

基于转录组测序的茄子 SSR 标记开发

魏明明, 陈钰辉, 刘富中, 张映, 连勇

(中国农业科学院蔬菜花卉研究所, 北京 100081)

摘要:应用 Trinity 软件对茄子转录组测序数据组装, 得到总长度为 47919660 bp 的 45404 条 Unigenes, MISA 从中检索到 8316 个 SSR 位点, 发生频率为 18.32%, 平均 5.63 kb 一个位点。SSR 位点中单碱基重复类型最多, 为 5372 个, 占到 64.60%; 其次为三碱基重复 1628 个, 占到 19.58%。三碱基重复中 AAG/CTT 是优势重复单元, 占三碱基重复数的 31.6%; 二碱基重复中 AG/CT 是优势重复单元, 占二碱基重复数的 42.3%。利用 Primer 3 设计引物, 共得到 858 对 SSR 引物, 随机选取 100 对引物对 17 份茄子材料进行扩增, 结果表明: 有 84 对可以扩增出条带清晰的片段, 有 47 对引物扩增片段为多态性片段。对 47 对多态性引物进行分析, 多态性信息含量范围为 0.10~0.64, 平均多态性信息含量为 0.32, UPGMA 聚类分析可将 17 份材料分为 3 类。以上结果表明, 基于茄子转录组测序开发的 SSR 标记可以为茄子的遗传多样性分析和遗传图谱构建提供更加丰富的标记来源。

关键词: 茄子; 转录组; SSR; 引物; 多态性

Development of SSR Markers for Eggplant with Transcriptome Sequencing Data

WEI Ming-ming, CHEN Yu-hui, LIU Fu-zhong, ZHANG Ying, LIAN Yong

(The Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081)

Abstract: The transcriptome sequences of eggplant were assembled by using Trinity program. The 45404 Unigenes were obtained with the length of 47919660 bp in total and 8316 SSR loci were detected from these Unigenes by using MISA software. The density of the SSR were 5.63 kb with the frequency of 18.32%. The type of single nucleotide SSR was the most abundant (5372) which counted for 64.60%, and tri-nucleotide (1628) counted for 19.58%. The tri-nucleotide repeat motifs of AAG/CTT were the predominant repeat types that counted for 31.6% in the tri-nucleotide repeat motifs. The dinucleotide repeat motifs of AG/CT was the predominant repeat types that counted for 42.3%. A total of 858 pairs of SSR primers were designed by using primer3 program. Among of them, the 100 pairs of SSR markers were randomly selected and verified by using 17 eggplant germplasms. The results showed that 84 pairs of primers were able to amplify PCR products, of which 47 pairs of primers produced polymorphic bands. Through analysis the 47 pairs of polymorphic primers, the result showed that the PIC values ranged between 0.10 and 0.64 with average of 0.32. The 17 eggplant germplasms could be divided into three phylogenetic groups with UPGMA method. The results indicated that the development of SSR markers based on the transcriptome sequencing of eggplant will provide more reliable markers for map structure and analysis of genetic polymorphism in eggplant.

Key words: eggplant; transcriptome; SSR; primer; polymorphism

收稿日期: 2016-01-28 修回日期: 2016-03-04 网络出版日期: 2016-10-17

URL: <http://www.cnki.net/kcms/detail/11.4996.S.20161017.1529.010.html>

基金项目: 国家科技支撑计划项目(2012BAD02B02); 国家“863”计划项目(2012AA100103); 中央级公益性科研院所基本科研业务费专项(1610092015002-01); 中国农业科学院创新工程项目; 农业部园艺作物生物学与种质创制重点实验室项目

第一作者研究方向为茄子标记开发及枯萎病抗性鉴定。E-mail: weimingming5566@126.com; 陈钰辉为并列第一作者

通信作者: 连勇, 研究方向为茄子优异基因资源挖掘。E-mail: lianyong@caas.cn

近年来,随着转录组测序技术(RNA-seq, RNA sequencing)的发展,通过高通量测定转录组 cDNA 的序列,揭示特定细胞或组织中表达的全部基因或表达序列标签(EST, expressed sequence tag),已在许多作物上得到广泛应用,并开发出了大量转录水平的分子标记。S. Roland 等^[1]通过甘薯转录组测序,利用 SSR 检索程序筛选得到 1621 个 SSR 位点;刘峰等^[2]利用测序获得的铃形椒 HDA149 和牛角椒 9704B 两个转录组数据和下载的辣椒 ESTs 数据构建 3 个 Unigenes 库,通过 SSR 检索程序筛选得到 17319 个 SSR 位点。简单重复序列(SSR, simple sequence repeat)为广泛存在于真核和原核生物基因组中的 1~6 个核苷酸串联重复单元^[3-4],具有多态性高、重复性好、共显性、易检测、操作简单、无放射、所用时间较短、覆盖面广等优点,在遗传多样性分析、遗传连锁图谱构建、基因定位及分子标记辅助育种等方面应用广泛^[5-7]。利用转录组数据开发的 SSR 标记既有 EST-SSR 标记的优点,同时其海量的数据为 SSR 标记的开发提供了比 EST-SSR 更全面的信息,能够提高遗传多样性和分子标记辅助育种的准确性,较经典的 SSR 标记开发方法减少了 cDNA 文库构建、SSR 克隆筛选、测序等繁琐程序,大大降低了工作量和开发成本。

茄子(*Solanum melongena* L.)是中国重要的果菜类蔬菜之一,具有重要的经济和营养价值。中国是世界第一茄子生产及消费的国家,年产量达 2770 万 t,占世界茄子总产量的 60% 以上^[8]。国内学者利用 SSR 标记对茄子进行了大量的研究,其中包括茄子重要性状的定位^[9-11]、遗传多样性分析^[12-13]、杂交品

种种纯度鉴定^[14]等方面。汪国平等^[15]和陈孟强等^[16]分别对番茄和马铃薯的 SSR 标记在茄子中是否具有通用性进行了分析。T. Nunome 等^[17-18]利用构建的茄子 gDNA 和 cDNA 富集文库先后共开发出 1054 个 SSR 标记,但可供使用的仅有 236 对。A. Stägel 等^[19]利用从 SGN 和 EMBL 数据库中检索获得的 3358 个茄子 EST 序列,也只开发出 55 个 SSR 标记。目前中国试验所用的 SSR 标记均来自日本 Nunome 实验室开发的标记^[20]以及其他茄科作物中的通用性标记,国内无标记开发的报道。茄子与其他茄科作物如番茄、辣椒、马铃薯等相比,在遗传多样性分析、种质资源鉴定、分子标记辅助育种方面有很大的差距。其中可供使用的分子标记过少,现有 SSR 标记多态性低是造成茄子分子研究进展缓慢的主要因素。本研究利用转录组测序的方法对茄子转录组进行深度高质量的测序,并且大规模的进行 SSR 位点搜索及 SSR 标记开发,发掘出一批扩增效率高、多态性好的引物,以期茄子的功能基因定位、遗传多样性分析、分子标记辅助育种提供标记基础,提高茄子育种效率。

1 材料与方法

1.1 植物材料及数据来源

茄子 SSR 扩增试验材料取自于 17 份来源不同的茄子(*Solanum melongena* L.)资源材料(表 1)。材料播种于中国农业科学院蔬菜花卉研究所廊坊实验基地,播种时间 2014 年春季。取对茄期植株幼嫩叶片 2 片(约 2 cm²),于 2 mL 离心管内用冷冻抽干机真空抽干 48 h,打样机粉碎后备用。

表 1 茄子 SSR 扩增试验材料

Table 1 The experimental materials of SSR amplification in eggplant

材料编号 Material number	名称 Name	来源 Source	材料编号 Material number	名称 Name	来源 Source
118	04Y-74	俄罗斯	127	51	俄罗斯
227	威十长茄	中国哈尔滨	231	紫金茄	中国
239	绿色圆茄	巴西	242	白云	中国广东
247	锦州腌渍小茄子	中国辽宁	249	绿茄王	中国哈尔滨
251	特大牛心茄	中国西安	252	白雪公主	中国西安
254	宇航绿长茄	中国西安	255	06-418	中国安徽
256	无籽黑茄	美国	260	33-22	荷兰
266	超九叶二号	中国安徽	267	华丰紫尊	中国广东
268	农夫长茄	中国广东			

SSR 标记开发的数据来源于本课题组对 10cw60 (*Solanum melongena* L.) 和 RMM620 (*Solanum melongena* L.) 测序得到的 Unigenes 数据库。其中 10cw60 为长筒果, 白色, 无叶刺, 花为浅紫色; RMM620 为圆球果, 白果有紫色斑纹, 有叶刺, 浅紫色花。材料于对茄期, 分别对顶部嫩叶、嫩茎、功能叶取样, 混合后取 2 g 包于锡箔纸内, 迅速放入液氮中速冻, $-80\text{ }^{\circ}\text{C}$ 冰箱保存备用。

1.2 DNA 提取

基因组 DNA 提取采用李荣华等^[21]改良的 CTAB 法。提取后的 DNA 在 1% 的琼脂糖凝胶上进行检测, 电压、电流和电泳时间分别为 120 V、220 mA 和 40 min。用 BioSpec-nano 检测 DNA 的浓度, 通过 $\text{OD}_{260}/\text{OD}_{230}$ 、 $\text{OD}_{260}/\text{OD}_{280}$ 比值检测 DNA 质量。

1.3 茄子转录组测序和 SSR 位点搜索

转录组测序由博奥生物公司进行 RNA 提取, 并在 HiSeqTM2500 平台上完成测序。获得的原始数据去除测序接头序列, 将 2 个样品数据合并, 用 Trinity 软件按照默认参数进行拼接。用 MISA 软件对得到的 Unigenes 进行扫描, 搜索 SSR 位点。单碱基重复大于 10 次, 二碱基重复大于 6 次, 三碱基、四碱基、五碱基及六碱基重复次数大于 5 次的承认为 SSR 位点。

1.4 SSR 引物设计

选择含有 SSR 位点的 Unigenes, 利用 Primer 3 批量设计引物。设计引物参数为 GC 含量 40% ~ 60%, 退火温度 50 ~ 60 $^{\circ}\text{C}$, 引物长度 18 ~ 25 bp, 目的片段预期 100 bp 以上。引物交由上海生工有限公司合成。

1.5 PCR 扩增和电泳检测

PCR 扩增采用 10 μL 的体系: 5 μL 2 \times Green Master Mix (Promega), 0.5 μL 上、下游引物, 2 μL DNA 模板, 2 μL ddH₂O, 并向体系中加入 15 μL 无菌液体石蜡油。PCR 反应程序: 94 $^{\circ}\text{C}$ 预变性 5 min; 94 $^{\circ}\text{C}$ 变性 30 s, 55 $^{\circ}\text{C}$ 退火 30 s, 72 $^{\circ}\text{C}$ 延伸 30 s, 35 个循环, 72 $^{\circ}\text{C}$ 延伸 8 min, 16 $^{\circ}\text{C}$ 保温。PCR 扩增完成后, 向 PCR 产物中加入 2 μL 的 6 \times loading buffer, 用 8% 的聚丙烯酰胺凝胶电泳检测, 电泳设置恒压 180 V, 时间设置为 2.5 h。

1.6 数据分析

采用人工读胶的方法, 在相同的迁移率位置上有带记为 1, 无带或不易分辨清楚的记为 0, 缺失数据记为 9, 建立原始矩阵。采用 NTSYS-pc 软件计算遗传相似系数 (GS, genetic similarity), 并利用遗传

相似系数按照非加权组平均法 (UPGMA) 构建系统发育树。采用 PowerMarker V3.25 软件计算每对引物扩增的等位基因位点数 (N_a , number of alleles), 基因多样性指数 (H , gene diversity), 期望杂合度 (H_e , expected heterozygosity), 多态性信息含量 (PIC, polymorphism information content)。

2 结果与分析

2.1 茄子转录组测序

茄子高代自交系 10cw60 与 RMM620 通过 HiSeqTM2500 平台进行转录组测序, 分别得到 6.87 G 和 5.16 G 的原始数据。将 2 个样品的数据合并, 用 Trinity 进行 de novo 拼接。在拼接后的序列中, 选最长的可变剪切序列作为该基因的 Unigenes。拼接后得到 45404 条 Unigenes, 总长度 47919660 bp。Unigenes 长度在 301 ~ 13043 bp 之间, 平均长度为 1055 bp, 其中 N50 为 1606 bp, 拼接质量较好。

2.2 茄子 SSR 位点的数量及重复类型分布特点

利用 MISA 软件对 45404 条 Unigenes 进行扫描, 在 7048 条 Unigenes 中检测到 8316 个 SSR 位点, 发生频率为 18.32%, 平均 5.63 kb 含有一个 SSR 位点。其中 5968 条 Unigenes 中含有一个 SSR 位点, 发生频率为 13.14%, 1080 条 Unigenes 中含有 2 个及 2 个以上的 SSR 位点, 发生频率为 2.38%。在所有的 SSR 位点中, 以单重复基元的 SSR 位点类型为主, 其中复合重复基元的 SSR 位点只有 418 个。

茄子转录组中 SSR 位点的类型丰富, 从单碱基到六碱基重复均有分布, 每个重复类型所含的数量相差很大。其中单碱基类型数量最多, 为 5372 个, 占总 SSR 位点的 64.60%; 其次为三碱基和二碱基类型, 分别为 1628 和 1236 个, 占 19.58% 和 14.86%; 四五六碱基最少, 共占 0.96%。

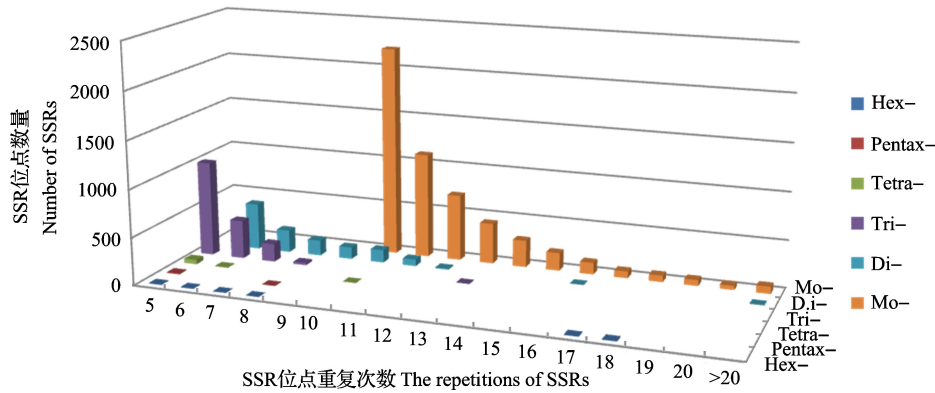
非复合单元的茄子转录组 SSR 位点的长度因重复类型不同而有所差异。最短的为 10 bp 的单碱基重复, 最长的为 108 bp 的六碱基重复。不同重复类型 SSR 位点的平均长度也不相同, 单碱基重复平均长度为 12 bp; 二、三、四、五、六碱基平均长度分别为 15 bp、17 bp、21 bp、27 bp 和 50 bp (表 2)。8316 个 SSR 位点中, SSR 位点长度分布不均匀, 长度在 10 ~ 15 bp 之间位点分布最多, 有 6665 个; 位点长度大于 25 bp 的最少, 仅有 20 个, 占 0.2%。

表 2 茄子转录组 SSR 位点类型、数量及长度
Table 2 Type, number and length of SSRs in eggplant

SSR 类型	数目	类型比例 (%)	序列长度(bp)	平均长度(bp)
Type of SSRs	Number	Percentage	Length	Average length
单碱基重复 Monucleotide repeat	5372	64.60	10 ~ 24	12
二碱基重复 Dinucleotide repeat	1236	14.86	12 ~ 48	15
三碱基重复 Trinucleotide repeat	1628	19.58	15 ~ 39	17
四碱基重复 Tetranucleotide repeat	55	0.66	20 ~ 40	21
五碱基重复 Pentanucleotide repeat	10	0.12	25 ~ 40	27
六碱基重复 Hexnucleotide repeat	15	0.18	30 ~ 108	50

茄子转录组 SSR 位点的重复次数从 5 ~ 24 次均有分布,其中以 10 次重复数量最多,有 2214 个位点,占总位点的 28.23%。不同类型的 SSR 位点中,单碱基重

复以 10 次重复为主;二碱基重复以 6 次重复为主;三、四、五、六碱基均以 5 次重复为主。各类型的 SSR 位点均随着重复次数的增加,位点数量减少(图 1)。



Mo - 代表单碱基重复;Di - 代表二碱基重复;Tri - 代表三碱基重复;Tetra - 代表四碱基重复;Pentax - 代表五碱基重复;Hex - 代表六碱基重复
Mo-means Monucleotide repeat,Di-means Dinucleotide repeat, Tri-means Trinucleotide repeat, Tetra-means Tetranucleotide repeat, Pentax-means Pentanucleotide repeat, Hex-means Hexnucleotide repeat

图 1 茄子不同类型 SSR 位点重复次数分布

Fig. 1 Distribution of various classes of SSR motifs with different numbers in eggplant

2.3 茄子 SSR 位点重复单元及频率

从 SSR 位点重复单元来看,在 8316 个 SSR 位点中共有 54 类重复单元,其中单碱基重复含 2 类,二碱基重复含 4 类,三碱基重复含 10 类,四碱基重复含 13 类,五碱基重复含 10 类,六碱基重复含 15 类。

从 SSR 位点重复单元出现频率来看,单碱基重复中,A/T 重复明显高于 C/G 重复数量。在二碱基

重复中,(AG/CT)_n 和 (AT/AT)_n 单元数量最多,分别为 523 和 428 个,占二碱基重复的 42.3% 和 34.6%;三碱基重复中,(AAG/CTT)_n 单元数量最多,有 515 个,所占比例为 31.6%,其次为(AAC/GTT)_n、(ATC/ATG)_n,分别为 293、201 个,占 18%、12.3%;四碱基重复中(AAAT/ATTT)_n 单元数量最多为 23 个,占 41.8%。二、三碱基不同重复单元比例见图 2。

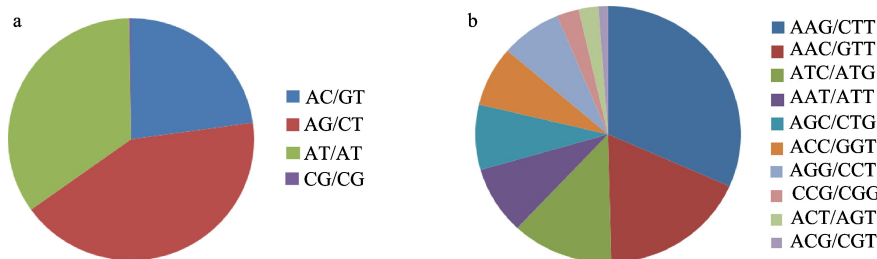


图 2 二碱基 (a) 和三碱基 (b) 中不同重复单元的分布

Fig. 2 Distribution of different repeat motifs in dinucleotide (a) and trinucleotide (b)

2.4 SSR 引物设计

去除单碱基重复和 SSR 位点长度小于 12 bp 的 SSR 位点,用 Primer 3 对筛选出的 SSR 位点的序列设计引物,最后共设计出 858 对 SSR 引物。设计后的引物与 T. Numome 等^[18]与 A. Stägel 等^[19]开发的 SSR 引物进行比对,并未发现重复引物。因此,858 对引物的重复类型比较丰富,从二碱基到六碱基重复均有分布。

2.5 SSR 引物的有效性验证和多态性分析

从 858 对引物中随机选取 100 对进行合成(表 3),

对 17 份茄子资源材料的基因组 DNA 进行 PCR 扩增。结果显示,有 16 对引物未扩增出产物,7 对引物扩增的目的片段比预期长,3 对引物扩增的目的片段比预期短,且扩增片段不特异;有 27 对引物扩增条带为单一条带,这些引物对本试验选择的茄子资源材料的遗传分析无贡献;有 47 对引物扩增出多态性条带,占设计引物的 47%,可以作为茄子遗传多样性分析的有效标记。图 3 为引物 EPSSR254、EPSSR194、EPSSR2、EPSSR45 在 17 份茄子资源材料中的多态性扩增结果。

表 3 100 对 SSR 引物信息

Table 3 The information of 100 SSR primers

引物名称 Primer name	重复基元 Repetitive unit	重复长度(bp) Repeated length	产物长度(bp) Product length	上游引物 Forward primer	下游引物 Reverse primer
EPSSR1	(AC)10	20	195	ATCTTCACCTGATTCGTCGTC	TGAAATCCAAGTACCGATTC
EPSSR2	(AG)8	16	198	CCTTACCCTCTCCATTTC	GTTCTTTGCTTGTGCTCT
EPSSR5	(AT)10	20	198	TTTTACCGACTACCCGAAA	TATGCTGGGATTCATCTCT
EPSSR7	(CT)9	18	178	GTCGATTCTAGCAAGGCTTA	CTACCACCTCCCATCGTCT
EPSSR9	(AC)8	16	198	GATTAATACCTGCGGAAGAA	ACCATGTGGATTCACAAAC
EPSSR10	(AG)9	18	164	CAAACACAAAGGAACAACCT	ATGTGGAGCATTTGAGTCAC
EPSSR11	(AC)8	16	155	GGTCAGTACCTTTTCCTCCT	CTTGAACCATCCATCGTAGT
EPSSR15	(AT)9	18	179	GAATTATTTATCGTCCCACC	CTATTGGATCGATTGCTCAT
EPSSR16	(AT)12	24	183	CACGGTGTGAAATAACATTG	CTAAACTCACGAACCTCCAC
EPSSR18	(CT)10	20	166	ATGGACATTGGAACAAGAAC	TGATCCATCAAAATCACTCC
EPSSR19	(CT)10	20	193	TTTCTCTCCAGAAGAAGCTCG	AAGGATGGTTTCAGAACAAC
EPSSR20	(CT)10	20	196	ACAGAGAAGTACGAGGCTGA	TCTATCCCGTCTTCTTTTGA
EPSSR29	(AT)8	16	174	GTGAACGCAGGATTTTAC	CAATGAAGTCCCTTTTGAAG
EPSSR31	(AC)8	16	197	GCTTCCACAAATTCCTAAAG	CCTCCCCTCTCTCTAAATA
EPSSR39	(AT)8	16	197	GGTTTTTCAAGATTATCCCC	GGAATCTGTTTGAGAGGGTT
EPSSR43	(AT)10	20	176	TCATTTTGTTCGGTCTATCC	CAACATTATCAACTCAACGG
EPSSR44	(AG)8	16	156	TGCTAATCTGAAACGAATCC	GTTTTCCACTACCTCAATCG
EPSSR45	(GT)8	16	173	CTTCTGTGAATCTTCGTG	AACTAATACTCTGACTCCGCC
EPSSR57	(AT)9	18	184	CCTAGCAAAGAATCATGAGC	AGAATGGAGCTCACAACAAC
EPSSR67	(AT)9	18	185	CAGGTCAGCTGGACTAAGAA	TTAATGGCAGAGTCCAGTT
EPSSR69	(AG)9	18	187	AAAATTCGGTTCACCACTC	GACAGCACTGTATGATGGAA
EPSSR70	(ACT)5	15	199	CAAAGGGTGAATAATGAGG	CTAGGGGTCTTGCTCTTTC
EPSSR71	(CGG)5	15	200	GAGGGGTGAAGATTTC	ATGGAAGTCTTAATGGCTA
EPSSR72	(AAG)5	15	180	GACGATTATCGAATGGAATG	GAATTGATCCTTCTGTGTC
EPSSR74	(AAG)5	15	185	ATTCTTGATGGCAGAGAC	CAAATTAACCTCCATCTTCC
EPSSR75	(AAC)5	15	151	AAGTTGGGTTGCTACTGAAA	GAGTTAAAGTGTGCTGAGA
EPSSR77	(CTT)6	18	168	GAAGATTTTCGATTCTGACCA	AATGAAGAAATCAGCCGTC
EPSSR78	(AAG)6	18	153	ATCCAAGAAAATCCCTTCAG	CCAAATGAAGAAACAGAGGA
EPSSR79	(ATG)5	15	159	GATTTGAACGAAGACGAAGA	CACAAACAGAAATGCCACTA
EPSSR80	(ACC)6	18	162	CACCATGGCAACTACTCTCT	CCTTGATAACAGGCTGAAGA
EPSSR81	(GGT)5	15	171	TGGATTATGTCGTAGGGATT	GTTATGATGCTTGGCTCTTC
EPSSR82	(AAC)5	15	192	TTTGCTACTTGCTGGATCTT	GCTGAAGGACTACTAGGGTTC

表 3(续)

引物名称 Primer name	重复基元 Repetitive unit	重复长度(bp) Repeated length	产物长度(bp) Product length	上游引物 Forward primer	下游引物 Reverse primer
EPSSR86	(GAA)5	15	168	TGATGTCCAAGCTAACAGTG	ATAATACTGGATCCGTGTCC
EPSSR87	(CCT)5	15	200	TGGTGTCTAGATGTTGTTG	TGGATGCAAATTAAGTCCTC
EPSSR88	(CTT)5	15	199	GTCTCTTCACCATGTTCCGAT	GTGAGGTAGTTAATGGCAGC
EPSSR89	(AAT)7	21	170	AGGAAAAAGGAGGACATGAT	ACAGCAAACCTTGCTTGTCTT
EPSSR91	(CGG)6	18	183	GATGAGACAAAGGTGGTTGT	TGCAAGTGATATAGCCACAG
EPSSR92	(ATG)5	15	187	CTCCCTTGTTGTTCTTTTTTC	AGGACAAGCACAGTGATAGC
EPSSR93	(AAG)5	15	177	GCAAGATCTAGAGCTGCAAT	GGGGTCCCTTCTTTGTGA
EPSSR94	(GTT)5	15	189	CCTTCCATGAGGTAGTGCTA	AGCCTAACAGCAGAGCATTAA
EPSSR95	(AAG)5	15	188	TAGCGGTTGGTAAAGAAAAAC	GGGAAATCTGAAATCCAAC
EPSSR97	(ATC)5	15	171	AAGTTGTCTCTTTTGGTGA	TGGACAGATTTGATGACAAC
EPSSR98	(GGT)5	15	200	GATTTTCTTGAAGTTGTGGG	CCACAATCTCTTCCATCA
EPSSR99	(ATT)5	15	150	ATCGAAAAGGAGTGGTCATAG	ATACTGCACTGGAAGCTCAT
EPSSR100	(ATC)5	15	187	CCTTGCTTTGACTTGTGAAG	GTGGAATCTTGCTGATGAC
EPSSR101	(AGT)7	21	199	CATGGGAGAAGAGCATAGAG	CACAGGCCCTTCTGTTCTTTA
EPSSR103	(AGG)6	18	200	CGCACTCCATTAACGACTAA	TGTAATCCATCCAGTTCCTT
EPSSR104	(AAT)6	18	200	GTTTGCAACAACATAGGTCC	AACACGATGCCAAACATAAC
EPSSR105	(CTT)6	18	198	AGGACAAATGGTTTCATCAG	ATTCTTGAAGGCATACATGG
EPSSR106	(AAC)5	15	179	TGCTAGACTTATCCCTCCA	ATTAGCTCCAGACACCGAC
EPSSR108	(AAC)5	15	197	GGGGAGAAAAGATGGAGATA	CFTTGCATTGTCTTGCTTTTC
EPSSR109	(ATC)5	15	184	GGTTGTGCATTCCTTTTTAC	GTTGGCAATGAGTTTCTCTT
EPSSR110	(AAG)5	15	146	ACCCTTCACACAAAGAGAGA	GCAATTTGGAGGAAAAGAG
EPSSR111	(AAC)5	15	193	GCCTAACACCAAAGAGGATA	GTCAACCTCCGATTTCTTT
EPSSR112	(ATC)6	18	200	GCAAATGCTAAGGAGAAGAT	GTTTCCACCAAAAAGTTGAAG
EPSSR114	(CTT)6	18	186	TAGGACAGAGGAAGAGCAAC	AGGAAGTGACACCCTTTTCT
EPSSR115	(CTT)6	18	162	AATGCTTGTTCAGAAAACCTC	GTTACATTGAGATCCTTCCG
EPSSR116	(AAG)5	15	185	ACCAACACCATCACACATTA	GAAAATTATCATCGGAGACC
EPSSR119	(GTT)5	15	189	CGTGTCTCTTAGCATTATCT	TGTCACCAAGACCCTTTCTA
EPSSR120	(CCT)5	15	167	ATCTTCACAAAACGAAACGAC	GTTCTCTGCCGTGTACTCTC
EPSSR121	(ACC)6	18	186	ACACCTTTACCACTAACCCA	TCTTGAAGTTGTTGGAGGAT
EPSSR124	(AAC)5	15	200	GTTGGATTACAGGTGTTTTT	TTGGTTTGCTCCAGAATAGT
EPSSR126	(AAG)5	15	200	ATAGGAGGATGTAGGGGA	CTATACCAGCAATCAGCACA
EPSSR127	(CTT)5	15	198	CCTCCTCTTCTTCTCCATT	GCTGCTTTAATGTGGTTTCT
EPSSR128	(GGT)5	15	198	AATCAATACCAGGAGGTGG	AGTGTCCCTTGAATTGCTTA
EPSSR129	(AGC)7	21	180	GACTGTACGAGTGCTTCTC	TGTTGCACTTGTGTATTGC
EPSSR130	(ATT)7	21	194	ATTGGAATTGTACACCAAGG	ATGCAGCTAAGCATGAAATC
EPSSR131	(AGG)6	18	192	CACATGCTATGTTCAAGAG	TCATCTGTTCAATGGACTCTC
EPSSR178	(AAC)5	15	156	CTAAGCCATTGATTAGCAGC	TCTGCTTCTCGACTATAGGG
EPSSR183	(GGT)5	15	177	CGAAAGATGGTGGTATACT	CAAATGGTAATATTCCTCG
EPSSR193	(CTG)5	15	180	CAAGAAGTTCATGGAGGAAG	CACATCTTTTTCATCTGGGTT
EPSSR217	(GAA)6	18	167	AACGAATAACAGCAGGTGAT	CCCCTGTTCTCTAAATCCA
EPSSR254	(GTT)5	15	186	GGTGGTCTAGTGGAAGACAA	CCACTGTAGTTTCTCCCAA
EPSSR259	(AAC)8	24	165	GACTCATAACCCAAGAACCA	TAAAGGACCAGCAAGAAGAA
EPSSR293	(AAT)6	18	168	TGCAAGATACCTACGACTGT	CAAGCATCAATAATAACCC
EPSSR304	(AGT)6	18	194	GGTGGTGTAAAAGATGACC	CCTTCTTTGCTTTCCATTAC
EPSSR308	(AGG)5	15	197	GGTTTTCTTCTTGGAATTGG	CGTATCGTTTAGCTCAATCC

表 3(续)

引物名称 Primer name	重复基元 Repetitive unit	重复长度(bp) Repeated length	产物长度(bp) Product length	上游引物 Forward primer	下游引物 Reverse primer
EPSSR322	(CTT)6	18	182	GACGAAATGTTCAACAGTGA	CAAGAATCGCCTTTTACACT
EPSSR330	(GGT)5	15	188	ACCATCACCATCACCATAAT	ATAGTATCCACCACCCCT
EPSSR356	(ATC)5	15	195	TCCAGTAGGAGGAGGATGTA	CCACTTCTTGATTTTGGAG
EPSSR360	(ATC)6	18	192	CTGGAAGAGAAAAGCAGAA	TGCTTGTGTTGAGGATGTA
EPSSR443	(ATC)5	15	192	GCTGTGCAATGTCATCTAAA	ATAGTTCATTATCAAGGCCG
EPSSR526	(GTT)6	18	174	AGCGAAGTTTTGGACTAGAA	CTAGTTTCCFTTACGGGCTT
EPSSR527	(AAG)6	18	161	ATGTTTGTGGGTTTCAGAC	CTTCCTCCCTTTTATAGTGT
EPSSR534	(GTT)6	18	194	GTAGAAAACATACTGGGACG	TCTTCAAACCAATCCTCAC
EPSSR560	(AAC)5	15	192	AAAAGACACTCAAAGCTGC	GGTAGCTGTAGATTTTGGG
EPSSR653	(ATC)5	15	200	TCTTCCTCCTCCTCTTCTTC	GACGATGATGCAAATGATG
EPSSR844	(ATAC)5	20	184	GGGCATTCTTCAAGAGTTTT	TACAGAACAAAGCCATCCTTT
EPSSR845	(GATG)5	20	199	CGGCAGAACTAGTTGGTATT	AACTGGCCATCATAAATGTC
EPSSR846	(TTTG)5	20	155	ATAGATGACCAGTGTTTGCC	TTGTGTCAGCAGTAGCAGAG
EPSSR848	(AAAT)5	20	190	CGGTTAAGATTAAAGGCAAC	AAAAGGGGAAAAAGTGTAG
EPSSR849	(TTGG)5	20	186	CTGCTGGA AAAAAGAAAACAAG	ATCGTTAATTTGGTCCACAG
EPSSR850	(TTTA)5	20	193	ATATGGAAGCTTTCCTGGTT	GGGAAGATAACGACTGTTGA
EPSSR852	(GGTGCT)6	36	158	ACTAGGTGGTGGATATGGTG	AACTATCACC AAAACGTGCAT
EPSSR853	(GGTGCT)6	36	179	GTAGCGTAAACGCAACTCC	GAGTAACCTGAACGCTTGAC
EPSSR854	(CCAGAA)17	102	192	GAAATCCTCAAGGACCTCTT	TTCCTTAGGACCTGCAATA
EPSSR855	(ATGGGT)5	30	162	CCAAATGGTAGGTTCTGGTA	TTCATAGGTCTGTTCATCCC
EPSSR856	(CACTAC)5	30	196	GCCAACAGTCTGAGATTCAT	TCAGATTGGAAGAAGTTGG
EPSSR857	(GGTCGA)6	36	197	GTCTGGTTATGCTCCATTTT	CATCATGGTATGGGTGATTA
EPSSR858	(CTGGAG)8	48	196	CCTCATCAGAAATGAAACCT	GGAAATGTGGAATCAAGAGA

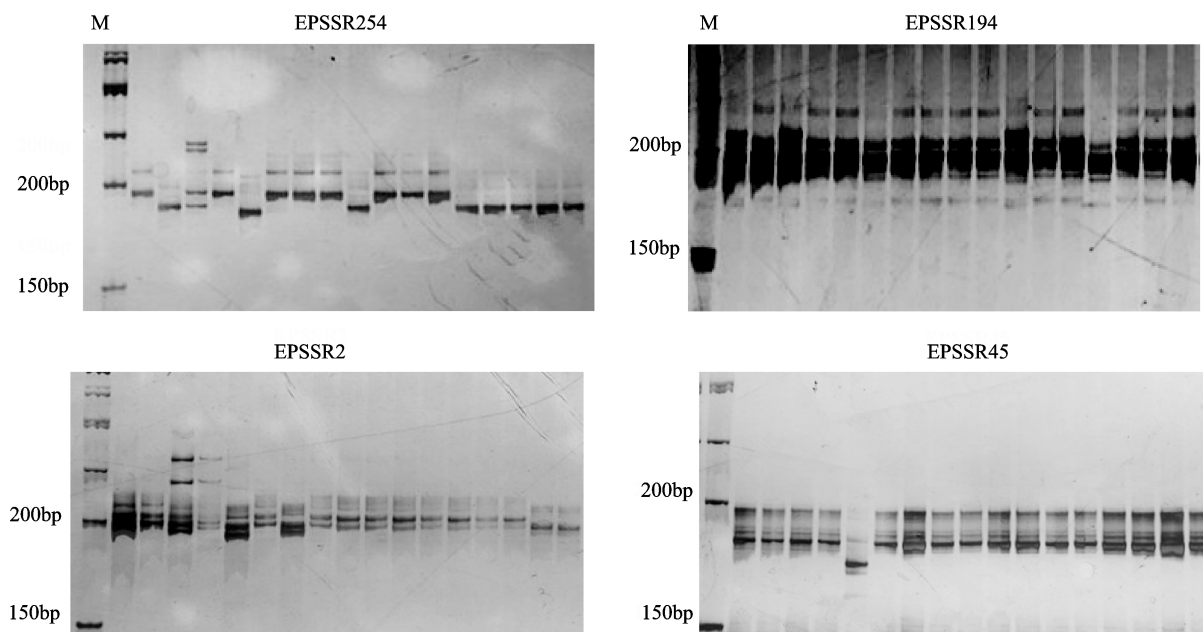


图 3 多态性引物 EPSSR254、EPSSR194、EPSSR2、EPSSR45 在 17 份材料中的扩增结果

Fig. 3 The amplification results of polymorphic primer EPSSR254, EPSSR194, EPSSR2, and EPSSR45 in 17 materials

对这 47 对多态性引物进行多态性分析(表 4),共检测到 139 个等位基因,平均每对引物检测到 2.95 个等位基因,多态性信息含量在 0.10~0.64 之间,平均多态性信息含量为 0.32。其中, PIC 含量大于 0.5 的高度多态性引物有 4 对, PIC 含量在 $0.25 < PIC < 0.5$ 之间的中度多态性引物有 28 对, PIC 含量小于 0.25 的引物为 15 对,引物多态性较高。基因多样性指数在 0.10~0.69 之间,其中引物 SSR254 反应材料的基因多样性为 0.69,且引物检测材料的杂合度在 0~0.94 之间,说明引物可以用于材料的遗传多样性分析。

对 17 份茄子资源材料进行遗传多样性分析。结果表明:在遗传距离为 0.70 时,可将 17 份材料按地域聚成 3 类(图 4)。第 1 类中 118、127、256、260 分别来源于俄罗斯、俄罗斯、美国、荷兰;第 2 类全部来源于中国;第 3 类 239 是茄子近缘野生种 *Solanum aethiopicum* gr. Gilo;说明获得的这些 SSR 引物可以用于茄子遗传多样性分析。

3 讨论

对茄子转录组测序拼接得到的 45404 条 Uni-genes 进行 SSR 搜索,共检测到 8316 个 SSR 位点,发生频率为 18.32%,在茄子转录组序列中平均 5.63 kb 包含一个 SSR 位点,同为茄科植物的辣椒发生频率为 7.83%^[2],茄子的发生频率要高于辣椒。同样茄子 SSR 位点的发生频率也高于禾本科植物的大麦(3.4%)、小麦(3.2%)、高粱(3.6%)、水稻(4.7%)^[22]。可以看出茄子的 SSR 位点出现频率比较丰富。但是出现频率要低于油棕(22.6%)、野三七(16.86%)^[23-24]。不同物种 SSR 位点的出现频率不同,这可能在于不同植物的基因组中,SSR 积累速度及进化速度会有所不同^[25],也可能与测序深度、拼接质量以及 SSR 位点检索的参数设置有关。

茄子 SSR 位点重复类型从单碱基到六碱基均有分布,其中单碱基重复数目最多,为 5372 个,占 64.60%,其次为三碱基重复为 1628 个,占 19.58%,二碱基重复数为 1236 个,占 14.86%。检索结果与辣椒、油棕的检索结果相同^[2,23]。相关报道表明在大多数物种的 SSR 位点中,一般三碱基频率最高,少数几种双子叶植物中二碱基重复为主要类型^[26-27]。其中大豆、番茄、玉米、小麦以三碱基重复为主^[28-29],而腊梅、茶树、野三七以二碱基重复为主^[25,30-31]。这种现象的出现,除了与物种本身的基因型有关外,与 SSR 检索参数的设置有一定关系,人为地改变 SSR 搜索标准,即可出现不同碱基重复

表 4 47 对引物多态性信息含量分析

Table 4 The analysis of polymorphic information in 47 SSR primers

引物名称 Primer name	等位基因数目 N_a	基因多样性 H	期望杂合度 H_e	多态性信息含量 PIC
EPSSR2	2.00	0.35	0.00	0.29
EPSSR5	3.00	0.55	0.00	0.45
EPSSR29	3.00	0.51	0.00	0.43
EPSSR31	4.00	0.51	0.00	0.47
EPSSR39	4.00	0.65	0.94	0.59
EPSSR43	3.00	0.29	0.00	0.27
EPSSR44	3.00	0.36	0.00	0.33
EPSSR45	3.00	0.20	0.00	0.19
EPSSR57	3.00	0.20	0.00	0.19
EPSSR67	4.00	0.50	0.00	0.45
EPSSR69	3.00	0.29	0.00	0.27
EPSSR75	3.00	0.43	0.00	0.37
EPSSR78	3.00	0.20	0.00	0.19
EPSSR82	3.00	0.29	0.00	0.27
EPSSR88	2.00	0.10	0.00	0.10
EPSSR89	3.00	0.55	0.17	0.45
EPSSR97	3.00	0.20	0.00	0.19
EPSSR99	2.00	0.10	0.00	0.10
EPSSR101	2.00	0.35	0.00	0.29
EPSSR110	3.00	0.20	0.00	0.19
EPSSR121	3.00	0.20	0.00	0.19
EPSSR127	2.00	0.10	0.00	0.10
EPSSR129	3.00	0.51	0.00	0.43
EPSSR130	4.00	0.56	0.33	0.51
EPSSR131	3.00	0.36	0.00	0.33
EPSSR178	4.00	0.67	0.00	0.60
EPSSR183	3.00	0.43	0.00	0.37
EPSSR193	3.00	0.36	0.00	0.33
EPSSR217	3.00	0.36	0.22	0.33
EPSSR254	2.00	0.69	0.06	0.64
EPSSR259	2.00	0.20	0.00	0.18
EPSSR293	2.00	0.28	0.00	0.24
EPSSR304	2.00	0.40	0.22	0.32
EPSSR308	2.00	0.28	0.00	0.24
EPSSR322	2.00	0.28	0.00	0.24
EPSSR330	4.00	0.30	0.11	0.28
EPSSR356	2.00	0.40	0.00	0.32
EPSSR360	3.00	0.51	0.00	0.43
EPSSR443	3.00	0.43	0.00	0.37
EPSSR526	3.00	0.36	0.00	0.33
EPSSR527	2.00	0.20	0.00	0.18
EPSSR534	3.00	0.20	0.00	0.19
EPSSR560	4.00	0.30	0.00	0.28
EPSSR653	5.00	0.42	0.06	0.40
EPSSR849	3.00	0.29	0.00	0.27
EPSSR855	2.00	0.46	0.06	0.35
EPSSR858	3.00	0.48	0.00	0.40

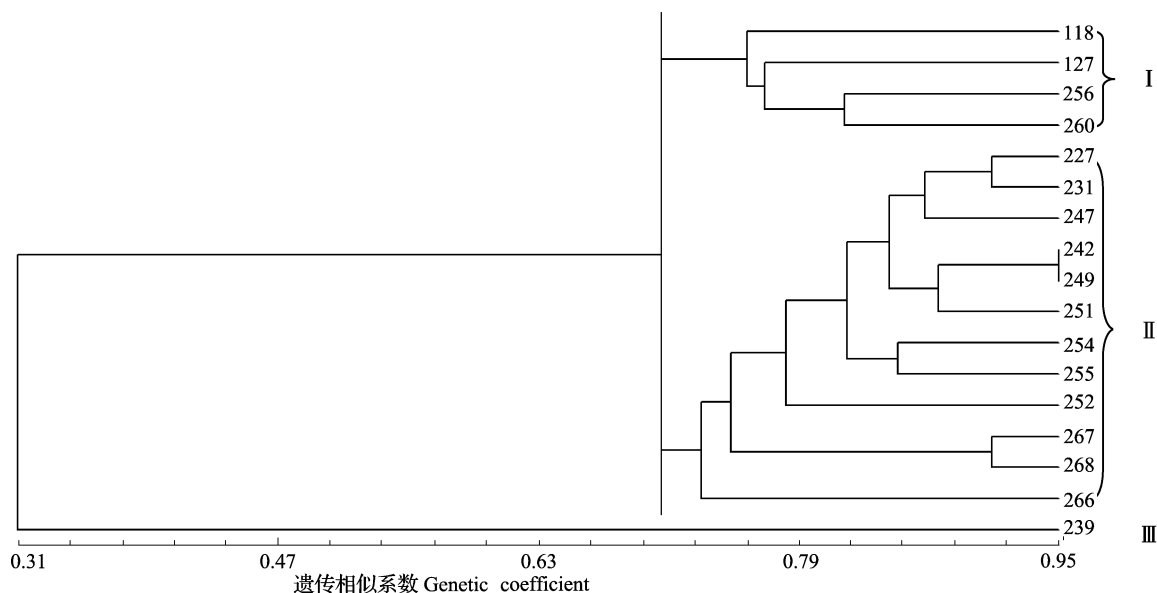


图4 供试茄子材料的 UMPGA 聚类结果

Fig. 4 Dendrogram of tested eggplant materials by UMPGA

占优势的情况^[32-33]。除单碱基外,三碱基中(AAG/CTT)_n数量最多,二碱基中,(AG/CT)_n重复数量最多,这与辣椒、杜仲报道相同^[2,34]。(CG/CG)_n重复数最少,这可能是由于双子叶和单子叶植物中CG位点重复量较低^[35],富含CG位点的区域较稳定,不易出现复制滑动,从而较少出现CG重复序列^[36]。相关报道表明大多数双子叶植物中常见的重复基序类型有AG、AT、AAG、AGG和AGC等^[37],这可能与编码蛋白质的密码子的偏好性有关。

SSR长度是影响SSR多态性的重要因素^[38],其中SSR长度<12bp多态性极低,SSR长度在12~20bp之间多态性中等,SSR长度>20bp多态性较高,且单核苷酸重复容易发生错配而导致测序失败^[24],因此本试验中去除12bp以下长度和单碱基重复序列进行引物设计,从8316个SSR位点中共设计了858对引物。

从858对引物中随机选取100对进行扩增,16对引物未扩增出产物,可扩增率为84%。其中7对引物的扩增片段比预期扩增片段大,可能与扩增片段插入内含子有关,3对引物的扩增片段比预期短,可能为非特异性目的片段。剩余74对引物中,47对引物具有多态性,占设计引物的47%。转录组测序开发的SSR标记是表达基因序列,受选择压力大,序列保守性强,导致转录组测序开发的SSR比基因组开发的SSR标记多态性低^[39]。但是茄子SSR位点较多,扩增率较高,多态性相对较高。47对引物将17份材料按照地域分布聚为3种类型,说

明这些引物可以用于茄子遗传多样性分析。本试验开发的SSR标记与K. Miyatake等^[20]相比标记数量较多,更适用于中国国内实验室的分子标记分析。利用转录组数据开发SSR标记与利用基因组测序开发的SSR标记^[40]相比,虽然基因组数据更全面,可开发出更多的SSR标记,但转录组测序数据是表达基因和外显子的序列,具有丰富的功能信息,一旦基因功能得到证实,可直接用于辅助选择育种。且来源于保守的表达序列,在中间具有很高的穿梭性,适合在种间进行遗传背景的比较。转录组数据是开发标记的理想资源,基于转录组开发的SSR标记除了具有一般分子标记的特点外,还具有信息量大、通用性好的特殊优势,并且因为来源于表达的基因组区域,更能直接反应表达基因的多态性^[41]。因此本试验通过转录组数据进行标记开发,可以发掘到比传统方法更多的SSR位点,得到更多的多态性SSR标记。

本研究利用茄子转录组数据开发的SSR标记具有实用性,目前茄子起步较晚,茄子可利用的标记较少。本研究结果为进一步开发茄子SSR标记提供了基础,这些标记为茄子遗传资源评价、绘制遗传图谱、品种改良和种质资源遗传多样性分析奠定了分子基础。

参考文献

- [1] Roland S, Luz R T, Omar P, et al. A sweet potato gene index established by de novo assembly of pyrosequencing and Sanger sequences and mining for gene-based microsatellite markers[J].

- BMC Genomics, 2010, 11(2):604-614
- [2] 刘峰,王运生,田雪亮,等. 辣椒转录组 SSR 挖掘及其多态性分析[J]. 园艺学报, 2012, 39(1):168-174
- [3] Morgante M, Olivieri A M. PCR-amplified microsatellites as markers in plant genetics[J]. Plant J, 1993, 13(1):175-182
- [4] Kalia R K, Rai M K, Kalia S, et al. Microsatellite markers: an overview of the recent progress in plants[J]. Euphytica, 2011, 177(3):309-334
- [5] Milea A, Neeta S, Harish P. Advances in molecular marker techniques and their applications in plant sciences[J]. Plant Cell Rep, 2008, 27(4):617-631
- [6] 罗冉,吴委林,张盼,等. SSR 分子标记在作物遗传育种中的应用[J]. 基因组学与应用生物学, 2010, 29(1):137-143
- [7] 程小毛,黄晓霞. SSR 标记开发及其在植物中的应用[J]. 中国农学通报, 2011, 27(5):304-307
- [8] FAO. 2011. Food and Agriculture Organization of the United Nations. <http://www.fao.org/>
- [9] 相丛超,张巍巍,罗双霞,等. 茄子果萼下颜色的遗传分析和 SSR 标记的相关性分析[J]. 河北农业大学学报, 2015, 38(4):50-55
- [10] 乔军. 茄子果实性状遗传研究及果型 QTL 定位[D]. 北京:中国农业科学院, 2011
- [11] 孙源文. 茄子枯萎病抗性遗传初步研究. [D]. 北京:中国农业科学院, 2012
- [12] 管志坤. 茄子种质资源遗传多样性 SSR 和 SRAP 分析[D]. 保定:河北农业大学, 2012
- [13] 冯英娜,柳李旺,刘卫东,等. 茄子 SSR 遗传多样性及其农艺性状的关联分析[J]. 江苏农业学报, 2014, 30(4):839-847
- [14] 王利英,乔军,石瑶,等. 茄子 SSR 多态性引物的筛选及品种纯度鉴定[J]. 华北农学报, 2012, 27(4):98-101
- [15] 汪国平,牛玉,汪文毅,等. 番茄 SSR 标记在茄子及其他茄科作物上的通用性分析[J]. 华南农业大学学报, 2014, 35(4):56-60
- [16] 陈孟强,张娅,庄泽彬,等. 马铃薯 SSR 标记在茄子上的通用性分析[J]. 中国园艺文摘, 2013(10):28-29
- [17] Nunome T, Suwabe K, Iketani H, et al. Identification and characterization of microsatellites in eggplant[J]. Plant Breeding, 2003, 122(3):256-262
- [18] Nunome T, Negoro S, Kono I, et al. Development of SSR markers derived from SSR enriched genomic library of eggplant (*Solanum melongena* L.) [J]. Theor Appl Genet, 2009, 119(6):1143-1153
- [19] Stügel A, Portis E, Toppino L, et al. Gene-based microsatellite development for mapping and phylogeny studies in eggplant[J]. BMC Genomics, 2008, 9(1):357
- [20] Miyatake K, Saito T, Negoro S, et al. Development of selective markers linked to a major QTL for parthenocarpy in eggplant (*Solanum melongena* L.) [J]. Theor Appl Genet, 2012, 124(8):1403-1413
- [21] 李荣华,夏岩石,刘顺枝,等. 改进的 CTAB 提取植物 DNA 方法[J]. 实验室研究与探索, 2009, 28(9):14-16
- [22] 李小白,向林,罗洁,等. 转录组测序 (RNA-seq) 策略及其数据在分子标记开发上应用[J]. 中国细胞生物学学报, 2013, 35(5):720-726
- [23] Kantety R V, La R M, Matthews D E, et al. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat [J]. Plant Mol Biol, 2002, 48(5):501-510
- [24] 周丽霞,肖勇,杨耀东. 油棕转录组 SSR 标记开发研究[J]. 广东农业科学, 2014, 41(14):136-143
- [25] 李翠婷,张广辉,马春花,等. 野三七转录组中 SSR 位点信息分析及其多态性研究[J]. 中草药, 2014, 45(10):1468-1472
- [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[J]. 基因组学与应用生物学, 2011, 30(5):513-520
- [27] Liang X, Chen X, Hong Y, et al. Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species [J]. BMC Plant Biol, 2009, 9(1):35
- [28] Kumpatla S P, Mukhopadhyay S. Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species [J]. Genome, 2005, 48(6):985-998
- [29] Cardle L, Ramsay L, Milbourne D, et al. Computational and experimental characterization of physically clustered simple sequence repeats in plants [J]. Genetics, 2000, 156:847-854
- [30] Varshney R K, Graner A, Sorrells M E. Genic microsatellite markers in plants: features and applications [J]. Trends Biotechnol, 2005, 23(1):48-55
- [31] 李响,杨楠,赵凯歌,等. 蜡梅转录组 EST-SSR 标记开发与引物筛选[J]. 北京林业大学学报, 2013, 35(1):25-32
- [32] 杨华,陈琪,韦朝领,等. 茶树转录组中 SSR 位点的信息分析[J]. 安徽农业大学学报, 2011, 28(6):882-886
- [33] Wei W, Qi X, Wang L, et al. Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers [J]. BMC Genomics, 2011, 12:451
- [34] Aggarwal R K, Hendre P S, Varshney R K, et al. Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species [J]. Theor Appl Genet, 2007, 114:359-372
- [35] 黄海燕,杜仲. EST-SSR 引物开发及遗传多样性研究[D]. 北京:中国林业科学研究院, 2013
- [36] Li S, Yin T, Wang M, et al. Characterization of microsatellites in the coding regions of the *Populus* genome [J]. Mol Breeding, 2011, 27(1):59-66
- [37] Zhang Z, Deng Y, Tan J, et al. A genome-wide microsatellite polymorphism database for the indica and japonica rice [J]. DNA Res, 2007, 14(1):37-45
- [38] Chagné D, Chaumeil P, Ramboer A, et al. Cross-species transferability and mapping of genomic and cDNA SSRs in pines [J]. Theor Appl Genet, 2004, 109(6):1204-1214
- [39] 李珊,周天华,赵桂仿,等. 蹄香表达序列标签资源的 SSR 信息分析[J]. 中草药, 2010, 41(3):464-468
- [40] 戚维聪,程计华,黄邦全,等. 基于海甘蓝 RNA-Seq 序列开发 EST-SSR 分子标记 [J]. 江苏农业学报, 2014, 30(5):997-1002
- [41] Hirakawa H, Shirasawa K, Miyatake K, et al. Draft genome sequence of eggplant (*Solanum melongena* L.): the representative solanum species indigenous to the old world [J]. DNA Res, 2014, 21(6):649-660