草珊瑚叶片转录组特征分析及活性成分合成相关基因挖掘

蓝祖栽，胡　营，闫志刚，李　刚，韦筱媚，韦桂丽，马梅冰，唐美琼*

广西壮族自治区药用植物园 国家中医药传承创新中心 西南濒危药材资源开发国家工程研究中心 广西药用资源保护与遗传改良重点实验室/广西壮族自治区中药资源智慧创制工程研究中心，广西 南宁　530023

摘　要：目的　探索草珊瑚 *Sarcandra glabra* 叶片活性成分生物合成分子基础，合理的开发利用草珊瑚资源。**方法**　采用 BGISEQ-500 测序平台测定草珊瑚叶片的转录组，对经过滤和 Trinity 组装得到的 unigene 进行生物信息学分析。**结果**　共获得了 75 862 个 unigenes，平均长度 980 bp，其中 42 369 个 unigenes 在 7 大公共数据库中得到成功注释，占总基因数的 55.85%。21 801 个 unigenes 被注释于细胞组分、分子功能和生物学过程 3 大类共 55 个 GO 条目京都基因与基因组百科全书（Kyoto encyclopedia of genes and genomes，KEGG）富集到 2 条可能与药效物质形成密切相关的通路，分别是萜类和聚酮化合物的代谢和其他次生代谢物生物合成。通过进一步挖掘发现，其中 47 个 unigenes 可能参与萜类化合物骨架生物合成，14 个 unigenes 可能参与倍半萜类化合物生物合成，62 个 unigenes 可能参与黄酮类化合物生物合成。另外，转录组序列中共检测到 19 690 个 SSR 位点，其中二核苷酸重复类型所占比例最高，达到了 46.63%；针对所有的 SSR 位点，共设计出 11 209 对引物。**结论**　很好地扩展了草珊瑚的转录组信息，为进一步解析草珊瑚活性成分生物合成途径提供了基因资源，促进草珊瑚分子育种的研究。

关键词：草珊瑚；转录组；次生代谢；生物合成；SSR

中图分类号：R286.12　　**文献标志码：**A　　**文章编号：**0253 - 2670(2024)14 - 4860 - 11

DOI: 10.7501/j.issn.0253-2670.2024.14.022

Transcriptome characteristic and discovery of related genes involved in active ingredients biosynthesis of *Sarcandra glabra* leaves

LAN Zuzai, HU Ying, YAN Zhigang, LI gang,WEI Xiaomei, WEI Guili, MA Meibing, TANG Meiqiong

Guangxi Engineering Research Center of TCM Resource Intelligent Creation, Guangxi Key Laboratory of Medicinal Resources Protection and Genetic Improvement, National Engineering Research Center for Southwest Endangered Medicinal Materials Resources Development，National Center for TCM Inheritance and Innovation, Guangxi Botanical Garden of Medicinal Plants, Nanning 530023, China

Abstract: Objective　To explore the molecular basis of active ingredients biosynthesis in the leaves of *Sarcandra glabra*, and exploit this resource reasonably. **Methods**　*S. glabra* leaf was analyzed by transcriptome sequencing with the BGISEQ-500 platform. After filtration and the Trinity assembly, the unigenes were analyzed by bioinformatics. **Results**　A total of 75 862 unigenes with an average length of 980 bp were obtained. Among them, 42 369 unigenes, accounting for 55.85% of the total, were successfully annotated in seven public databases. A total of 21 801 unigenes were categorized into 55 GO terms among three categories: biological process, cellular component and molecular function. KEGG enrichment analyses shown that there were two pathways, namely “metabolism of terpenoids and polyketides” and “biosynthesis of other secondary metabolites”, may be closely related to the formation of pharmacodynamic basis. Moreover, it was discovered that 47 unigenes, 14 unigenes and 62 unigenes might participate in terpenoid backbone, sesquiterpenoids and flavonoid biosynthesis, respectively. In addition, a total of 19 690 SSR sites were detected in the transcriptome sequences, of which the highest proportion of di-nucleotides was 46.63%. A total of 11 209 primer pairs were designed for all SSR sites. **Conclusion**　This work extend the transcriptome information of *S. glabra*, which will provide gene resources for further study the biosynthesis pathway of active ingredients in *S. glabra* and facilitate molecular breeding in this species.

Key words: *Sarcandra glabra* (Thunb.) Nakai; transcriptome; secondary metabolism; biosynthesis; SSR

收稿日期：2024-02-03

基金项目：广西科技重大专项（桂科 AA22096021）；广西创新驱动发展专项（桂科 AA18242040）；广西药用植物保育人才小高地自主研究项目（GXYYXGD202201）

作者简介：蓝祖栽（1966—），男，研究员，研究方向为药用植物资源学和植物栽培学研究。E-mail: zuzai@126.com

***通信作者：**唐美琼（1984—），高级工程师，研究方向为药用植物遗传育种。E-mail: tangmeiqiong2006@163.com

草珊瑚 *Sarcandra glabra* (Thunb.) Nakai 为金粟兰科草珊瑚属多年生常绿半灌木植物，别名九节茶、九节风、接骨木等，药材名肿节风，具有清热凉血、活血消斑、祛风通络等功效，民间多用于跌打损伤和风湿关节炎的治疗[1]。草珊瑚最早以“接骨木”之名记载于唐代《本草拾遗》，明代《汝南圃史》中首次出现“九节茶”一名，后以九节茶之名当作药用开始载于清代《生草药性备要》，历版《中国药典》中以肿节风之名收录，草珊瑚一名出现于现代著作《云南植物志》中[2]。陈超志等[2]、余意等[3]先后通过名称、原植物、产地以及性味功效等对草珊瑚进行综合考证，一致认为历代本草书中对其名记载比较杂乱，但对植物形态描述以及跌打损伤、清凉热血方面的功效记载基本一致，为规范药材来源，建议名称以草珊瑚或九节茶为宜。现代药理研究表明，倍半萜类、香豆素类、黄酮类、有机酸类和挥发油类等化合物是草珊瑚重要的药效物质基础，其抗肿瘤、抗炎、抗氧化、消炎镇痛等多种药理活性已经得到了证实[4-5]。基于草珊瑚中生物活性多样的化合物及丰富的药理活性，现以草珊瑚为主要原料开发生产了复方草珊瑚含片、草珊瑚注射液、肿节风片、清热消炎宁胶囊等多种药品。可见，选育和培育这些活性成分（次生代谢物质）含量高的草珊瑚种质，对于草珊瑚产业的可持续发展极其重要。为此，围绕着草珊瑚种质资源遗传多样性研究方面，如随机扩展多态性（random amplified polymorphic，RAPD）[6]、简单序列重复区间扩增多态性分子标记（amplified fragment length polymorphism，AFLP）[7]、简单重复序列间扩展（inter-simple sequence repeat，ISSR）[8]等 DNA 分子标记技术已相继开展。近年来，开始有关于草珊瑚转录组学[9]、代谢组学[10]、叶绿体基因组测序[11]等相关的研究报道，但由于基因组信息的缺乏，草珊瑚功能基因的挖掘及更深层次的研究尤为薄弱，严重制约了草珊瑚分子育种和遗传改良等工作的开展。

基因通过转录和翻译代谢酶来调控化合物的合成，次生物质的合成代谢途径上涉及大量关键酶基因，而转录组是连接基因组遗传信息和蛋白组生物功能的中间纽带，可以更加全面的揭示生物个体特定时期或特定组织中全局基因的表达情况，是研究基因表达的重要手段，被越来越多的应用于药用植物次生代谢物质的合成及调控机制研究[12]。此外，基于转录组数据的筛选可以获得大量的简单重复序列（simple sequence repeat，SSR）、单核苷酸多态性（single nuceotide polymorphism，SNP）和插入缺失标记（insertion-delection，lnDel）等分子标记位点信息，这些位点可能与特定功能的基因紧密连锁，更易与表型性状相联系，可进一步应用于植物的品种鉴定、种群关系及分子育种设计等方面[13-15]。SSR 因具有数量多、易检测、多样性信息量大、共显性遗传等特点而成为群体遗传学、遗传育种和系统进化等研究的有效手段。随着测序技术的发展，基于转录组数据筛选 SSR 位点在动植物中也逐步被应用[16-17]。但迄今为止，草珊瑚中有关这方面的研究还鲜有报道。

叶片是草珊瑚的重要药用部位，主治骨折[18]。研究表明，黄酮类物质和迷迭香酸主要集中在草珊瑚的叶片部位[19-20]，Wang 等[21]最新的研究报道了从叶片中新分离获得 6 种新的倍半萜类化合物，这些研究为叶片药用价值的进一步开发提供了重要信息，因此，有必要进一步加强对草珊瑚叶片部位的遗传基础研究。车苏容等[21]对五年生草珊瑚根和叶片进行转录组比较分析，发现差异显著的关键基因显著富集于苯丙烷类、萜类及类黄酮等次生代谢途径；Xu 等[9]对幼苗期草珊瑚和海南草珊瑚叶片进行转录组比较分析，发现谷胱甘肽 S 转移酶（glutathione S-transferase，GST）的最高功能同源基因数目最多。这些研究表明，转录组在草珊瑚功能基因的挖掘中具有可行性，但与其药用次生代谢合成调控相关的基因仍需要进一步挖掘。因此，本研究采用高通量测序技术探讨草珊瑚叶片转录组的特征，以期获得更多与黄酮类和萜类代谢途径相关的信息，进一步丰富草珊瑚次生代谢产物生物合成的基因资源，同时为草珊瑚资源的合理采集、精准利用及分子育种设计研究提供参考和方向。

1 材料与仪器

1.1 材料

草珊瑚样品取自广西壮族自治区药用植物园，经韦筱媚高级工程师鉴定为金粟兰科植物草珊瑚 *S. glabra* (Thunb.) Nakai，选取 3 株三年生单株草珊瑚叶片混合后立即用液氮冷冻并储存于−80 ℃冰箱备用。

1.2 仪器

BGIseq-500 型测序仪（深圳华大基因股份科技有限公司），Centrifuge 5430 R 型高速冷冻离心机（Eppendorf 公司，德国），Agilent 2100 Bioanalyzer 和 Nanodrop 2000 型紫外分光光度计（Agilent 公司，美国），T100 型 PCR 仪器（BIO-RAD 公司，美国）。

2 方法

2.1 RNA 提取与检测

草珊瑚叶片 RNA 根据试剂盒（EastepTM Super Total RNA Extraction Kit，批号 LS1040）说明书进行提取。RNA 的浓度和纯度通过 Nanodrop 2000 检测，完整性通过琼脂糖凝胶电泳评估。纯度、浓度和完整性经检测合格后的 RNA 用于转录组测序。

2.2 转录组测序

首先用带有 Oligod T 的磁珠富集有 polyA 尾巴 mRNA 的方法对 total RNA 进行处理；接着用打断 buffer 把获得的 RNA 进行片段化处理，并用随机的 N6 引物进行反转录后再合成 cDNA 二链，形成双链 DNA；随后把合成的双链 DNA 末端补平，并进行 5’端磷酸化、3’端形成突出 1 个“A”的粘末端，再连接 1 个 3’端有凸出“T”的鼓泡状的接头；连接产物通过特异的引物进行 PCR 扩增；PCR 产物热变性成单链，再用一段桥式引物将单链 DNA 环化得到单链环状 DNA 文库；上机测序。转录组测序工作委托深圳华大基因科技有限公司进行，采用 BGISEQ-500 平台进行高通量测序。

2.3 转录组数据组装及基因功能注释与分类

测序得到的原始数据经过滤处理获得 clean reads 后，先采用 Trinity（v2.5.1）软件对 clean reads 进行组装，再使用 Tgicl 软件将组装的转录本进行聚类去冗余后得到 unigene。分别使用 Blastn（v2.2.23）软件对 unigene 进行核酸序列数据库（nucleotide sequence database，Nt）注释，Blastx（v0.8.31）软件对 unigene 进行非冗余蛋白库（non-redundant protein sequence database，Nr）、真核生物蛋白相邻类的聚簇（clusters of orthologous groups for eukaryotic complete genomes，KOG）、京都基因与基因组百科全书（Kyoto encylopedia of genes and genomes，KEGG）以及 Swiss-prot 蛋白质序列数据库（Swiss-prot protein sequence database，Swiss-prot）注释，Blast2GO（v2.5.0）软件以及 Nr 注释结果进行基因本体论（gene ontology，GO）注释，并对注释到 GO 3 个方面[生物学过程（biological process，BP）、细胞组分（cellular component，CC）和分子功能（molecular function，MF）]的下一级分类注释的基因和，KEGG 注释下的相关基因所参与的代谢通路进行分析。

2.4 基因表达量的计算

使用 Bowite2（v2.2.5）软件将 clean read 比对到 unigenes，然后使用 RSEM（v1.2.12）软件计算样品的基因表达量，最终以 FPKM（Fragments per kilobase of exon model per million mapped fragments）形式输出。

2.5 转录组 SSR 位点的检测

使用 MISA（v1.0）软件对 unigene 进行检测，检测的标准：单碱基重复数≥12，双碱基重复数≥6，三碱基重复数≥5，以此类推，重复单元最多有 6 个碱基，2 个微卫星之间的距离小于 100 bp 时组成一个复合微卫星。然后使用 Primer3（v2.2.2）软件对检测到的 SSR 进行引物设计，对得到的引物通过以下方式进行筛选：（1）引物不存在 SSR；（2）将获得的引物比对到 uigene 序列，引物的 5’端允许有 3 个碱基的错配，3’端允许有 1 个碱基的错配；（3）去掉比对到不同的 unigene 上的引物，筛选唯一匹配的引物；（4）使用 SSRFinder 校验 SSR。最后使用产物序列来寻找 SSR，并筛选出与 MISA 检验结果相同的 SSR 产物。采用 Excel 软件对 SSR 各类型数量、比例和序列分布等进行统计分析。

3 结果与分析

3.1 转录组测序与组装

通过 BGISEQ-500 测序平台对草珊瑚叶片 RNA 进行测序，共得到 49 079 576 个 raw reads（SRA：SRR22671437），经去除接头污染、未知碱基 N 含量大于 5%以及低质量的 reads 后，最终获得 43 967 184 个 clean reads，总核苷酸数目约为 6.6 Gb，Q_{20}、Q_{30} 值分别为 97.06%和 89.09%。经过 *De novo* 拼接后，得到 contig 共有 130 947 个，总长度约为 96 Mb，以长度为 200～1 500 bp 的数量最多；最终获得的 unigene 数目为 75 862 个，总长度约为 74 Mb，GC 量为 41.59%（表 1）。对 unigene 的长度分布特征进行分析后发现，在总 unigene 中，所占比例最大为 300 bp，共 21 695 个（28.6%）。而 1 000～2 000 bp 的有 16 656 个，约占 22%；大于 2 000 bp 的有 10 125 个，约占 13.3%（图 1）。总体来说，本研究中的转录组测序和组装效果较理想，unigenes 的整体长度分布均匀，符合后续分析要求。

3.2 基因功能注释与分类

为获得草珊瑚叶片转录本基因的功能注释信息，将组装得到的 75 862 个 unigenes 基因分别在 7 大数据库：Nr、Nt、Swiss-prot、KEGG、KOG、蛋白质家族（Protein Family，Pfam）和 GO 中进行比对。结果表明，共有 42 369 个 unigenes 基因至少在

表 1 RNA 测序数据和质量控制信息

Table 1 RNA sequencing data and corresponding quality control information

序列种类	200～1 500 bp	＞1 500～3 000 bp	＞3 000 bp	N50	N90	GC/%	总长度/bp	平均长度/bp
contigs	112 801	13 812	4334	1466	254	41.24	94 957 902	725
unigenes	59 780	12 012	4070	1747	362	41.59	74 370 453	980

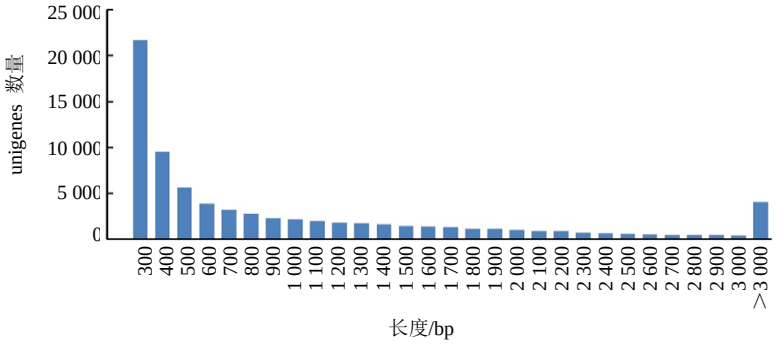


图 1 Unigene 的长度分布

Fig. 1 Length distribution of unigenes

一个数据库中获得注释，占总数的 55.85%；在 7 大数据库中均获得注释的 unigene 共有 10 966 个，占总数的 14.46%；Nr 数据库得到成功注释的基因有 39 416 个，获得的信息最多；GO 数据库仅有 21 801 个基因得到成功注释，获得的信息最少（表 2）。

基于 GO 分析，对所得到的 unigene 的潜在功能进行分类。21 801 个 unigenes 分为 BP、CC 和 MF 3 大类，55 个 GO 条目，其富集 unigene 数量从高到低排列前 30 的条目已在图 2 中列出。在生物学过程方面，可分成 26 类，参与细胞过程（cellular

表 2 Unigene 功能注释结果

Table 2 Results of unigenes annotation

数据库	Nr	Nt	Swissport	KEGG	KOG	Pfam	GO	Intersection	Overall
基因数	39 416	26 594	27 104	30 881	29 937	26 319	21 801	10 966	42 369
占比/%	51.96	35.06	35.73	40.71	39.46	34.69	28.745	14.46	55.85

Intersection 表示被 7 大数据库均注释上的 unigene，Overall 表示被七大数据库中任意一个数据库注释上的 unigene。

Intersection represents successful annotated in all of seven databases, Overall represents successful annotated at least in one of seven databases.

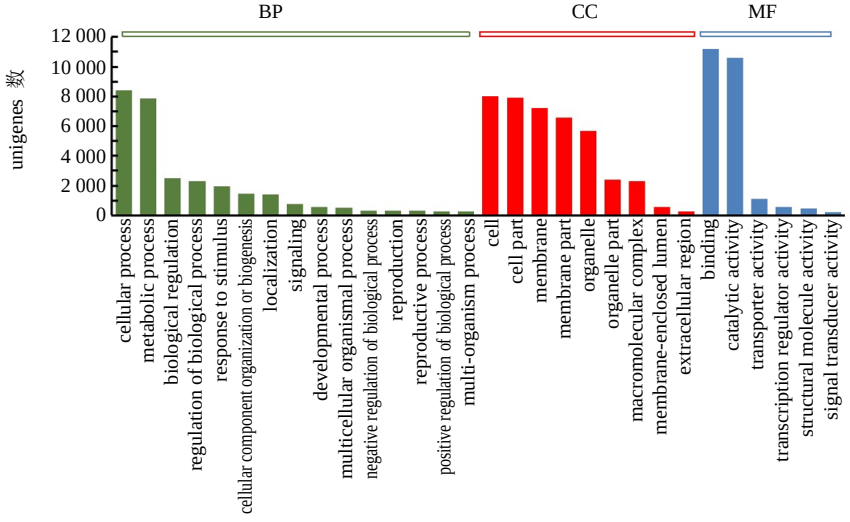


图 2 草珊瑚 unigene 的 GO 功能分类

Fig. 2 Functional classification of *S. glabra* unigene by GO enrichment analysis

process）和代谢过程（metabolic process）的数量最多，分别为 8 407 和 7 880 个；在细胞组分中，共分为 16 类，较多的 unigene 参与细胞（cell）、细胞器（cell part）和细胞膜（membrane），数量均达到 7 000 个以上；在分子功能方面，共注释到 13 类，主要富集于结合（binding）和催化因子（catalytic activity），数量分别为 11 201 个和 10 593 个。

KEGG 功能注释结果表明，共有 32 311 个 unigenes 被注释到 138 条 KEGG 途径中，涉及到代谢（metabolism）、细胞过程（cellular process）、环境信息处理（environment information process）、遗传信息处理（genetic information process）和生物系统（organismal systems）5 个大类（图 3）。代谢大类中共有 11 个通路，其中富集在全局和总览图（global and overview maps）通路中的 unigene 数量最多，有 7 301 个；其次为碳水化合物代谢（carbohydrate metabolism）、翻译（translation）和折叠，分类和降解（folding, sorting and degradation）等通路，分别有 3 375、2 719、2 331 个 unigenes。

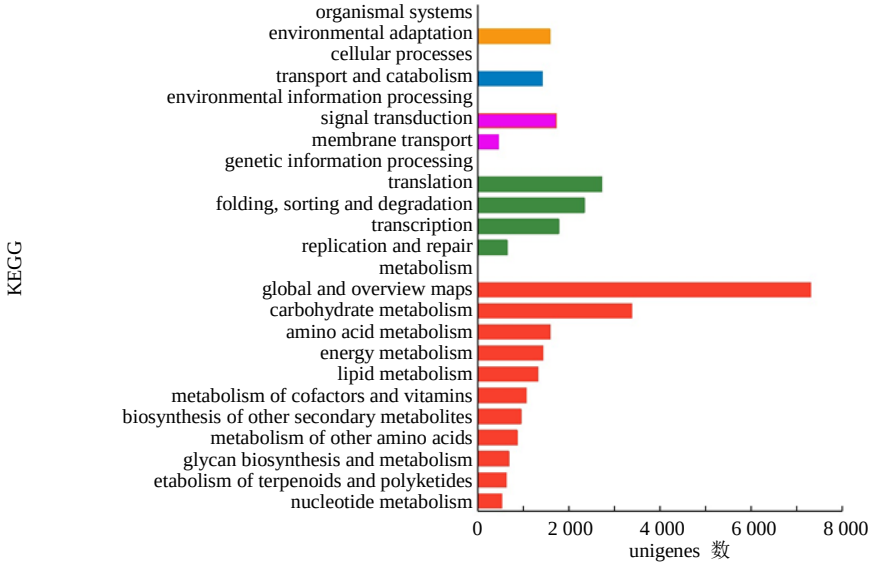


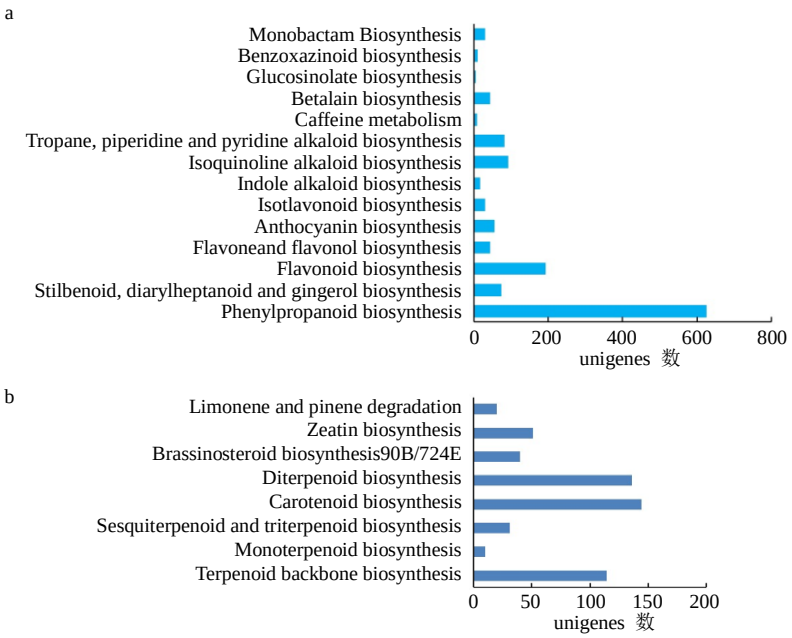
图 3　草珊瑚 unigene 的 KEGG 富集分析

Fig. 3　KEGG pathway enrichment analysis of *S. glabra* unigene

基于草珊瑚叶片具有药用作用，本研究进一步对可能与药效物质基础形成密切相关的其他次生代谢产物生物合成（biosynthesis of other secondary metabolites）和萜类和聚酮化合物的代谢（metabolism of terpenoids and polyketides）2 条通路在 Level 3 层级上进行了 KEGG 分析。在其他次生代谢生物合成途径中共有 14 个通路，包含苯丙烷、黄酮、花青素等多种具有药用作用的次生代谢物质，其中富集在苯丙烷生物合成（phenylpropanoid biosynthesis）的 unigene 数量最多，有 626 个；其次为类黄酮生物合成（flavonoid biosynthesis），有 193 个（图 4-a）。在萜类和聚酮化合物的代谢途径中共有 8 个通路，全部是萜类相关物质的合成途径，包括单萜类、倍半萜和三萜、二萜类和类胡萝卜素等多种类型的萜类物质生物合成途径，其中富集在类胡萝卜素生物合成（carotenoid biosynthesis）上的 unigene 数量最多，有 144 个；其次为二萜类生物合成（diterpenoid biosynthesis），有 136 个（图 4-b）。

3.3　活性成分合成相关基因挖掘

3.3.1　萜类合成途径相关基因分析　萜类化合物是广泛存在于植物中的一类代谢产物，在植物的生长发育、环境适应和防御反应等方面发挥着重要作用，药用植物中已有不少萜类化合物被开发应用，如紫杉醇、青蒿素，说明萜类化合物在药物开发中具有广阔的前景。目前，已经在植物中明确萜类化合物的合成大体上分为 3 个阶段：中间体异戊烯基焦磷酸及其双键异构体二甲基烯丙基焦磷酸的生成—直接前体物质的生成—萜类生成及其修饰，而萜类合酶（terpene synthase）、脱氧-*D*-木酮糖-5-磷酸还原异构酶（1-deoxy-*D*-xylulose-5-phosphate reductoisomerase）、法呢基焦磷酸合酶（farnesyl pyrophosphate synthase）等在萜类物质合成过程中扮演着非常重要的作用[22]。萜类化合物也是草珊瑚的重要化合物之一，根据 KEGG 富集分析结果



a-其他次生代谢产物合成途径；b-萜类和聚酮化合物的代谢。

a-biosynthesis of other secondary metabolites; b-metabolism of terpenoids and polyketides.

图 4　与药效物质形成密切相关的通路

Fig. 4　Analysis of KEGG pathway related to pharmacodynamic basis formation

（图 3、4-b），在萜类和聚酮化合物的代谢途径中，共涉及到萜类骨架、单萜类、二萜类、油菜素类固醇、类胡萝卜素、玉米素以及倍半萜和三萜生物等合成途径。结合 unigene 注释结果，经过进一步统计，共鉴定到 61 个 unigenes 可能参与萜类骨架及倍半萜生物合成相关途径（表 3），其中与萜类骨架合成相关的 47 个，与倍半萜类合成相关的 14 个。在本研究中，鉴定到的脱氧-*D*-木酮糖-5-磷酸还原异构酶 unigene 数目最多，共 9 个（Unigene664_00902、Unigene749_00902、Unigene1170_00902、Unigene1659_00902、Unigene705_00902、Unigene421_00902、Unigene1792_00902）。直接前体物质生成之后，由萜类合酶催化相应的前体物质形成各类萜类化合物[22]，本研究中共鉴定到 3 个参与萜类骨架生物合成通路的 unigenes（Unigene11377_00902、Unigene45924_00902、Unigene34117_00902）。在草珊瑚已知的萜类物质中，目前以倍半萜化合物最多，主要有金粟兰内酯、草珊瑚内酯、银线草醇等[5]。在倍半萜和三萜合成途径中，鲨烯合酶（squalene synthase，SQS）和萜类环化酶（terpenoidcyclases，TCS）是 2 个重要的酶[23]，本研究中鉴定到 1 个鲨烯合酶 unigene（Unigene23860_00902）。这些信息为今后开展草珊瑚萜类物质生物合成相关基因的克隆提供了候选资源。

3.3.2　黄酮类物质生物合成途径分析　黄酮类物质是草珊瑚重要的活性物质基础，目前已经从草珊瑚中分离出 40 余种黄酮类化合物，包括山柰酚、槲皮素、芦丁、金丝桃苷、新落妇苷等[5]。基于本转录组测序结果，根据 KEGG 富集分析及基因注释的相关分析，本研究共筛选出 62 个 unigenes 可能参草珊瑚叶片重要药效成分黄酮类物质的生物合成。结合其它物种中已有的研究，本研究对草珊瑚黄酮类物质生物合成途径进行了初步推导，如图 5 所示，由苯丙烷生物合成途径产生的类黄酮合成途径的重要前体物质对香豆酰辅酶 A（*p*-coumaroyl-CoA），首先在查耳酮合酶（chalcone synthase）的作用下将对香豆酰辅酶 A 转化为柚皮素查耳酮（naringenin chalcone），再由查耳酮异构酶（chalcone isomerase）转化成柚皮素（naringein），进一步经柚皮素 3-双加氧酶（naringenin 3-dioxygenase）作用后生成香橙素（dihydrokaempferol），再经黄酮合酶（flavonol synthase）催化后得到山柰酚（kaempferol），后在不同的糖基转移酶作用下得到含槲皮素在内的多种不同产物。其中，山柰酚经类黄酮 3’,5’-羟化酶（flavonoid 3’,5’-hydroxylase）催化后生成槲皮素（quercetin），再经过不同的酶作用后，最终产生不

表 3 萜类骨架及倍半萜合成的相关 unigenes

Table 3 Unigenes related to terpenoid backbone and sesquiterpenoids biosynthesis

骨架	名称	数量	FPKM（Fragments Per）
萜类骨架	dehydrodolichyl diphosphate synthase	2	1.60～5.85
	terpene synthase	3	0.92～1.13
	(*E*)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase	7	0.74～111.33
	hydroxymethylglutaryl-CoA synthase	1	50.72
	diphosphomevalonate decarboxylase	1	10.41
	1-deoxy-*D*-xylulose-5-phosphate reductoisomerase	3	0.88～1.82
	1-deoxy-*D*-xylulose-5-phosphate synthase	4	0.36～89.65
	protein farnesyltransferase subunit beta	2	1.33～5.74
	geranyl diphosphate synthase	3	0.36～2.39
	protein-S-isoprenylcysteine *O*-methyltransferase	1	11.30
	NAD^+-dependent farnesol dehydrogenase	3	5.83～17.64
	2-*C*-methyl-*D*-erythritol 4-phosphate cytidylyltransferase	2	2.77～140.76
	acetyl-CoA C-acetyltransferase	1	1.33
	geranylgeranyl pyrophosphate synthase 11	2	45.91～129.16
	prenylcysteine oxidase/farnesylcysteine lyase	1	49.89
	solanesyl diphosphate synthase	2	1.58～18.26
	isoprene synthase	1	1.00
	isopentenyl-diphosphate delta-isomerase	1	16.09
	phosphomevalonate kinase	1	14.97
	farnesyl pyrophosphate synthase	1	9.22
	tricyclene synthase EBOS	1	0.26
	farnesol kinase	1	9.73
	myrcene synthase	1	1.00
倍半萜	(3*S*,6*E*)-nerolidol synthase	5	1.65～21.16
	alpha-humulene synthase-like isoform X1	4	0.65～14.01
	thioredoxin 1	1	1.07
	valencene synthase-like	1	1.60
	squalene synthase	1	100.92
	valencene/7-*epi*-alpha-selinene synthase	1	0.74
	hypothetical protein	1	17.64

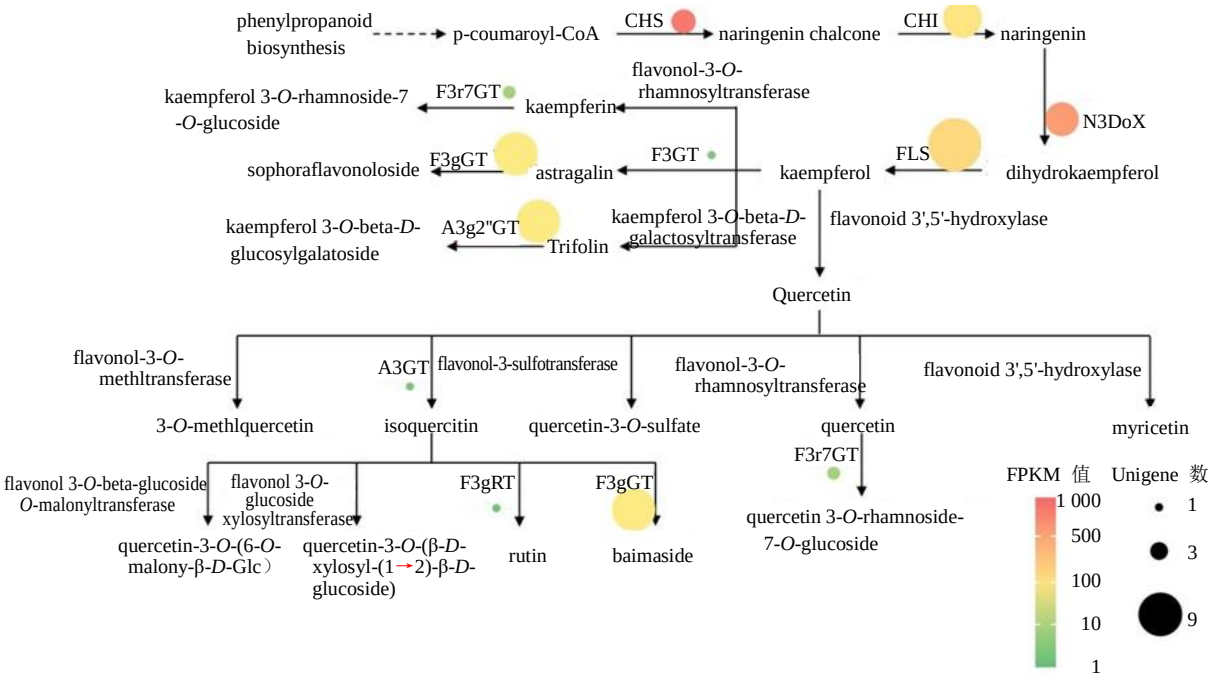
同的产物。山柰酚、槲皮素和芦丁具有抗炎、抗氧化、清热解毒等功效，是草珊瑚重要的药效基础。据此推测，对香豆酰辅酶 A 同样也是草珊瑚叶片中山柰酚、槲皮素和芦丁等物质生物合成的重要底物。

3.4 Unigene 的 TFs 预测

转录因子（transcription factor，TF）在植物的生长发育、次生代谢产物合成及响应外界环境等过程中起着重要的调控作用。根据草珊瑚转录组组装结果，本研究通过对具有转录因子能力的 unigene 进行了预测，结果显示，共 1 508 个 unigenes 被注释为转录因子，它们分布于 56 个转录因子家族。按基因数目由高至低的顺序，排列前 20 的转录因子家族如图 6 所示，排列前 3 的基因家族为 MYB、MYB-related 和 bHLH，分别有共 205、155、92 个转录因子。

3.5 SSR 位点分析

3.5.1 草珊瑚转录组中 SSR 位点的数量与分布根据草珊瑚叶片转录组测序组装结果，经检测后共得到 19 690 个 SSR 位点，长度位于 12～76 bp。SSR 类型丰富，单核苷酸～六核苷酸重复单元类型均存在，其中二核苷酸为优势重复类型，共 9 182 个位点，约占总数的 46.63%；依次为单核苷酸类型，共 4 882 个位点，约占总数的 24.79%、三核苷酸类型，共 4 557 个位点，约占总数的 23.14%；最少的为四



CHS-查耳酮合酶；CHI-查耳酮异构酶；N3DoX-柚皮素 3-加氧酶；FLS-黄酮醇合酶；F3GT-黄酮醇 3-*O*-葡糖基转移酶；F3r7GT-黄酮醇-3-*O*-*L*-鼠李糖苷-7-*O*-葡糖基转移酶；F3gGT-黄酮醇-3-*O*-葡萄糖苷/半乳糖苷葡糖基转移酶；A3g2"GT-花青素 3-*O*-葡萄糖苷 2"-*O*-葡糖基转移酶；A3GT-花青素 3-*O*-葡糖基转移酶；F3gRT-黄酮醇-3-*O*-葡萄糖苷 *L*-鼠李糖基转移酶；红色字体部分为本研究中注释得到的 unigene，黑色加粗字体为从草珊瑚叶片中已经鉴定出的黄酮类物质；气泡的颜色代表 FPKM 值（选择 1 个 FPKM 值最大的 unigene 作图），气泡的大小代表 unigene 的数量。

CHS-chalcone synthase; CHI-chalcone isomerase; N3DoX-naringenin 3-dioxygenase; FLS-flavonol synthase; F3GT-flavonol 3-*O*-glucosyltransferase; F3r7GT-flavonol-3-*O*-*L*-rhamnoside-7-*O*-glucosyltransferase; F3gGT-flavonol-3-*O*-glucoside/galactoside glucosyltransferase; A3g2"GT-anthocyanidin 3-*O*-glucoside 2"-*O*-glucosyltransferase; A3GT-anthocyanidin 3-*O*-glucosyltransferase; F3gRT-flavonol-3-*O*-glucoside *L*-rhamnosyltransferase; The red words represents the unigenes annotated in this study, the black bold words represents flavonoids that have been identified from *S. glabra* leaves, the color of the bubbles represents FPKM value (the unigene with the highest FPKM value was used to plot), the size of the bubbles represents unigene number.

图 5　草珊瑚黄酮类物质生物合成途径

Fig. 5　Flavone biosynthesis pathway in *S. glabra*

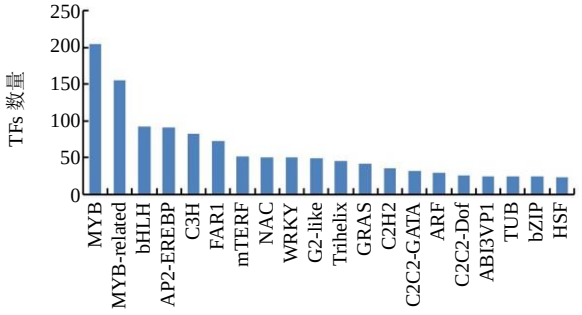


图 6　转录因子家族的 TF 数量

Fig. 6　TF number of each TF family

核苷酸类型，共 227 个位点，约占总数的 1.15%（表 4）。进一步对重复基序类型进行统计发现，出现频率最高的 5 种基序依次为：AG/CT、A/T、AT/AT、AAG/CTT、ATC/ATG（图 7）。这些 SSR 位点的组成特征和分布，可为进一步开发草珊瑚的微卫星分子标记、遗传多样性分析和亲缘关系鉴定奠定了基础。

3.5.2　SSR 引物设计与筛选　利用 Primer 3 对 19 690 个 SSR 位点的 unigenes 设计引物，通过去除不符合条件的引物，以及将产物序列利用 SSRFinder 校检后，最终筛选出 11 209 对引物（表 5）。

4　讨论

草珊瑚是多年生药材，在人工栽培条件下，一般种植三年后采收的草珊瑚药材质量较好[5]。草珊瑚叶片富含黄酮类、萜类及迷迭香酸等化合物，而根中富含异喹皮啶类物质，可根据需要对草珊瑚资源进行合理的采集，如需以抗菌、消炎和抗氧化为目的时，应以地上部为主；如以抗癌为目的时，应以地下部为主[24]。因此，揭示草珊瑚特定药用部位活性成分积累的相关信息，有利于提升其根、茎和叶资源的精准利用。但迄今为止，鲜有关于草珊瑚次生代谢物质生物合成与调控相关的报道。

表 4　草珊瑚转录组中 SSR 重复基序分布

Fig. 4　Distribution of SSR repeat motif in *S. glabra* transcriptome

重复基序数目	SSR 位点数/个						总数量
	单核苷酸	二核苷酸	三核苷酸	四核苷酸	五核苷酸	六核苷酸	
4	0	0	0	0	234	416	650
5	0	0	2 152	161	54	60	2 427
6	0	2 004	953	50	12	41	3 060
7	0	1 286	594	7	5	16	1 908
8	0	996	482	4	1	1	1 484
9	0	781	98	3	0	1	883
10	0	500	110	0	0	0	610
11	0	629	65	0	0	1	695
12	1 334	729	44	2	0	0	2 109
13	856	214	16	0	0	0	1 086
14	668	233	17	0	0	0	918
15	505	134	181	0	0	0	650
16	422	170	1	0	0	0	600
17	281	170	3	0	0	0	452
18	200	186	0	0	0	0	389
19	163	160	0	0	0	0	323
20	126	133	0	0	0	0	259
＞20	191	612	3	0	0	0	806
总数量	4 882	9 182	4 557	227	306	536	19 690

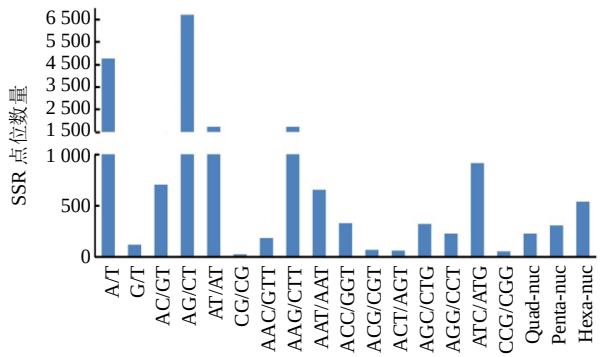


图 7　SSR 中核苷酸基序分布

Fig. 7　Distribution of nucleotide motif in SSR

药用植物相关数据是众多环境因素和遗传因素构成复杂网络[25]，药用植物的活性成分多为次生代谢产物，涉及到大量的合成通路及基因表达调控过程，通过转录组测序分析，可以快速全面的了解基因的表达信息，为深入研究有效活性成分的生物合成途径及形成的分子机制提供基础数据[12]。本研究通过对草珊瑚叶片进行转录组测序，共获得了 43 967 184 个 clean reads，经组装后得到 75 862 个 unigenes，进一步对基因功能进行注释分类发现，大量 unigenes 富集在细胞过程、代谢过程、细胞、细胞器、细胞膜、结合因子和催化因子活性等 GO 条目，以及全局和总览图、碳水化合物代谢、翻译、折叠，分类和降解等 KEGG 途径。这与车苏容等[26]对草珊瑚叶和根进行转录组测序的结果相类似。本研究进一步对萜类和聚酮化合物的代谢和其它次生代谢生物合成途径进行分析发现，大量的 unigene 涉及到萜类骨架、单萜类、二萜类、倍半萜和三萜等萜类物质以及类黄酮、花青素、异黄酮、黄酮和黄酮醇等黄酮类物质的生物合成途径，这暗示着草珊瑚叶片中富含萜类和黄酮类成分，这与先前对草珊瑚不同部位化学成分和活性物质成分研究结果一致[4]。根据其他物种中对黄酮类物质生物合成途径的相关报道结合本测序分析结果，首次对草珊瑚叶片中的黄酮类物质山柰酚、槲皮素和芦丁生物合成途径进行了推导，这不仅为草珊瑚萜类和黄酮类化合物的生物合成途径研究奠定基础，也为草珊瑚的遗传育种提供了重要的基因资源。下一步可结合 qRT-PCR 技术、蛋白组学或代谢组学等，对不同类型、不同生长年限、不同部位草珊瑚进行差异性比较分析，筛选出萜类、黄酮类活性物质生物合成和积累途径中的关键基因。

微卫星标记在遗传学、生态学和分类学等研究中起着重要的作用，近年来，随着测序成本的降低及测序技术的发展，基于转录组数据筛选微卫星标

表 5 草珊瑚 SSR 引物序列（部分引物序列示例）

Table 5 SSR primer sequences in *S. glabra* (example of partial primer sequences)

重复基元及类型	上游引物（5’-3’）	下游引物（5’-3’）	产物长度/bp
TC(2*10)	TACGCAGATTCTGAGGCTTTTAG	AACAATCCCTACCATTCCATTTT	138
TGA(3*12)	ACATTTGGACTTCGGATTTGTAA	ATTTGGGATCCGCTTATAATTTG	143
GAT(3*7)	TACAAATTATAAGCGGATCCCAA	GCCTCAAATAGATGGGACAAGTA	116
AG(2*12)	TAAGAGACGATTTGACTTTCCGA	GCAATAACTCCCACAAACAAGAC	98
TTTGA(5*5)	AGATCGCATTGGAACAGACTAGA	GACAACATAAATTAACAGCCCCA	109
TC(2*11)	CTCCATTGAAGAGAGATTGGTTG	CTGTCGACCCTATTGATTTTCAG	142
AAAT(4*5)	AAATAACGCAATTAGAGGGAAAAA	TTTTGTCTTTTCCGATAGATTTGTA	160
ATC(3*6)	TCGTAAATTGTATGGGCAGTTCT	CCCAAATAGATGGGATAAGGCTA	128
AG(2*11)	TTTGATACCAACTTGATTTTCGG	GCCAATAGGGACTCTAACCCTAA	130
CCA(3*5)	GAACTGCAAACTGTGTACCAGAA	CAATAGATGTGACACGAAGGACA	154
TCT(3*7)	CTCTTCTTTGGCTTTCTTTGAGA	TGATAGAGGGTGTGAACTGTGAA	147
CGTT(4*5)	GATCCCTTTGGAAGATTGAAGAT	AGGTTCAATCCCAGCTTACTGAT	148
AG(2*12)	TACTGCATCACAAGAAAAACCAA	TGGATGTTAGTTGGAGATGTTCC	112
TC(2*20)	TGTATTTCCATCTCCAGAGTACGA	AGTGCTCTCATTCAATGGTTGTT	142

记在大麦[27]、花生[28]、棉花[29]等植物中均有报道。本研究基于转录组测序首次对草珊瑚叶片中的 SSR 位点分布特征进行了分析，从草珊瑚叶片中共获得了 19 690 个 SSR 位点，且类型丰富，存在单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸六种类型的重复基序，并以二核苷酸重复基序为优势基序，这与 Xu 等[9]的研究结果存在差异，推测其原因可能与材料的生长期有关，本研究仅取材为三年生草珊瑚叶片，而 Xu 等[9]的研究是幼苗期的草珊瑚叶片进行分析的结果。

综上，本研究获得了草珊瑚叶片的转录组信息，为草珊瑚叶片的药用价值分析提供了重要信息，也进一步丰富了草珊瑚的分子标记资源，为草珊瑚的分子育种和功能基因挖掘奠定了基础。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 罗育, 黄春喜, 吴耀生, 等. 3 种 DNA 分子标记法联合鉴别草珊瑚及其混伪品 [J]. 中草药, 2020, 51(3): 733-740.

[2] 陈超志, 李书渊. 九节茶的本草考证 [J]. 中药材, 2015, 38(12): 2628-2631.

[3] 余意, 车苏容, 卢伟, 等. 闽产道地药材草珊瑚的本草考证 [J]. 安徽农学通报, 2019, 25(12): 42-44.

[4] 徐艳琴, 刘小丽, 黄小方, 等. 草珊瑚的研究现状与展望 [J]. 中草药, 2011, 42(12): 2552-2559.

[5] Zeng Y L, Liu J Y, Zhang Q, *et al*. The traditional uses, phytochemistry and pharmacology of *Sarcandra glabra* (Thunb.) Nakai, a Chinese herb with potential for development: Review [J]. *Front Pharmacol*, 2021, 12: 652926.

[6] 张志勇, 何平. 药用植物草珊瑚 RAPD 扩增条件优化 [J]. 广西植物, 2009, 29(4): 455-458.

[7] 唐美琼, 韦荣昌, 姚绍嫦, 等. 广西草珊瑚种质资源遗传多样性的 AFLP 分析 [J]. 中草药, 2012, 43(7): 1398-1402.

[8] 魏艺聪, 林培玲, 陈莹, 等. 采用 ISSR 标记分析草珊瑚的 DNA 指纹图谱与其品质相关性 [J]. 中草药, 2014, 45(11): 1620-1624.

[9] Xu Y Q, Tian S Y, Li R Q, *et al*. Transcriptome characterization and identification of molecular markers (SNP, SSR, and indels) in the medicinal plant *Sarcandra glabra* spp [J]. *Biomed Res Int*, 2021, 2021: 9990910.

[10] Xie D J, Chen L Y, Zhou C C, *et al*. Transcriptomic and metabolomic profiling reveals the effect of LED light quality on morphological traits, and phenyl-propanoid-derived compounds accumulation in *Sarcandra glabra* seedlings [J]. *BMC Plant Biol*, 2020, 20(1): 476.

[11] Wang W, Zou P S, Liu G F, *et al*. Characterization of the complete chloroplast genome sequence of *Sarcandra glabra* (Chloranthales) [J]. *Mitochondrial DNA B Resour*, 2020, 5(1): 864-865.

[12] 张晓萌, 李健春, 王琼, 等. 转录组测序技术在中医药领域的应用 [J]. 中国现代中药, 2016, 18(8): 1084-1087.

[13] Varshney R K, Graner A, Sorrells M E. Genic microsatellite markers in plants: Features and applications [J]. *Trends Biotechnol*, 2005, 23(1): 48-55.

[14] Li Y C, Korol A B, Fahima T, *et al*. Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review [J]. *Mol Ecol*, 2002, 11(12): 2453-2465.

[15] Kashi Y, King D G. Simple sequence repeats as advantageous mutators in evolution [J]. *Trends Genet*,

2006, 22(5): 253-259.

[16] 李敏, 王青, 陈晨, 等. 基于转录组测序的温带臭虫 SSR 和 SNP 位点分析 [J]. 山西农业大学学报: 自然科学版, 2019, 39(4): 52-57.

[17] 林恩文, 林榕榕, 陈钦常, 等. 龙眼全基因组和转录本序列 SSR 位点的鉴定 [J]. 福建农林大学学报: 自然科学版, 2022, 51(4): 493-501.

[18] 贾敏如, 李星炜. 中国民族药志要 [M]. 北京: 中国医药科技出版社, 2005: 365-366.

[19] 励娜, 杨荣平, 王宾豪, 等. 不同部位草珊瑚总黄酮的含量测定 [J]. 重庆中草药研究, 2007(2): 16-17.

[20] 尹培培, 卢楠, 谢汝根, 等. 肿节风野生与栽培品不同部位迷迭香酸含量比较 [J]. 中国实验方剂学杂志, 2011, 17(21): 97-99.

[21] Wang Y Y, Li Q R, Chi J, *et al*. Sesquiterpenoids from the leaves of *Sarcandra glabra* [J]. *Chin J Nat Med*, 2022, 20(3): 215-220.

[22] 梁宗锁, 方誉民, 杨东风. 植物萜类化合物生物合成与调控及其代谢工程研究进展 [J]. 浙江理工大学学报: 自然科学版, 2017, 37(2): 255-264.

[23] Lee M H, Jeong J H, Seo J W, *et al*. Enhanced triterpene and phytosterol biosynthesis in *Panax ginseng* overexpressing squalene synthase gene [J]. *Plant Cell Physiol*, 2004, 45(8): 976-984.

[24] 徐艳琴, 刘小丽, 黄小方, 等. 草珊瑚的研究现状与展望 [J]. 中草药, 2011, 42(12): 2552-2559.

[25] 秦双双, 韦坤华, 胡营, 等. 药用植物 4.0 研究进展与展望 [J]. 中国现代中药, 2023, doi: 10.13313/ j.issn. 1673-4890.20221109001.

[26] 车苏容, 张家源, 卢伟, 等. 转录组测序法研究草珊瑚叶和根的基因差异表达 [J]. 福建农业学报, 2020, 35(6): 598-610.

[27] 徐婷婷, 汪巧玲, 邹淑琼, 等. 基于高通量测序的大麦 In Del 标记开发及应用 [J]. 作物学报, 2020, 46(9): 1340-1355.

[28] 刘宇, 闫彩霞, 李春娟, 等. 花生栽培种 InDel 有效标记筛选与评估 [J]. 核农学报, 2020, 34(2): 256-264.

[29] 陆海燕, 陈璐, 王显生, 等. 基于高通量测序的棉花 InDel 标记开发及其应用 [J]. 棉花学报, 2019, 31(4): 297-306.

[责任编辑 时圣明]