

## 濒危珍稀植物半枫荷的转录组分析

叶兴状<sup>1</sup> 刘丹<sup>1</sup> 罗佳佳<sup>1</sup> 范辉华<sup>2</sup> 张国防<sup>1</sup> 刘宝<sup>1\*</sup> 陈世品<sup>1</sup>

(1. 福建农林大学林学院, 福州 350002; 2. 福建省林业科学研究院, 福州 350012)

**摘要** 为加强中国特有濒危植物半枫荷资源的保护与利用工作, 采用高通量测序平台 Illumina HiSeq 2500 对其进行转录组测序, 将得到的数据过滤后进行 de novo 组装并聚类去冗余, 获得 77 629 个 Unigenes, 通过九大功能数据库比对、分析、注释, 最终有 45 293 个 Unigenes 获得注释信息; 其中在 KOG 按功能分为 25 个子类, 获得 25 253 个功能注释信息, GO 功能注释可分为细胞组分、生物学过程、分子功能 3 大类, 分别可细分为 22、26、17 亚类(共 65 个亚类); 与 KEGG 数据库对比, 共发现 286 条代谢通路, 其中发现可能与半枫荷药用活性成分相关的次生代谢产物生物合成的 177 条途径。根据组装结果预测出 88 个基因家族共 1 547 个编码转录因子的 Unigenes, 发现控制药效合成的转录因子家族。另外, 根据注释结果检测到 12 579 个 SNP 多态位点和预测出 57 671 个 CDS 位点。本研究首次对半枫荷转录组进行分析, 为深入开展半枫荷分子生物学研究提供基础数据来源。

**关键词** 半枫荷 转录组分析 基因注释 药用活性成分 代谢通路

中图分类号: Q949.751.4 文献标志码: A doi: 10.7525/j.issn.1673-5102.2019.02.014

## Transcriptome Analysis for Rare and Endangered Plants of *Semiliquidambar cathayensis*

YE Xing-Zhuang<sup>1</sup> LIU Dan<sup>1</sup> LUO Jia-Jia<sup>1</sup> FAN Hui-Hua<sup>2</sup> ZHANG Guo-Fang<sup>1</sup> LIU Bao<sup>1\*</sup> CHEN Shi-Pin<sup>1</sup>

(1. College of Forestry, Fujian Agriculture and Forestry University, Fuzhou 350002; 2. Fujian Research Institute of Forestry, Fuzhou 350012)

**Abstract** In order to strengthen the conservation and utilization of China's endangered plant, *Semiliquidambar cathayensis*, a high-throughput sequencing platform, Illumina HiSeq 2500, was used to sequence the transcriptome of the plants. The obtained data was filtered and assembled by de novo. Then, 77 629 unigenes were obtained through classification and redundancy removal. By comparison analysis, annotation of the 9 functional databases, 45 293 unigenes get annotation information, which was divided into 25 sub-categories by function in KOG with 25 253 functional annotation information obtained. The functional annotation of GO could be divided into such three parts as cell components, biological processes, and molecular functions, which is composed of 2 226, and 17 subclasses respectively (65 subclasses in total). With the comparison of KEGG database, 286 metabolic pathways are found, among which there are 177 pathways to biosynthesis of secondary metabolites, which may be related to the medicinal active components of *S. cathayensis*. Based on the assembling result, 1 547 unigenes encoding the transcription factors are predicted in 88 gene families. Thus, a family of transcription factors controlling pharmacokinetic synthesis are discovered. Based on the annotation results of *S. cathayensis*, 12 579 SNP polymorphic sites are detected and 57 671 CDS sites are predicted. This study is the first time to analyze the transcriptional sequences of *S. cathayensis*, providing a basic data for the in-depth study on molecular biology in *S. cathayensis*.

基金项目: 福建省第三轮种业创新与产业化工程—特色树种香叶树、半枫荷种质创新与产业化工程项目(ZYCX-LY-2017002)

第一作者简介: 叶兴状(1991—), 男, 硕士研究生, 主要从事半枫荷遗传多样性研究。

\* 通信作者: E-mail: #0511779@qq.com

收稿日期: 2018-11-19

Foundation item: The Third Round of Seed Industry Innovation and Industrialization Project in Fujian Province—the Germplasm Innovation and Industrialization Project of the special species *Lindera communis* and *Semiliquidambar cathayensis*(ZYCX-LY-2017002)

First author introduction: YE Xing-Zhuang(1991—), male, master student, the main research area is the study of genetic diversity of *Semiliquidambar cathayensis*.

\* Corresponding author: E-mail: #0511779@qq.com

Received date: 2018-11-19

**Key words** *Semiliquidambar cathayensis* transcriptome analysis ;gene annotation ;medicinal active ingredient ; metabolic pathways

转录组( Transcriptome )能够在缺乏基因资源的条件下,运用转录组测序技术获得物种的生长和代谢规律,并揭示其生物学特性与基因内在关联,同时可获得物种绝大多数的转录产物<sup>[1]</sup>。转录组测序不仅能够发掘得到新的代谢通路及功能基因,还是开展 EST-SSR 分子标记的前提,而 EST-SSR 分子标记是目前最重要的群体遗传多样性研究方法之一。虽然第三代高通量测序已问世,但第二代高通量测序相较第三代测序具有成本低、操作简便、运行时间短等特点,是一种性价比高的基因序列研究手段。近年来已有多种药用植物完成转录组测序,如郑少华<sup>[2]</sup>等对淫羊藿( *Epimedium brevicornu* )进行转录组分析,找到类黄酮素代谢通路;Roberto A Barrero 等<sup>[3]</sup>对狼毒( *Stellera chamaejasme* )进行转录组分析,提出了前列腺素合成上游途径的新见解,为批量产生这种化合物奠定理论基础,将可用于 HIV 研究或治疗患者;李响等<sup>[4]</sup>对冬虫夏草( Chinese Caterpillar Fungus )进行转录组分析,开发了一种用于合成主要药用化合物(如虫草素)的模型;Xu Guo 等<sup>[5]</sup>对铁皮石斛( *Dendrobium officinale* )茎进行转录组分析,揭示了代表参与生物碱骨架合成的 25 个基因及 69 个独特序列。

半枫荷( *Semiliquidambar cathayensis* )属金缕梅科( Hamamelidaceae )半枫荷属( *Semiliquidambar* )是我国特有的濒危国家二级保护植物<sup>[6]</sup>,半枫荷在《中国生物多样性红色名录—高等植物卷》中被评为 VU 级<sup>[7]</sup>,具有枫香属( *Liquidambar* )和蕈树属( *Altingia* )两属的综合性状特征,异型叶是其重要形态特征,叶子兼具有枫香( *Liquidambar formosana* )和阿丁枫( *Altingia chinensis* )特征,半枫荷嫩叶部分紫红,极为美观,是南方为数不多的变色树种,具有良好的园林绿化应用前景。此外其材质优良、树干通直、旋刨性良好、材质优良,可作旋刨制品及家具,不仅是珍贵用材树种<sup>[8-9]</sup>,而且药用价值极高,其根、枝、叶,都可入药,有祛风、除湿、活血、止痛、消肿等功效,可治疗腰肌劳损、跌打淤积、风湿性关节炎,且外伤常用其叶止血<sup>[10-12]</sup>。目前半枫荷的研究处于起步阶段,主要涉及种群生态特征和药理研究等领域,鲜有涉及分子生物学领域,胡君等<sup>[13]</sup>在四川省发现了半枫荷新分布,并调查半枫荷种群生境特征;廖娜等<sup>[14]</sup>

发现半枫荷中的多酚具有抗菌和显著抗氧化能力;周光雄等<sup>[15]</sup>首次从金缕半枫荷中分离得齐墩果酸、3-羰基齐墩果酸等 9 种有抗炎药效的化合物,为半枫荷药理研究提供基础数据;此外孙静<sup>[16]</sup>等还发现半枫荷根的活性成分齐墩果酸对病毒性肝炎的抗原具有抑制活性的作用。由于半枫荷用途广、经济价值高,再加上天然更新困难,对生境要求苛刻,同时因根可入药,树根常被药农挖断,人为破坏严重,目前仅在贵州、湖南、福建、广西、江西、广东等省份零星分布<sup>[11-12,17]</sup>,已濒临灭绝,亟需保护。

鉴于此,本研究选择福建省沙县、顺昌野生半枫荷做为研究材料,借助高通量测序平台 Illumina HiSeq 2500 首次对半枫荷叶进行转录组测序,建立半枫荷基因文库,然后利用生物信息学的方法对半枫荷进行功能注释,代谢途径、CDS 预测及 SNP 检测等分析,以期初步了解半枫荷基因表达、分布情况,如参与药效合成的酶的基因分布与表达情况,为半枫荷遗传多样性研究提供基础数据,同时也为今后半枫荷功能基因的研究奠定基础。

## 1 材料与方法

### 1.1 实验材料

本实验选取半枫荷叶为实验材料。于 2018 年 3 月份到福建省顺昌县、沙县各选取 1 棵半枫荷,采集一年生叶子放入预先准备的冰盒中带回实验室,将叶片擦干去除污渍,后用锡箔纸将叶片包好做好标记,并用液氮处理 10~20 分钟,再用超低温零下 80℃ 冰箱保存,用于半枫荷引物开发及遗传多样性分析实验。

### 1.2 方法

#### 1.2.1 半枫荷总 RNA 的提取

将采集的 2 棵半枫荷叶样品混合在一起,然后用生工生物(上海)有限公司的 EZ-10DNA away RNA Mini-prep Kit 试剂盒提取总 RNA,使用杭州奥盛仪器有限公司的 Nano-200 微量分光光度计测定 RNA 的浓度和质量,如果 OD260NM/OD280NM 的范围为 1.8~2.2,28S/18S > 2,RNA 完整值(RIN) > 8.5,OD260NM/OD230NM > 1.8,则认为提取的 RNA 符合建库要求。

#### 1.2.2 半枫荷测序与拼接组装

采用 Illumina Hi Seq2500 的高通量测序技术对

半枫荷的转录组进行测序,测序得到的原始数据,里面含有带接头的、低质量的序列。为了保证信息分析质量,对原始数据过滤,得到 Clean 数据,参数如下:

- ①去除带 N 碱基的序列;
- ②去除 reads 中的接头序列;
- ③从 reads 3'到 5'方向开始去除低质碱基(Q 值 < 20);
- ④使用滑窗法去除 reads 尾部质量值在 20 以下的碱基(窗口大小为 5 bp);
- ⑤去除 reads 长度小于 35nt 的 reads 本身及其配对 reads。
- ⑥从 reads 5'到 3'方向开始去除低质碱基(Q 值 < 20);

将 reads 回帖到全部转录本,并归一化后得到 reads 在转录本上的平均分布,发现有少许偏移,但未发生明显偏移,符合建库要求。然后将 Clean 数据利用 Trinity<sup>[18]</sup>软件(参数 min\_kmer\_cov 2,其余默认)进行拼接组装。建库过程中的 PCR 扩增会产生冗余序列(duplicate reads)指的是碱基排列完全相同的序列,如果建库扩增时不均匀,可能会导致部分基因冗余序列过高,造成冗余序列分布的异常,使用 RSeQC 软件去除冗余序列,得到 Unigenes。

### 1.2.3 基因功能注释

运用 NCBI Blast + 将转录本与 CDD、KOG、NR、NT、GO、Swissprot、PFAM、KEGG、TrEMBL 等 9 个数据库进行比对,得到其功能注释信息。根据转录本与 Swissprot、TrEMBL 的注释结果得到 GO 功能注释信息。利用 KAAS<sup>[19]</sup>与基因和蛋白质数据库进行比对,得到转录本 KEGG 注释信息。利用以上数据库的分析结果,筛选 Unigenes 最优比对片段作为该 Unigenes 的 CDS,使用 getorf<sup>[20]</sup>检测 Unigenes 的开放阅读框(Open Reading Frame, ORF),采用 hmmsearch<sup>[21]</sup>将 ORF 比对到转录因子蛋白结构域(数据来源于 PlantTFDB),然后根据 PlantTFDB 描述的转录因子家族特征对 Unigenes 进行能力鉴定。最后使用 BCFTools<sup>[22]</sup>根据 Mapping 结果找出可能的单核苷酸多态性(Single Nucleotide Polymorphisms, SNP)位点,进行 SNP 分析。

## 2 结果与分析

### 2.1 半枫荷转录本组装与 CDS 预测分析

通过使用软件将 clean 数据 De novo 组装成转录本,共得到 127 249 个 Transcripts。对半枫荷转录本长度进行分析(图 1A,表 1),经组装、去冗长等

处理得到 77 629 个 Unigenes,200 ~ 500 bp 范围内的 Unigenes 最多,为 36 630 条,其次是 300 ~ 400 bp、400 ~ 500 bp 和大于 2 000 bp。最短长度为 201 bp,最长 10 975 bp, N<sub>50</sub>长度 1 105 bp,总长度 48 512 368 bp,Unigene 平均长度为 624.93 bp,其中长度分布在 200 ~ 500 bp 的 Unigenes 占比 67.88%,分布在 500 ~ 1 000 bp 的有 11 296 个,占比 14.55%,长度大于 1 000 bp 的有 13 635 个,占比 17.57%。同时,为了更好的发掘半枫荷功能基因和便于遗传指纹图谱的构建,本研究对半枫荷编码序列(Coding Sequence, CDS)进行预测,获取 57 671 个 Unigenes 作为 CDS。CDS 序列 1 000 nt 以上的有 8 050 个,仅占 13.96%,100 ~ 500 nt 的序列有 40 608 个,占 70.41%,其中 200 ~ 300 nt 的序列有 20 615 个,占总 CDS 序列的 37.75%,500 ~ 1 000 nt 的序列仅占 15.63%。将图 1B 与图 1A 对比,序列分布更加集中于 100 ~ 600 nt。

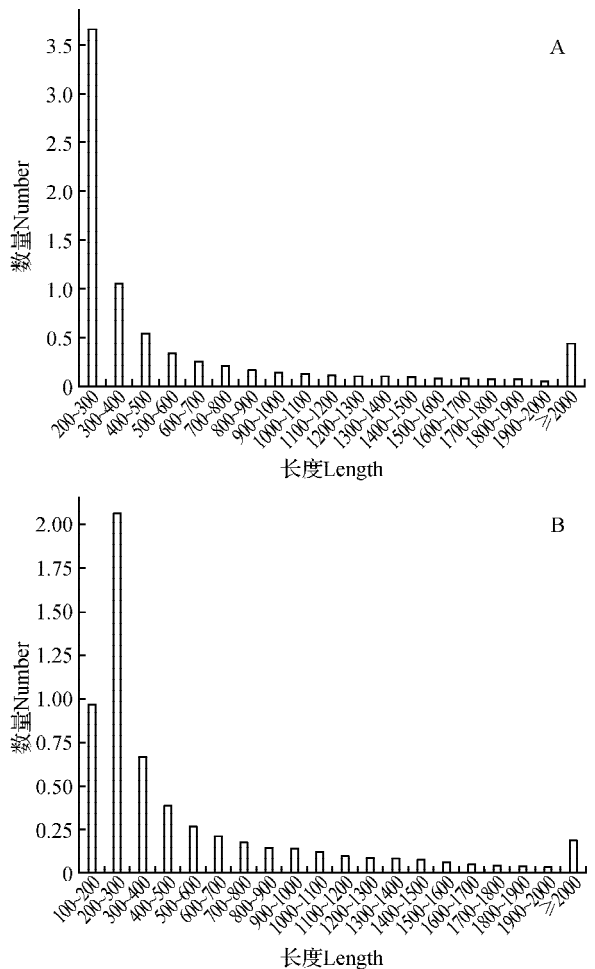


图 1 半枫荷 Unigene 的长度和 CDS 长度的分布图  
Fig. 1 Distribution of unigene lengths and CDS length for *S. cathayensis*

表 1 拼接结果统计图  
Table 1 Statistical graph of splicing results

项目 Item	数量 No.	≥500 bp	≥1 000 bp	N <sub>50</sub>	N <sub>90</sub>	最长值 Max len	最小值 Min Len	总长度 Total len	平均长度 Average len
Transcript	127 249	57 280	32 144	1 305	285	10 795	201	97 669 341	767.55
Unigene	77 629	24 931	13 635	1 105	245	10 795	201	48 512 368	624.93

## 2.2 功能注释及 COG 分类

将 *de novo* 组装得到的 77 629 个基因序列分别与 CDD、KOG、NR、NT、GO、Swissprot、PFAM、KEGG、TrEMBL 这 9 个数据库进行比对,成功注释到 9 个数据库中(表 2);在 9 个数据库中至少 1 个数据库注释成功的 Unigenes 数为 45 293,占总 Unigenes 数的 58.35%;共有 2 692 个 Unigenes 与 9 个数据库都能匹配成功,占总 Unigenes 数的 3.47%(表 2);而只能注释到 NR、KEGG、COG、Swissprot 蛋白库的 Unigenes 分别只有 6 971、0、120、3 362 个(图 2);此外,还有 32 336(41.65%)个 Unigenes 序列未能获得功能注释信息,这有可能是测序技术不同引起的。

表 2 Unigene 注释结果统计表

Table 2 Statistics of annotation results

数据库 Database	Unigenes 数量 Number of unigenes	百分比 Percentage (%)
注释到 CDD Annotated in CDD	29 998	38.64
注释到 KOG Annotated in KOG	22 802	29.37
注释到 NR Annotated in NR	39 076	50.34
注释到 NT Annotated in NT	24 962	32.16
注释到 PFAM Annotated in PFAM	21 129	27.22
注释到 Swissprot Annotated in Swissprot	36 239	46.68
注释到 TrEMBL Annotated in TrEMBL	38 996	50.23
注释到 GO Annotated in GO	39 293	50.62
注释到 KEGG Annotated in KEGG	4 029	5.19
注释到至少一个数据库 Annotated in at least one database	45 293	58.35
注释到所有数据库 Annotated in all database	2 692	3.47
Unigenes 总量 Total Unigenes	77 629	100

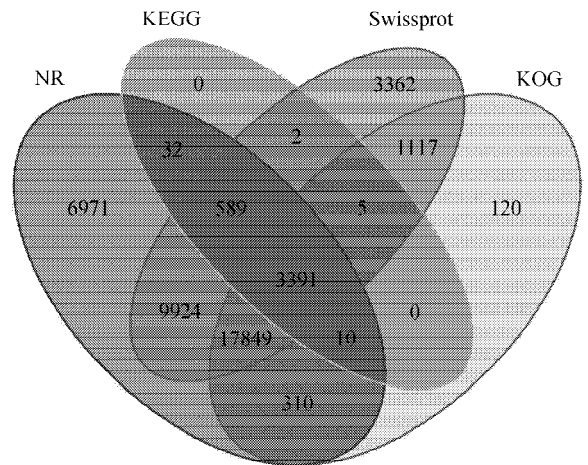


图 2 半枫荷 Unigene 的功能注释维恩图

Fig. 2 Venn diagram functional annotation distribution of unigenes of transcriptome for *S. cathayensis*

## 2.3 KOG 功能注释

将组装得到的 Unigenes 比对到 KOG 蛋白质库中,半枫荷 KOG 功能注释获信息结果显示细分为 25 个子类,共有 22 802 个 Unigenes 被注释(占 Unigenes 总数的 29.37%),其各个子类的基因表达丰度差异较大,共包括 25 253 个功能注释信息,基本涵盖了半枫荷大部分的生命活动。其中,信号传递机制,翻译后修饰、蛋白质折叠和分子伴侣,一般功能基因最多,分别有 2 858(11.32%)、2 699(10.69%)和 2 552(10.11%)。而核结构和细胞活性较少,分别仅有 55(0.22%)和 6(0.02%)个。另有 971(占 3.85%)个功能注释信息未明确其生物学功能(图 3)。

## 2.4 NR 功能注释

将获得的半枫荷 Unigenes 比对到 NR 库中( $E \leq 1e-5$ ),NR 功能注释获信息结果显示有 39 076 个 Unigenes 被注释(占 50.34%),且与其它物种已知基因序列具有不同程度的匹配性。如注释物种分布图所示(图 4):注释为葡萄(*Vitis vinifera*)相关基因的序列最多,达 10 111 条,占 NR 库中被注释 Unigenes 的 25.89%,其次,序列同源性大于 2% 的近缘物种还有大麦亚种(*Hordeum vulgare*)

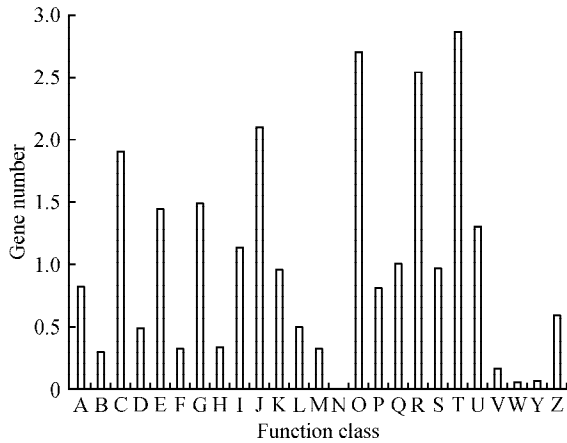


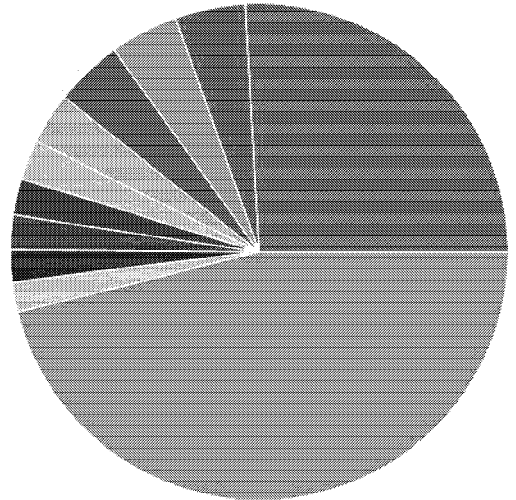
图3 半枫荷转录组 Unigenes 的 KOG 功能分布图

A. RNA 加工和修饰 ;B. 染色体结构和动力学 ;C. 能源产生与转化 ;D. 细胞周期调控 细胞分裂 染色体分离 ;E. 氨基酸转运和代谢 ;F. 核酸转运和代谢 ;G. 碳水化合物转运和代谢 ;H. 辅酶转运和代谢 ;I. 脂类转运和代谢 ;J. 翻译 核糖体结构和生物发生 ;K. 转录 ;L. 复制 重组和修饰 ;M. 细胞壁/细胞膜生物发生 ;N. 细胞活性 ;O. 翻译后修饰 蛋白翻转 伴侣 ;P. 无机离子转运和代谢 ;Q. 次生代谢物合成 转运和代谢 ;R. 只有一般功能预测 ;S. 未知功能 ;T. 信号传递机制 ;U. 细胞间运输 分泌物和囊泡运动 ;V. 防御机制 ;W. 细胞外结构 ;Y. 核结构 ;Z. 细胞骨架

Fig. 3 KOG functional annotation distribution of unigenes of transcriptome for *S. cathayensis*

A. RNA processing and modification ; B. Chromatin structure and dynamics ; C. Energy production and conversion ; D. Cell cycle control ,cell division ,chromosome partitioning ; E. Amino acid transport and metabolism ; F. Nucleotide transport and metabolism ; G. Carbohydrate transport and metabolism ; H. Coenzyme transport and metabolism ; I. Lipid transport and metabolism ; J. Translation ,ribosomal structure and biogenesis ; K. Transcription ; L. Replication ,recombination and repair ; M. Cell wall/membrane/envelope biogenesis ; N. Cell motility ; O. Posttranslational modification ,protein turnover ,chaperones ; P. Inorganic ion transport and metabolism ; Q. Secondary metabolites biosynthesis ,transport and catabolism ; R. General function prediction only ; S. Function unknown ; T. Signal transduction mechanisms ; U. Intracellular trafficking ,secretion and vesicular transport ; V. Defense mechanisms ; W. Extracellular structures ; Y. Nuclear structure ; Z. Cytoskeleton

subsp. *Vulgare* )、荷花( *Nelumbo nucifera* )、可可( *Theobroma cacao* )、枣( *Ziziphus jujuba* )、麻风树( *Jatropha curcas* )、蓖麻( *Ricinus communis* )、甜橙( *Citrus sinensis* )、毛果杨( *Populus trichocarpa* )和梅( *Prunus mume* ) ,其余 46.07% 的注释 Unigenes 分布于其他 648 个物种中。由于缺乏半枫荷基因组信息和技术局限性 ,仍有 29( 0.074% )个 Unigenes 在 NR 库中未能获得注释。



■ 葡萄 *Vitis vinifera*(10111)  
 ■ 大麦亚种 *Hordeum vulgare* subsp. *vulgare*(1788)  
 ■ 荷花 *Nelumbo nucifera*(1732)  
 ■ 可可 *Theobroma cacao*(1723)  
 ■ 枣 *Ziziphus jujuba*(1290)  
 ■ 麻风树 *Jatropha curcas*(1023)  
 ■ 蓖麻 *Ricinus communis*(922)  
 ■ 甜橙 *Citrus sinensis*(879)  
 ■ 毛果杨 *Populus trichocarpa*(812)  
 ■ 梅 *Prunus mume*(765)  
 ■ Other(18002)

图4 半枫荷转录组 Unigenes 的 NR 注释物种分布图

Fig. 4 NR annotated species distribution of Unigenes of transcriptome for *S. cathayensis*

## 2.5 GO 功能注释及分类

根据转录本与 Swissprot、TrEMBL 的注释结果得到 GO 功能注释信息。综合描述半枫荷叶片中相关基因的生物学特征 ,从宏观上了解其表达基因的功能分布情况 ,对获得相应 GO 条目的 Unigenes 进行统计分析(图 5) ,共有 39 293 个 Unigenes 获得 317 475 个 GO 注释 ,平均每条 8.08 个。把获得 GO 注释的 Unigenes 划分为细胞组分、分子功能和生物学过程三个类型 ,分别各获得 133 668 ( 42.10% )、53 236( 16.77% )和 130 571( 41.13% )个注释 ,可细分为 65 个子类 ,其中 ,细胞、细胞组分、结合、细胞器、催化活性、细胞过程、代谢过程和单一有机体过程等获得注释较多。

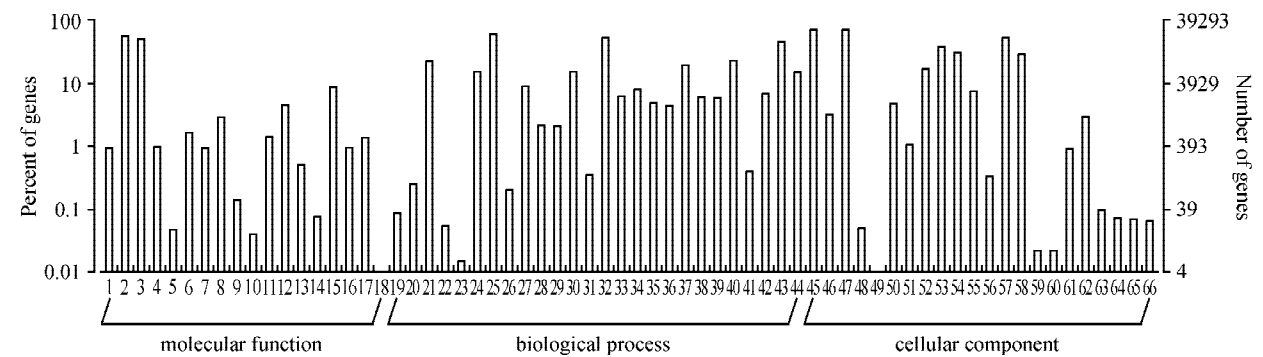
## 2.6 KEGG 代谢通路分析

将半枫荷的 Unigenes 序列映射到 KEGG 代谢库(  $E \leq 1e-5$  )根据注释信息对其可能参与或涉及的代谢途径进行统计分析。结果表明 4 029 条 Unigenes 获得注释( 5.19% ) ;其参与的代谢通路可归为 5 大类别、33 个子类。由图 6 可知 5 种代谢通路大类中 ,代谢相关的通路获得 2 786 个注释 ,占 41.99% ,其次是遗传信息处理和生物系统

分别获得 1 562( 23. 55% )、1 036( 15. 62% ),再次是细胞过程与环境信息处理分别获得 650( 9. 80% )、600( 9. 04% )。进一步细分为 33 子类代谢途径,其中,翻译获得注释最多,为 722( 17. 92% )其次为信号传导、折叠排序和退化、碳水化合物代谢。共有 8 923 条次 Unigenes 归入 286 条代谢途径,按数量排列,靠前的有 15 个代谢通路(表 3)。核糖体代谢途径注释到的 Unigenes 数量最多,有 389( 4. 36% )条 Unigenes 其次为碳代谢( 221, 2. 48% )、氨基酸的生物合成( 193, 2. 16% )和内质网蛋白处理( 171, 1. 92% )等途径。

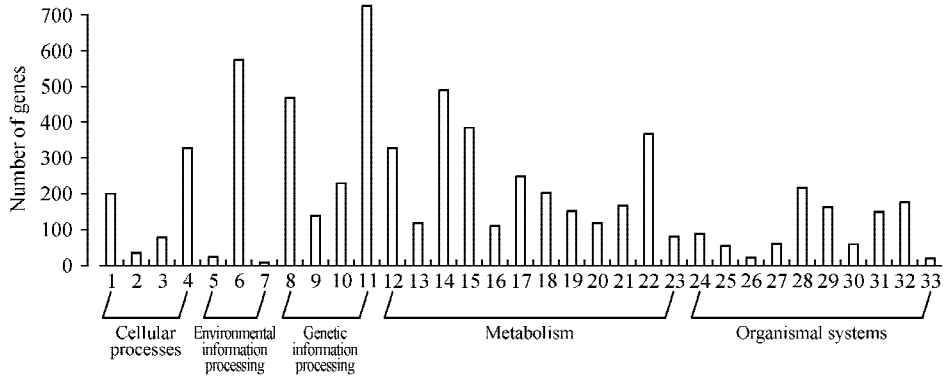
在半枫荷的转录组数据中,有 64 条 Unigenes 映射到苯丙素生物合成( ko00940 );有 65 条 Uni-

genes 映射到萜类物质生物合成( ko00904, ko00902, ko00909, ko00900 );有 27 条 Unigenes 映射到黄酮类物质生物合成( ko00941, ko00942, ko00944, ko00965 );有 2 条 Unigenes 映射到硫甙生物合成( ko00966 ),可合成糖苷类物质;有 18 条 Unigenes 映射到了合成生物碱类( ko00232、ko00950)通路;有 1 条 unigene 映射到了黄酮和黄酮醇的生物合成( ko00944 )通路,有 1 条 Unigene 映射到了异黄酮的生物合成( ko00943 )通路(表 4)。半枫荷中多次次生代谢途径及相关 Unigenes 的发现,说明其次生代谢生物合成途径的复杂性,通过对半枫荷化学成分的分析研究,为后续从半枫荷中分离新化合物提供了线索,也为阐述半枫荷药效物质基础提供了理论依据。



**图 5 半枫荷转录组 Unigenes 的 GO 功能分类统计图** 1. 抗氧化活性 2. 结合 3. 催化活性 4. 电子载体活动 5. metallochaperone activity 6. 金属伴活动分子功能调节器 7. 分子转导活性 8. 核酸结合转录因子活性 9. 营养库活性类 10. 蛋白质标记 11. 信号转导活性 12. 结构分子活性 13. 转录因子活动与蛋白质结合 14. 传译调活活性 15. 转运蛋白活性 16. 受体活性 17. 酶调节活性 18. 受体调节活性 19. 运转状态 20. 生物粘附 21. 生物调节 22. 细胞聚集 23. 细胞杀伤 24. 细胞成分组织或生物合成 25. 细胞过程 26. 解毒; 27. 发育过程 28. 生长 29. 免疫系统过程 30. 定位 31. 运转 32. 代谢过程 33. 多组织进程 34. 多细胞生物过程 35. 生物过程负调控; 36. 生物过程的积极调节 37. 生物过程的调节 38. 繁殖 39. reproductive process 40. 生殖过程对刺激的反应 41. 节律过程 42. 信号传导 43. 单一有机体过程 44. 建立斑点定位 45. 细胞 46. 细胞连接 47. 细胞部分 48. 细胞外基质 49. 胞外基质成分 50. 胞外区域 51. 胞外区域部分 52. 大分子复合体 53. 膜 54. 膜部分 55. 膜封闭腔 56. 拟核 57. 细胞器 58. 细胞器部分 59. 其他有机体 60. 其他有机体部分 61. 超分子纤维 62. 共质体 63. 突触 64. 突触部分 65. 病毒 66. 病毒部分

**Fig. 5 GO functional classification of unigenes of transcriptome for *S. cathayensis*** 1. antioxidant activity ;2. binding ;3. catalytic activity ;4. electron carrier activity ;5. metallochaperone activity ;6. molecular function regulator ;7. molecular transducer activity ;8. nucleic acid binding transcription factor activity ;9. nutrient reservoir activity ;10. protein tag ;11. signal transducer activity ;12. structural molecule activity ;13. transcription factor activity ;14. protein binding ;14. translation regulator activity ;15. transporter activity ;16. receptor activity ;17. enzyme regulator activity ;18. receptor regulator activity ;19. behavior ;20. biological adhesion ;21. biological regulation ;22. cell aggregation ;23. cell killing ;24. cellular component organization or biogenesis ;25. cellular process ;26. detoxification ;27. developmental process ;28. growth ;29. immune system process ;30. localization ;31. locomotion ;32. metabolic process ;33. multi-organism process ;34. multicellular organismal process ;35. negative regulation of biological process ;36. positive regulation of biological process ;37. regulation of biological process ;38. reproduction ;39. reproductive process ;40. response to stimulus ;41. rhythmic process ;42. signaling ;43. single-organism process ;44. establishment of localization ;45. cell ;46. cell junction ;47. cell part ;48. extracellular matrix ;49. extracellular matrix component ;50. extracellular region ;51. extracellular region part ;52. macromolecular complex ;53. membrane ;54. membrane part ;55. membrane-enclosed lumen ;56. nucleoid ;57. organelle ;58. organelle part ;59. other organism ;60. other organism part ;61. supramolecular fiber ;62. symplast ;63. synapse ;64. synapse part ;65. virion ;66. virion part



**图 6 半枫荷转录组 Unigenes 的 KEGG 功能注释分布统计图** 1. 细胞生长和死亡 2. 细胞运动 3. 细胞群体 4. 运输和分解代谢 5. 膜运输 6. 信号转导 7. 信号分子和相互作用 8. 折叠排序与退化 9. 复制和修复 10. 转录 11. 翻译 12. 氨基酸代谢 13. 其他次级代谢产物的生物合成 14. 碳水化合物代谢 15. 能量代谢 16. 多糖合成与代谢 17. 类脂(化合)物代谢作用 18. 辅助因子和维生素的代谢 19. 其他氨基酸的代谢 20. 萜类化合物和聚酮化合物的代谢 21. 核苷酸代谢 22. 概观 23. 异生素生物降解和新陈代谢 24. 老化; 25. 循环系统 26. 进化 27. 消化系统 28. 内分泌系统 29. 环境适应 30. 排泄系统 31. 免疫系统 32. 神经系统 33. 感官系统

**Fig. 6 KEGG functional annotation distribution of unigenes of transcriptome for *S. cathayensis*** 1. Cell growth and death; 2. Cell motility; 3. Cellular community; 4. Transport and catabolism; 5. Membrane transport; 6. Signal transduction; 7. Signaling molecules and interaction; 8. Folding sorting and degradation; 9. Replication and repair; 10. Transcription; 11. Translation; 12. Amino acid metabolism; 13. Biosynthesis of other secondary metabolites; 14. Carbohydrate metabolism; 15. Energy metabolism; 16. Glycan biosynthesis and metabolism; 17. Lipid metabolism; 18. Metabolism of cofactors and vitamins; 19. Metabolism of other amino acids; 20. Metabolism of terpenoids and polyketides; 21. Nucleotide metabolism; 22. Overview; 23. Xenobiotics biodegradation and metabolism; 24. Aging; 25. Circulatory system; 26. Development; 27. Digestive system; 28. Endocrine system; 29. Environmental adaptation; 30. Excretory system; 31. Immune system; 32. Nervous system; 33. Sensory system

**表 3 半枫荷中 Unigenes 数量最多的 16 个代谢通路**

**Table 3 Top sixteen metabolic pathways involved in *S. cathayensis* unigenes**

代谢通路 Pathway	Unigenes 数量 Number of unigenes	Unigene 比例 Percent of unigenes(%)	代谢通路编号 Pathway ID
核糖体 Ribosome	389	4.36	ko03010
碳代谢 Carbon metabolism	221	2.48	ko01200
氨基酸合成 Biosynthesis of amino acids	193	2.16	ko01230
内质网蛋白加工 Protein processing in endoplasmic reticulum	171	1.92	ko04141
剪接 Spliceosome	158	1.77	ko03040
RNA 转运 RNA transport	154	1.73	ko03013
氧化磷酸化 Oxidative phosphorylation	150	1.68	ko00190
嘌呤代谢 Purine metabolism	136	1.52	ko00230
植物激素信号转导 Plant hormone signal transduction	128	1.43	ko04075
胞吞作用 Endocytosis	110	1.23	ko04144
泛素介导的蛋白水解 Ubiquitin mediated proteolysis	108	1.21	ko04120
植物病原体相互作用 Plant-pathogen interaction	108	1.21	ko04626
淀粉和蔗糖代谢 Starch and sucrose metabolism	100	1.12	ko00500
糖酵解/葡萄糖生成 Glycolysis/Gluconeogenesis	99	1.11	ko00010
mRNA 监测通路 mRNA surveillance pathway	96	1.08	ko03015
嘧啶代谢 Pyrimidine metabolism	94	1.05	ko00240

## 2.7 TF 编码能力及 SNP 检测

本研究共预测出 88 个基因家族共 1 547 个编码转录因子(Transcription Factor, TF)的 Unigenes (图 7) 另有 57 个转录因子未知具体基因家族(其中锌指蛋白 C2H2 类 TF 家族占比最多,为 151 个(9.76%),其次为 AP2/ERF-ERF、MYB-related 和 bHLH 家族 TF,分别占 75(4.85%)、69(4.46%)和 68(4.40%)个 TF,而 HRT、SOH1、STAT、ULT 所占比例最少,均为 1(0.065%)个。同时预测出同源异型盒(Homeobox, HB)家族 TF,包括 HB-HD-ZIP、HB-PHD、HB-KNOX、HB-WOX、HB-BELL、HB-other;可能参与调控药效成分合成有关的 TF 有 AP2/ERF-ERF、WRKY、bHLH、bZIP、MYB 等;还发现可能参与调节种子发育、萌发有关的 TF 有 bZIP、MYB、MYB-related、C2C2-Dof、NAC、NF-YB、NF-YA、NF-YC、MADS-M-type、MADS-MIKK;与逆境抗性相关 TF 主要有:bZIP、AP2/ERF-ERF、AP2/ERFWRKY、HB 类、MYB、MYB-related 和 HSF。这些 TF 功能的发现表明半枫荷转录调控的复杂性、交叉性和多样性,也为后续研究半枫荷濒危原因、药效药理研究、遗传多样性、指纹图谱、连锁图谱构建奠定基础。

表 4 半枫荷药用活性成分相关的次生代谢产物生物合成途径

Table 4 Biosynthetic pathway of secondary metabolites involved in medicinal active ingredient of *S. cathayensis*

代谢通路 Pathway	Unigenes 数量 Number of unigenes	药效成分 Effective components	代谢通路编号 Pathway ID
咖啡因代谢 Caffeine metabolism	2	生物碱 Alkaloids	ko00232
萜类骨架生物合成 Terpenoid backbone biosynthesis	43	萜类 terpene	ko00900
单萜类生物合成 Monoterpenoid biosynthesis	4	萜类 terpene	ko00902
二萜生物合成 Diterpenoid biosynthesis	11	萜类 terpene	ko00904
倍半萜类化合物和三萜类化合物的生物合成 Sesquiterpenoid and triterpenoid biosynthesis	7	萜类 terpene	ko00909
苯丙烷类生物合成 Phenylpropanoid biosynthesis	64	苯丙素类 Phenylpropanoid	ko00940
类黄酮生物合成 Flavonoid biosynthesis	21	黄酮类 flavonoids	ko00941
花青素生物合成 Anthocyanin biosynthesis	1	黄酮类 flavonoids	ko00942
异黄酮生物合成 Isoflavonoid biosynthesis	1	异黄酮类 Isoflavones	ko00943
黄酮和黄酮醇的生物合成 Flavone and flavonol biosynthesis	1	黄酮类 flavonoids	ko00944
异喹啉生物碱生物合成 Isoquinoline alkaloid biosynthesis	16	生物碱类 Alkaloids	ko00950
硫代葡萄糖苷生物合成 Glucosinolate biosynthesis	2	糖苷类 glycosides	ko00966
泛醌和其他萜类化合物-醌生物合成 Ubiquinone and other terpenoid-quinone biosynthesis	32	醌类 Quinonoids	ko00130
吲哚生物碱生物合成 Indole alkaloid biosynthesis	1	生物碱 Alkaloids	ko00901
二苯乙烯类、二芳基庚烷类和姜辣素生物合成 Stilbenoid ,diarylheptanoid and gingerol biosynthesis	11	生物碱 Alkaloids	ko00945
甜菜碱生物合成 Betalain biosynthesis	4	生物碱 Alkaloids	ko00965
聚酮化合物糖单元生物合成 Polyketide sugar unit biosynthesis	1	聚酮类 Polyketides	ko00523
花生四烯酸代谢 Arachidonic acid metabolism	11	有机酸及酚类 Organic acids and phenols	ko00590

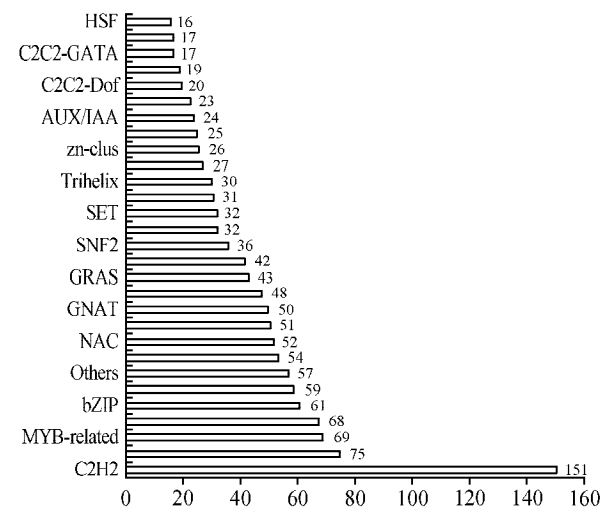
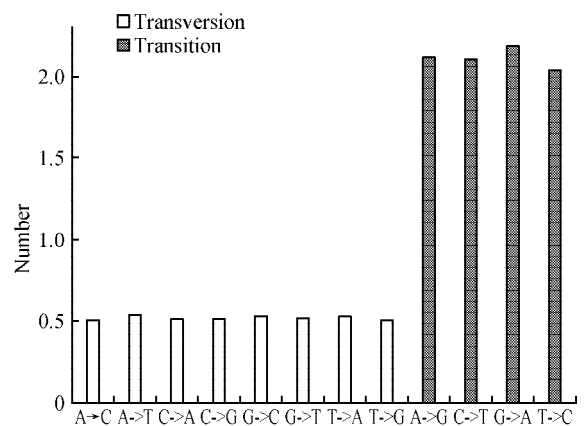


图 7 半枫荷转录组 Unigene 主要的转录因子家族分类图

Fig. 7 The main transcription factor family classification of unigenes of transcriptome for *S. cathayensis*

此外,本研究检测出 125 917 个单核苷酸多态性(Single Nucleotide Polymorphisms, SNP)位点(图 8),分析发现半枫荷 Unigenes 序列上 SNP 分布不均匀,转换突变类型(蓝色)的 SNP 数量比

图 8 半枫荷转录组 Unigene 的 SNP 变异类型分布图  
Fig. 8 SNP variant type distribution of unigenes of transcriptome for *S. cathayensis*

颠换突变(红色)的 2 倍略多。其中,转换突变类型(A→G、C→T、G→A、T→C)有 84 445 个,占 67.06%,颠换突变类型(A→C、A→T、C→A、C→G、G→C、G→T、T→A、T→G)有 41 472 个,占 32.94%。半枫荷 SNP 分析,结合 Unigenes 的表达量及表型,可在 mRNA 层面的基因型差异区别于其它物种,通过构建半枫荷 SNP 图谱,对了解



其亲属关系和物种起源进化关系具有重要的生物学意义。

### 3 讨论

本研究采用 RNA-Seq 技术首次对半枫荷叶片进行了高通量转录组测序分析,拼接组装后获得 77 629 条 Unigene,  $N_{50}$  长度为 1 105, 平均长度为 624.93 bp, 均比川芎 (*Ligusticum chuanxiong*)<sup>[23]</sup>、鹰嘴豆 (*Cicer arietinum*)<sup>[24]</sup>、葫芦巴 (*Trigonella foenum-graecum*)<sup>[25]</sup> 更长, 说明本研究所得序列中长片段较多, 拼接效果较好。半枫荷 SSR 出现频率为 13.74%、比同是中药材的野三七 (*Panax stipuleanatus*)<sup>[26]</sup> (16.82%)、半夏 (*Pinellia ternata*)<sup>[27]</sup> (16.24%) 更低, 但比党参 (*Codonopsis pilosula*)<sup>[28]</sup> (12.22%)、莲雾<sup>[29]</sup> (*Syzygium samarangense*) (9.27%)、川芎 (*Ligusticum chuanxiong*)<sup>[23]</sup> (8.11%) 更高, 其中有 13 635 条 Unigenes 的长度在 1 kb 以上, 进一步说明半枫荷 SSR 位点分布较密集, 可开发位点多。

所有的 Unigene 与 CDD、KOG、NR、NT、GO、Swissprot、PFAM、KEGG、TrEMBL 等 9 个数据库进行比对, 共有 45 293 条 (58.35%) 成功获得功能注释, 但仍有 32 337 条 (41.65%) 序列定位不清楚, 得到功能注释基因相对较少, 这与香樟 (*Cinnamomum camphora*)<sup>[30]</sup>、紫背天葵 (*Begonia fimbriatipula*)<sup>[31]</sup>、柳树 (*Salix babylonica*)<sup>[32]</sup> 均出现大量未获得注释 Unigenes 的结果相似, 可能是由于这些 Unigenes 本身为为非编码序列, 或是技术上的限制及已测序的半枫荷近缘种较少有关<sup>[33]</sup>。成功注释到 NR 数据库中的 Unigenes 中, 注释为葡萄同源种的有 25.89%, 远高于其他 648 个被注释物种, 出现这种情况, 可能是由于葡萄与半枫荷的进化史和生活史较为接近的缘故, 亦可能因为葡萄具有参考基因组。而与半枫荷能比对上仅有 6 个同科植物, 共有 43 条 Unigenes, 其中枫香 (*Liquidambar formosana*)、北美枫香 (*Liquidambar styraciflua*)、缺萼枫香 (*Liquidambar acalycina*)、马蹄荷 (*Exbucklandia populnea*)、北美金缕梅 (*Hamamelis virginiana*)、红花檵木 (*Loropetalum chinense*), 分别有 31、8、1、1、1 和 1, 之所以会出现这种情况, 可能是由于金缕梅科植物转录组、基因组信息严重缺乏, 也可能是因为本次测序得到的 Unigenes 中半枫荷特有的新功能基因较多导致未能获得相应匹配。

对拼接组装获得的 45 293 条 Unigenes 进行了

代谢途径分析和功能分类, 并根据 KEGG 数据库分析其可能参与或涉及的代谢通路, 共获得 4 029 (5.19%) 个 Unigenes 注释, 注释类别可细分为 5 个代谢通路大类、33 类代谢途径, 其中定位到代谢相关通路的基因数最多, 约占总注释量的 42%, 进一步证实半枫荷具有较强的代谢活动能力。值得一提的是, 本研究共发现 286 个具体 KEGG 代谢途径分支参与半枫荷碳水化合物代谢、生命活动代谢、次生物质代谢和遗传信息处理过程中, 其中 232 个 Unigenes 参与到合成药效次生代谢产物中, 包括生物碱类、萜类、苯丙素类、黄酮类、醌类和酚类等, 基于以上半枫荷药用代谢产物合成途径, 初步构建了半枫荷活性成分基因调控网络。我们知道评价中药品质的优劣, 不止是看某一成分含量的高低, 而是成分与成分间相互影响的整体特征, 因此本研究的结果及后续将要开展的 EST-SSR 分子标记为半枫荷药效质量评价标准提供基础数据。另外, 本研究用半枫荷叶片做转录组分析, 而半枫荷根药效比叶更大, 可以预测半枫荷根相关药效次级代谢产物代谢强度更大, 但需进一步验证。本研究根据组装结果预测出 88 个基因家族共 1 547 个编码转录因子的 Unigenes, 这些家族在半枫荷生长发育过程中发挥着不同的作用, 其中  $C_2H_2$  锌指蛋白基因家族数量最多, 可能参与激素的信号传导、生长发育、抗病信号途径、细胞分化、胁迫应答等相关活动<sup>[34]</sup>。此外, 由于条件限制, 本研究仅仅是发现了转录因子与药效成分合成有关, 但具体调控的靶基因和作用机制还不清楚, 需要进一步研究。本研究通过转录组测序分析, 共检测出 125 917 个 SNP 多态位点, 包括 84 445 个转换突变类型, 41 472 个颠换突变类型, 即转换突变类型约占 SNP 的三分之二, 与杜玮南<sup>[35]</sup>等得出的结论一致。通过 SNP 研究将为今后开展半枫荷遗传图谱构建、物种起源进化、遗传多样性分析、品种鉴定以及分子标记辅助选择育种等研究提供重要参考。

通过生物信息学方法对半枫荷进行 CDS 预测、功能注释、代谢相关通路、SNP 检测和 TF 编码能力预测, 综合 KOG、GO 和 KEGG 三大功能数据库的注释结果, 表明半枫荷转录、复制和翻译等基因表达丰富度较高, 具有较强的代谢活动和遗传信息处理能力, 同时找出半枫荷药效物质可能相关的次生代谢产物生物合成途径及转录因子。所得转录组信息可作为金缕梅科植物基因组序列的

重要组成部分,进一步丰富了该科植物的基因信息库,为进一步研究药用植物半枫荷的代谢途径,以及该物种的分子遗传育种和基因工程提供基础数据,也将为保护和利用这一濒危药用植物资源奠定基础。

## 参 考 文 献

- Lander E S, Linton L M, Birren B, et al. Initial sequencing and analysis of the human genome[ J ]. Nature, 2001, 409 ( 6822 ) : 860 - 921.
- Zeng S H, Xiao G, Guo J, et al. Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* ( Sieb. et Zucc. ) Maxim[ J ]. BMC Genomics, 2010, 11 ( 1 ) : 94.
- Barrero R A, Chapman B, Yang Y F, et al. De novo assembly of *Euphorbia fischeriana* root transcriptome identifies prostratin pathway related genes[ J ]. BMC Genomics, 2011, 12 ( 1 ) : 600.
- Xiang L, Li Y, Zhu Y J, et al. Transcriptome analysis of the *Ophiocordyceps sinensis* fruiting body reveals putative genes involved in fruiting body development and cordycepin biosynthesis[ J ]. Genomics, 2014, 103 ( 1 ) : 154 - 159.
- Guo X, Li Y, Li C F, et al. Analysis of the *Dendrobium officinale* transcriptome reveals putative alkaloid biosynthetic genes and genetic markers[ J ]. Gene, 2013, 527 ( 1 ) : 131 - 138.
- 于永福. 中国野生植物保护工作的里程碑——《国家重点保护野生植物名录(第一批)》出台[ J ]. 植物杂志, 1999 ( 5 ) : 3.  
Yu Y F. National key protected wild plants list ( first batch ) [ J ]. Plants, 1999 ( 5 ) : 3.
- 环境保护部, 中国科学院. 中国生物多样性红色名录高等植物卷[ M ]. 北京: 环境保护部, 中国科学院, 2013.  
Ministry of Environmental Protection, Chinese Academy of Sciences. China biodiversity red list higher plant volumes [ J ]. Beijing: Ministry of Environmental Protection, Chinese Academy of Sciences, 2013.
- 傅立国. 中国植物红皮书: 稀有濒危植物[ M ]. 北京: 科学出版社, 1991.  
Fu L G. Red book of Chinese plants: rare and endangered plants [ M ]. Beijing: Science Press, 1991.
- 赵厚涛, 宋培浪, 韩国营, 等. 国家Ⅱ级保护植物半枫荷的最新研究进展[ J ]. 北方园艺, 2010 ( 21 ) : 210 - 212.  
Zhao H T, Song P L, Han G Y, et al. The latest research progress on *Semiliquidambar cathayensis* H. T. Chang [ J ]. Northern Horticulture, 2010 ( 21 ) : 210 - 212.
- 林华. 半枫荷药材质量标准的规范化研究[ D ]. 广州: 广州中医药大学, 2008.  
Lin H. Study on standardization in medicinal quality criteria of *Pterospermum heterophyllum* [ D ]. Guangzhou: Guangzhou University of Chinese Medicine, 2008.
- 张颖, 俞喆, 牛蓓, 等. 药用半枫荷植物资源研究[ J ]. 中国农学通报, 2008, 24 ( 8 ) : 432 - 434.  
Zhang Y, Yun Z, Niu B, et al. Studies on the Resources of Medicinal Plants Ban Fenghe [ J ]. Chinese Agricultural Science Bulletin, 2008, 24 ( 8 ) : 432 - 434.
- 《全国中草药汇编》编写组. 全国中草药汇编. 上册[ M ]. 北京: 人民卫生出版社, 1975.  
Compilation Group of the National Compilation of Chinese Herbal Medicines. National compilation of Chinese herbal medicine ( first volume ) [ M ]. Beijing: People's Medical Publishing House, 1975.
- 胡君, 赵中国, 刘庆, 等. 四川省国家Ⅱ级保护植物新记录——半枫荷[ J ]. 四川林业科技, 2018, 39 ( 1 ) : 63 - 65.  
Hu J, Zhao Z G, Liu Q, et al. A new record of the national secondary protection species——*Semiliquidambar cathayensis* H. T. Chang in Sichuan Province [ J ]. Journal of Sichuan Forestry Science and Technology, 2018, 39 ( 1 ) : 63 - 65.
- 廖娜, 韦柳斌, 金旺, 等. 金缕半枫荷多酚提取及其抗氧化抗菌活性研究[ J/OL ]. 广西植物. [ 2018 - 11 - 12 ]. <http://kns.cnki.net/kcms/detail/45.1134.Q.20180522.1744.008.html>.  
Liao N, Wei L B, Jin W, et al. Extraction technology and the antioxidant and antibacterial activities of polyphenols from *Semiliquidambar cathayensis* [ J/OL ]. Guihaia. [ 2018 - 11 - 12 ]. <http://kns.cnki.net/kcms/detail/45.1134.Q.20180522.1744.008.html>.
- 周光雄, 杨永春, 石建功, 等. 金缕半枫荷化学成分研究[ J ]. 中草药, 2002, 33 ( 7 ) : 589 - 591.  
Zhou G X, Yang Y C, Shi J G, et al. Studies on chemical constituents from *Semiliquidambar cathayensis* [ J ]. Chinese Traditional and Herbal Drugs, 2002, 33 ( 7 ) : 589 - 591.
- 孙静, 郝雪凌, 崔向珍, 等. 半枫荷抗乙型肝炎病毒的药理活性研究[ J ]. 时珍国医国药, 2014, 25 ( 10 ) : 2391 - 2393.  
Sun J, Zheng X L, Cui X Z, et al. A study of inhibition effects on hepatitis B virus of *Semiliquidambar cathayensis* in vitro [ J ]. Lishizhen Medicine and Materia Medica Research, 2014, 25 ( 10 ) : 2391 - 2393.
- 谢金兰, 王运昌, 范剑明. 浅谈半枫荷的资源开发和保护[ J ]. 农技服务, 2007, 24 ( 12 ) : 85.  
Xie J L, Wang Y C, Fan J M. On the development and protection of *Semiliquidambar cathayensis* resources [ J ]. Services of Agricultural Technology, 2007, 24 ( 12 ) : 85.

18. Haas B J ,Papanicolaou A ,Yassour M ,et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis[ J ]. Nature Protocols 2013 8( 8 ) :1494 - 1512.
19. Moriya Y ,Itoh M ,Okuda S ,et al. KAAS :an automatic genome annotation and pathway reconstruction server[ J ]. Nucleic Acids Research 2007 35( S2 ) :W182 - W185.
20. Rice P ,Longden I ,Bleasby A. EMBOSS :the European molecular biology open software suite[ J ]. Trends in Genetics 2000 16( 6 ) :276 - 277.
21. Mistry J ,Finn R D ,Eddy S R ,et al. Challenges in homology search :HMMER3 and convergent evolution of coiled-coil regions[ J ]. Nucleic Acids Research 2013 41( 12 ) :e121.
22. Quinlan A R ,Hall I M. BEDTools :a flexible suite of utilities for comparing genomic features[ J ]. Bioinformatics , 2010 26( 6 ) :841 - 842.
23. 袁灿 ,彭芳 ,杨泽茂 ,等. 川芎转录组 SSR 分析与 EST-SSR 标记的开发[ J ]. 中国中药杂志 ,2017 42( 17 ) :3332 - 3340.  
Yuan C ,Peng F ,Yang Z M ,et al. EST-SSR identification , markers development of *Ligusticum chuanxiong* based on *Ligusticum chuanxiong* transcriptome sequences[ J ]. China Journal of Chinese Materia Medica 2017 42( 17 ) :3332 - 3340.
24. Garg R ,Patel R K ,Tyagi A K ,et al. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification[ J ]. DNA Research 2011 18( 1 ) :53 - 63.
25. Patel S ,Shah D B ,Panchal H J. De novo RNA seq assembly and annotation of *Trigonella foenum-graecum* L. ( SRR066197 ) [ J ]. Legume Genomics and Genetics , 2014 5( 7 ) :1 - 7.
26. 李翠婷 ,张广辉 ,冯春花 ,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.  
Li C T ,Zhang G H ,Ma C H ,et al. Analysis on SSR loci information in transcriptome of *Panax vietnamensis* var. *fusicidiscus* and its polymorphism[ J ]. Chinese Traditional and Herbal Drugs 2014 45( 10 ) :1468 - 1472.
27. 王森 ,张震 ,姜倪皓 ,等. 半夏转录组中的 SSR 位点信息分析[ J ]. 中药材 2014 37( 9 ) :1566 - 1569.  
Wang S ,Zhang Z ,Jiang N H ,et al. SSR information in transcriptome of *Pinellia ternata* [ J ]. Journal of Chinese Medicinal Materials 2014 37( 9 ) :1566 - 1569.
28. 王东 ,曹玲亚 ,高建平. 党参转录组中 SSR 位点信息分析[ J ]. 中草药 2014 45( 16 ) :2390 - 2394.  
Wang D ,Cao L Y ,Gao J P. Data mining of simple sequence repeats in *Codonopsis pilosula* transcriptome[ J ]. Chinese Traditional and Herbal Drugs 2014 45( 16 ) :2390 - 2394.
29. 魏秀清 ,许玲 ,章希娟 ,等. 莲雾转录组 SSR 信息分析及其分子标记开发[ J ]. 园艺学报 ,2018 45( 3 ) :541 - 551.  
Wei X Q ,Xu L ,Zhang X J ,et al. Analysis on SSR information in transcriptome and development of molecular markers in wax apple[ J ]. Acta Horticulturae Sinica 2018 45( 3 ) :541 - 551.
30. Jiang X M ,Wu Y F ,Xiao F M ,et al. Transcriptome analysis for leaves of five chemical types in *Cinnamomum camphora* [ J ]. Hereditas 2014 36( 1 ) :58 - 68.
31. 张少平 ,洪建基 ,邱珊莲 ,等. 紫背天葵高通量转录组测序分析[ J ]. 园艺学报 2016 43( 5 ) :935 - 946.  
Zhang S P ,Hong J J ,Qiu S L ,et al. Sequencing and analysis of the transcriptome of *Gynura bicolor* [ J ]. Acta Horticulturae Sinica 2016 43( 5 ) :935 - 946.
32. 郑纪伟. 柳树转录组高通量测序及 SSR 标记开发研究 [ D ]. 南京 :南京林业大学 2013.  
Zheng J W. Research on transcriptome sequencing and SSR markers development of *Salix* [ D ]. Nanjing :Nanjing Forestry University 2013.
33. 蔡年辉 ,邓丽雨 ,许玉兰 ,等. 基于高通量测序的云南松转录组分析[ J ]. 植物研究 2016 36( 1 ) :75 - 83.  
Cai N H ,Deng L L ,Xu Y L ,et al. Transcriptome analysis for *Pinus yunnanensis* based on high throughput sequencing [ J ]. Bulletin of Botanical Research 2016 36( 1 ) :75 - 83.
34. 赵丽娟 ,易小娅 ,曾幼玲. 植物逆境相关 C<sub>2</sub>H<sub>2</sub> 型锌指蛋白的研究进展[ J ]. 分子植物育种 2016( 3 ) :578 - 585.  
Zhao L J ,Yi X Y ,Zeng Y L. Research progress on plant stress related C<sub>2</sub>H<sub>2</sub> type zinc finger protein[ J ]. Molecular Plant Breeding 2016( 3 ) :578 - 585.
35. 杜玮南 ,孙红霞 ,方福德. 单核苷酸多态性的研究进展 [ J ]. 中国医学科学院学报 2000 22( 4 ) :392 - 394.  
Du W N ,Sun H X ,Fang F D. The research development of single nucleotide polymorphism[ J ]. Acta Academiae Medicinae Sinicae 2000 22( 4 ) :392 - 394.