first three longest SSR motifs in the five Gramineae plants

项目 Item	甘蔗 <i>S. spp</i>	高粱 <i>S. bicolor</i>	玉米 <i>Z. mays</i>	水稻 <i>O. sativa</i>	二穗短柄草 <i>B. distachyon</i>
单核苷酸 Mono-nucleotide	(T)75 (T)63 (G)49	(A)71 (A)59 (A)53	(A)88 (A)85 (A)83	(C)51 (A)49 (A)48	(A)49 (C)45 (A)43
二核苷酸 Di-nucleotide	(TA)71 (TG)69 (TA)55	(AT)280 (AT)276 (AT)270	(AC)1366 (AC)910 (AT)178	(AC)170 (AT)104 (AT)100	(AT)312 (AT)182 (AT)158
三核苷酸 Tri-nucleotide	(TGT)123 (ATT)59 (TTA)56	(ACT)366 (AAT)327 (AAT)318	(ACC)291 (AAT)207 (ACT)132	(AAT)165 (AAT)147 (AAT)126	(AAT)225 (AAT)171 (AAT)138
四核苷酸 Tetra-nucleotide	(TTAT)23 (ACAT)25 (ATGT)26	(ACAT)524 (AGAT)388 (ACAT)260	(ACAT)196 (ACAT)144 (AAAG)100	(ACAT)132 (ACAT)96 (ACAT)96	(ACAT)196 (ACAT)180 (ACAT)180
五核苷酸 Penta-nucleotide	(CTTTT)29 (TTTTG)25 (AATAT)24	(AATAT)740 (AATAT)430 (AATAT)315	(AATAT)115 (ACTAT)115 (AATAT)85	(AATAT)55 (AATAT)55 (AATAT)55	(AGATC)100 (ACGCC)75 (AGATG)65
六核苷酸 Hexa-nucleotide	(ATTGTC)43 (TTTTTG)32 (TTATAT)16	(AAATAT)390 (AGATAT)366 (AAATAT)294	(AATAGT)198 (AATAGT)72 (AACCAT)66	(ACCTAT)90 (ACATAT)78 (ACATAT)78	(AACAGC)90 (ACTGAT)78 (AGAGAT)66

SSR 序列都出现在高粱中, 甘蔗次之。同时, 5 个禾本科植物都有一个最长 ACAT 重复基序, 其余类型除了有 1 个 C/G 外, 都由 A/T 组成。在所有的五核苷酸重复类型中, AATAT 重复基序最多, 占到 9 个, 其余类型只含有 1~3 个 G/C, 说明大多数长的五核苷酸重复类型都是由 A/T 组成。在六核苷酸重复类型中, 玉米、水稻和二穗短柄草的重复序列长度都远远低于高粱和甘蔗。

2.3 甘蔗 SSR 重复基序种类及频率特征分析

在甘蔗栽培种基因组中, 单、二、三核苷酸重复基序占比达到 76.31%以上, 因此分析其优势重复基序和各种类型 SSR 序列的碱基组成, 对于进一步筛选和验证多态性 SSR 引物具有重要作用。如图 1 所示, 从基序结构出现的频率上看, A/T 和 G/C 各占 84.78%和 15.22%, 分别是甘蔗栽培种基因组 SSR 中单核苷酸出现频率最高和最低的结构; AT/TA(31.51%)和 CG/GC(6.32%)分别是二核苷酸中出现最多和最少的基序结构; 三核苷酸出现频率最多的结构是 TGT/ACA(16.04%), 其次是 CGC/GCG、CCG/GGC 和 GCC/CGG 基序结构类型, 分别占 12.02%、11.31%、10.67%, 合计占 34.00%; 最少的则是 TCA/AGT, 占 0.35%。甘蔗和 4 种禾本科植物 SSR 基序都有碱基的偏好性和规律, 在单核苷酸中, 甘蔗、高粱和水稻 A/T 的比例也都高于 G/C 的比例, 分别为 84.78%、65.60%和 64.40%, 但在玉米和二穗短柄草的结果完全相反; 在二核苷酸中, 甘蔗(31.51%)和其他 4 个物种(22.6%~54.2%)都是 AT/TA 重复比例最高, 而 GC/CG 重复的比例最低, 基本都在 3.1%~6.6%之间; 在三核苷酸类型中, CGC/GCG、CCG/GGC 和 GCC/CGG

三者合计在所有物种中都最高, 且以水稻(44.75%)最高, 甘蔗(34.0%)次之, 玉米(12.74%)最低^[26]。

2.4 SSR 标记的开发及其扩增效率和多态性验证

根据 Pan^[12]筛选到 21 对多态性丰富的 SSR 引物对应的重复基序类型, 它们以 TG 和 AG 重复基序为主, 本研究选择以 TG 和 AG 基序类型以及重复次数分别在 TG(11~69)、AG(23~38)之间的 SSR 位点, 对其进行引物设计和合成, 共计 50 对 SSR 引物, 对 4 个不同甘蔗属材料(栽培种 R570, 栽培种 ROC1, 热带种 LA purple 和割手密种 SES208)进行 SSR 扩增和多态性筛选。共有 45 对引物能够扩增出清晰的扩增条带, 其余的 5 对引物没有扩增条带或者扩增产物量较弱, 其中 35 对引物在 4 个材料上呈现多态性(表 5), 多态率为 70% (35/50), 其中 TG 重复类型的引物有 28 对, AG 重复类型的引物有 7 对。图 2 显示了部分引物筛选扩增结果。为了进一步验证本研究鉴定到 SSR 引物的多态性, 选用 20 对多态性较高的 SSR 引物, 对我国 50 年来育成 163 个甘蔗品种的 18 个骨干亲本(它们的血缘来自热带种、割手密种、大茎野生种和印度种的 2~4 个种, 且具有较低的共祖系数)、2 个甘蔗祖先种(割手密种 SES 208 和热带种 LA purple)和 4 个重要甘蔗栽培种(LCP85-384、R570、ROC16 和 ROC22)进行遗传多样性分析。结果 20 对引物在 24 个甘蔗实验材料上呈现多态性, 共扩增得到等位基因 95 个, 每对扩增出 1~7 个等位基因(表 5), 平均每一对引物扩增出 4.75 个等位基因。图 3 展示了其中 FAFUR-S22 引物在 24 个供试甘蔗材料上的 PCR 扩增电泳图谱。

图 1 1~3 核苷酸重复基序类型及数量

Fig. 1 Number and types of mono-, di-, and tri-nucleotide repeat motifs

图 2 7 对不同 SSR 引物在 4 个甘蔗属材料上扩增的电泳图

Fig. 2 Electrophoretic patterns of seven pairs of SSR primers amplified in four *Saccharum* clones

1~4: FAFUR-S44; 5~8: FAFUR-S45; 9~12: FAFUR-S46; 13~16: FAFUR-S47; 17~20: FAFUR-S48; 21~24: FAFUR-S49; 25~28: FAFUR-S50。4 个扩增产物为 SES208 (1, 5, 9, 13, 17, 21, 25)、LA purple (2, 6, 10, 14, 18, 22, 26)、ROC16 (3, 7, 11, 15, 19, 23, 27)和 R570 (4, 8, 12, 16, 20, 24, 28); M: 50 bp DNA ladder (3421A)。

1~4: FAFUR-S44; 5~8: FAFUR-S45; 9~12: FAFUR-S46; 13~16: FAFUR-S47; 17~20: FAFUR-S48; 21~24: FAFUR-S49; 25~28: FAFUR-S50。The amplification products of four samples were SES208 (1, 5, 9, 13, 17, 21, 25), LA purple (2, 6, 10, 14, 18, 22, 26), ROC16 (3, 7, 11, 15, 19, 23, 27), and R570 (4, 8, 12, 16, 20, 24, 28). M: 50 bp DNA ladder (3421A)。

基于上述 20 对 SSR 引物所给出的 95 种等位基因类型, 对 24 份甘蔗属材料的 UPGMA 聚类分析, 供试材料之间的遗传相似系数分布在 0.40~0.82 之间(图 4), 在遗传相似性系数为 0.525 时, 可将 24 个甘蔗材料分成 5 种类型, 第 1 种类型包含甘蔗栽培种 Co 1001 和 Co 419; 第 2 种类型有 19 个甘蔗材料; 第 3 种类型为 1 个甘蔗材料热带种类型 LA purple; 第 4 类型为 1 个甘蔗栽培种材料 CP28-11; 第 5 种类型为 1 个割手密种 SES 208。需要注意的是, 在相似性系数为 0.4 时, 割手密种 SES208 与

他甘蔗栽培种和热带种(LA purple)较早分开, 表明割手密种与甘蔗栽培种具有较远的遗传关系。根据张琼等^[29]分析结果, CP28-11 具有热带种(0.5)、割手密(0.125)和印度种(0.375)的血缘关系, 遗传关系介于割手密和热带种之间。本研究中, 热带种(LA purple)在相似性系数为 0.525 时与其他栽培种分开, 接着是印度种 Co 1001 和 Co 419 在相似性系数为 0.551 时与其他栽培种分开, 表明印度种亲缘关系介于热带种与甘蔗栽培种之间, 也具有较丰富的遗传多样性。

图 3 SSR 引物(FAFUR-S22)在 24 个甘蔗材料上扩增的电泳图

Fig. 3 Electrophoretic patterns of SSR primers FAFUR-S22 in twenty four tested *Saccharum* clones

1: Co 1001; 2: Co 419; 3: CP28-11; 4: CP49-50; 5: CP67-412; 6: CP72-1210; 7: F108; 8: NCo310; 9: ROC1; 10: 川 73-219; 11: 桂糖 11 号; 12: 华南 56-12; 13: POJ2878; 14: 科 5; 15: 崖城 71-374; 16: 粤农 73-204; 17: 云蔗 65-225; 18: 华南 56-21; 19: LCP85-384; 20: R570; 21: ROC16; 22: ROC22; 23: LA purple; 24: SES208; M: 50 bp DNA ladder (3421A).

1: Co 1001; 2: Co 419; 3: CP28-11; 4: CP49-50; 5: CP67-412; 6: CP72-1210; 7: F108; 8: NCo310; 9: ROC1; 10: C73-219; 11: Guitang 11; 12: Huanan 56-12; 13: POJ2878; 14: Ke 5; 15: Yacheng 71-374; 16: Yuenong 73-204; 17: Yunzhe 65-225; 18: Huanan 56-21; 19: LCP85-384; 20: R570; 21: ROC16; 22: ROC22; 23: LA purple; 24: SES208; M: 50 bp DNA ladder (3421A).

表 5 具有扩增多态性的甘蔗 SSR 引物信息表

Table 5 Primers information of sugarcane SSR with polymorphic amplification

引物名称 Primer name	重复基序 Motif	左引物序列 Left-primer (5'-3')	退火温度 T_m (°C)	右引物序列 Right-primer (5'-3')	退火温度 T_m (°C)	产物大小 Product size (bp)	PIC
FAFUR-S1	(TG)69	TCATACCCATTGGAAGAAGC	60.5	GTTATGTTGCCGTGCCAAGT	59.8	278	0.85
FAFUR-S3	(TG)39	TAGCCTTTGGTCGTTCTTGG	58.2	AATGCTTCATCCATAGGGGA	59.3	259	0.84
FAFUR-S7	(TG)32	GCCTGGGGAACATATGCTGTA	59.1	CAAGCATTGAAGTTGCCAAA	59.0	254	0.61
FAFUR-S12	(TG)24	CGTCAGTTGCTCAGCTCTTG	58.0	CCCTGGGAAGAAGAGGTAGG	58.6	223	0.69
FAFUR-S17	(TG)19	AATGATGTTTCGCCTGATCC	60.2	ACCAACACAACCTCGTACCC	60.1	166	0.77
FAFUR-S18	(TG)19	CCACATTCTTCGACCCTGTT	59.8	CCATCCTGCGAACTAACCAT	59.7	183	0.71
FAFUR-S22	(TG)18	AGGGCACGAGGTATTGCTTA	58.9	AACCGGTCAAATCACACACA	59.2	179	0.68
FAFUR-S24	(TG)17	ATCTTTCGGCATCAACTTGG	60.1	GCTTCAAGCCATCTGTCTCC	60.3	274	0.73
FAFUR-S32	(TG)13	CAACGAATTCACCTTGACACA	60.0	TCATGGCTATTGTGGTCTGG	60.4	207	0.61
FAFUR-S33	(TG)13	CTCCTCTGTCACCCAGCACT	58.8	GATCACCCCAGATCCAGAGA	59.6	179	0.75
FAFUR-S34	(TG)13	TGCTGATTATGTGCTGCCTC	58.5	CACGCCTAGGGCATAAGAGA	58.4	222	0.67
FAFUR-S36	(TG)12	AGGCATGGGAATTTCTCTCC	60.1	GGCCTCTCTTTAGTGCAGGA	59.8	265	0.77
FAFUR-S38	(TG)12	GACACCCACCACAGGACTTT	60.3	CCCTCCCCAATCCTATCAGT	60.1	198	0.66
FAFUR-S40	(TG)11	GCTGATGTTTGGTCATGTGG	61.0	TGCAGACTCAGAAGTAGCCG	60.5	246	0.86
FAFUR-S41	(TG)11	TGTTTCAGGCACTGTTTGG	60.9	AGCAATGTGTTCTCCATCCA	60.5	261	0.73
FAFUR-S42	(AG)38	CGGCACAAGTAAATGCAAGA	59.7	AGTACTGCCAACAAGGCAGG	58.3	230	0.85
FAFUR-S43	(AG)34	CTTGAGCTCGTAGCCTCCTC	60.3	GCCTCTGCTGTCTGCTCTCT	59.6	267	0.92
FAFUR-S44	(AG)31	AGTGCAGGTTGGCTTCTGT	60.2	GGGGATTCCAAGTCTCAACA	59.8	206	0.82
FAFUR-S47	(AG)25	GTACCAGCCCCAAAACCTGGA	59.8	TTGTCACTGGGAACACGGTA	60.1	280	0.73
FAFUR-S49	(AG)23	TTCTCCGTCAACTGTCATGC	59.6	TCCTACGGAGGGAAATCAAAA	60.2	273	0.81
FAFUR-S4 [#]	(TG)33	CGACTGGAAGAAGATCGAGG	58.2	GAGGTACTGCATGCCCAAAT	60.1	185	-
FAFUR-S5 [#]	(TG)33	CTTCTCCAGTAGCCGAGT	59.3	TCTCGAATTCGCAAGGAACT	57.9	257	-
FAFUR-S6 [#]	(TG)32	GGAAGGAGGAGATGGAAAGG	59.4	CGCAACACGTACACACACAC	59.6	245	-
FAFUR-S9 [#]	(TG)26	GTTTCTTCTCGAGGGGAG	57.9	AATGCTGGGATCGAAGTTTG	60.2	213	-
FAFUR-S15 [#]	(TG)20	TGCTATCTCCTGCTTGACACA	60.2	GCCTCACACACACACACACA	59.4	268	-
FAFUR-S16 [#]	(TG)19	TGCTTGCTAGCTTGCGACTA	60.4	ACAAC TAGGCCATCAGTGGG	59.7	268	-
FAFUR-S19 [#]	(TG)19	AGCCCAACAGAAATACGCAC	60.6	GGGCTCACTCAAAAACCAAAA	58.8	269	-

(续表 5)

引物名称	重复基序	左引物序列	退火温度	右引物序列	退火温度	产物大小	PIC
Primer name	Motif	Left-primer (5'-3')	$T_m(^{\circ}\text{C})$	Right-primer (5'-3')	$T_m(^{\circ}\text{C})$	Product size (bp)	
FAFUR-S20 [#]	(TG)18	TCGATTGGAGTCTTCAGCAA	59.9	CCCATGAGATTGTATTCGGC	60.3	269	-
FAFUR-S21 [#]	(TG)18	TGCACTGTTTAAATTCCCCC	60.3	AAATCTCCCTTCATGATGCC	58.8	229	-
FAFUR-S25 [#]	(TG)17	TCGTAGAAGCACTTCAGGGAG	58.8	CCAAAATAAGGCCATCGAAA	60.1	162	-
FAFUR-S26 [#]	(TG)17	CTTTGTCCCCCTTCTCCATCC	57.9	TCTCGAAGTCGCAAGGAACT	60.5	185	-
FAFUR-S28 [#]	(TG)16	TGGCTCACTGAAAATCTCCC	61.1	TGTGTGGCAAGATAAGAAGGG	60.3	250	-
FAFUR-S29 [#]	(TG)15	TGCTGATTATGTGCTGCGTC	59.6	ATCGATCACACACCTAGGGC	59.5	234	-
FAFUR-S46 [#]	(AG)25	ATCGATCCTGGGGTAGCTTT	58.4	TTTCTCTGCAAGACTGCAA	58.7	262	-
FAFUR-S48 [#]	(AG)23	TTCCAGATTCTTTCCACGG	60.3	GTCACCTGGGAACCTACCCCT	59.6	257	-

[#]表示引物对的多态性相对较低。# indicates that the primer pairs have relatively low polymorphism.

图 4 基于 SSR 分子标记的 24 份甘蔗属材料的 UPGMA 聚类分析
Fig. 4 UPGMA dendrogram of twenty four *Saccharum* clones based on SSR markers

3 讨论

SSR 分子标记具有多态性高、重复性好、操作简便及共显性等优点，被广泛应用于甘蔗指纹图谱构建^[11]、遗传多样性分析^[12-14]、遗传连锁图谱构建及 QTL 定位等方面^[15-18]。但目前尚未完全破译甘蔗栽培种基因组，对于 SSR 标记开发仅仅停留在 BAC 文库和表达序列标签(EST)水平。本研究首次利用已经完成的甘蔗栽培种单倍体基因组数据，分析 SSR 在基因组分布频率、基序类型及序列特征等相关信息，对于甘蔗基因组遗传研究和分子标记开发提供重要数据支撑。

3.1 甘蔗栽培种基因组 SSR 位点的特征分析

本研究从甘蔗栽培种 R570 的 4660 个 BAC 文库序列组装的单倍体基因组(累计总长为 382 Mb，预测到 25,316 个编码蛋白基因)中，共发现 27,241 个 SSR 位点，在基因组中 1~6 核苷酸重复基序中，单核苷酸重复基序占比最高，

达到 40.67%，其次是三核苷酸(占 23.67%)、二核苷酸重复基序(占 11.97%)。这与拟南芥^[30]、水稻^[30]、小麦^[31]、玉米^[3]、棉花^[31]、马铃薯^[32]、苹果^[15]和葡萄^[33]的基因组 SSR 位点中存在大量的单、二、三核苷酸的结果基本一致；与高粱^[26,34]、根瘤菌和秀丽线虫^[35]的研究结果不同，它们的基因组 SSR 位点以四、五、六核苷酸重复基序为主。一般情况下，短重复基序占多数表明物种进化水平相对较高^[36]，而长重复基序占多数的物种具有较低的突变频率或较短的进化时间^[37]。甘蔗与高粱在八至九百万年前由共同祖先分化而来，高粱进化相对较慢，保持了相对完整的祖先基因组构成^[22]。本研究结果进一步说明，与高粱相比，甘蔗在进化与分类地位中处于相对较高水平，其基因组也经历了较长的进化时间或具有较高的突变频率。但是，甘蔗栽培种基因组遗传基础复杂、杂合度高，尚未完成基因组测序，本研究仅依据高粱基因组为参考，利用

BAC 文库序列拼装一套甘蔗单倍体基因组作为 SSR 挖掘的模板^[25], 基因组数据并不完整。因此, 分析结果尚无法与其他物种上的 SSR 数据比较, 但是也初步证明了与大多数禾本科植物具有相似的核苷酸基序组成。

在不同物种的 SSR 位点信息特征研究方面, SSR 基序的结构组成, 尤其是 1~3 核苷酸重复基序具有明显的偏好性和规律性, 但 4~6 核苷酸重复基序, 由于基序类型呈现指数上升(分别是 2,521,020 和 4092), 碱基偏好性和规律性表现的不明显。本研究结果显示, 甘蔗栽培种基因组序列中, 单核苷酸重复基序类型以 A/T 出现的次数最多(占 84.78%), 这与水稻、高粱^[26]、苹果^[15]和烟草^[38]的研究结果一致, 但是与二穗短柄草、玉米 G/C 基序结构出现次数不一致^[26]; 而二核苷酸重复基序类型则以 AT/TA 出现的次数最多(占 31.51%), CG/GC(占 6.32%)最少, 该结果与高粱、烟草、苹果的研究结果一致, 但是在其他禾本科植物小麦^[39]、二穗短柄草、水稻和玉米中不一致^[26], 它们是以 AG 重复类型最多; 三核苷酸重复基序类型以 TGT/ACA (16.04%) 出现的次数最多(占 16.04%), 其次是含有 CGC/GCG 等 6 类不同组合基序类型(合计占 34.0%), 与拟南芥(AAG/TTC)、烟草(AAC/TTG)、苹果(AAC/GTT)基本相似, 与大麦、小麦、玉米、水稻、高粱、黑麦等其他禾本科作物有所不同, 它们以 CCG/GCC 出现次数最多, 而 ATT/TAA 最少。综合以上, 甘蔗的 1~3 核苷酸重复基序均以 A 和 T 核苷酸构成的基序为主要类型, 这与真核生物基因组 SSR 位点分析结果基本一致^[36], 产生这样的结果有 2 种可能原因, 一是甲基化的 C 残基转变为 T^[40]; 二是在 DNA 复制过程中产生 SSR 位点, 基于 A、T 结构类型基序比 G、C 类型需要较少能量, 导致富含 A、T 碱基结构类型比 G、C 类型容易产生^[38], 但是随着核苷酸重复类型从 3 增至 6 时, 这种氢键的能量优势就相对不明显了。

3.2 甘蔗栽培种基因组 SSR 的多态性分析和应用

近年来, 随着基因组测序技术进步和测序成本的降低, 已经完成许多生物基因组测序, 这些测序结果对于相应物种基因组 SSR 标记开发和应用提供了数据支撑, 还可以利用所开发的 SSR 标记对未测序的近缘物种进行遗传分析, 加快其群体结构分析、遗传多样性研究、亲缘关系鉴定、重要性状的功能基因 QTL 定位及关联标记开发与应用。本研究通过分析 Pan 等^[12]筛选和鉴定的 21 对多态性丰富的 SSR 引物的重复基序类型, 选择了 50 对 SSR 引物, 它们由 TG(41 对)和 AG(9 对)基序重复类型组成。此外, 利用 4 个甘蔗属材料进行 PCR 扩增和多态性分析, 共有 45 对引物扩增到预期的扩增片段, 其中 TG 和 AG 基序类型的分别有 37 对(扩增效率 90.24%)和 8 对, 扩增效率为 88.89%, 远高于 Cordeiro 等^[41]开发甘蔗 EST-SSR 引物的 60%, 也高于 Oliveira 等^[14]开发的甘蔗基因组 SSR 的 70%、Bushman 等^[42]针对多年生黑麦草开发的基因组和 EST-SSR 的 75 %、Fernandez 等^[43]获得的甜瓜基因组和 EST 的 79 %和 Hwang 等^[44]发掘的西瓜 EST-SSR 的

79%。此外, 在我们设计的 50 对 SSR 引物中, 有 35 对呈现出明显的多态性, TG 和 AG 基序类型分别有 28 对和 7 对, 且 AG 基序的多态性(77.78%)高于 TG 类型(68.29%)。

本研究还在对 SSR 标记初步扩增分析的基础上, 进一步利用 24 个甘蔗重要亲本材料验证上述 20 对 SSR 引物的多态性。结果 20 对引物扩增出 95 种不同等位基因, 每对引物扩增 1~8 个等位基因, 平均每对引物扩增 4.75 个。与 Pinto 等^[45]的 6.04 个、Oliveira 等^[14]的 7.55 个、Marconi 等^[46]的 6.0 个相比, 本研究开发的甘蔗 SSR 引物扩增出的等位基因偏少, 主要原因应该是不同研究所采用的电泳分离方法存在差异, 当将本研究鉴定的 SSR 引物应用于甘蔗遗传研究时, 建议进一步优化 PCR 扩增条件, 同时采用分辨率更高的电泳分离方法, 比如变性聚丙烯酰胺凝胶电泳、毛细管电泳等, 以便更准确和高效地对目标材料进行扩增和多态性分析。同时, 20 对 SSR 引物扩增的多态性信息量(PIC)的变异幅度在 0.61~0.92 之间, 平均值为 0.78, 说明本研究筛选和鉴定的 SSR 引物所扩增出来的条带多态性丰富。特别需要指出的是, 川 73-219 和粤农 73-204 在相似性系数为 0.9 时没有分开, 说明两者亲缘关系非常近, 需要用更多 SSR 引物鉴定和分析。

综上所述, 本研究利用大小约为 382 Mb 的甘蔗栽培种 R570 的单倍体基因组, 开发了 27,241 个 SSR 位点, 平均每 BAC 片段有 6.29 个。通过分析和比较其他物种基因组序列上 SSR 位点、类型及结构等特征, 设计了 50 对 SSR 引物, 分别利用 4 个甘蔗属近缘种(R570、ROC1、LA purple 和 SES 208)和 24 个甘蔗重要杂交亲本材料, 进行扩增效率和多态性分析, 获得了 20 对引物扩增条带清晰、多态性条带比率高, 具有较好应用前景的 SSR 标记。这 20 对引物在 24 个甘蔗杂交亲本中扩增出 95 种不同等位基因, 每对引物扩增 1~8 个等位基因之间, 平均为 4.75 个。本研究建立的 SSR 引物筛选和鉴定方法, 对于开发高质量多态性 SSR 标记具有积极的借鉴作用; 所开发的 20 对具有良好多态性的 SSR 标记, 能够为甘蔗及其近缘种的品系鉴定、遗传多样性分析、遗传图谱构建及重要性状的遗传机制解析等提供了分子标记支撑。

References

- [1] 刘燕群, 李玉萍, 梁伟红, 宋启道, 秦小立, 叶露. 国外甘蔗产业发展现状. 世界农业, 2015, (8): 147-152.
Liu Y Q, Li Y P, Liang W H, Song Q D, Qin X L, Ye L. Current situation of sugarcane industr in the world, *World Agric*, 2015, (8): 147-152 (in Chinese with English abstract).
- [2] FAOSTAT: http://www.fao.org/faostat/zh/#data?tdsourcetag=s_pctim_aiomsg.
- [3] 彭绍光. 甘蔗育种学. 北京: 农业出版社, 1990. pp 4-5.
Peng S G. Sugarcane Genetic Breeding. Beijing: Agricultural Press, 1990. pp 4-5 (in Chinese).
- [4] Hermann S R, Aitken K S, Jackson P A, George A W, Piperidis N, Wei X, Kilian A, Detering F. Evidence for second division restitution as the basis for $2n+n$ maternal chromosome transmission

- in a sugarcane cross. *Euphytica*, 2012, 187: 359–368.
- [5] Piperidis G, Piperidis N, D'Hont A. Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Mol Genet Genomics*, 2010, 284: 65–73.
 - [6] Wang H B, Chen P H, Yang Y Q, D'Hont A, Lu Y H. Molecular insights into the origin of the brown rust resistance gene *Brul* among *Saccharum* species. *Theor Appl Genet*, 2017, 130: 2431–2443.
 - [7] Smith D, Devey M E. Occurrence and inheritance of microsatellites in *Pinus radiata*. *Genome*, 1994, 37: 977–983.
 - [8] Morgante M, Hanafey M, Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat Genet*, 2002, 30: 194–200.
 - [9] Levinson G, Gutman G A. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol*, 1987, 4: 203–221.
 - [10] Michael K, Eva-Maria D, Ugo R, Nicoletta C, Uta-Dorothee I, Manfred K, Wolfgang R M. Haplotype studies support slippage as the mechanism of germline mutations in short tandem repeats. *Electrophoresis*, 2004, 25: 3344–3348.
 - [11] Liu X L, Ma L, Chen X K, Ying X M, Cai Q, Liu J Y, Wu C W. Establishment of DNA fingerprint identity for sugarcane cultivars in Yunnan, China. *Acta Agron Sin*, 2010, 36: 202–210.
 - [12] Pan Y B. Highly polymorphic microsatellite DNA markers for sugarcane germplasm evaluation and variety identity testing. *Sugar Technol*, 2006, 8: 246–256.
 - [13] Ram K S, Satya N J, Suhail K, Sonia Y, Nandita B, Saurabh R, Vasudha B, Sanjay K D, Raman K, Sushil S. Development, cross-species/genera transferability of novel EST-SSR markers and their utility in revealing population structure and genetic diversity in sugarcane. *Gene*, 2013, 524: 309–329.
 - [14] Oliveira K M, Pinto L R, Marconi T G, Mollinari M, Ulian E C, Chabregas S M, Falco M C, Burnquist W, Garcia A A, Souza A P. Characterization of new polymorphic functional markers for sugarcane. *Genome*, 2009, 52: 191–209.
 - [15] 关玲, 章镇, 王新卫, 薛华柏, 刘艳红, 王三红, 乔玉山. 苹果基因组 SSR 位点分析与应用. *中国农业科学*, 2011, 44: 4415–4428.
Guan L, Zhang Z, Wang X W, Xue H B, Liu Y H, Wang S H, Qiao Y S. Evaluation and application of the SSR loci in apple genome. *Sci Agric Sin*, 2011, 44: 4415–4428 (in Chinese with English abstract).
 - [16] 刘新龙, 毛钧, 陆鑫, 马丽, Karen S A, Jackson P A, 蔡青, 范源洪. 甘蔗 SSR 和 AFLP 分子遗传连锁图谱构建. *作物学报*, 2010, 36: 177–183.
Liu X L, Mao J, Lu X, Ma L, Karen S A, Jackson P A, Cai Q, Fan Y H. Construction of molecular genetic linkage map of sugarcane based on SSR and AFLP markers. *Acta Agron Sin*, 2010, 36: 177–183 (in Chinese with English abstract).
 - [17] Andru S, Pan Y B, Thongthawee S, Burner D M, Kimbeng C A. Genetic analysis of the sugarcane (*Saccharum* spp.) cultivar 'LCP 85-384': I. linkage mapping using AFLP, SSR, and TRAP markers. *Theor Appl Genet*, 2011, 123: 77–93.
 - [18] Yang X, Islam M S, Sood S, Maya S, Hanson E A, Comstock J, Wang J. Identifying quantitative trait loci (QTLs) and developing diagnostic markers linked to orange rust resistance in sugarcane (*Saccharum* spp.). *Front Plant Sci*, 2018, 9: 350.
 - [19] Shamshad U H, Kumar P, Singh R K, Verma K S, Bhatt R, Sharma M, Kachhwaha S, Kothari S L. Assessment of functional EST-SSR markers (Sugarcane) in cross-species transferability, genetic diversity among Poaceae plants, and bulk segregation analysis. *Genet Res Int*, 2016, 2016: 7052323.
 - [20] Jianping W, Bruce R, Simone M, Qingyi Y, Jan E M, Haibao T, Cuixia C, Fares N, Graham W, John B. Microcollinearity between autopolyploid sugarcane and diploid sorghum genomes. *BMC Genomics*, 2010, 11: 261.
 - [21] Laurent G, Paulo A. Sugarcane genomics: depicting the complex genome of an important tropical crop. *Curr Opin Plant Biol*, 2002, 5: 122–127.
 - [22] Jannoo N, Grivet L, Chantret N, Garsmeur O, Glaszmann J C, Arruda P, D'Hont A. Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. *Plant J*, 2007, 50: 574–585.
 - [23] Paterson A H, Bowers J E, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A. The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, 2009, 457: 551–556.
 - [24] Garsmeur O, Charron C, Bocs S, Jouffe V, Samain S, Couloux A, Droc G, Zini C, Glaszmann J, Van Sluys M. High homologous gene conservation despite extreme autopolyploid redundancy in sugarcane. *New Phytol*, 2011, 189: 629–642.
 - [25] Garsmeur O, Droc G, Antonise R, Grimwood J, Potier B, Aitken K, Jenkins J, Martin G, Charron C, Hervouet C. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat Commun*, 2018, 9: 2638–2638.
 - [26] 郑燕, 张耿, 吴为人. 禾本科植物微卫星序列的特征分析和比较. *基因组学与应用生物学*, 2011, 30: 513–520.
Zheng Y, Zhang Z, Wu W R. Characterization and comparison of microsatellites in Gramineae. *Genom Appl Biol*, 2011, 30: 513–520 (in Chinese with English abstract).
 - [27] Martins W, de Sousa D, Proite K, Guimarães P, Moretzsohn M, Bertoli D. New softwares for automated microsatellite marker development. *Nucleic Acids Res*, 2006, 34: e31.
 - [28] Smith J S, Chin E C, Shu H, Smith O S, Wall S J, Senior M L, Mitchell S E, Kresovich S, Ziegler J. An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. *Theor Appl Genet*, 1997, 95: 163–173.
 - [29] 张琼, 齐永文, 张垂明, 陈勇生, 邓海华. 我国大陆甘蔗骨干亲本亲缘关系分析. *广东农业科学*, 2009, (10): 44–48.
Zhang Q, Qi Y W, Zhang T M, Chen Y S, Deng H H. Pedigree analysis of genetic relationship among core parents of sugarcane in mainland China. *Guangdong Agric Sci*, 2009, (10): 44–48 (in Chinese with English abstract).
 - [30] Lawson M J, Zhang L. Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genome Biol*, 2006, 7: R14.
 - [31] Kantety R V, La Rota M, Matthews D E, Sorrells M E. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol*, 2002, 48: 501–510.
 - [32] Tang J, Baldwin S J, Jacobs J M, Linden C G, Voorrips R E, Leunissen J A, van Eck H, Vosman B. Large-scale identification

- of polymorphic microsatellites using an in silico approach. *BMC Bioinformatics*, 2008, 9: 374.
- [33] 蔡斌, 李成慧, 姚泉洪, 周军, 陶建敏, 章镇. 葡萄全基因组 SSR 分析和数据库构建. *南京农业大学学报*, 2009, 32(4): 28–32.
- Cai B, Li C H, Yao Q H, Zhou J, Tao J M, Zhang Z. Analysis of SSRs in grape genome and development of SSR database. *J Nanjing Agric Univ*, 2009, 32(4): 28–32 (in Chinese with English abstract).
- [34] 陆景标, 李杰勤, 卢杰, 詹秋文. 高粱非编码区 SSR 引物设计以及 e-PCR 的验证. *种子*, 2010, 29(9): 1–6.
- Lu J B, Li J Q, Lu J, Zhan Q W. Design of SSR primers and verification of e-PCR in non-coding regions of sorghum genome. *Seed*, 2010, 29(9): 1–6 (in Chinese with English abstract).
- [35] 高亚梅, 韩毅强, 汤辉, 孙东梅, 王彦杰, 王伟东. 根瘤菌基因组内简单重复序列的分析. *中国农业科学* 2008, 41: 2992–2998.
- Gao Y M, Han Y Q, Tang H, Sun D M, Wang Y J, Wang W D. Analysis of simple sequence repeats in Rhizobium genomes. *Sci Agric Sin*, 2008, 41: 2992–2998 (in Chinese with English abstract).
- [36] Toth G, Gaspari Z, Jurka J. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res*, 2000, 10: 967–981.
- [37] Harr B, Schlotterer C. Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide under representation. *Genetics*, 2000, 155: 1213–1220.
- [38] 董治军, 焦芳婵, 肖炳光. 普通烟草及其祖先种基因组 SSR 位点分析. *中国农业科学*, 2015, 48: 2108–2117.
- Tong Z J, Jiao F C, Xiao B G. Analysis of SSR loci in *Nicotina tabacum* genome and its two ancestral species genome. *Sci Agric Sin*, 2015, 48: 2108–2117 (in Chinese with English abstract).
- [39] Yu J K, La Rota M, Kantety R V, Sorrells M E. EST derived SSR markers for comparative mapping in wheat and rice. *Mol Genet Genomics*, 2004, 271: 742–751.
- [40] Schorderet D F, Gartler S M. Analysis of CpG suppression in methylated and nonmethylated species. *Proc Natl Acad Sci USA*, 1992, 89: 957–961.
- [41] Cordeiro G M, Casu R, McIntyre C L, Manners J M, Henry R J. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant Sci*, 2001, 160: 1115–1123.
- [42] Bushman B S, Larson S R, Mott I W, Cliften P F, Wang R R, Chatterton N J, Hernandez A G, Ali S, Kim R W, Thimmapuram J. Development and annotation of perennial Triticeae ESTs and SSR markers. *Genome*, 2008, 51: 779–788.
- [43] Fernandez I, Eduardo I, Blanca J, Esteras C, Pico B, Nuez F, Arus P, Garcia J, Monforte A J. Bin mapping of genomic and EST-derived SSRs in melon (*Cucumis melo* L.). *Theor Appl Genet*, 2008, 118: 139–150.
- [44] Hwang J Y, Ahn S G, Youl Oh J, Choi Y W, Kang J S, Park Y H. Functional characterization of watermelon (*Citrullus lanatus* L.) EST-SSR by gel electrophoresis and high resolution melting analysis. *Sci Hortic*, 2011, 130: 715–724.
- [45] Pinto L R, Oliveira K M, Ulian E C, Garcia A A, de Souza A P. Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. *Genome*, 2004, 47: 795–804.
- [46] Marconi T G, Costa E A, Miranda H R, Mancini M C, Cardoso C B, Oliveira K M, Pinto L R, Mollinari M, Garcia A A, Souza A P. Functional markers for gene mapping and genetic diversity studies in sugarcane. *BMC Res Notes*, 2011, 4: 264.