

徐萍, 张莹, 于梦雯, 等. 重楼全长转录组测序及生物信息学分析[J]. 江苏农业学报, 2025, 41(7): 1412-1421.

doi: 10.3969/j.issn.1000-4440.2025.07.017

## 重楼全长转录组测序及生物信息学分析

徐萍<sup>1</sup>, 张莹<sup>1</sup>, 于梦雯<sup>1</sup>, 徐洪高<sup>1</sup>, 涂振华<sup>1</sup>, 章晓叶<sup>1</sup>, 张洁<sup>1</sup>, 欧征刚<sup>2</sup>,  
陈佳<sup>3</sup>, 郑国伟<sup>1</sup>

(1. 云南中医药大学中药学院, 云南 昆明 650500; 2. 普拉底乡人民政府农业综合服务中心, 云南 怒江 673501; 3. 云南中医药大学民族医药学院, 云南 昆明 650500)

**摘要:** 本研究旨在通过对重楼全长转录组数据的获取和分析, 为系统挖掘重楼皂苷合成基因奠定基础。采用 PacBio Sequel 平台的第三代测序技术, 对皂苷含量差异显著的 3 种重楼属植物云南重楼 [*Paris polyphylla* Smith var. *yunnanensis* (Franch.) Hand.-Mazz.]、球药隔重楼 (*Paris fargesii* Franch.) 和长柱重楼 [*Paris forrestii* (Takht.) H. Li] 的根茎混合样品进行深度测序, 对所得数据进行全面的生物信息学分析。共获得 42 771 个转录本, 其中 40 945 个 (95.73%) 转录本成功比对到 KEGG、NR、Swiss-Prot、TrEMBL、KOG、GO 和 Pfam 等 7 大功能数据库。成功识别出与胆固醇合成相关的 48 个基因, 特异性筛选出参与重楼皂苷修饰的 2 类关键酶基因: 120 个细胞色素 P450 酶基因和 53 个 UDP-糖基转移酶基因。此外, 还检测到 2 327 个转录因子, 分别属于 71 个转录因子家族。在 MISA 分析中共发现 27 451 个微卫星序列 (SSR), 以 SSR 之间有重叠的复合类型最多, 占全部重复类型的 32.88%。综上, 本研究借助高通量全长转录组测序, 系统地解析了重楼根茎组织的转录谱, 并成功挖掘出与重楼皂苷合成密切相关的候选基因及调控元件, 为后续深入探讨重楼皂苷的生物合成分子机制奠定了基础。

**关键词:** 重楼; 全长转录组; 功能注释; 皂苷

**中图分类号:** R282.71 **文献标识码:** A **文章编号:** 1000-4440(2025)07-1412-10

## Full-length transcriptome sequencing and bioinformatics analysis of *Paris*

XU Ping<sup>1</sup>, ZHANG Xuan<sup>1</sup>, YU Mengwen<sup>1</sup>, XU Honggao<sup>1</sup>, TU Zhenhua<sup>1</sup>, ZHANG Xiaoye<sup>1</sup>,  
ZHANG Jie<sup>1</sup>, OU Zhenggang<sup>2</sup>, CHEN Jia<sup>3</sup>, ZHENG Guowei<sup>1</sup>

(1. College of Traditional Chinese Medicine, Yunnan University of Chinese Medicine, Kunming 650500, China; 2. Agricultural Comprehensive Service Center of Puladi Township People's Government, Nujiang 673501, China; 3. College of Ethnic Medicine, Yunnan University of Chinese Medicine, Kunming 650500, China)

**收稿日期:** 2024-09-04

**基金项目:** 国家自然科学基金地区科学基金项目 (82160718); 云南省科技厅基础研究专项面上项目 (202301AT070257); 云南省兴滇英才支持计划青年人才专项 (YNWR-QNBJ-2020-257、XDYC-QNRC-2022-0277); 云南省中医药基础研究联合专项 (202101AZ070001-057)

**作者简介:** 徐萍 (2000-), 女, 云南昭通人, 硕士研究生, 研究方向为中药材有效成分积累机制。(E-mail) 2633822198@qq.com

**通讯作者:** 陈佳, (E-mail) 394626580@qq.com; 郑国伟, (E-mail) gwzhengk@163.com

**Abstract:** The purpose of this study is to lay a foundation for the systematic exploration of saponin synthesis genes in *Paris* by obtaining the full-length transcriptome data of *Paris*. The third-generation sequencing technology on PacBio Sequel platform was used. In this study, the rhizome mixed samples of *Paris polyphylla* Smith var. *yunnanensis* (Franch.) Hand.-Mazz., *Paris fargesii* Franch. and *Paris forrestii* (Takht.) H. Li with significant differences in saponin content were deeply sequenced, and then the obtained data were comprehensively

analyzed by bioinformatics. A total of 42 771 transcripts were obtained, of which 40 945 (95.73%) were successfully matched to seven functional databases, namely KEGG, NR, Swiss-Prot, TrEMBL, KOG, GO and Pfam. Through enrichment analysis of KEGG pathway, 48 genes related to cholesterol synthesis were successfully identified, and two kinds of key enzyme genes involved in saponin modification were specifically screened: 120 cytochrome P450 enzyme genes and 53 UDP-glucosyltransferase genes. Additionally, 2 327 transcription factors were detected, belonging to 71 transcription factor families. A total of 27 451 microsatellite sequences (SSR) were found by MISA analysis, there were the most coincident types with overlapping SSR, accounting for 32.88% of all repeat types. With the help of high-throughput full-length transcriptome sequencing technology, the transcriptome of rhizome tissue of *Paris* was systematically analyzed, and the candidate genes and regulatory elements closely related to saponin synthesis were successfully excavated. These findings lay the groundwork for further study on the molecular mechanism of saponin biosynthesis in *Paris*.

**Key words:** *Paris*; full-length transcriptome; function annotation; saponin

重楼属(*Paris* L.)植物为藜芦科(Melanthiaceae)多年生草本植物。目前,中国有30种重楼属植物,占全球所有重楼属植物的81%<sup>[1]</sup>,其中云南重楼[*Paris polyphylla* Smith var. *yunnanensis* (Franch.) Hand.-Mazz.]和七叶一枝花[*P. polyphylla* Smith var. *chinensis* (Franch.) Hara]收载于2020年版《中国药典》,作为药用重楼的2种基原植物<sup>[2]</sup>;球药隔重楼(*P. fargesii* Franch.)则收载于《四川省藏药材标准》(2014年版)<sup>[3]</sup>。研究发现,长柱重楼[*P. forrestii* (Takht.) H. Li]和云南重楼、七叶一枝花具有相似的治疗功效<sup>[4]</sup>。因此,随着重楼药材资源的减少,在民间长柱重楼常作为重楼药材的替代品入药。重楼除了具有清热解毒、消肿止痛以及凉肝定惊的功效外<sup>[5]</sup>,还具有抗炎、抗氧化及抗肿瘤等药理活性<sup>[6]</sup>。甾体皂苷(重楼皂苷)为重楼的主要活性成分,约占总活性化合物的80%<sup>[7]</sup>。重楼皂苷主要分为薯蓣皂苷类和偏诺皂苷类,其中薯蓣皂苷类包括重楼皂苷I、重楼皂苷II、重楼皂苷III、重楼皂苷V和纤细薯蓣皂苷,偏诺皂苷类包括重楼皂苷VI、重楼皂苷VII、重楼皂苷H<sup>[7]</sup>。在不同重楼属植物中,皂苷的含量具有差异,云南重楼和长柱重楼主要含薯蓣皂苷类,球药隔重楼主要含偏诺皂苷类<sup>[8]</sup>。目前,从延龄草苷和偏诺皂苷-3-O-葡萄糖苷合成重楼皂苷V和重楼皂苷VI的途径已经得到了清晰的阐述<sup>[9]</sup>,但其余皂苷的合成通路仍在探索中,转录组测序对于找到重楼皂苷合成的关键基因具有重要的意义。

二代测序技术是转录组研究的主要方法,具有高通量、高灵敏度、应用范围广等优势<sup>[10]</sup>,然而其读取的DNA片段的长度较短,在转录本异构体的组装过程中存在较多的嵌合体错误,容易丢失可变剪切

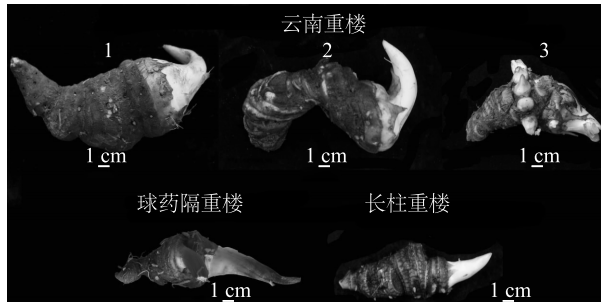
等重要信息,对后续的转录组数据分析造成影响<sup>[11]</sup>。随着三代单分子测序技术的兴起和应用,以PacBio为代表的三代全长转录组测序可以实现无需组装即可获得高质量的单个全长转录本序列,可以更真实地反映测序物种的转录组信息<sup>[12]</sup>。目前,全长转录组测序已广泛应用于药用植物有效成分合成相关基因的筛选中,如利用PacBio测序技术在艾(*Artemisia argyi*)中筛选到了12种与萜类化合物生物合成相关的候选基因<sup>[13]</sup>;地黄(*Rehmannia glutinosa*)的全长转录组测序结果显示有14种酶被注释到毛蕊花糖苷的合成途径中<sup>[14]</sup>;通过对滇黄精(*Polygonatum kingianum*)进行全长转录组测序筛选到6个与多糖合成相关的INV基因<sup>[15]</sup>;利用牛津纳米孔技术获得了甘青蒿(*Artemisia tangutica*)全长转录组数据,并筛选到9个与萜类化合物生物合成相关的候选基因<sup>[16]</sup>;Xu等<sup>[17]</sup>利用PacBio iso-seq和DNB-seq对毛茛(*Ranunculus japonicus*)进行全长转录组测序,筛选出9种类黄酮生物合成关键酶以及22种参与类黄酮和苯丙烷生物合成的MYB转录因子。本研究拟利用不同重楼属植物皂苷种类不同这一特点,对3种重楼属植物进行全长转录组测序,以便更好地发掘可能与重楼皂苷合成相关的基因,为重楼皂苷的体外合成提供一定的科学依据。

## 1 材料与方法

### 1.1 试验材料

从腾冲(23.92°N, 102.41°E, 2 110.9 m)、临沧(24.64°N, 99.76°E, 1 671.8 m)、文山(23.00°N, 104.00°E, 1 360.0 m)、红河(23.93°N, 102.41°E, 2 130.3 m)以及怒江(27.71°N, 98.72°E, 2 216.4 m)种植基地分别采集3种皂苷含量(高、中、低)的云

南重楼 [*Paris polyphylla* Smith var. *yunnanensis* (Franch.) Hand.-Mazz.]、球药隔重楼 (*P. fargesii* Franch.) 和长柱重楼 [*P. forrestii* (Takht.) H. Li] (图 1), 根茎洗净, 切片后立即放入液氮中速冻, 储存于  $-80\text{ }^{\circ}\text{C}$  冰箱中备用。



1: 低皂苷含量云南重楼; 2: 中等皂苷含量云南重楼; 3: 高皂苷含量云南重楼。

图 1 不同种重楼表型

Fig.1 Phenotypes of different species of *Paris*

## 1.2 RNA 的提取

参照试剂盒说明书, 分别提取各样品总 RNA, 使用 Agilent 2100 系统 (美国加利福尼亚州安捷伦科技公司产品) 和 Qubit RNA 检测试剂盒 (美国加利福尼亚州生命技术公司产品) 评估 RNA 的完整性和浓度。取适量每个样品的高质量 RNA, 将这些 RNA 样本合并, 形成一个 RNA 混合样本, 用于全长转录组测序。

## 1.3 文库构建和全长转录组测序

使用 SMARTer PCR cDNA Synthesis Kit 试剂盒从纯化的总 RNA 中合成全长 cDNA, 并通过 PCR 扩增以获得更多的双链 cDNA 模板。使用 AMPure XP 系统纯化 PCR 产物, 并在 Agilent Bioanalyzer 2100 系统上评估 SMRT bell 文库质量, 随后使用 PacBio 测序仪进行全长转录组测序。

## 1.4 SMRT 测序及数据处理

使用 SMRT link 软件对所得下机数据进行分析, 基于同一聚合酶读长拆分得到的子读长, 经过自我纠错形成一个环形共识序列 (CCS)。根据序列是否包含 5' 端引物、3' 端引物、polyA 尾来确定是否为全长非嵌合 (FLNC) 序列。通过 FLNC 聚类去冗余分析, 得到一致性转录本序列, 随后利用 arrow 算法对一致性序列进行校正, 以获得高质量的共识序列<sup>[18]</sup>, 利用 CD-HIT 软件进行聚类去冗余, 得到最终

全长转录本序列。

## 1.5 编码序列预测和基因功能注释

使用 TransDecoder 对去冗余后得到的转录本进行编码序列 (CDS) 预测, 得到和已知蛋白质库有同源性的编码框以及可信度最高的蛋白质编码框。使用 Diamond 软件<sup>[19]</sup> 将去冗余后的单基因序列与数据库 KEGG、NR、Swiss-Prot、TrEMBL、KOG、GO、Pfam 比对, 进行功能注释。

## 1.6 长链非编码 RNA 预测

对上述 7 大数据库 (KEGG、NR、Swiss-Prot、TrEMBL、KOG、GO、Pfam) 均未注释到的转录本, 使用编码潜能预测软件 CNCI<sup>[20]</sup>、CPC2<sup>[21]</sup>、PLEK<sup>[22]</sup> 进行编码潜能预测, 将 3 个软件 (CNCI、CPC2 及 PLEK) 一致预测为非编码 (Non-coding) 的序列视为潜在的长链非编码 RNA (LncRNA)。

## 1.7 微卫星序列和转录因子的鉴定

使用 MISA 软件<sup>[23]</sup> 鉴定转录本中的微卫星序列 (SSR), 并使用 Shi 等<sup>[24]</sup> 描述的方法分析重复基序类型的特征。基于转录因子 (TF) 家族的 Pfam 搜索结果, 使用 hmmscan 比对方式鉴定转录因子。

## 1.8 重楼皂苷合成基因的挖掘

依据前期关于重楼皂苷生物合成途径的研究报道<sup>[9, 25]</sup>, 从基因功能注释 KEGG 途径中筛选 2 个与重楼皂苷合成密切相关的候选代谢通路 (甾体生物合成途径、萜类骨架生物合成途径)<sup>[26-27]</sup>, 从候选途径中查找相关酶, 再根据酶查找酶的相应基因, 其中合成胆固醇的上游基因<sup>[28-29]</sup> 包括: *GPPs*、*FPPs*、*SQE*、*CAS*、*SMT*、*SMO*、*CPII*、*CYP51G*、*C14-R*、*8, 7 SI* 和 *CD-SD*, 从胆固醇到重楼皂苷主要由细胞色素 P450 酶基因 (*CYPs*) 和 UDP-糖基转移酶基因 (*UGTs*) 催化<sup>[30]</sup>。

## 2 结果与分析

### 2.1 PacBio 单分子实时测序数据分析

采用 PacBio SMRT 测序技术共获得了 92.11 Gb 数据。去除低质量序列和接头序列后, 得到 26 043 008 个子读长, 平均长度为 3 465 bp (表 1)。为了提供更准确和可靠的序列, 序列进行自我纠错后获得 639 807 个 CCS, 平均长度为 3 791 bp, 通过筛选最终得到了 578 320 个具有 poly A 的 FLNC 序列, 聚类去除冗余和序列校正后, 共获得 51 109 个共识序列。利用 CD-HIT 软件去冗余后得到 42 771 个最

终全长转录本,用于进一步功能注释,长度范围为118~12 195 bp,  $N50$  ( $N50$  指将所有序列按长度从长到短排序后,累加长度达到总序列长度一半时,所对应的序列长度)为4 073 bp,平均转录本长度为3 660 bp(表1),18.74%的转录本长度小于2 500 bp,55.82%的转录本长度为2 500~4 500 bp(图2)。

表1 重楼 PacBio SMRT 测序结果

Table 1 Summary of PacBio SMRT sequencing of *Paris*

类型	指标	数值	
聚合酶读长	数据量大小 (Gb)	92.11	
	数量	818 625	
	平均长度 (bp)	112 518	
	$N50$ (bp)	182 670	
子读长	数据量大小 (Gb)	90.21	
	数量	26 043 008	
	平均长度 (bp)	3 465	
	$N50$ (bp)	3 825	
环形共识序列	数量	639 807	
	平均长度 (bp)	3 791	
	最大长度 (bp)	14 700	
	最小长度 (bp)	80	
	$N50$ (bp)	4 039	
	全长非嵌合序列	全长序列数量	582 793
		全长非嵌合序列数量	578 816
具有 polyA 的全长非嵌合序列数量		578 320	
平均长度 (bp)		3 681	
最大长度 (bp)		14 623	
最小长度 (bp)		50	
$N50$ (bp)		3 937	
共识序列		数量	51 109
	平均长度 (bp)	3 657	
	最大长度 (bp)	12 196	
	最小长度 (bp)	118	
	$N50$ (bp)	4 039	
同源异构体	数量	42 771	
	平均长度 (bp)	3 660	
	最小长度 (bp)	118	
	最大长度 (bp)	12 195	
	$N50$ (bp)	4 073	

$N50$  指将所有序列按长度从长到短排序后,累加长度达到总序列长度一半时,所对应的序列长度。

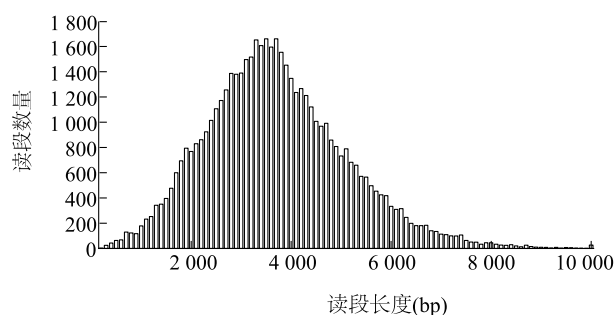


图2 同源异构体长度分布

Fig.2 Isoform size distribution

## 2.2 转录本功能注释

在42 771个转录本中,有40 945 (95.73%)个转录本成功比对到 KEGG、NR、Swiss-Prot、TrEMBL、KOG、GO 和 Pfam 数据库并进行了注释。分别注释了35 021个、40 488个、32 834个、40 525个、29 603个、36 425个、34 927个转录本(表2)。

表2 数据库注释结果统计

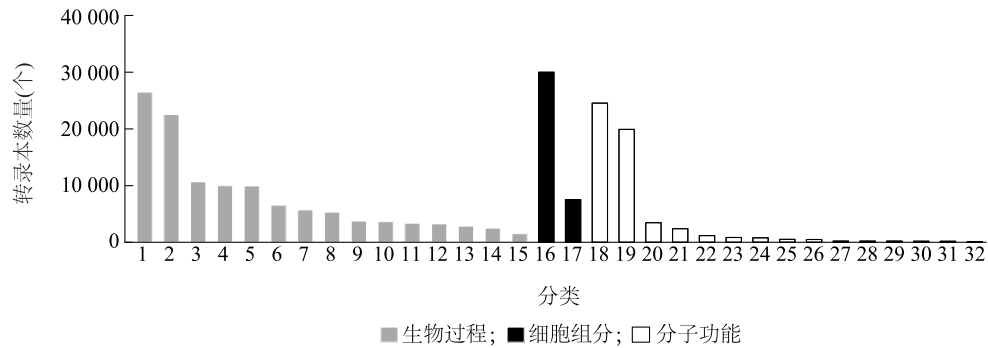
Table 2 Statistics of database annotation results

数据库	基因数量	占比 (%)
KEGG	35 021	81.88
NR	40 488	94.66
Swiss-Prot	32 834	76.77
TrEMBL	40 525	94.75
KOG	29 603	69.21
GO	36 425	85.16
Pfam	34 927	81.66
单基因总数	42 771	100.00

GO 分析结果显示,36 425个单基因被聚类为44个GO项,图3列出了32项,其中72.88% (26 548个)的单基因富集在细胞过程中,82.49% (30 047个)富集在细胞解剖实体中,67.50% (24 587个)富集于结合中。

KOG 注释结果显示比对为“一般功能预测”的重楼转录本最多(8 837个),其次为“信号转导机制”,转录本最少的为“细胞运动”,仅35个转录本(图4)。

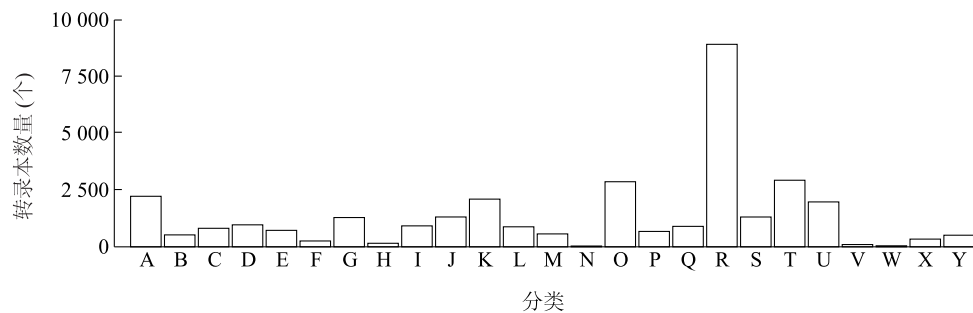
KEGG 富集分析将35 021个高质量转录本分为144个信号通路,涉及5个类别:代谢、有机系统、遗传信息处理、环境信息处理和细胞过程,其中次生代谢物的生物合成、淀粉和蔗糖代谢显著富集(图5)。



1:细胞过程;2:代谢过程;3:生物调节;4:刺激影响;5:生物过程调节;6:发育过程;7:多细胞生物过程;8:定位;9:生殖;10:生殖过程;11:信号;12:生物过程负调节;13:种间相互作用;14:生物过程正调节;15:生长;16:细胞解剖实体;17:含蛋白质的复合物;18:结合;19:催化活性;20:ATP(腺嘌呤核苷三磷酸)依赖的活动催化活性;21:转运活性;22:转录调节活性;23:结构分子活性;24:分子功能调节剂活性;25:翻译调节活性;26:分子传感器活性;27:小分子传感器活性;28:分子接头活性;29:细胞骨架运动活性;30:分子载体活性;31:蛋白质折叠伴侣;32:抗氧化活性。

图 3 GO 注释结果

Fig.3 GO annotation results



A:RNA 处理和修饰;B:染色体结构与动态;C:能量生产和转换;D:细胞周期控制、细胞分裂、染色体分割;E:氨基酸运输和代谢;F:核苷酸转运和代谢;G:碳水化合物运输和代谢;H:辅酶运输和代谢;I:脂质运输和代谢;J:翻译、核糖体结构和生物合成;K:转录;L:复制、重组和修复;M:细胞壁/膜/胞膜生物发生;N:细胞运动;O:翻译后修饰、蛋白质周转、分子伴侣;P:无机离子转运和代谢;Q:次级代谢物生物合成、转运和分解代谢;R:一般功能预测;S:未知功能;T:信号转导机制;U:细胞内运输、分泌和囊泡转运;V:防御机制;W:细胞外结构;X:核结构;Y:细胞骨架。

图 4 KOG 注释结果

Fig.4 KOG annotation results

NR 注释结果(图 6)显示,与重楼全长转录组基因序列匹配度最高的基因主要来源于海枣(24.45%)、油棕(24.29%)和椰子(6.94%)。

### 2.3 LncRNA 预测结果

LncRNA 指长度大于 200 核苷酸(nt)的非编码 RNA,在转录、剪接和翻译过程中发挥重要作用<sup>[31]</sup>,使用 CNCI、CPC2 和 PLEK 软件进行潜能编码预测,分别预测出 952 个、1 449 个和 781 个 LncRNA。在 3 种预测方法中,1 637 个 LncRNA 中仅有 340 个为 3 个软件共同预测到的 LncRNA(图 7)。

### 2.4 转录因子检测

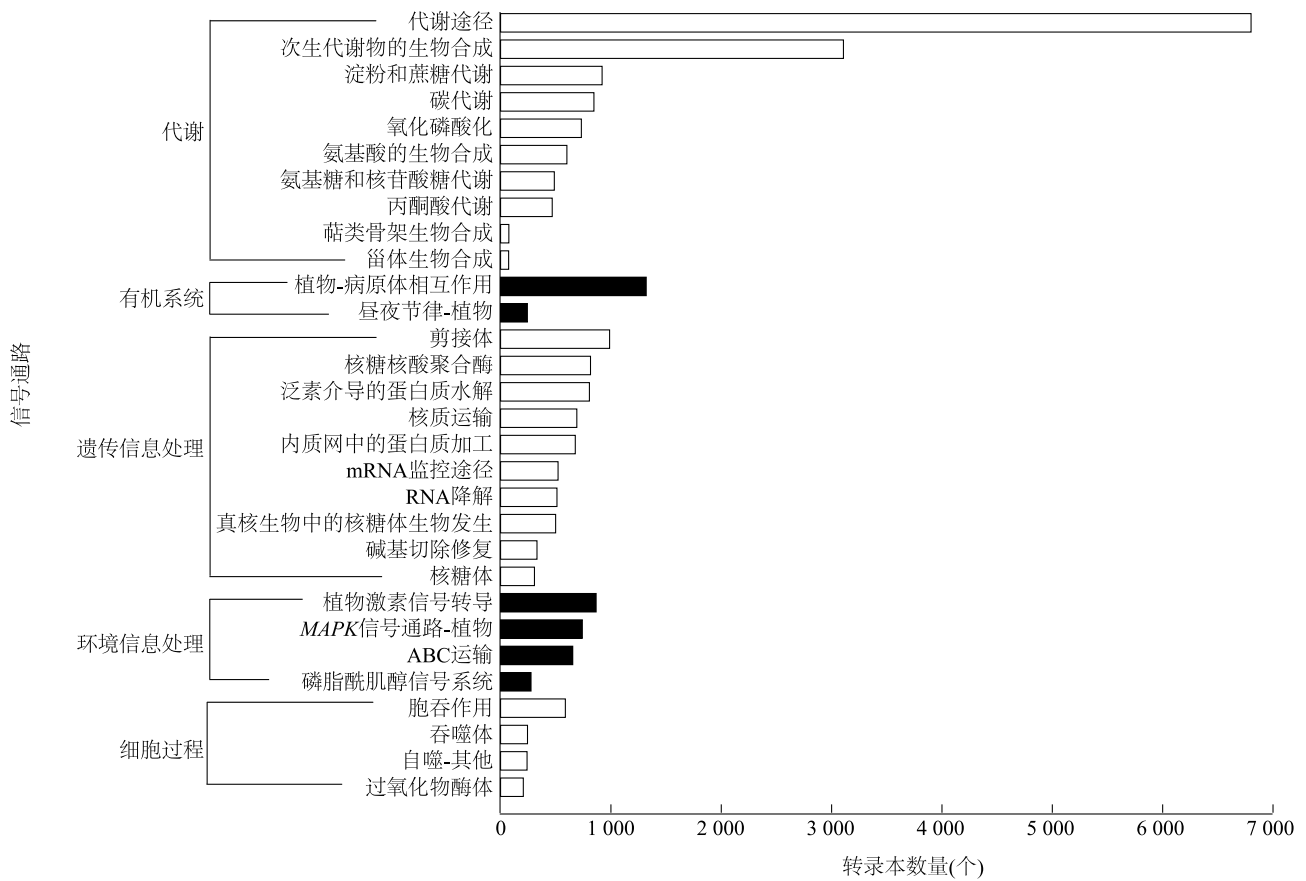
本研究共鉴定了 2 327 个转录因子(TF),分为 71 个家族,其中 SNF2(281 个)、FAR1(144 个)和

SET(131 个)家族数量最多,另外 MYB、AP2/ERF、bHLH、WRKY 和 bZIP 等转录因子可能参与了甾体皂苷的合成<sup>[26,32]</sup>,在本研究中分别鉴定了 58 个 MYB、19 个 AP2/ERF、24 个 bHLH、16 个 WRKY 和 32 个 bZIP 转录因子(图 8)。

### 2.5 SSR 检测

使用 MISA 工具共鉴定出 27 451 个 SSR,其中单碱基重复 7 899 个(28.77%),2 个碱基重复 5 372 个(19.57%),3 个碱基重复 3 314 个(12.07%),4 个碱基重复最少,仅占 0.16%(45 个)(表 3)。

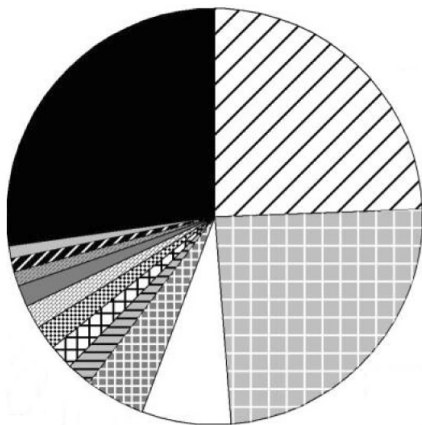
对所有 SSR 位点的统计分析结果表明,A/T 的单核苷酸重复基序数量最多,占有单核苷酸重复基序数量的 87.26%(12 829 个),而 C/G 仅占



MAPK: 丝裂原活化蛋白激酶; ABC: 三磷酸腺苷结合盒式蛋白。

图5 KEGG 注释结果

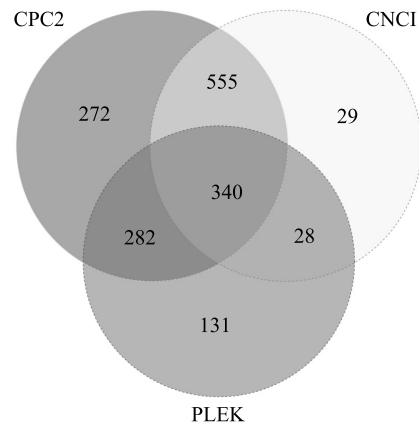
Fig.5 KEGG annotation results



- ▨ 海枣 (*Phoenix dactylifera*) (24.45%);
- ▩ 油棕 (*Elaeis guineensis*) (24.29%);
- 椰子 (*Cocos nucifera*) (6.94%); ▨ 参薯 (*Dioscorea alata*) (1.97%);
- ▨ 石刁柏 (*Asparagus officinalis*) (4.62%);
- ▨ 非洲圆薯蕷 (*Dioscorea cayenensis* subsp. *rotundata*) (1.95%);
- ▨ 葡萄 (*Vitis vinifera*) (1.88%); ▨ 凤梨 (*Ananas comosus*) (1.81%);
- ▨ 马来亚野蕉 (*Musa acuminata* subsp. *malaccensis*) (1.71%);
- ▨ 穴芭蕉 (*Musa troglodytarum*) (1.09%);
- ▨ 雷公藤 (*Tripterygium wilfordii*) (1.07%);
- ▨ 欧洲栓皮栎 (*Quercus suber*) (0.99%); ▨ 其他 (27.24%)

图6 NR 注释结果

Fig.6 NR annotation results

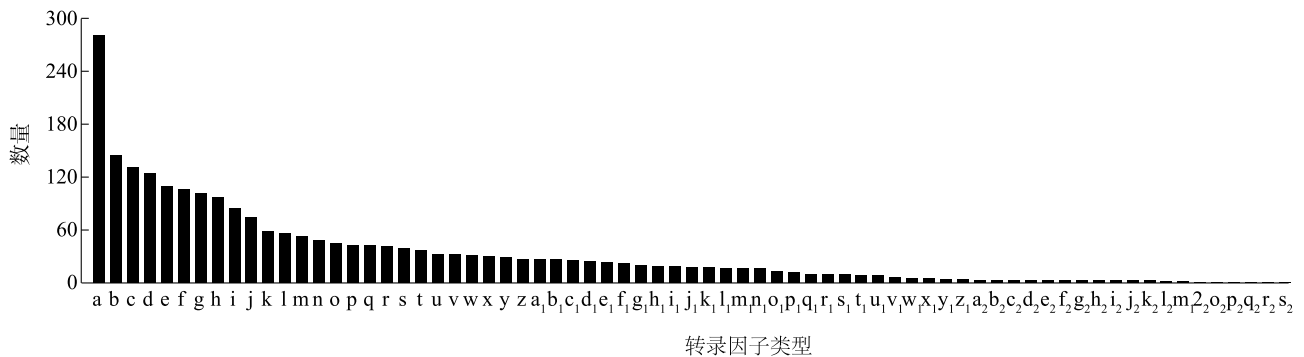


CNCI、CPC2 和 PLEK 为 3 种编码潜能预测软件。

图7 长链非编码 RNA (LncRNA) 预测结果

Fig.7 Prediction results of long non-coding RNA (LncRNA)

12.74% (1 873 个) (图 9a)。在二核苷酸重复基序中, AG/CT 的二核苷酸重复基序最多 (11 556 个, 63.81%), 其次是 AA/TT (4 434 个, 24.49%), 而 CG/CG 最少, 仅占 0.14% (25 个) (图 9b)。



a:SNF2;b:FAR1;c:SET;d:PHD;e:其他;f:C3H;g:IWS1;h:B3-ARF;i:Jumonji;j:GRAS;k:MYB;l:SBP;m:B3;n:MYB-related;o:C2H2;p:TRAF;q:SWI/SNF-BAF60b;r:AUX/IAA;s:CAMTA;t:DDT;u:LUG;v:bZIP;w:EIL;x:ARID;y:NAC;z:Trehelix;a<sub>1</sub>:HB-BELL;b<sub>1</sub>:C2C2-GATA;c<sub>1</sub>:HB-HD-ZIP;d<sub>1</sub>:bHLH;e<sub>1</sub>:TAZ;f<sub>1</sub>:HB-other;g<sub>1</sub>:RWP-RK;h<sub>1</sub>:RB;i<sub>1</sub>:AP2/ERF-AP2;j<sub>1</sub>:SWI/SNF-SWI3;k<sub>1</sub>:GARP-G2-like;l<sub>1</sub>:NF-X1;m<sub>1</sub>:WRKY;n<sub>1</sub>:CPP;o<sub>1</sub>:PLATZ;p<sub>1</sub>:mTERF;q<sub>1</sub>:zn-clus;r<sub>1</sub>:HSF;s<sub>1</sub>:BES1;t<sub>1</sub>:TUB;u<sub>1</sub>:LOB;v<sub>1</sub>:GNAT;w<sub>1</sub>:HB-PHD;x<sub>1</sub>:GARP-ARR-B;y<sub>1</sub>:Tify;z<sub>1</sub>:GeBP;a<sub>2</sub>:ULT;b<sub>2</sub>:STAT;c<sub>2</sub>:Pseudo ARR-B;d<sub>2</sub>:MBF1;e<sub>2</sub>:MADS-M-type;f<sub>2</sub>:HMG;g<sub>2</sub>:CSD;h<sub>2</sub>:C2C2-Dof;i<sub>2</sub>:BSD;j<sub>2</sub>:BBR-BPC;k<sub>2</sub>:Alfin-like;l<sub>2</sub>:NF-YA;m<sub>2</sub>:LIM;n<sub>2</sub>:VOZ;o<sub>2</sub>:SRS;p<sub>2</sub>:HB-WOX;q<sub>2</sub>:HB-KNOX;r<sub>2</sub>:DBP;s<sub>2</sub>:C2C2-YABBY。

图 8 转录因子的类型分布

Fig.8 Distribution of transcription factor types

表 3 微卫星序列 (SSR) 的统计分析

Table 3 Statistical analysis of microsatellite sequences (SSRs)

SSR 类型	数量	占比 (%)
c	1 567	5.71
c*	9 027	32.88
p1	7 899	28.77
p2	5 372	19.57
p3	3 314	12.07
p4	45	0.16
p5	49	0.18
p6	178	0.65

c 为复杂重复类型;c\* 为 SSR 之间有重叠的复合类型;p1 为单碱基重复;p2 为 2 个碱基重复;p3 为 3 个碱基重复;p4 为 4 个碱基重复;p5 为 5 个碱基重复;p6 为 6 个碱基重复。

### 2.6 重楼皂苷合成途径相关酶及其相应基因的挖掘

重楼属植物的主要活性成分为甾体皂苷,这些复杂的化合物来源于植物体内的萜类化合物<sup>[33]</sup>。通过 KEGG 代谢通路分析发现,73 个转录本参与了甾体生物合成途径,75 个转录本参与了萜类骨架生物合成(图 5),进一步分析发现,其中包含了 48 个与胆固醇合成相关的基因(表 4),除此之外还筛选到了 120 个 CYPs 基因以及 53 个 UGTs 基因(表 4)。

## 3 讨论

重楼属植物在民间被广泛用于治疗疗疮痈肿、蛇虫咬伤以及惊风抽搐等<sup>[1]</sup>,其主要活性成分为重

楼皂苷,具有抗炎、抗氧化和抗肿瘤等药理作用<sup>[34]</sup>。重楼皂苷主要通过甲羟戊酸(MVA)途径和 2-C-甲基-D-赤藓醇-4-磷酸(MEP)途径合成前体物质<sup>[33]</sup>,再由一系列细胞色素 P450 酶对前体分子结构进行特定位置的羟基化修饰,最终在糖基转移酶的催化下生成各类重楼皂苷<sup>[35]</sup>,但重楼皂苷的具体合成通路仍在探索中。鉴于重楼属植物体内所含重楼皂苷的种类与含量呈现出显著的种间差异<sup>[8]</sup>,故从这些具有特定皂苷组成特征的植物中系统地挖掘和鉴定与重楼皂苷生物合成相关的基因,以实现更为精确的目标基因识别,并优化合成路径,为提升药用植物品质及发现新药源提供科学支撑。采用转录组测序可以有效地预测药用植物的关键基因<sup>[36]</sup>。本研究通过对富含重楼皂苷 I 和重楼皂苷 II 的云南重楼、富含重楼皂苷 III 的长柱重楼和富含重楼皂苷 VI、重楼皂苷 VII、重楼皂苷 H 的球药隔重楼<sup>[8]</sup>根茎混合样品进行全长转录组测序,对序列信息进行分析。通过测序共获得了 42 771 个转录本,比 Hua 等<sup>[30]</sup>所测得的转录本数量多 2 896 个,进一步证明以不同重楼属植物作为试验材料,有利于筛选与重楼皂苷合成相关的基因。其中,95.73% 的转录本成功比对到 7 大数据库并进行了注释,通过 KEGG 富集通路筛选到 48 个与胆固醇合成相关的基因,同时还分别筛选到了 120 个 CYPs 基因以及 53 个 UGTs 基因,对这些与功能相关的基因进行深入的差异表达分析,不

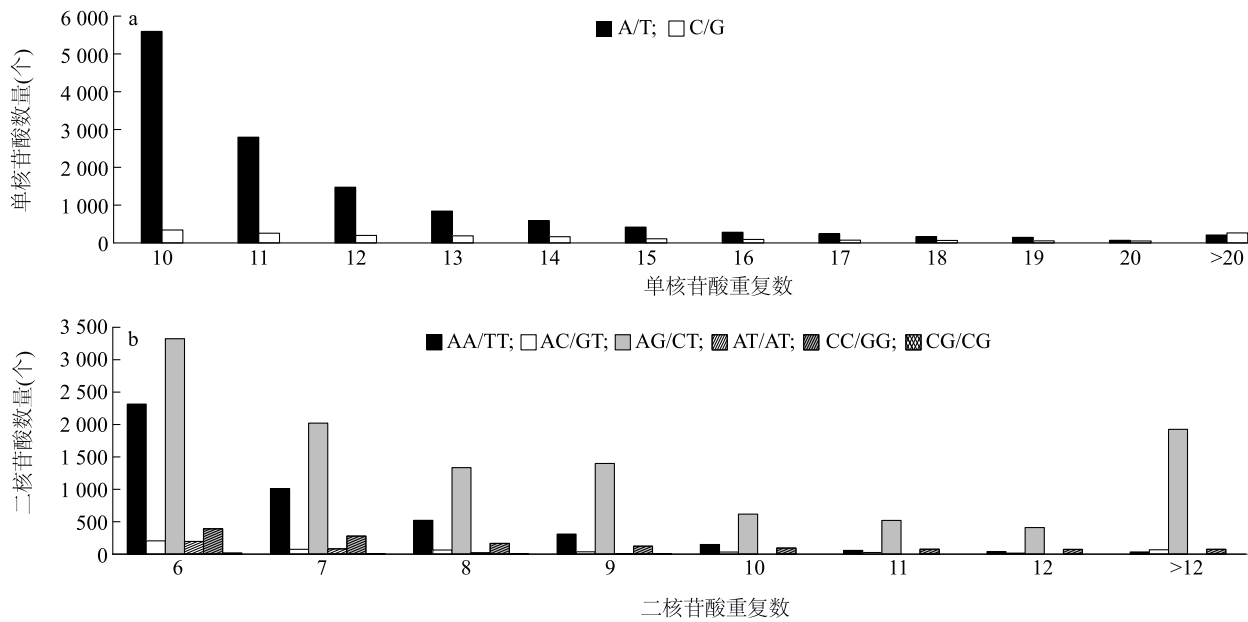


图9 单核苷酸和二核苷酸重复基序的类型和数量

Fig.9 The types and numbers of mononucleotide and dinucleotide repeat motifs

表4 重楼皂苷合成途径相关酶及其相应基因

Table 4 Enzymes and corresponding genes involved in saponin biosynthetic pathways in *Paris*

酶	基因	转录本数量(个)	通路号
香叶基焦磷酸合酶	<i>GPPs</i>	1	ko00900
法尼基焦磷酸合酶	<i>FPPs</i>	1	ko00900
鲨烯合酶	<i>SQE</i>	4	ko00100
环阿屯醇酶	<i>CAS</i>	6	ko00100
甾醇 24-C-甲基转移酶	<i>SMT</i>	2	ko00100
甾醇-4 $\alpha$ -甲基氧化酶	<i>SMO</i>	1	ko00100
环丙基甾醇异构酶	<i>CPII</i>	1	ko00100
甾醇 14 $\alpha$ -脱甲基酶	<i>CYP51G</i>	18	ko00100
甾醇 C-14 还原酶	<i>C14-R</i>	5	ko00100
甾醇 8,7 异构酶	<i>8,7 SI</i>	8	ko00100
甾醇 C-5(6) 去饱和酶	<i>CD-SD</i>	1	ko00100
细胞色素 P450 酶	<i>CYPs</i>	120	ko00996、ko00943、ko00941、ko00940、ko00909、ko00906、ko00905、ko00904、ko00902、ko00592、ko00402、ko00130、ko00073、ko00071
UDP-糖基转移酶	<i>UGTs</i>	53	ko00999、ko00944、ko00941、ko00908、ko00460

ko00900: 萜类骨架生物合成; ko00100: 甾体生物合成途径; ko00996: 各种生物碱的生物合成; ko00943: 异黄酮生物合成; ko00941: 类黄酮生物合成; ko00940: 类苯丙酸生物合成; ko00909: 倍半萜和三萜生物合成; ko00906: 类胡萝卜素生物合成; ko00905: 油菜素类固醇生物合成; ko00904: 二萜生物合成; ko00902: 单萜生物合成; ko00592:  $\alpha$ -亚麻酸代谢; ko00402: 苯并噁生物合成; ko00130: 泛醌和其他萜类醌生物合成; ko00073: 角质、木栓质和蜡的生物合成; ko00071: 脂肪酸降解; ko00999: 各种植物次生代谢产物的生物合成; ko00944: 黄酮和黄酮醇生物合成; ko00941: 类黄酮生物合成; ko00908: 玉米素生物合成; ko00460: 氨基酸代谢。

仅能够为解析重楼植物次生代谢产物的合成途径、积累动态及调控机理提供关键的理论依据,同时也有助于为将来针对重楼甾体皂苷类化合物生物合成相关基因的定向克隆工作储备有价值的候选基因。

转录因子在调节植物次生代谢产物积累、生理代谢和胁迫方面发挥着重要作用<sup>[37]</sup>。迄今,已在 MYB、WRKY、bHLH、bZIP 和 AP2/ERF 转录因子家族中发现了多种具有重要功能的转录因子成员<sup>[38]</sup>。

据推测,这些转录因子通过参与调控植物体内生物合成基因的表达来调节这些特殊代谢物的生物合成<sup>[39]</sup>,部分转录因子被认为与萜类和甾体化合物合成有关<sup>[40-43]</sup>。在本研究中共鉴定了2 327个转录因子,分别属于71个家族,其中包含58个MYB、19个AP2/ERF、24个bHLH、16个WRKY和32个bZIP。

综上,本研究通过重楼全长转录组测序及生物信息学分析,充实了重楼基因信息,为解析重楼活性成分代谢调控及关键基因筛选提供了重要参考。

## 4 结论

通过对重楼样本进行全长转录组测序并进行数据分析,成功获取了42 771个去冗余后的独立转录本序列。其中40 945个转录本成功比对至包括KEGG在内的7大生物信息学数据库,并完成了注释。这些单基因在细胞解剖实体、一般功能预测和次生代谢物的生物合成中占据较高比例。在NR数据库中,重楼与海枣和油棕的基因序列匹配度最高。本研究还预测识别出了27 451个微卫星序列(SSR),揭示了丰富的遗传变异信息源;同时鉴定出1 637个潜在的长链非编码RNA(LncRNA),以及2 327个可能参与调控基因表达的转录因子。筛选了149个与重楼皂苷合成相关的关键候选转录因子,鉴定了与胆固醇合成直接相关的基因48个,以及120个细胞色素P450酶基因和53个UDP-糖基转移酶基因。以上结果极大地拓宽了重楼的公共转录组数据库,并为鉴定参与重楼皂苷和其他次生代谢产物生物合成的候选基因提供了宝贵的资源。

### 参考文献:

- [1] 张翔宇,陈晓芳,柳敏,等.重楼属药用植物资源分布及少数民族应用研究[J].中国野生植物资源,2023,42(1):103-109,116.
- [2] 中华人民共和国国家药典委员会.中国药典(一部)[M].北京:中国医药科技出版社,2020:271-272.
- [3] 四川省食品药品监督管理局.四川省藏药材标准:2014年版[M].成都:四川科学技术出版社,2014:152-156.
- [4] 周姣姣,陶爱恩,何正春,等.不同生长年限长柱重楼根茎中主要次生代谢产物的累积变化[J].黑龙江农业科学,2018(3):121-125.
- [5] 王宇飞,江媛,杨成金,等.滇重楼化学成分、药理作用和临床应用研究进展[J].中草药,2022,53(23):7633-7648.
- [6] 张艺博,张慧中,付京,等.重楼总皂苷的现代研究进展与展望[J].中国实验方剂学杂志,2024,30(1):232-243.
- [7] GAO X Y, ZHANG X, CHEN W, et al. Transcriptome analysis of *Paris polyphylla* var. *yunnanensis* illuminates the biosynthesis and accumulation of steroidal saponins in rhizomes and leaves [J]. *Phytochemistry*, 2020, 178: 112460.
- [8] 李龙星,李海峰,王成军.9种重楼属植物叶中甾体皂苷成分差异的研究[J].中国民族民间医药,2023,32(19):12-18,29.
- [9] CHEN Y G, YAN Q, JI Y H, et al. Unraveling the serial glycosylation in the biosynthesis of steroidal saponins in the medicinal plant *Paris polyphylla* and their antifungal action [J]. *Acta Pharmaceutica Sinica B*, 2023, 13(11): 4638-4654.
- [10] ZHU Y X, HUANG Y F, WEI K H, et al. Full-length transcriptome analysis of *Zanthoxylum nitidum* (Roxb.) DC [J]. *Peer J*, 2023, 11: e15321.
- [11] 秦少伟,闫伟伟,李才林,等.基于三代测序技术的植物信号通路相关基因在胡杨异形叶发生中的功能分析[J].分子植物育种,2023,21(20):6688-6697.
- [12] ZHANG C C, REN H D, YAO X H, et al. Full-length transcriptome analysis of pecan (*Carya illinoensis*) kernels [J]. *G3: Genes, Genomes, Genetics*, 2021, 11(8): jkab182.
- [13] CUI Y P, GAO X Q, WANG J S, et al. Full-length transcriptome analysis reveals candidate genes involved in terpenoid biosynthesis in *Artemisia argyi* [J]. *Frontiers in Genetics*, 2021, 12: 659962.
- [14] 周延清,邵露蓉,李慧敏,等.地黄全长转录组分析及毛蕊花糖苷生物合成下游途径关键酶基因的挖掘与解析[J].植物生理学报,2022,58(1):197-206.
- [15] 米琪,赵艳丽,徐萍,等.滇黄精全长转录组测序及生物信息学分析[J].药学报,2024,59(6):1864-1872.
- [16] 余静雅,夏铭泽,张发起.青藏高原特有植物甘青蒿的全长转录组分析[J].中国野生植物资源,2023,42(12):17-24.
- [17] XU J Y, SHAN T Y, ZHANG J J, et al. Full-length transcriptome analysis provides insights into flavonoid biosynthesis in *Ranunculus japonicus* [J]. *Physiologia Plantarum*, 2023, 175(4): e13965.
- [18] OUYANG H L, WANG X Y, ZHENG X L, et al. Full-length SMRT transcriptome sequencing and SSR analysis of *Bactrocera dorsalis* (Hendel) [J]. *Insects*, 2021, 12(10): 938.
- [19] BUCHFINK B, XIE C, HUSON D H. Fast and sensitive protein alignment using DIAMOND [J]. *Nature Methods*, 2015, 12(1): 59-60.
- [20] JU Z G, LIANG L, ZHENG Y Q, et al. Full-length transcriptome sequencing and RNA-seq analysis offer insights into terpenoid biosynthesis in *Blumea balsamifera* (L.) DC [J]. *Genes*, 2024, 15(3): 285.
- [21] KANG Y J, YANG D C, KONG L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features [J]. *Nucleic Acids Research*, 2017, 45(W1): W12-W16.
- [22] JIAO N B, XU J R, WANG Y, et al. Genome-wide characterization of post-transcriptional processes related to wood formation in *Dalbergia odorifera* [J]. *BMC Genomics*, 2024, 25(1): 372.
- [23] FENG Y Z, ZHAO Y, ZHANG J J, et al. Full-length SMRT transcriptome sequencing and microsatellite characterization in *Paulow-*

- nia catalpifolia*[J]. Scientific Reports,2021,11(1):8734.
- [24] SHI Z Y, ZHAO W Q, LI Z A, et al. Development and validation of SSR markers related to flower color based on full-length transcriptome sequencing in *Chrysanthemum* [J]. Scientific Reports, 2022,12(1):22310.
- [25] XIANG Q A, ZHANG Q J, KANG F M, et al. Identification of genes involved in biosynthesis of spirostane-type polyphyllin in *Paris fargesii* based on transcriptome analysis[J]. Chinese Traditional and Herbal Drugs,2023,54(20):6798-6812.
- [26] LI Y, YANG H, LI Z H, et al. Advances in the biosynthesis and molecular evolution of steroidal saponins in plants[J]. International Journal of Molecular Sciences,2023,24(3):2620.
- [27] HOU L X, ZHANG F R, YUAN X C, et al. Comparative transcriptome analysis reveals key genes for polyphyllin difference in five *Paris* species [J]. Physiologia Plantarum, 2022, 174 ( 6 ): e13810.
- [28] YIN X, LIU J, KOU C X, et al. Deciphering the network of cholesterol biosynthesis in *Paris polyphylla* laid a base for efficient diosgenin production in plant chassis [J]. Metabolic Engineering, 2023,76:232-246.
- [29] GUO S Y, YIN Y, LEI T, et al. A cycloartenol synthase from the steroidal saponin biosynthesis pathway of *Paris polyphylla* [J]. Journal of Asian Natural Products Research, 2021, 23 ( 4 ): 353-362.
- [30] HUA X, SONG W, WANG K Z, et al. Effective prediction of biosynthetic pathway genes involved in bioactive polyphyllins in *Paris polyphylla*[J]. Communications Biology,2022,5(1):50.
- [31] WANG X Y, FAN H, WANG B S, et al. Research progress on the roles of lncRNAs in plant development and stress responses [J]. Frontiers in Plant Science,2023,14:1138901.
- [32] CHENG Q, ZENG L Q, WEN H, et al. Steroidal saponin profiles and their key genes for synthesis and regulation in *Asparagus officinalis* L. by joint analysis of metabolomics and transcriptomics [J]. BMC Plant Biology,2023,23(1):207.
- [33] 张 雪,王希付,赵荣华,等. 药用植物甾体皂苷生物合成途径研究进展[J]. 中国实验方剂学杂志,2020,26(14):225-234.
- [34] 王晓菲,李 鹏,杨 玥,等. 重楼皂苷抗炎、抗氧化及抗肿瘤作用机制研究进展[J]. 药物评价研究,2023,46(12):2699-2708.
- [35] 何 敏,郭思远,尹 艳,等. 滇重楼甾体糖基转移酶的克隆及功能表征[J]. 中国中药杂志,2023,48(14):3774-3785.
- [36] HE Y L, CHEN H, ZHAO J, et al. Transcriptome and metabolome analysis to reveal major genes of saikosaponin biosynthesis in *Bupleurum chinense*[J]. BMC Genomics,2021,22(1):839.
- [37] HUSSAIN A, KHAN M I, ALBAQAMI M, et al. *CaWRKY30* positively regulates pepper immunity by targeting *CaWRKY40* against *Ralstonia solanacearum* inoculation through modulating defense-related genes [J]. International Journal of Molecular Sciences,2021, 22(21):12091.
- [38] MERAJ T A, FU J Y, ALI RAZA M, et al. Transcriptional factors regulate plant stress responses through mediating secondary metabolism [J]. Genes,2020,11(4):346.
- [39] SHI M, ZHANG S W, ZHENG Z Z, et al. Molecular regulation of the key specialized metabolism pathways in medicinal plants [J]. Journal of Integrative Plant Biology,2024,66(3):510-531.
- [40] 肖 亮,陈瑞兵,吴 宇,等. AP2/ERF 转录因子调控次生代谢产物生物合成的研究进展 [J]. 中国中药杂志,2020,45(22):5412-5420.
- [41] LI J, YU H W, LIU M L, et al. Transcriptome-wide identification of WRKY transcription factors and their expression profiles in response to methyl jasmonate in *Platycodon grandiflorus* [J]. Plant Signaling & Behavior, ,2022,17(1):2089473.
- [42] HAN H, WANG C N, YANG X Y, et al. Role of bZIP transcription factors in the regulation of plant secondary metabolism [J]. Planta,2023,258(1):13.
- [43] LI Y L, WANG J, LI L Y, et al. Natural products of pentacyclic triterpenoids:from discovery to heterologous biosynthesis [J]. Natural Product Reports,2023,40(8):1303-1353.

(责任编辑:陈海霞)