

大青转录组测序及生物信息学分析

刘敏¹, 黄炜忠², 何孟璐¹, 梅瑜¹, 王继华¹

(1. 广东省农作物遗传改良重点实验室/广东省农业科学院作物研究所, 广东广州 510640;

2. 广东罗浮山国药股份有限公司, 广东惠州 516100)

摘要:【目的】利用高通量测序技术获得大青的转录组信息特征。【方法】通过高通量测序平台 Illumina HiSeq™ 2500 对大青进行转录组测序, 采用 Trinity 软件 *de novo* 组装获得 Unigene, 并基于序列同源性对 Unigene 进行功能注释, 得到大青转录组的遗传信息。【结果】测序数据经过质控后共获得 26 394 223 个高质量的 Reads, 通过 *de novo* 组装获得 100 191 个 Unigene, N50 长度为 1 055 bp, 平均长度 724.4 bp。其中 59 690 个 (59.58%) Unigene 在 NR、SwissProt、KOG、GO、KEGG 数据库中均得到注释。其中 KEGG 数据库注释到 38 260 个 Unigene, 涉及 136 条代谢通路。在大青转录组中共鉴定到 407 个 Unigene 参与萜类化合物的生物合成, 165 个 Unigene 涉及类黄酮生物合成, 29 个 Unigene 参与黄酮和黄酮醇生物合成, 37 个 Unigene 参与异黄酮生物合成, 同时, 还鉴定到 210 类转录因子。MISA 分析发现 6 680 个 Unigene 包含 8 640 个简单重复序列。【结论】利用高通量测序技术和生物信息分析获得了大青的转录组信息特征, 这些数据可为后期开展功能基因鉴定、解析黄酮类化合物次生代谢途径及其调控机制奠定研究基础。

关键词: 大青; 萜类; 黄酮; 转录组; 简单重复序列 (SSR)

中图分类号: R282.5

文献标志码: A

文章编号: 1007-3213(2022)01-0177-07

DOI: 10.13359/j.cnki.gzxbtcm.2022.01.030

Transcriptome Sequencing and Bioinformatic Analysis of *Clerodendron cyrtophyllum* Turez

LIU Min¹, HUANG Wei-Zhong², HE Meng-Lu¹, MEI Yu¹, WANG Ji-Hua¹

(1. Key Laboratory of Crops Genetics and Improvement of Guangdong Province/Crop Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou 510640 Guangdong, China; 2. Guangdong Luofushan Sinopharm Co., Ltd., Huizhou 516100 Guangdong, China)

Abstract: Objective To obtain the transcriptome information characteristics of *Clerodendron cyrtophyllum* Turez by high-throughput sequencing technology. **Methods** The transcriptome sequencing of *Clerodendron cyrtophyllum* Turez was performed by high-throughput sequencing platform Illumina HiSeq™ 2500. Unigenes were *de novo* assembled by Trinity software, and then the function annotation of unigenes was carried out based on sequence homology to obtain the genetic information of *Clerodendron cyrtophyllum* Turez transcriptome. **Results** After quality control of sequencing data, a total of 26 394 223 high quality reads were obtained, 100 191 unigenes were obtained through *de novo* assembly, and the N50 length was 1 055 bp, with an average length of 724.4 bp. A total of 59 690 (59.58%) unigenes were annotated in NR, SwissProt, KOG, GO and KEGG databases. Among them, 38 260 unigenes were annotated in the KEGG database, involving 136 metabolic pathways. A total of 407 unigenes were identified in the transcriptome of *Clerodendron cyrtophyllum* Turez involved in terpenoid biosynthesis, 165 unigenes involved in flavonoid biosynthesis, 29 unigenes involved in flavone and flavonol biosynthesis, 37 unigenes involved in isoflavonoid biosynthesis, and 210 transcription factors (TFs) were also identified. MISA

收稿日期: 2021-04-15

作者简介: 刘敏(1989-), 男, 硕士, 助理研究员; E-mail: 279437871@qq.com

通讯作者: 王继华(1979-), 男, 博士, 研究员; E-mail: wangjihua@gdaas.cn

基金项目: 以农产品为单元的广东省现代农业产业技术体系创新团队建设(南药产业)(编号: 2019KJ148); 科技创新战略专项资金(高水平农科院建设)-金颖之星(编号: R2019PY-JX003); 广东省博罗县南药产业园项目(2019)

analysis showed that 6 680 unigenes contained 8 640 simple sequence repeats (SSRs). **Conclusion** High-throughput sequencing technology and bioinformatic analysis are used to obtain the transcriptome information characteristics of *Clerodendron cyrtophyllum* Turez, and these data will lay a foundation for later research on functional gene identification and analysis of secondary metabolic pathways of flavonoids and their regulatory mechanisms.

Keywords: *Clerodendron cyrtophyllum* Turez; terpenoid; flavonoid; transcriptome; simple sequence repeats (SSR)

大青 (*Clerodendron cyrtophyllum* Turez) 是马鞭草科大青属 (*Verbenaceae clerodendron* L.) 灌木或小乔木植物, 又名大青、大青木、山大青和羊咪青等, 多生于海拔 1 700 米以下的平原、山地林下或溪谷旁, 分布于两湖、两广及云贵高原等地, 朝鲜、越南、马来西亚也有分布^[1-2]。大青全株皆可入药, 具有清热解毒、祛风除湿的功效, 主治乙脑、流脑、外感热病热盛烦渴、咽喉肿痛、口疮、黄疸、热毒痢、急性肠炎、痈疽肿毒、衄血、血淋、外伤出血等病症。现代药理学研究表明, 大青具有抑菌抗炎、抗病毒、利尿、镇痛降压的作用, 是多种重要的中成药及保健品的主要成分^[3-4]。大青叶在我国南方也常作为食疗药材, 其干燥叶也是凉茶的主要原料之一, 在广东多被制作成大青叶茶等保健食品^[5]。

据《本草纲目》考证, 民间药用大青为马鞭草科大青属植物大青, 但其种类繁多。现根据地域区分就有广东大青、江西大青、广西大青、浙江大青等不同品种, 这些不同品种之间的形态特征极其相似^[6]。而现代中药材的资源鉴定多以形态鉴定及气味辨别为主, 但不同大青的种类大多形态及气味极其相近, 极难辨别鉴定区分; 且不同品种之间虽然化学成分相近, 但部分药效成分的含量却差别极大^[7]。随着生物学技术的发展, 基于高通量测序技术的转录组和生物信息学分析不受基因组信息的限制, 在解析药用活性成分合成途径、挖掘相关功能基因、开发中药分子标记等方面得到了广泛的应用^[8-12]。目前, 许多药用植物的转录组学研究已经开展, 如金银花^[12]、地黄^[13]、甘草^[14]、茯苓^[15]等, 并鉴定到大量药效活性物质的合成途径相关基因, 并开发了可靠的分子标记。目前, 大青的研究多集中于化学成分鉴定及药理活性研究等方面, 而关于大青有效化学成分的合成代谢途径关键基因的挖掘与鉴定尚未见报道。因

此, 本研究开展大青的转录组测序及生物信息学分析, 以期为进一步解析其有效活性物质的合成代谢通路, 挖掘其调控关键基因, 开发分子标记提供基础数据, 现将研究结果报道如下。

1 材料与方 法

1.1 实验材料 供试材料种植于广东省农业科学院作物研究所特色作物与南药资源圃。选取健康的大青植株的根、茎、叶, 迅速用锡箔纸包裹并在液氮中冷冻, 存储于-80℃冰箱中备用。

1.2 大青总 RNA 的提取 用 TRIzol (上海生工) 方法分别提取大青根、茎、叶总 RNA, 采用 Qubit 2.0 RNA 检测试剂盒 (Fluorometer Life Tech Invitrogen) 分别对大青的根、茎、叶总 RNA 质量进行检测和定量, 选取光密度 OD (260 nm)/OD (230 nm) 值大于 1.8 的 RNA 样品等量混合, 建库。

1.3 大青转录组测序与拼接组装 将检测合格的 RNA 委托北京百迈客生物科技有限公司采用 Illumina HiSeq™ 2500 的高通量测序平台对大青的转录组进行测序。测序获得原始数据后, 按照标准分析流程 (通过 FastQC 软件进行质量评估, 再使用 Trimmomatic 软件去掉接头和低质量序列) 对数据进行处理, 得到 Clean Data 用于 *de novo* 组装, 然后采用 Trinity 软件将转录本拼接组装获得 Unigene。

1.4 基因功能注释 采用 Blastx 程序将组装的 Unigene 与保守域数据库 (Conserved Domain Database, CDD), 非冗余蛋白序列 (NCBI non-redundant protein sequences, NR), 核酸序列数据库 (NCBI non-redundant nucleotide sequences, NT), 蛋白结构域预测 (Protein Families Database of Alignments and Hidden Markov Models, PFAM), 真核生物蛋白质同源簇数据库/蛋白质聚类 (eu Karyotic ortholog groups/clusters of orthologous groups, KOG/COG), SwissProt, TrEMBL, KEGG 直系同源数据库 (KEGG

ortholog database, KO), 基因本体论(Gene Ontology, GO)等多个数据库比对, 获得Unigene功能得到注释信息、GO功能注释信息。

1.5 简单重复序列(SSR)、单核苷酸多态性(SNP)检测和引物设计 采用微卫星识别工具(Microsatellite Identification Tool, MISA)软件对大青转录组所有Unigene存在的SSR微点进行鉴定及分析, 并使用Primer 3软件(<http://primer3.sourceforge.net/releases.php>)设计SSR引物。

2 结果与分析

2.1 大青转录组测序与 *de novo* 组装 采用Illumina HiSeq™ 2500高通量测序平台进行大青转录组测序, 获得Raw Reads后, 通过软件去除接头盒低质量序列后, 共得到26 394 223条Clean Reads, 长度为7 859 604 312 bp, GC含量为45.5%, Q20达98.17%, Q30达94.58%。该结果表明, 大青转录组数据质量较高, 能够开展后续生物信息学分析。利用Trinity软件将大青Clean Reads进行*de novo*组装成转录本, 共得到172 984条转录本, 共计183 501 164 bp, 平均长度1 060.8 bp, N50值为1 726 bp, 其中序列长度大于500 bp的有105 429个, 占总序列数的60.94%, 序列长度大于1 000 bp的有64 345个, 占总序列数的37.19%; 去除冗余处理后得到100 191个Unigene, 平均长度724.4 bp, N50值为1 055 bp, 其中序列长度大于500 bp的Unigene数有42 108个, 占总序列数的42.03%, 序列长度大于1 000 bp的Unigene数有18 421个, 占总序列数的18.39%。见表1、图1。从组装结果来看, 转录本和Unigene的N50值均大于平均值, 表明大青转录组的测序深度大, 组装效果较好。

2.2 大青转录组功能注释及其分类 使用Blast软件将组装的Unigene与九大生物信息学数据库进行比对分析, 得到大青转录组中的基因功能、代谢通路和转录因子等数据。通过比对分析, 共有59 690(59.58%)个Unigene至少在一个数据库中获

表1 大青转录组测序的结果

Table 1 Results of transcriptome sequencing of *Clerodendron cyrtophyllum* Turez

| | 转录本 总数/条 | ≥500 bp/ [条(%)] | ≥1 000 bp/ [条(%)] | N50/bp | 平均 长度/bp |
|------------|-------------|--------------------|----------------------|--------|-------------|
| Transcript | 172 984 | 105 429(60.94) | 64 345(37.19) | 1 726 | 1 060.8 |
| Unigene | 100 191 | 42 108(42.03) | 18 421(18.39) | 1 055 | 724.4 |

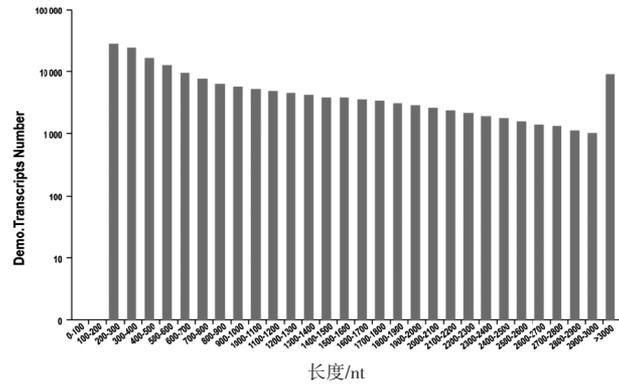


图1 大青Unigene序列长度分布

Figure 1 Sequence length distribution of unigenes of *Clerodendron cyrtophyllum* Turez

得信息注释, 长度介于300 bp到1 000 bp之间的Unigene有31 333个, 长度大于1 000 bp的Unigene有16 325个。除COG数据库外, 在其他八大数据库中能注释的大青Unigene数目均在30%以上。其中: 在NR和TrEMBL数据库中注释到的Unigene数目最多, 分别达到58 952(58.84%)个和58 311(58.2%)个; 在COG数据库中注释到的Unigene数目最少, 仅为13 668(13.64%)个; 49 278(49.18%)个Unigene在GO数据库中得到注释; 38 260(38.19%)个Unigene在KEEG数据库中得到注释; 32 453(32.39%)个Unigene在KOG数据库中得到注释; 尚有40 501(40.42%)个Unigene尚未得到注释。见表2。结果表明, 大青转录组获得的数据可以用于后续开展相应分析。

2.2.1 NR功能注释 大青的Unigene在NR数据库中的比对结果表明, 其与芝麻(*Sesamum indicum*)、玫瑰木属(*Rhodamnia argentea*)、紫花风铃木(*Handroanthus impetiginosus*)的基因组有一定的相似度。其中与芝麻的匹配度最高, 共有11 140条Unigene与之相匹配, 占整个NR数据库注释Unigene的18.9%, 其次分别为玫瑰木属和紫花风铃木, 分别占比为18.41%和10.25%, 见图2。从物种注释数据可以看出, 大青的Unigene序列全部可以与植物相匹配, 且与唇形目(管状花目)物种的序列相似度最高。但由于马鞭草科大青属植物的参考基因组数据比较少, 因此并未从比对数据库中发现同科属的比对结果。

2.2.2 KOG功能分类 将组装的大青Unigene与KOG数据库比对分析, 结果表明共有32 453条(32.39%)Unigene被注释到25个KOG功能分类

表2 大青Unigene注释的统计结果
Table 2 Statistical results of annotated unigenes of *Clerodendron cyrtophyllum* Turez

| 数据库 | 基因数/个 | (300 bp≤长度<1 000 bp)/条 | ≥1 000 bp/条 | 比例/% |
|-----------|--------|------------------------|-------------|-------|
| COG | 13 668 | 6 487 | 5 706 | 13.64 |
| GO | 49 278 | 25 697 | 14 012 | 49.18 |
| KEGG | 38 260 | 19 671 | 11 803 | 38.19 |
| KOG | 32 453 | 16 785 | 9 930 | 32.39 |
| PFAM | 34 907 | 17 349 | 13 242 | 34.84 |
| Swissprot | 39 148 | 20 387 | 12 202 | 39.07 |
| TrEMBL | 58 311 | 30 624 | 16 256 | 58.20 |
| NOG | 50 197 | 26 214 | 14 466 | 50.10 |
| NR | 58 952 | 30 969 | 16 270 | 58.84 |
| 合计 | 59 690 | 31 333 | 16 325 | 59.58 |

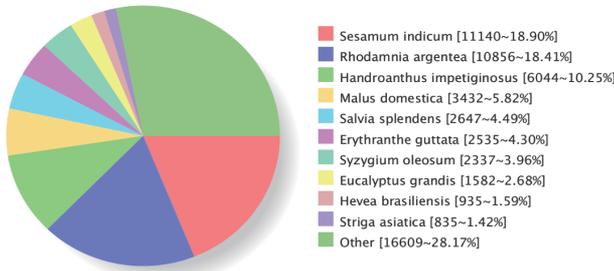


图2 大青Unigene的NR注释物种相似度分布
Figure 2 Similarity distribution of NR annotated species of unigenes of *Clerodendron cyrtophyllum* Turez

中。其中涉及一般功能预测的基因最多，共有6 359个(占19.59%)，其后依次是翻译后修饰，蛋白质转运，伴侣基因(3 945个，12.16%)，信号转导机制基因(3 382个，10.42%)，细胞内运输，分泌和囊泡运输(2 072个，6.38%)，碳水化合物运输与代谢(2 110个，6.5%)及翻译，核糖体结构与生物起源(1 935个，5.96%)，仅有11个基因注释到细胞运动性，占比0.03%。此外，还有1 862个Unigene被注释到未知功能。共有1 252个Unigene被注释到次生代谢物生物合成、运输和分解代谢，占总基因的比例为3.86%，这表明这些基因可能与大青中次生代谢物的生物合成、运输和积累相关。见图3。

2.2.3 GO功能注释 将组装的大青Unigene与GO数据库比对分析，结果显示共有49 278个(49.18%)Unigene被注释到43个GO功能分类中。其中，生物过程注释到的Unigene最多，共有76 354个，占76.2%，其中注释较多的功能分别为

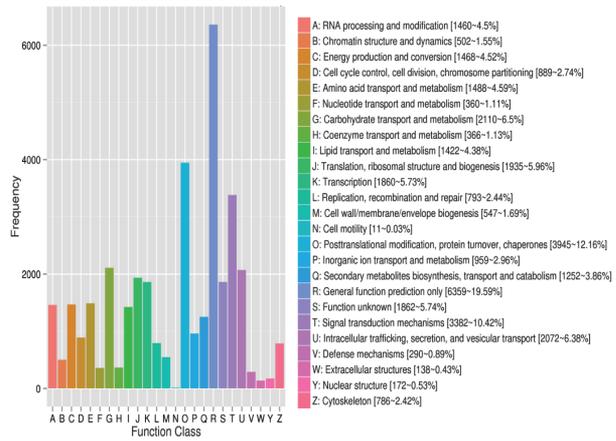


图3 大青Unigene的KOG功能分类
Figure 3 KOG functional classification of unigenes of *Clerodendron cyrtophyllum* Turez

细胞过程(26 146个，26.1%)，代谢过程(22 970个，22.93%)，生物调节(8 166个，8.15%)，定位(5 359个，5.35%)及刺激响应(4 681个，4.67%)。注释到分子功能中的Unigene数目为55 108个，占比为55%，其中注释到较多功能的依次为结合(25 094个，25.05%)，催化活性(22 737个，22.69%)及转运活性(2 509个，2.5%)。注释到细胞组分中的Unigene的数目为47 555个，占比为47.46%，其中注释到较多的功能依次为细胞结构物质(27 133个，27.08%)，细胞内物质(15 645个，15.62%)及蛋白质复合物(4 771个，4.76%)。见图4。

2.2.4 KEGG代谢通路分析 大青中主要含有萜类、黄酮类、酚酸类及氨基酸等药效活性成分，但大青缺少基因组信息作为参考，因此，对其基

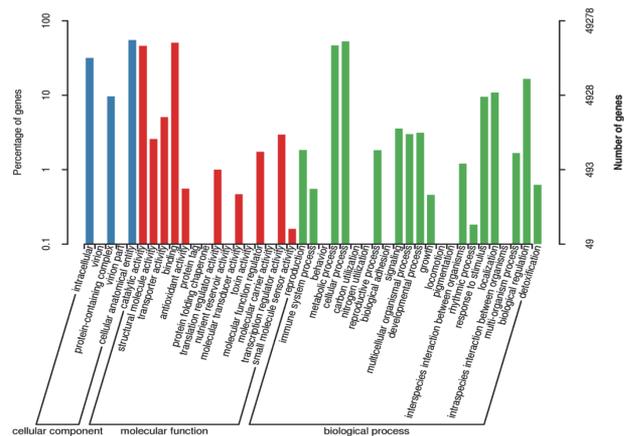


图4 大青Unigene的GO功能分类
Figure 4 GO functional classification of unigenes of *Clerodendron cyrtophyllum* Turez

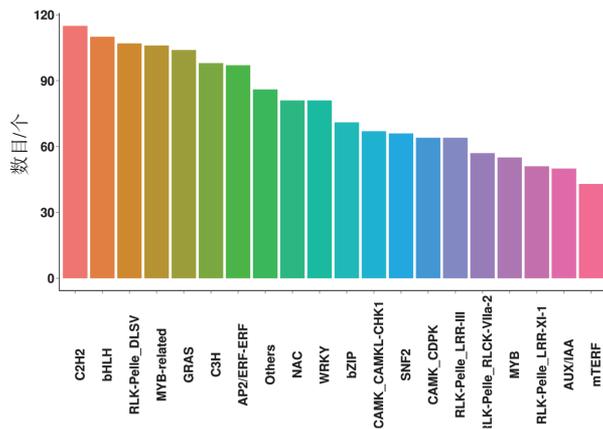


图6 大青转录因子分类

Figure 6 Classification of transcription factors of *Clerodendron cyrtophyllum* Turez

双碱基重复(2 195个, 25.4%), 三碱基重复(745个, 8.6%), 混合碱基重复(500个, 5.7%)和四碱基重复(42个, 0.4%); 最少的为五碱基重复和六碱基重复, 分别仅检测到5个和5个位点, 分别占0.06%和0.06%。根据8 640个SSR的位点和Unigene序列, 使用Premier 3.0设计出相应的SSR扩增引物, 为采用分子标记辅助育种技术的大青遗传改良、群体结构的研究提供了基础数据。见图7。

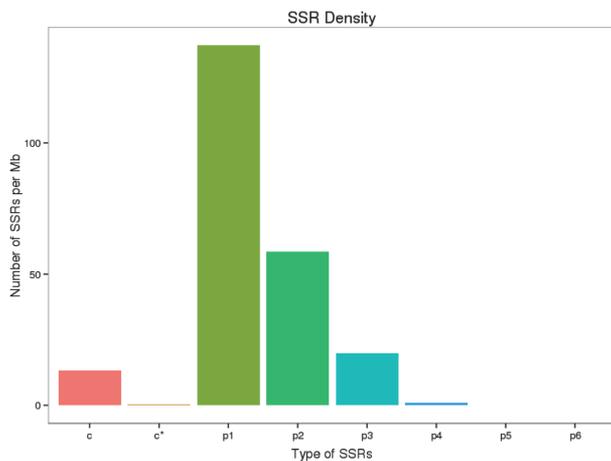


图7 大青Unigene的SSR位点分析

Figure 7 SSR locus analysis of unigenes of *Clerodendron cyrtophyllum* Turez

3 讨论

近年来, 本草基因组学、转录组学技术的快速发展使得更多药用植物的基因资源得以被保护及利用。通过转录组数据的分析研究, 可挖掘重要植物功能基因和代谢通路, 进一步构建中药指

纹图谱, 为天然药物来源新途径、种质资源鉴定、保存及良种选育提供分子基础^[6]; 其次, 通过对代谢通路关键基因的研究, 可为中药资源活性成分的生物合成与调控提供新的思路和方法; 第三, 通过转录水平的调节, 提高药用成分的产量与活性, 寻找最佳中药状态, 为重要的良种选育、规范种植、质量控制提供技术支撑^[17]。

目前, 对大青化学成分分析和植株药理作用的研究已有一定的经验, 但对大青的研究多限于整体水平实验, 还未见关于其基因组、转录组学的研究报道, 这制约了大青次生代谢产物的生物合成途径解析, 不能充分挖掘其新的药用价值以使其资源得以充分利用^[18-19]。本研究采用Illumina HiSeq 2500的高通量测序平台对大青的不同组织进行转录组测序并构建大青的Unigene库。通过*de novo*组装共获得100 191条Unigene, N50为1 055 bp, 长度大于1 000 bp的Unigene有18 421个, 占总序列数的18.39%。N50的长度是评价转录组组装质量的重要指标, 大青转录组组装的N50长度与茶树(1 081 bp)的长度相近, 但比中药黄芩(797.64 bp)要长, 说明大青的转录组组装序列质量较高, 能够满足后续数据分析的要求^[20-21]。本研究通过与多个公共数据库进行比对分析获得Unigene的功能注释, 结果表明, 大青组装的Unigene共有59 690个(59.58%), Unigene在至少1个数据库中得到注释, 其中以NR数据库中比对到的Unigene最多, 共有58 952个Unigene得到注释, 其结果显示与管状花目物种芝麻的相似度最高。这可能是由于马鞭草科大青属植物的种类少、研究基础薄弱导致大青转录组Unigene并未比对到同科属物种。另外, 还有40 501个(40.42%)Unigene没有在一个数据库中获得注释, 可能是由于组装的Unigene序列太短而缺乏保守区域, 或者数据库中缺乏大青基因组、转录组信息导致未能匹配得到注释。在KEGG数据库中共注释到38 260个Unigene参与136个代谢途径, 包括萜类、类胡萝卜素、生物碱类、黄酮类等次级代谢的生物合成途径。其中, 407个Unigene参与萜类化合物的生物合成, 231个Unigene参与黄酮类化合物的生物合成, 386个Unigene参与苯丙烷素的生物合成。这些数据为进一步解析大青中活性成分的生物合成途径提供了研究基础。SSR标记操作简单、重复性好。转录组

数据中包含着大量SSR位点,随着高通量测序技术的快速发展,利用这些位点开发的功能EST-SSR标记在群体遗传学及标记辅助育种研究中显得更加方便快捷。随着大量药用植物的转录组数据产生及功能基因得到注释,也必然为药用植物的功能EST-SSR标记开发奠定了基础^[22]。本研究利用MISA软件共检测到8 640个SSR位点,其中单碱基重复最为丰富(5 135个,59.4%),有很大一部分SSR位点的Unigene在各大数据库中得到功能注释,这也为大青后续功能EST-SSR分子标记开发提供了数据基础。

本研究对大青转录组进行了初步的探究,弥补了大青基因组信息的不足,为解析大青次级代谢物质合成通路及分子生物学方面的研究奠定了基础。大青SSR位点的发掘,可为其分子标记的开发、遗传多样性分析、种质资源鉴定与优选、分子标记育种等提供理论基础,为利用分子手段鉴定、区分大青品种及评价其质量提供了依据。

参考文献:

- [1] 宋立人. 中华本草[M]. 上海: 上海科学技术出版社, 1999: 5944-5945.
- [2] 薛漓, 饶伟文. 路边青的鉴别研究[J]. 中草药, 2004, 35(4): 455-457.
- [3] 李艳. 大青根化学成分的研究[D]. 沈阳: 沈阳药科大学, 2008.
- [4] 王鹏. 大青的化学成分研究[D]. 上海: 上海中医药大学, 2011.
- [5] 曾晓春, 吴映明, 凌敏, 等. 广东大青镇咳平喘祛痰作用的实验研究[J]. 中医药学刊, 2004, 22(5): 916-916.
- [6] 中国科学院中国植物志编辑委员会. 中国植物志: 第65(1)卷[M]. 北京: 科学出版社, 1998.
- [7] 王继华, 蔡时可, 梅瑜, 等. 溪黄草转录组测序及生物信息学分析[J]. 广州中医药大学学报, 2020, 37(8): 1568-1574.
- [8] YANG L, DING G, LIN H, et al. Transcriptome analysis of medicinal plant *Salvia miltiorrhiza* and identification of genes related to tanshinone biosynthesis[J]. PLoS One, 2013, 8(11): e80464.
- [9] ZHAO S, TUAN P A, LI X, et al. Identification of phenylpropanoid biosynthetic genes and phenylpropanoid accumulation by transcriptome analysis of *Lycium chinense* [J]. BMC Genomics, 2013, 14(1): 802.
- [10] GAO W, SUN H X, XIAO H, et al. Combining metabolomics and transcriptomics to characterize tanshinone biosynthesis in *Salvia miltiorrhiza*[J]. BMC Genomics, 2014, 15: 73.
- [11] ALEX V M, MICHELE F, JACOB P, et al. CathaCyc, a metabolic pathway database built from *Catharanthus roseus* RNA-Seq data[J]. Plant Cell Physiol, 2013, 54(5): 637-685.
- [12] YUAN Y, SONG L, LI M, et al. Genetic variation and metabolic pathway intricacy govern the active compound content and quality of the Chinese medicinal plant *Lonicera japonica* Thunb. [J]. BMC Genomics, 2012, 13: 195.
- [13] SUN P, SONG S H, ZHOU L L, et al. Transcriptome analysis reveals putative genes involved in iridoid biosynthesis in *Rehmannia glutinosa* [J]. Int J Mol Sci, 2012, 13(10): 13748-13763.
- [14] RAMILOWSKI J A, SAWAI S, SEKI H, et al. *Glycyrrhiza uralensis* transcriptome landscape and study of phytochemicals [J]. Plant Cell Physiol, 2013, 54(5): 697-710.
- [15] SHU S, CHEN B, ZHOU M, et al. *De novo* sequencing and transcriptome analysis of *Wolfiporia cocos* to reveal genes related to biosynthesis of triterpenoids [J]. PLoS One, 2013, 8(8): e71350.
- [16] 孙国祥, 胡玥珊, 张春玲, 等. 构建中药数字化指纹图谱研究[J]. 药物分析杂志, 2009(1): 160-169.
- [17] 张铁军. 中药质量认识与质量评价[J]. 中草药, 2011(1): 11-19.
- [18] 刘时乔, 周福军, 高贤, 等. 大青属常见药用植物的研究进展[J]. 药物评价研究, 2012, 34(6): 469-473.
- [19] 吴明美, 王利勤, 华燕. 大青属植物的化学成分及药理作用研究进展[J]. 化学工程与装备, 2011(4): 112-116.
- [20] 庞丹丹, 张亚真, 徐礼羿, 等. 基于转录组测序的茶树儿茶素合成调控相关基因的挖掘[J]. 分子植物育种, 2019, 17(17): 5604-5612.
- [21] 王德富, 成媛媛, 杨锋, 等. 黄芩高通量转录组测序数据组装和分析[J]. 山西农业科学, 2016, 44(8): 1065-1072.
- [22] 袁媛, 黄璐琦. 中药资源转录组分析操作指南[M]. 上海: 上海科学技术出版社, 2016.

【责任编辑: 侯丽颖】