网络出版时间:2025-09-10 DOI:10.13207/j.jnwafu.2026.03.004

余甘子转录组与 SSR 信息分析及分子标记开发

王建超,张小艳,谢丽雪,张立杰,李韬
(福建省农业科学院 果树研究所,福建 福州 350013)

[摘 要] 【目的】明确余甘子转录组 SSR 位点的分布特征,开发余甘子 SSR 分子标记技术,为余甘子的亲缘关系鉴定、遗传育种和创新利用提供理论依据。【方法】以 6 份表型差异较大的余甘子叶片为材料,通过高通量测序技术获得余甘子叶非参转录组数据,分析其转录组和 SSR 位点分布特征,设计 SSR 引物并进行有效性和多态性筛选,最终开发余甘子 SSR 分子标记技术。【结果】余甘子叶转录组测序共获得 65 074 条基因,平均长度为 1 064 bp,71.77% 的基因成功获得功能注释;所有基因中共获得 18 245 个 SSR 位点,发生频率为 21.68%,含有 192 种重复基元,其中 A/T 重复基元数目最多;多态性 SSR 位点共 2 714 个,占比为 14.88%,表明 SSR 位点可用于遗传多样性分析的潜力较大。对 18 245 个含有 SSR 位点的基因进行 SSR 引物设计,共设计出 12 488 对 SSR 引物,成功率为 73.88%,其中 7 602 对引物的重复基元为单核苷酸,占比为 60.87%。随机挑选 154 对引物进行 SSR 位点有效性和多态性分析,筛选出 10 对扩增条带清晰且多态性较好的 SSR 引物,这 10 对 SSR 引物重复基元分别为二核苷酸 7 个,三核苷酸 1 个和五核苷酸 2 个,扩增产物大小为 127~270 bp,平均为 220.5 bp。【结论】多态性 SSR 位点用于余甘子遗传多样性分析的潜力较大,筛选出 10 对扩增条带清晰且多态性较好的 SSR 引物。

[关键词] 余甘子;转录组;SSR;引物筛选;分子标记
[中图分类号] S667.9 [文献标志码] A [文章编号] 1671-9387(2026)03-0035-11

Analysis of *Phyllanthus emblica* transcriptome and SSR information and molecular marker development

WANG Jianchao, ZHANG Xiaoyan, XIE Lixue, ZHANG Lijie, LI Tao
(*Fruit Research Institute, Fujian Academy of Agricultural Sciences, Fuzhou, Fujian* 350013, *China*)

Abstract: 【Objective】 This research aims to clarify the distribution characteristics of SSR loci in the transcriptome of *Phyllanthus emblica* and develop the SSR molecular marker technology of *P. emblica*, to provide a theoretical basis for the identification of genetic relationships, genetic breeding, and the innovative utilization of *P. emblica*. 【Method】 Six leaf samples of *P. emblica* exhibiting significant phenotypic differences were utilized as materials. Non-parametric transcriptome data for these leaves were obtained through high-throughput sequencing technology, followed by an analysis of the transcriptome and SSR locus characteristics. SSR primers were then designed and evaluated for their effectiveness and polymorphism. 【Result】 The results indicated that a total of 65 074 genes were identified through transcriptome sequencing of *P. emblica* leaves, with an average length of 1 064 bp. Among them, 71.77% of the genes were successfully annotated. Additionally, a total of 18 245 SSR loci were identified across all genes, showing a frequency of

[收稿日期] 2024-12-20 [修回日期] 2025-03-31
[基金项目] 国家热带植物种质资源库-余甘子种质资源分库项目(NTPGRC2024-011);福建省属公益类科研院所专项(2022R1028004,2024R1027002);福建省人民政府与中国农业科学院农业高质量发展超越"5511"协同创新工程项目(XTCXGC2021019-GSS01)
[作者简介] 王建超(1988—),男,河南南阳人,助理研究员,博士,主要从事特色果树种质资源保护与生理生化研究。E-mail:447327289@qq.com
[通信作者] 李韬(1969—),男,福建福州人,副研究员,主要从事果树栽培与分子生物学研究。E-mail:Leetao06@163.com

21.68%, which included 192 repeat motifs, with A/T repeat motifs being the most abundant. Among all loci, 2 714 polymorphic SSRs were detected, accounting for 14.88%, suggesting that SSR loci had great potential for genetic diversity analysis. A total of 12 488 pairs of SSR primers were su-ccessfully designed for the 18 245 genes containing SSR loci, with a success rate of 73.88%. Among all the primers, 7 602 pairs of primers had the repeat motif of mononucleotide, accounting for 60.87%. Subsequently, 154 pairs of primers were randomly selected for validation and polymorphism analysis. Ultimately, 10 pairs of SSR primers exhibiting clear amplified bands and good polymorphism were identified. The repeat motifs of these 10 pairs consisted of 7 dinucleotides, 1 trinucleotide, and 2 pentanucleotides. The size of the amplified products ranged from 127 to 270 bp, with an average of 220.5 bp.【Conclusion】Polymorphic SSR loci hold significant potential for genetic diversity. 10 pairs of SSR primers, which exhibited clear amplified bands and substantial polymorphism were selected.

Key words: *Phyllanthus emblica*; transcriptome; SSR; primers screening; molecular marker

余甘子(*Phyllanthus emblica*)属大戟科(Euphorbiaceae)叶下珠属(*Phyllanthus*)多年生灌木或乔木[1],分布于东经 70°～122°,北纬 1°～29°的热带或亚热带地区,包括中国、印度、泰国、尼泊尔、缅甸、斯里兰卡、马来西亚和印度尼西亚等国。我国余甘子分布于福建、云南、广西、广东、海南、台湾、四川和贵州以及香港特别行政区的亚热带和热带地区,地域分布广大,种质资源储量和遗传多样性丰富,但90%以上仍处于野生或半野生状态,亟需对各地的余甘子地方品种和野生资源进行收集保护。国家热带植物种质资源库-余甘子种质资源分库是由科技部、财政部资助建设的国家级余甘子种质资源保护平台,以余甘子的收集保护、评价和创新利用为主要任务,但余甘子各分布区命名繁杂,同名异种或同种异名的现象普遍存在,严重影响种质资源收集、保存和育种工作的正常进行。开展遗传多样性研究并及时摸清种质间的遗传关系,对其资源保护、遗传育种和创新利用具有重要意义。

遗传多样性是生物保护学研究的核心,也是植物遗传研究和性状改良的关键资源。随着现代分子生物技术的发展,分子标记技术已广泛应用于植物遗传多样性和亲缘关系鉴定[2]。目前,国内外学者先后应用余甘子种质资源开展了随机扩增多态性DNA标记(random amplified polymorphic DNA, RAPD)[3-8]、相关序列扩增多态性(sequence-related amplified polymorphism, SRAP)[9]和简单重复序列间扩增(inter-simple sequence repeat, ISSR)[10-11]等DNA分子标记技术,但简单重复序列(simple sequence repeat, SSR)在余甘子种质资源的鉴定方面鲜有报道,有较广阔的研究空间。SSR广泛存在于基因组序列中,具有重复性好、数量丰富、共显性和操作简单等特点。本研究筛选叶片表型变异较大的6份余甘子叶片为材料,开展非参转录组测序、转录组分析和微卫星特征分析,明确余甘子转录组SSR位点的分布特征;继而设计SSR引物并进行有效性和多态性筛选,开发余甘子SSR分子标记技术,为余甘子的亲缘关系鉴定、遗传育种和创新利用提供理论依据。

1 材料与方法

1.1 试验材料

试验材料来源于福建省农业科学院果树研究所国家热带植物种质资源库-余甘子种质资源分库(119°19′59" E, 26°7′48" N, 海拔 20 m)。转录组测序材料为6份叶片表型变异较大的余甘子叶,叶片表型如图1所示,其来源分别为中国福建(1)、云南(2)、广西(3)、广东(4)及泰国(5)和缅甸(6)。SSR引物筛选所用DNA模板为随机挑选的8份余甘子材料。所有材料均为余甘子嫩叶,采摘后立即置于液氮冷冻,并于−80 ℃保存备用。



图1 用于转录组测序的6份余甘子叶

Fig. 1 Six *Phyllanthus emblica* leaves for transcriptome sequencing

1.2 试验方法

1.2.1 余甘子转录组测序 RNA的提取与检测。

余甘子叶样品 RNA 提取参考 KUMAR[12] 的方法；采用 Nano Photometer Spectrophotometer 检测提取的 RNA 纯度（OD_{260}/OD_{280} 与 OD_{260}/OD_{230} 比值），采用 Agilent 2100 bio analyzer 精确检测 RNA 的完整性。

文库构建与质检。将检测合格的 RNA 以 New England Biolabs（NEB）普通和链特异性建库，质检使用 Qubit 2.0 Fluorometer 进行定量，稀释文库至 1.5 ng/μL，再使用 Agilent 2100 bio analyzer 对文库的 insert size 进行检测，insert size 符合预期，再利用逆转录定量聚合酶链反应（reverse transcription-polymerase chain reaction，RT-PCR）对文库有效浓度进行准确定量（文库有效浓度高于 2 nmol/L）。质检合格后的文库按照有效浓度及目标下机数据量 pooling 后再进行 Illumina 测序（苏州帕诺米克生物医药科技有限公司完成）。

1.2.2　余甘子转录组 SSR 位点鉴别及引物设计　使用 MISA 软件对拼接的 Unigene 序列进行 SSR 检测筛选，以单核苷酸≥10 次、二核苷酸≥6 次、三核苷酸≥5 次、四核苷酸≥5 次、五核苷酸≥5 次、六核苷酸≥5 次的最少重复次数为搜索标准。采用 Primer 3（v 2.3.5），以引物长度 18～27 bp、退火温度 57～63 ℃、上下游引物温度差不大于 2 ℃、预期扩增产物大小 100～280 bp 为主要参数，设计 SSR 位点序列引物。

1.2.3　基因功能注释　得到拼接转录本后，取相似度＞30%，且期望值 $E<10^{-5}$ 的转录本进行非冗余蛋白库（non-redundant protein sequence database，NR）、核酸序列数据库（nucleotide sequence database，NT）、蛋白质家族分析与建模（protein family analysis and modeling，PFAM）、同源蛋白簇数据库（clusters of orthologous groups from complete genomes，KOG/COG）、蛋白质序列数据库（Swiss-Prot protein sequence database，Swiss-prot）、京都基因和基因组百科全书同源性分析数据库（Kyoto encyclopedia of genes and genomes orthology，KO）、基因本体数据库（gene ontology，GO）和京都基因和基因组百科全书（Kyoto encyclopedia of genes and genomes，KEGG）的基因功能比对，获得 Unigene 的功能信息。

1.2.4　余甘子 SSR 多态引物筛选　采用李金璐等[13] 改良的 CTAB 法提取余甘子基因组 DNA，选择质量浓度为 100～200 ng/μL、OD_{260}/OD_{280} 为 1.83～2.14、1%琼脂糖凝胶电泳条带完整性较好的 DNA 样品，用微量紫外分光光度计检测 DNA 纯度，将 DNA 母液质量浓度稀释为 20 ng/μL，－20 ℃保存备用。

从 20 bp 以上的 SSR 引物中随机挑选 154 对（包括六核苷酸、五核苷酸、四核苷酸、三核苷酸和二核苷酸等 5 种重复基元）进行引物有效性验证和多态性筛选。以随机选取的 8 份余甘子材料混合 DNA 为模版，分别利用 154 对引物进行 PCR 扩增。PCR 反应体系（20 μL）：上、下游引物各 0.4 μL，DNA 模板（20 ng/μL）1 μL，10×Buffer 2 μL，dNTPs 1.6 μL，Easy *Taq* DNA Polymerase 0.2 μL，用 ddH_2O 补足至 20 μL。PCR 扩增程序：95 ℃预变性 5 min；94 ℃变性 45 s，55 ℃退火 1 min，72 ℃延伸 1 min，30 个循环；最后 72 ℃延伸 1 min。扩增产物用 1%琼脂糖凝胶电泳进行初检，初步筛选出条带清晰、能扩增出预期大小特异条带的引物。符合初筛条件的引物分别以随机选取的 8 份余甘子 DNA 为模板，进行 PCR 扩增（方法同上），扩增产物采用 1%琼脂糖凝胶电泳，筛选出扩增条带清晰、多态性较高且稳定性好的引物为多态性引物[14]。

1.3　数据分析

使用 Gene Marker 软件包确认 SSR 等位基因，用 Cor 通过 POPGENE 32 软件计算遗传参数来评价引物质量。采用 Microsoft Excel 2016 软件统计余甘子 SSR 位点数、出现频率、分布平均距离、重复单元类型及占比、重复单元碱基组成、基元重复次数以及序列长度等。

2　结果与分析

2.1　余甘子转录组测序结果

2.1.1　RNA 测序及序列拼接组装　余甘子转录组测序结果经过原始数据过滤（去除带接头序列（reads），去除无法确定的碱基信息，去除低质量序列）、测序错误率检查和 GC 含量分布检测，获得有效序列（clean reads）。由表 1 可知，6 份余甘子样品共获得 23 452 178～23 992 653 条原始序列（raw reads），过滤后获得 21 992 228～23 084 071 条有效序列，GC 含量为 42.88%～43.37%，Q20 为 97.70%～98.08%，Q30 为 93.33%～94.22%，表明测序质量较好。

经 Trinity（v 2.5.1）拼接，从头组装获得转录本 185 875 条，核苷酸总数为 218 579 058 bp，长度为 301～16 964 bp，平均长度为 1 176 bp，N50 长度为 1 601 bp（表 2）。其中，长度为 301～500 bp 的核苷

酸有 42 749 条,占比为 23.00%;501～1 000 bp 的有 59 546 条,占比为 32.04%;1 000～2 000 bp 的有 55 233 条,占比为 29.72%;2 000 bp 以上的有 28 339 条,占比为 15.25%。序列经拼接和去冗长后,组装获得 65 074 条基因(Unigene),核苷酸总数为 69 213 477 bp,长度为 301～16 964 bp,平均长度为 1 064 bp,N50 长度为 1 525 bp(表 2)。其中,长度为 300～500 bp 的核苷酸有 20 440 条,占比为 31.41%;长度为 501～1 000 bp 的核苷酸有 20 992 条,占 35.25%;1 000～2 000 bp 的有 14 923 条,占 22.93%;2 000 bp 以上的有 8 719 条,占 13.40%。

表 1 余甘子转录组测序结果

Table 1 Transcriptome sequencing results of *Phyllanthus emblica*

样品 Sample	原始序列/条 Raw reads	有效序列/条 Clean reads	有效数据/Gb Clean bases	错误率/% Error rate	Q20/%	Q30/%	GC 含量/% GC content
1	23 992 653	23 084 071	6.93	0.02	98.08	94.22	43.19
2	23 936 822	23 030 251	6.91	0.02	98.06	94.20	43.37
3	23 645 540	22 593 632	6.78	0.03	97.81	93.55	43.05
4	23 870 065	22 987 681	6.90	0.03	98.01	94.07	43.06
5	23 452 178	22 613 389	6.78	0.03	97.70	93.33	43.02
6	22 813 421	21 992 228	6.60	0.03	97.92	93.87	42.88

表 2 组装获得转录本和基因的数量与长度分布

Table 2 Quantity and length distribution of unigenes and transcript obtained through assembly

项目 Item	最小长度/bp Min length	平均长度/bp Mean length	中间长度/bp Median length	最大长度/bp Max length	N50/bp	N90/bp	核苷酸总数/bp Total nucleotides	总数 Total
转录本 Transcripts	301	1 176	895	16 964	1 601	559	218 579 058	185 875
基因 Unigenes	301	1 064	720	16 964	1 525	472	69 213 477	65 074

2.1.2 基因的功能注释 由表 3 可知,成功注释到 NR、NT、KO、Swiss-prot、PFAM、GO、KOG 数据库的基因序列分别有 40 741,24 233,17 265,33 323,28 849,28 844 和 11 083 条,分别占总量的 62.60%,37.23%,26.53%,51.20%,44.33%,44.32%和 17.03%。其中成功注释到 NR 数据库的基因序列最多,其次是 Swiss-prot 数据库;至少成功注释到 1 个数据库的基因序列有 46 707 条,占比为 71.77%;未能成功注释的基因序列有 18 367 条,占比为 29.23%;成功注释到 7 个数据库的基因序列有 4 086 条,占比为 6.27%。由图 2 可知,仅能注释到 NT、NR、KOG 和 PFAM 数据库的基因序列分别有 1 136,7 455,202 和 3 条。

表 3 基因注释结果统计

Table 3 Statistics of unigene annotation results

项目 Item	基因序列/条 Number of unigenes	占比/% Percentage
注释到 NR 数据库 Annotated in NR	40 741	62.60
注释到 NT 数据库 Annotated in NT	24 233	37.23
注释到 KO 数据库 Annotated in KO	17 265	26.53
注释到 Swiss-Prot 数据库 Annotated in Swiss-Prot	33 323	51.20
注释到 PFAM 数据库 Annotated in PFAM	28 849	44.33
注释到 GO 数据库 Annotated in GO	28 844	44.32
注释到 KOG 数据库 Annotated in KOG	11 083	17.03
注释到上述 7 个数据库 Annotate to the above 7 databases	4 086	6.27
至少注释到 1 个数据库 Annotated in at least one database	46 707	71.77
合计 Total	65 074	100.00

2.1.3 基因 NR 功能注释 通过与 NR 数据库比对注释,可以获取本物种基因序列与近缘物种基因序列的相似性以及本物种基因的功能信息。NR 注释物种分布如图 3 所示。由图 3 可知,物种分布占比由大到小依次为橡胶树(*Hevea brasiliensis*,15.3%)、麻风树(*Jatropha curcas*,9.4%)、毛果杨(*Populus trichocarpa*,9.0%)、蓖麻(*Ricinus communis*,8.9%)、木薯(*Manihot esculenta*,8.9%)。

2.1.4 基因 GO 功能分类 GO 分类将基因的生物学特征分为生物过程、细胞组分和分子功能,按

GO 功能分类将余甘子的 28 844 条基因划分为 43 个类别，其中生物过程 26 个，细胞组分 5 个，分子功能 12 个（图 4）。

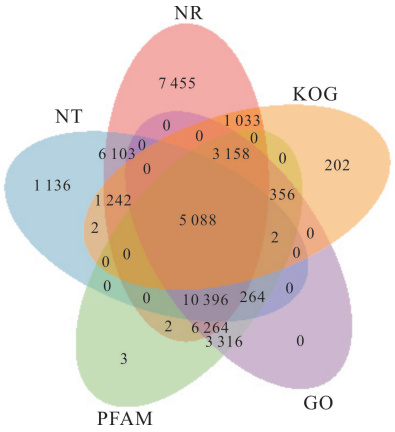


图 2　基因数据库功能注释结果

Fig. 2　Database function annotation results of unigene

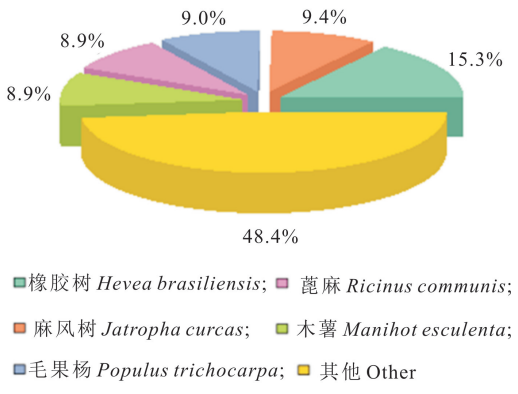


图 3　基因 NR 注释物种分布

Fig. 3　NR annotation species distribution of unigene

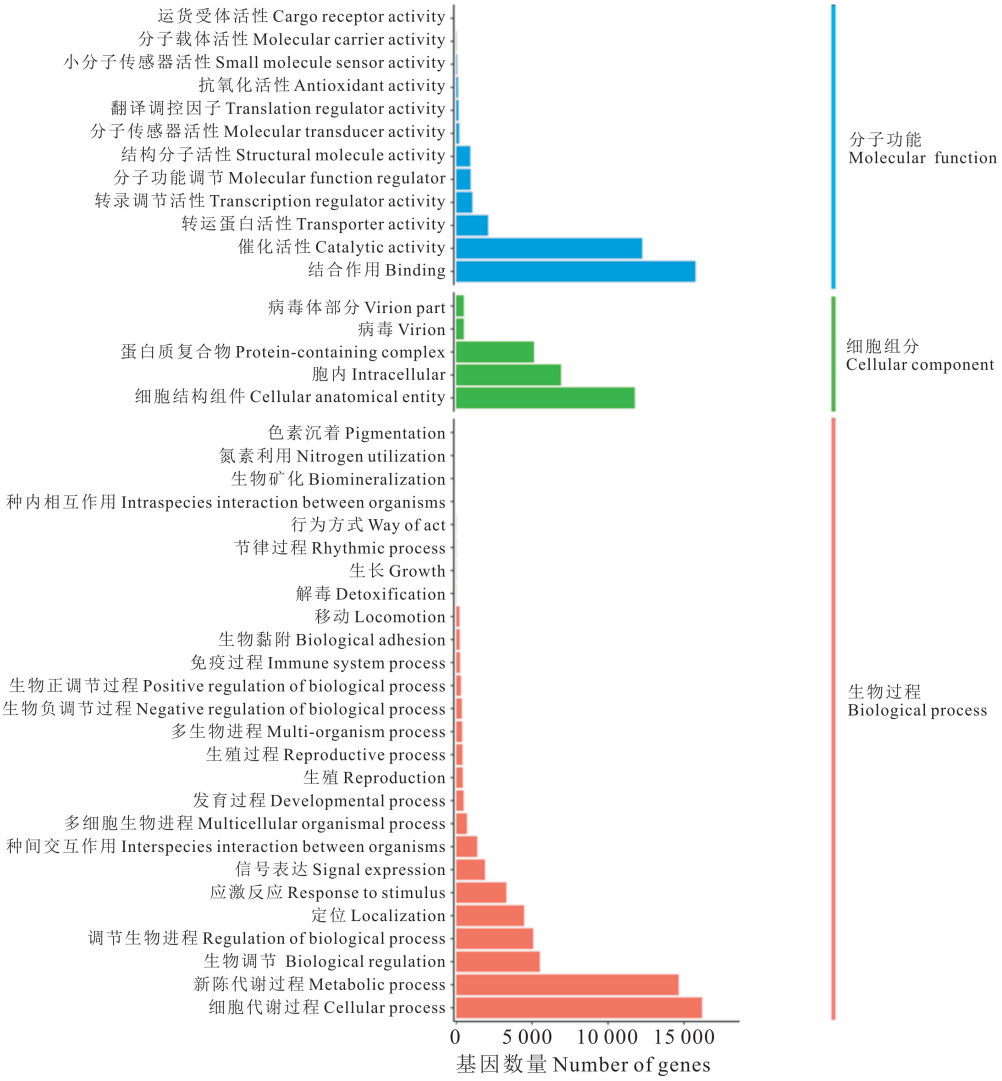


图 4　基因 GO 功能分类

Fig. 4　GO function classification of unigene

2.1.5　基因 KOG 功能分类　通过比对 KOG 功能分类，将基因归为 25 类（图 5）。其中注释基因数量

较多的有主要功能预测,翻译后修饰、蛋白质折叠和伴侣蛋白,翻译、核糖体结构和生物起源,基因占比分别为 13.38%,13.13%和 9.46%;注释基因数量较少的有核酸结构、抵御机制、胞外结构和细胞运动,其基因占比分别为 0.67%,0.55%,0.52%和 0.17%。

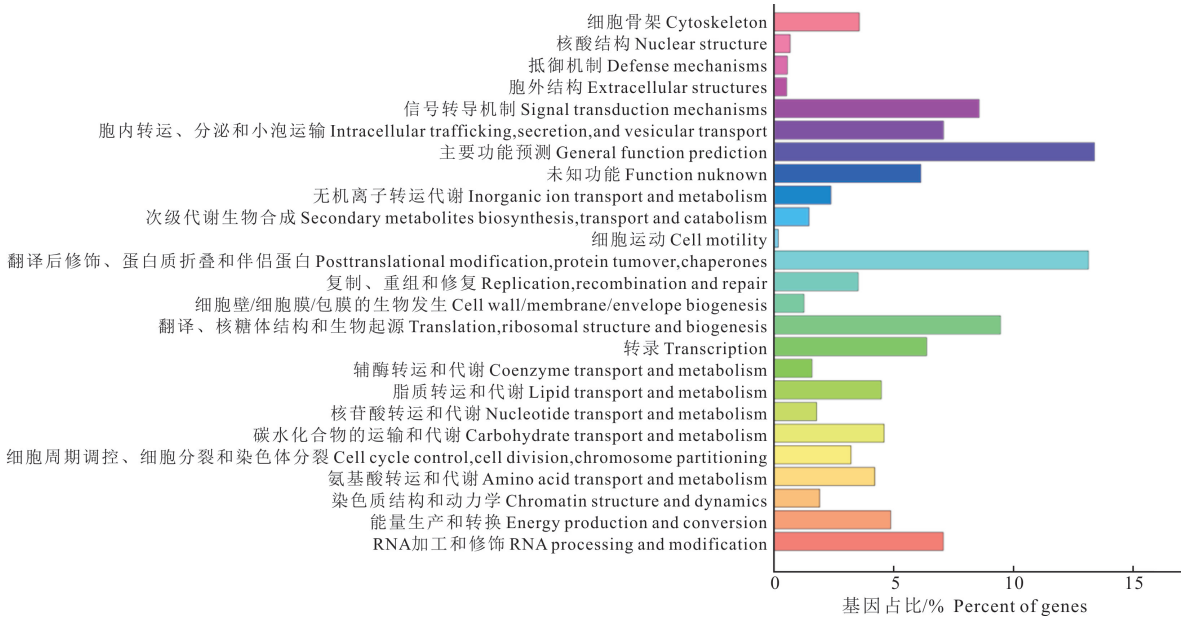
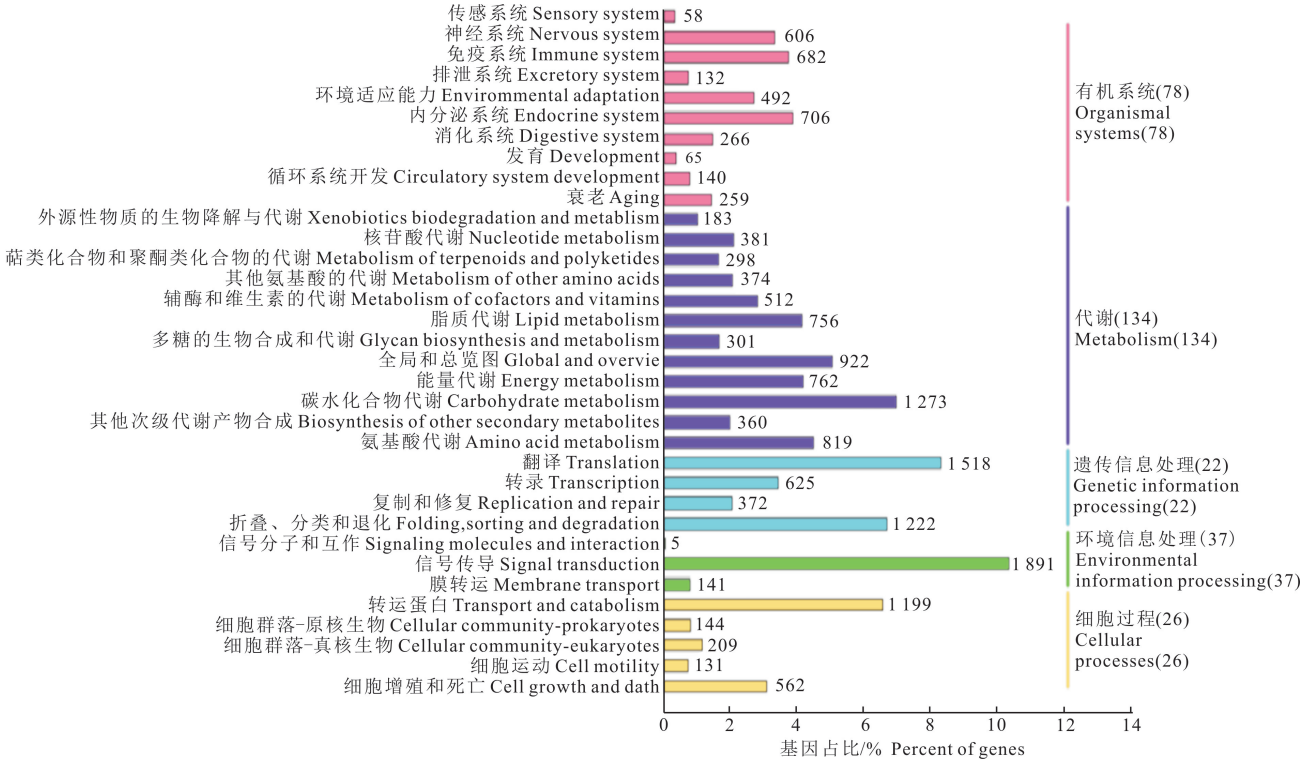


图 5 基因 KOG 功能分类

Fig. 5 KOG function classification of unigene

2.1.6 基因 KEGG 注释 为了解基因产物在细胞中的代谢途径及功能,将基因比对到 KEGG 数据库,结果将获释基因参与的代谢通路分为 5 个分支,34 个亚类,298 条代谢通路(图 6)。由图 6 可知,



柱后数据代表基因数量;括号中数据代表代谢通路数量

The data on the rightside of the column represent the number of unigene;The data in parentheses represent the number of pathway

图 6 基因 KEGG 功能分类

Fig. 6 KEGG function classification of unigene

由图 6 可知，细胞过程注释到 26 条代谢通路，共注释基因 2 245 条；环境信息处理注释到 37 条代谢通路，注释基因 2 037 条；遗传信息处理注释到 22 条代谢通路，注释基因 3 737 条；代谢共注释 134 条代谢通路，注释基因 6 941 条；有机系统注释 78 条代谢通路，共注释基因 3 406 条。

2.2　余甘子转录组 SSR 特征分析

2.2.1　SSR 位点数量与分布特点　余甘子转录组长度为 69 213 477 bp，在 65 074 条基因条带中共获得 18 245 个 SSR 位点，位于 14 110 条基因上，其发生频率为 21.68%（含有 SSR 位点的基因占基因总数的百分比），平均每 15.05 kb 含有 1 个 SSR 位点，SSR 出现频率为 28.04%（SSR 总数占基因总数的百分比），SSR 分布密度为 0.264 个/kb（SSR 总数占基因长度的百分比）。此外，仅含有 1 个 SSR 位点的基因为 11 015 条，含有 2 个或 2 个以上 SSR 位点的基因为 3 095 条。

由表 4 可知，单核苷酸至六核苷酸 6 种重复类型的 SSR 在余甘子转录组中均存在，其中重复类型最多为单核苷酸，占比为 64.49%，平均长度为 12.47 bp；二核苷酸和三核苷酸重复类型占比分别为 19.79%和 14.60%；四核苷酸、五核苷酸和六核苷酸重复类型的 SSR 数量比较少，占比分别为 0.76%，0.31%和 0.05%，平均长度分别为 21.36，26.79 和 34.20 bp。

表 4　余甘子转录组 SSR 重复类型分布

Table 4　Distribution of SSR repeat types in the transcriptome of *Phyllanthus emblica*

重复类型 Repeat type	SSR 数量 Amount of SSR	占比/% Percent	平均长度/bp Average length	总长度/bp Total length
单核苷酸 Mononucleotide	11 766	64.49	12.47	146 672
二核苷酸 Dinucleotide	3 611	19.79	19.85	71 676
三核苷酸 Trinucleotide	2 664	14.60	19.31	51 438
四核苷酸 Tetranucleotide	138	0.76	21.36	2 948
五核苷酸 Pentanucleotide	56	0.31	26.79	1 500
六核苷酸 Hexanucleotide	10	0.05	34.20	342
合计 Total	18 245			274 576

2.2.2　SSR 重复基元碱基组成　在余甘子转录组 18 245 个 SSR 位点中，有 192 种重复基元，其中单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸重复基元分别为 4，12，60，56，50 和 10 种。在考虑到碱基互补的情况下，余甘子转录组 18 245 个 SSR 位点中，单碱基、二碱基、三碱基、四碱基、五碱基和六碱基重复基元数分别为 2，4，10，20，23 和 10 种（图 7）。

从分布频率来看，单核苷酸重复基元 A/T 数目最多（11 738 个），占总 SSR 的 64.34%，占单核苷酸重复类型的 99.76%；二核苷酸重复基元 AG/CT 最多（2 467 个），占总 SSR 数的 13.52 %，占二核苷酸重复类型的 66.93%；三核苷酸重复类型中，AAG/CTT 和 AAT/ATT 为主要重复基元，分别有 571 和 455 个，分别占三核苷酸重复类型的 21.43%和 17.08%；四核苷酸重复基元以 AAAT/ATTT 最多，占四核苷酸重复类型的 31.88%；重复基元 AAACC/GGTTT 占五核苷酸重复类型的 14.29%；六核苷酸重复基元 10 种，均占 10%，分布平均。在余甘子转录组中，出现频率较高的重复基元为 A/T（64.34%）、AG/CT（13.52 %）、AAG/CTT（3.13%）、AC/GT（2.69%）和 AAT/ATT（2.49%）。

2.2.3　SSR 重复次数与长度　余甘子转录组 SSR 位点碱基重复次数为 5～84，其中碱基重复 10 次的 SSR 位点最多（4 210 个），占总 SSR 位点的 23.07%；其次为碱基重复 11，12，6 和 5 次的 SSR 位点数，分别为 2 314，1 642，1 634 和 1 470 个。碱基重复 5～22 次的 SSR 位点共有 17 602 个，占总 SSR 位点的 96.48%；碱基重复≥22 次的 SSR 位点仅占总 SSR 位点的 3.52%。

SSR 长度是决定多态性的重要因素之一，按照 SSR 长度可将多态性分为 3 种类型，分别是低多态性 SSR（<12 bp）、中等多态性 SSR（12～19 bp）和高多态性 SSR（≥20 bp）[15]。余甘子 SSR 的长度为 10～84 bp，长度≥20 bp 的高多态性 SSR 位点有 2 714 个（占比为 14.88%），其中单核苷酸重复、二核苷酸重复和三核苷酸重复的高多态性 SSR 位点分别有 378，1 365 和 768 个，分别占所有高多态性 SSR 位点的 13.93%，50.29%和 28.30%（表 5）。

2.3　余甘子 SSR 分子标记的开发

2.3.1　SSR 引物设计　对 18 245 个含 SSR 位点的基因进行 SSR 引物设计，共设计 12 488 对 SSR 引物，成功率为 73.88%。这 12 488 对 SSR 引物的

PCR 扩增产物中,重复类型为单核苷酸的 SSR 引物有 7 602 对,占比为 60.87%;重复类型为二核苷酸的 SSR 引物有 2 062 对,占比为 16.51%;重复类型为三核苷酸的 SSR 引物有 1 887 对,占比为 15.11%。

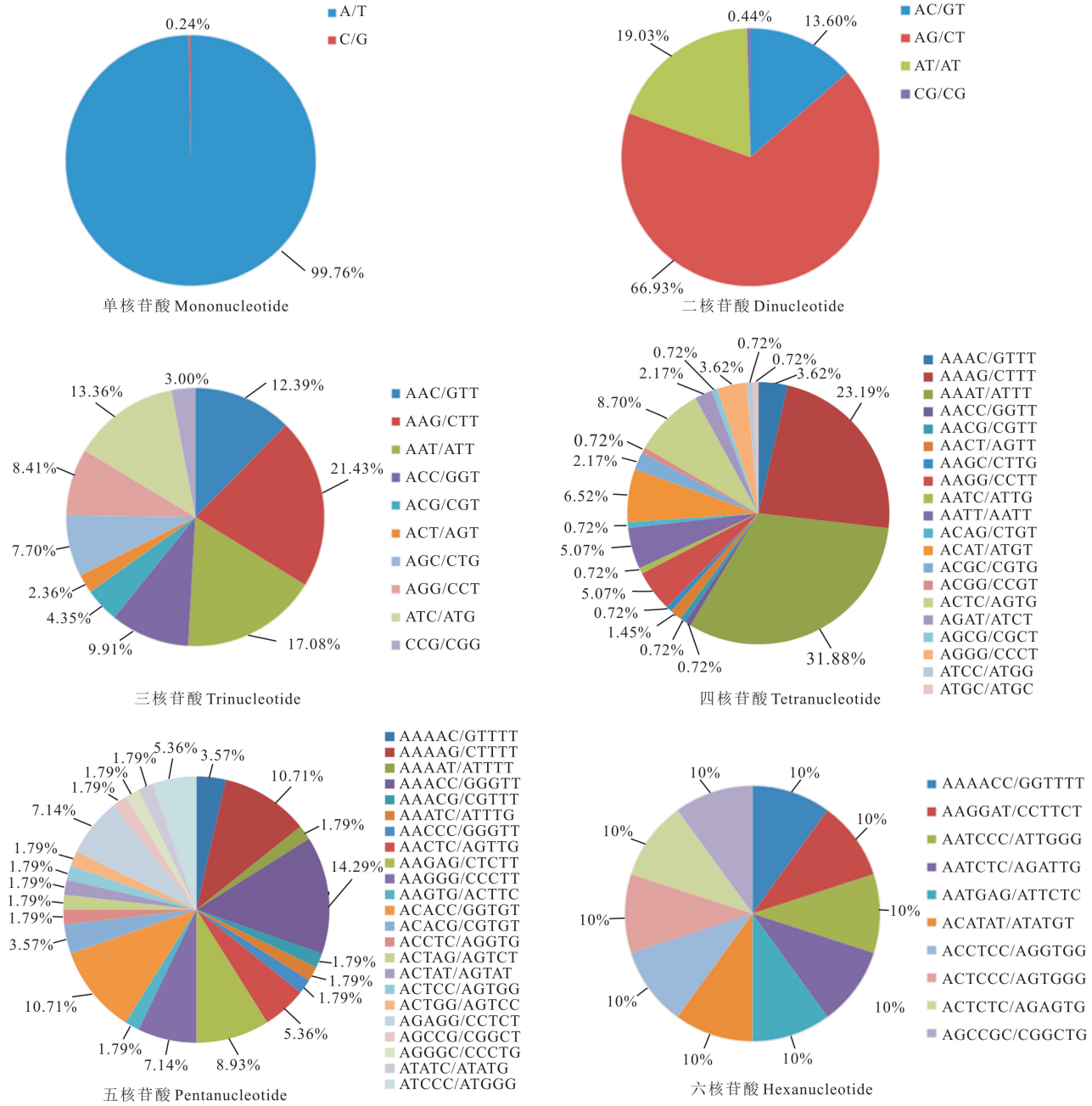


图 7 SSR 不同类型重复基元碱基组成

Fig. 7 Proportion of each element of different repeat types of SSR

表 5 余甘子 SSR 位点长度分析

Table 5 Frequency of SSR in each length range

SSR 长度范围 SSR length range	数量 Amount	占比/% Percentage	SSR 长度范围 SSR length range	数量 Amount	占比/% Percentage
<12 bp	6 031	33.06	≥20 bp	2 714	14.88
12~19 bp	9 500	52.07	三核苷酸 Trinucleotide	768	4.21
≥20 bp	2 714	14.88	四核苷酸 Tetranucleotide	35	0.19
单核苷酸 Mononucleotide	378	2.07	五核苷酸 Pentanucleotide	56	0.31
二核苷酸 Dinucleotide	1 365	7.48	六核苷酸 Hexanucleotide	10	0.05

2.3.2　SSR 引物筛选　随机挑选 154 对 SSR 引物进行有效性和多态性验证。首先以随机挑选的 8 份余甘子混合基因组 DNA 为模板，利用 154 对随机引物进行 PCR 扩增，经 1%琼脂糖凝胶电泳检测，初步筛选扩增特异条带与预期产物片段大小相符的 SSR 引物 45 对。

以初筛的 45 对 SSR 引物分别对 8 份余甘子基因组 DNA 进行 PCR 扩增，经 1%琼脂糖凝胶电泳检测，筛选出扩增条带清晰且多态性较好的 SSR 引物 10 对，其详细信息见表 6。由表 6 可知，10 对 SSR 引物重复基元为二核苷酸 7 个，三核苷酸 1 个和五核苷酸 2 个。其中二核苷酸重复基元有 3 种类型，分别为 CT、TC 和 AG；三核苷酸重复基元为 TTC；五核苷酸重复基元为 TCCCT 和 GAAAA。10 对 SSR 引物的 PCR 扩增产物长度为 127～270 bp，平均为 220.5 bp。其中引物 59 的 PCR 扩增产物长度最小，引物 118 的 PCR 扩增产物长度最大。

表 6　10 对多态性 SSR 引物信息

Table 6　Information of 10 pairs of polymorphic SSR primer

引物编号 Primer number	重复基元 Repeat motif	正向引物序列(5′→3′) Forward primer sequence	反向引物序列(5′→3′) Reverse primer sequence	产物长度/bp Size of product
4	TCCCT	AGAAGAAGACAGCGACCAAA	GGAGATTAGGTGGCGGGATG	236
6	GAAAA	TTGATTGTTCGTTGGTCGGC	AGAAGGGAAAAGTTGGGACGA	238
59	TTC	AGTCTGCCATCCGATGAAGC	GCCTCTGGCCTTAAACCTGA	127
79	CT	CGAAGCAATTGCCCAAACCT	GGGGCAAAGAAGTTCCCTGA	254
86	AG	ACGGAGAGGGAACACAAACC	TGAGGGTCAAAAAGGGTTCTCA	255
90	AG	ATATGATGGCCACCGTCACC	AGGTGTCTCTCTCTTGGACTCA	237
118	AG	GCAGCAAAGAGAGCTTCTTCA	GAGGTTGTAAGAGGGGCTTGT	270
127	TC	AGCACTACACACCAACCTTCT	GTGGCGGTCTATGGTTTTGC	265
136	CT	TGGTTTGTAGCGAACATGCTG	CAGCACAACAAGAGGCCAAC	137
138	CT	GCCACTTGAAAAGGTGAAGCA	GAGGACGATGTGGAGTTCCC	186

3　讨　论

本研究利用叶片表型变异较大的 6 份余甘子叶片样品进行转录组测序，经拼接和去冗长后，组装到 65 074 条基因，核苷酸总数为 69 213 477 bp，平均长度为 1 064 bp，N50 长度为 1 525 bp，均大于赵琼玲等[16]对 4 份余甘子果实转录组的测序结果和刘雄芳等[17]对云南干热河谷余甘子叶片转录组的测序结果。N50 越高表明测序质量越可靠，组装效果越好，本研究测序组装的质量和长度均满足了余甘子转录组分析的要求，可利用本次转录组数据开展转录组和微卫星定位分析。本研究利用 NR、NT、PFAM、KOG、Swiss-prot、KO 和 GO 共 7 个数据库的基因功能比对，成功注释的基因序列为 46 707 条，占比为 71.77%，其中注释到 NR 数据库的基因序列有 40 741 条，占比为 62.60%。通过与 NR 数据库比对注释，可以获取本物种基因序列与近缘物种基因序列的相似性以及本物种基因的功能信息。本研究结果表明，15.3%的基因注释为橡胶树(*Hevea brasiliensis*)，9.4%为麻风树(*Jatropha curcas*)，这与前人的研究结果麻风树 14.09%[18]有差异；此外，注释基因占比较高的还有毛果杨(*Populus trichocarpa*)、蓖麻(*Ricinus communis*)、木薯(*Manihot esculenta*)等。利用 KEGG 进一步注释余甘子基因，分析其产物在细胞中可能的代谢途径，结果表明基因参与的代谢通路分为 5 个分支，34 个亚类，298 个代谢通路，涉及代谢各通路的基因数据可为余甘子抗性机理探索、余甘子特殊代谢途径相关基因挖掘等研究提供基础数据。

本研究在余甘子 65 074 条基因中获得 18 245 个 SSR 位点，SSR 位点出现频率为 28.04%，SSR 位点分布密度为 0.264 SSRs/kb，高于前人的 0.182 SSRs/kb[18]。SSR 多态性是判断是否能够开发余甘子 SSR 分子标记的重要依据，影响 SSR 多态性的重要因素是重复次数和序列长度[19]。SSR 重复次数越多，则多态潜力越高[20]，重复次数≥12 则表现较高多态性[21]；SSR 序列长度≥20 bp 时为高多态性，长度为 12～19 bp 时中等多态性，长度＜12 bp 时低多态性[15]。本研究余甘子转录组 SSR 序列长度为 10～84 bp，长度≥20 bp 的高多态性 SSR 位点有 2 714 个，占比为 14.88%，这与碧桃(*Prunus persica* Duplex)[22]、短丝木犀(*Osmanthus serrulatus*)[23]等大多数植物相当，这些高多态性的 SSR 位点在余甘子分子标记开发中具有较大应用价值。在高多态性 SSR 位点中，单核苷酸重复、二核苷酸重复和三核苷酸重复的 SSR 位点共有 2 511 个，占所

有高多态性 SSR 位点的 92.52%。低级重复单元越多,表明该物种变异频率越高,进化水平越高[24]。

根据 18 245 条含有 SSR 位点的基因进行引物设计,共设计 12 488 对 SSR 引物,随机挑选 154 对进行 PCR 扩增,其中 45 对引物扩增的特异性条带与预期产物大小相符,为有效扩增引物。再利用初筛的 45 对 SSR 引物分别对 8 份余甘子基因组 DNA 进行 PCR 扩增,筛选 10 对 SSR 引物存在明显多态性,占比为 22.22%。该比率低于较多植物的多态比率[14,25],可能与采用琼脂糖凝胶筛选引物,未采用非变性聚丙烯酰胺凝胶有关。据报道,琼脂糖凝胶和聚丙烯酰胺凝胶进行 SSR 产物分离的结果基本一致,但因琼脂糖凝胶分辨率较低,导致多态性比率降低[26]。

4 结 论

对 6 份表型变异较大余的甘子叶片进行转录组分析,获得 65 074 条基因,平均长度为 1 064 bp,且 71.77%的基因成功获得功能注释信息。从 65 074 条基因条带中共获得 18 245 个 SSR 位点,位于 14 110 条基因上,其发生频率为 21.68%,多态性 SSR 位点有 2 714 个,用于遗传多样性分析的潜力较大,该研究结果为开发余甘子 SSR 分子标记提供了重要的数据信息。经有效性和多态性分析,从随机挑选的 154 对 SSR 引物中筛选出扩增条带清晰且多态性较好的 SSR 引物 10 对,其重复基元分别为二核苷酸 7 个,三核苷酸 1 个和五核苷酸 2 个,扩增产物大小为 127~270 bp。

[参考文献]

[1] 中国科学院中国植物志委员会. 中国植物志:余甘子 *Phyllanthus emblica* Linn.[M]. 北京:科学出版社,2004:87.
Chinese Flora Committee, Chinese Academy of Sciences. Flora reipublicae popularis sinicae: *Phyllanthus emblica* Linn.[M]. Beijing: Science Press, 2004:87.

[2] 胡文舜,陈秀萍,郑少泉. 龙眼 EST-SSR 标记开发及无患子科 5 个属种质遗传多样性分析[J]. 园艺学报,2019,46(7):1359-1372.
HU W S, CHEN X P, ZHENG S Q. EST-SSR markers developed from *Dimocarpus longan* and their application in genetic diversity analysis of five genera of Sapindaceae[J]. Acta Horticulturae Sinica, 2019, 46(7):1359-1372.

[3] 熊仪俊. 余甘子不同生态型特征与分化初步研究[D]. 北京:中国林业科学研究院,2003.
XIONG Y J. Preliminary study on the characteristics and differentiations of ecological types of *Phyllanthus emblica*[D]. Beijing: Chinese Academy of Forestry, 2003.

[4] 蔡英卿,赖钟雄,桑庆亮,等. 余甘子基因组 DNA 提取及 RAPD 反应条件优化[J]. 福建农林大学学报(自然科学版),2003,32(1):89-92.
CAI Y Q, LAI Z X, SANG Q L, et al. Genomic DNA extraction and optimization of RAPD analytic conditions of *Phyllanthus emblica* Linn.[J]. Journal of Fujian Agriculture and Forestry University (Natural Science Edition), 2003, 32(1):89-92.

[5] 蔡英卿,赖钟雄,陈义挺,等. 福建余甘子遗传资源的 RAPD 分析[J]. 热带作物学报,2007,28(2):74-79.
CAI Y Q, LAI Z X, CHEN Y T, et al. RAPD analysis of emblic (*Phyllanthus emblica* L.) genetic resources in Fujian Province[J]. Chinese Journal of Tropical Crops, 2007, 28(2):74-79.

[6] DNYANESHWAR W, PREETI C, KALPANA J, et al. Development and application of RAPD-SCAR marker for identification of *Phyllanthus emblica* Linn.[J]. Biological and Pharmaceutical Bulletin, 2006, 29(11):2313-2316.

[7] CHAURASIA A K, SUBRAMANIAM V R, KRISHNA B, et al. RAPD based genetic variability among cultivated varieties of Aonla (Indian Gooseberry, *Phyllanthus emblica* L.)[J]. Physiology and Molecular Biology of Plants, 2009, 15(2):169-173.

[8] THILAGA S, NAIR R R, KANNAN M R, et al. RAPD markers for screening shoot gall maker (*Betousa stylophora* Swinhoe) tolerant genotypes of amla (*Phyllanthus emblica* L.)[J]. Journal, Genetic Engineering & Biotechnology, 2017, 15(2):323-330.

[9] 郭林榕,周平,陈志峰. 22 份余甘子核心种质资源遗传多样性的 SRAP 分析[J]. 热带作物学报,2014,35(7):1382-1387.
GUO L R, ZHOU P, CHEN Z F. Analysis of genetic diversity of 22 *Phyllanthus emblica* germplasms with SRAP markers[J]. Chinese Journal of Tropical Crops, 2014, 35(7):1382-1387.

[10] 王建超,何银莺,黄旭萍,等. 响应面法优化余甘子种质资源 ISSR 反应体系[J]. 福建农业学报,2021,36(7):759-765.
WANG J C, HE Y Y, HUANG X P, et al. Response surface optimization of ISSR-PCR reaction for genetic study on *Phyllanthus emblica*[J]. Fujian Journal of Agricultural Sciences, 2021, 36(7):759-765.

[11] 刘晓生,郑道序,周春娟,等. 潮汕余甘子种质资源遗传多样性与亲缘关系的 ISSR 分析[J]. 中国南方果树,2014,43(1):18-22.
LIU X S, ZHENG D X, ZHOU C J, et al. Analysis of genetic diversity and genetic relationship of *Phyllanthus emblica* L. germplasm in Chaoshan area with ISSR[J]. South China Fruits, 2014, 43(1):18-22.

[12] KUMAR A, SINGH K. Isolation of high quality RNA from *Phyllanthus emblica* and its evaluation by downstream applications[J]. Molecular Biotechnology, 2012, 52(3):269-275.

[13] 李金璐,王硕,于婧,等. 一种改良的植物 DNA 提取方法[J]. 植物学报,2013,48(1):72-78.
LI J L, WANG S, YU J, et al. A modified CTAB protocol for plant DNA extraction[J]. Chinese Bulletin of Botany, 2013, 48(1):72-78.

[14] 方智振,叶新福,周丹蓉,等. '芙蓉李'转录组 SSR 信息分析

与分子标记开发[J]. 果树学报，2016，33(4)：416-424.
FANG Z Z，YE X F，ZHOU D R，et al. Analysis on SSR information in 'Furongli' plum in 'Furongli' plum transcriptome and development of molecular markers in *Prunus salicina* Lindl. [J]. Journal of Fruit Science，2016，33(4)：416-424.

[15] TEMNYKH S，DECLERCK G，LUKASHOVA A，et al. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.)：frequency，length variation，transposon associations，and genetic marker potential[J]. Genome Research，2001，11(8)：1441-1452.

[16] 赵琼玲，张永辉，罗会英，等. 余甘子转录组分析[J]. 中国农学通报，2018，34(20)：63-69.
ZHAO Q L，ZHANG Y H，LUO H Y，et al. Transcriptome analysis of *Phyllanthus emblica* L. [J]. Chinese Agricultural Science Bulletin，2018，34(20)：63-69.

[17] 刘雄芳，李太强，李正红，等. 云南干热河谷地区余甘子转录组分析[J]. 林业科学研究，2018，31(5)：1-8.
LIU X F，LI T Q，LI Z H，et al. Transcriptome analysis for *Phyllanthus emblica* distributed in dry-hot valleys in Yunnan，China[J]. Forest Research，2018，31(5)：1-8.

[18] 刘雄芳，李太强，张序，等. 重要食药同源植物余甘子转录组微卫星特征分析[J]. 植物研究，2019，39(2)：294-302.
LIU X F，LI T Q，ZHANG X，et al. Characteristic analysis of microsatellites in the transcriptome of *Phyllanthus emblica*：an important edible and medicinal plant[J]. Bulletin of Botanical Research，2019，39(2)：294-302.

[19] SHAROPOVA N. Plant simple sequence repeats：distribution，variation，and effects on gene expression[J]. Genome，2008，51(2)：79-90.

[20] THAO D V，YAMASHITA M，WATANABE A，et al. Development of tetranucleotide microsatellite markers in *Pinus kesiya* Royle ex Gordon [J]. Conservation Genetics Resources，2013，5(2)：405-407.

[21] WEBER J L. Informativeness of human (dC-dA) n · (dG-dT) n polymorphisms[J]. Genomics，1990，7(4)：524-530.

[22] 马秋月，廖卓毅，张得芳，等. 碧桃花瓣转录组微卫星特征分析[J]. 南京林业大学学报(自然科学版)，2015，39(3)：34-38.
MA Q Y，LIAO Z Y，ZHANG D F，et al. Deep sequencedbased transcriptome analysis of microsatellites in peach (*Prunus persica* cv. duplex) flowers[J]. Journal of Nanjing Forestry University (Natural Sciences Edition)，2015，39(3)：34-38.

[23] 陈林，李龙娜，杨国栋，等. 特有植物短丝木犀(*Osmanthus serrulatus*)转录组微卫星特征分析[J]. 分子植物育种，2016，14(4)：959-965.
CHEN L，LI L N，YANG G D，et al. Microsatellites characteristics of transcriptomic sequences from *Osmanthus serrulatus* [J]. Molecular Plant Breeding，2016，14(4)：959-965.

[24] HARR B，SCHLÖTTERER C. Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times，which cause their genome-wide under representation[J]. Genetics，2000，155(3)：1213-1220.

[25] 蒋路园，吴文丽，张恺恺，等. 南方红豆杉转录组 SSR 位点分析及其分子标记开发[J]. 中草药，2024，55(3)：928-936.
JIANG L Y，WU W L，ZHANG K K，et al. Transcriptome sequencing and development of SSR molecular markers of *Taxus chinensis* var. *mairei*[J]. Chinese Traditional and Herbal Drugs，2024，55(3)：928-936.

[26] 叶春秀，李有忠，庄振刚，等. 两种电泳方法在分析新陆早系列品种 SSR 标记结果中的比较与应用[J]. 分子植物育种，2014，12(5)：909-913.
YE C X，LI Y Z，ZHUANG Z G，et al. Comparation and application of the two electrophoresis on SSR of earliness cultivars in upland cotton in Xinjiang[J]. Molecular Plant Breeding，2014，12(5)：909-913.

（责任编辑：陈丽）