

# 基于混合测序策略的两品种当归中类黄酮调控基因差异分析

徐丽<sup>1</sup>, 朱田田<sup>1,2,3\*</sup>, 晋玲<sup>1,2,3</sup>, 康舒淇<sup>1</sup>, 张明惠<sup>1</sup>, 王圆圆<sup>1</sup>

(1. 甘肃中医药大学, 兰州 730000; 2. 西北中藏药省部共建协同创新中心, 兰州 730000;  
3. 甘肃省珍稀中药资源评价与保护利用工程研究中心, 兰州 730000)

**[摘要]** 目的:比较2个不同茎叶色当归品种岷归1号与岷归2号转录水平差异。方法:以2种颜色当归的新鲜叶片(带叶柄)和上端茎为材料,采用混合测序策略,应用全长转录组技术构建当归无参全长转录本文库,利用RNA-seq技术对两品种进行差异基因表达分析,再利用公共数据库对差异基因的生物学功能进行注释和精细分类,筛选调控当归茎叶色差异的主要候选基因。结果:当归转录本的测序结果良好,测序数据质量较高,当归全长转录本的34 528条序列在非冗余蛋白(NR)、京都基因与基因组百科全书(KEGG)、SwissProt及KOG数据库中分别注释到33 947、33 241、29 150和22 601条。对两品种当归差异表达基因(DEGs)进行精细分类,具有生物学功能和分子功能的705个DGEs可分为11类,主要富集在初级代谢(17.87%)、逆境响应(14.47%)、次级代谢(11.49%)等功能上,与颜色有关的差异表达基因主要集中在类黄酮生物合成途径上。结论:两品种当归茎叶颜色差异的主要原因可能与调控类黄酮的生物合成基因的表达差异有关,可为后续进行功能验证,进一步明确与当归主要药效成分之间的联系奠定基础。

**[关键词]** 当归; 全长转录组; RNA测序技术; 类黄酮生物合成; 差异表达基因

**[中图分类号]** R284.2;R289;R22;R2-031;R33 **[文献标识码]** A **[文章编号]** 1005-9903(2023)01-0141-07

**[doi]** 10.13422/j.cnki.syfjx.20221514

**[网络出版地址]** <https://kns.cnki.net/kcms/detail/11.3495.R.20220914.1634.002.html>

**[网络出版日期]** 2022-09-15 14:24

## Differential Analysis of Flavonoid-regulated Genes in Two Cultivars of *Angelica sinensis* Based on Hybrid Sequencing Strategy

XU Li<sup>1</sup>, ZHU Tiantian<sup>1,2,3\*</sup>, JIN Ling<sup>1,2,3</sup>, KANG Shuqi<sup>1</sup>, ZHANG Minghui<sup>1</sup>, WANG Yuanyuan<sup>1</sup>

(1. Gansu University of Chinese Medicine, Lanzhou 730000, China;

2. Northwest Collaborative Innovation Center for Traditional Chinese Medicine (TCM), Lanzhou 730000, China;

3. Engineering Research Center for Evaluation, Protection, and Utilization of Rare TCM Resources, Gansu Province, Lanzhou 730000, China)

**[Abstract]** **Objective:** To compare the transcriptional levels of two cultivars (Mingui 1 and Mingui 2) with different stem and leaf colors. **Method:** The fresh leaves with petioles and the upper stems from *Angelica sinensis* with two colors were selected as the material, and the hybrid sequencing strategy was used. The technology at the transcriptome was used to build the non-reference full-length transcript library of *A. sinensis*, and the RNA-seq technology was used to analyze differentially expressed genes of the two cultivars, reusing a public database for biology function annotation and fine classification of differentially expressed genes. The main candidate genes regulating color differences between stems and leaves of *A. sinensis* were screened out. **Result:**

**[收稿日期]** 2022-05-03

**[基金项目]** 甘肃省科技计划项目(20JR5RA182);道地药材生态种植及质量保障项目(国中医药科技[2020]153号);甘肃省教育科技创新项目(GSSYLXM-05);甘肃中医药大学科学研究与创新基金项目(2021KCZD-4);西北中藏药省部共建协同创新中心开放基金(Xbzzy202207)

**[第一作者]** 徐丽,在读硕士,从事中药资源保护、评价与可持续利用研究,E-mail:2024756481@qq.com

**[通信作者]** \*朱田田,副教授,硕士生导师,从事中药资源评价与分子生药学研究,Tel:0931-5161169,E-mail:ztt0935@163.com

The sequencing results of *A. sinensis* transcripts were good and the quality of the sequencing data was high. The 34 528 full-length transcripts were annotated into 33 947, 33 241, 29 150, and 22 601 in the Non-redundant Protein Sequence Database (NR), Kyoto Encyclopedia of Genes and Genomes (KEGG), SwissProt, and Clusters of Orthologous Groups for Eukaryotic Complete Genomes (KOG), respectively. The 705 differentially expressed genes of the two cultivars with biological and molecular functions were divided into 11 categories, which were mainly enriched in the primary metabolism (17.87%), stress response (14.47%), and secondary metabolism (11.49%). The differentially expressed genes related to colors were mainly concentrated in the flavonoid biosynthesis pathway. **Conclusion:** The main reason for the color differences of the stems and leaves in two cultivars of *A. sinensis* may be related to the expression differences of genes regulating flavonoid biosynthesis, which lays a foundation for subsequent functional verification and further clarification of the relationship with the main pharmacodynamic components of *A. sinensis*.

**[Keywords]** *Angelica sinensis*; full-length transcriptome; RNA sequencing technology; flavonoid biosynthesis; differentially expressed genes

当归为伞形科植物当归的干燥根<sup>[1]</sup>,是甘肃省重要道地药材之一<sup>[2]</sup>。目前,岷归1号为甘肃省主栽品种,岷归2号的种植面积近年来也不断扩大。两个当归品种的表型差异主要表现为前者叶深绿色、叶柄和茎呈淡紫红色;后者叶柄、叶脉和茎均为绿色。前期研究表明,这种表型差异主要是由类黄酮成分代谢差异所决定的<sup>[3]</sup>。类黄酮类化合物是植物体内广泛存在的一种成分,除了决定植物色泽与品质,还参与植物的发育,具有抗氧化、抗紫外线、免受微生物侵害及提高植物的抗寒、抗病性功能<sup>[4-5]</sup>。在生产上,岷归1号具有品质优良、特征显著、农艺综合性状优良、高产等特点<sup>[6]</sup>;岷归2号的质量综合指标显著优于2020年版《中华人民共和国药典》规定标准,具有高产、稳定、抗病虫能力强、早期抽薹率低、抗逆性广等特点<sup>[7]</sup>。

许多学者也对不同品种当归的代谢物及质量差异进行了分析,表明其在药效成分含量上存在一定差异。荔淑楠等<sup>[8]</sup>利用代谢组学技术对五个当归品种进行分析,共鉴定出38个显著差异代谢物;王明伟等<sup>[9]</sup>研究表明当归药材品种不同,其主要药效成分含量也有不同。阿魏酸为2020年版《中华人民共和国药典》规定的当归药材含量控制指标成分之一,近年来,绿原酸也被认为是与当归质量相关的药效成分<sup>[10]</sup>。研究表明,当归中黄酮类与绿原酸、阿魏酸成分均属于苯丙烷代谢产物,通过莽草酸途径生成<sup>[11]</sup>,因此,推测当归中有效成分的合成可能受到类黄酮成分的影响,但其调控机制尚不明确。

转录组学是功能基因组的重要研究部分,能在生物体的整体转录水平上通过系统地研究基因转录图谱,从而揭示复杂的生物学通路和性状调控网

络的分子机制<sup>[12]</sup>。二代转录组技术,即RNA测序技术(RNA-seq)已广泛地用于分析表达谱<sup>[13]</sup>、鉴定差异表达基因<sup>[14]</sup>和分析剪切变异体<sup>[15]</sup>等,但受到读长的限制,在转录本组装过程中会存在较多的嵌合体,并且不能准确地得到完整转录本的信息<sup>[16]</sup>。近年来,第三代全长转录组测序技术(Iso-Seq)开始兴起,该技术弥补了RNA-seq技术的不足,在不需知道检测物种全基因组的情况下,仍可以直接得到样本全长转录本信息,但不足之处是不能完成差异表达基因的分析<sup>[17]</sup>。

因此,本研究采用混合测序策略,应用全长转录组技术构建当归无参全长转录本文库,再利用RNA-seq技术对紫茎当归和绿茎当归进行差异基因分析,同时发挥两种转录组测序技术的优势,以保证检测到的基因结构更准确、序列更完整、序列表达量更为准确。

## 1 材料

实验材料选取具有岷归1号和岷归2号的功能叶(带叶柄)和上端茎,见图1。所有植株均采集于定西市农业科学研究院种子繁育基地,位于定西市渭源县会川镇上葱沟村(海拔2 507 m;35°2'39"N,104°1'55"E)。所有样品擦洗干净后置于液氮带回实验室于-80℃超低温冰箱中保存,用于二代转录组学分析;全长转录组实验材料为二代转录组学分析剩余材料的混合样本,委托广州基迪奥生物科技有限公司进行三代全长转录组和二代转录组测序检测分析。所有样品均由甘肃中医药大学中药资源教研室晋玲教授鉴定为伞形科植物当归*Angelica sinensis*的茎和叶。



图1 岷归1号和岷归2号当归植株形态

Fig. 1 Plant morphology of Mingui 1 and Mingui 2 *Angelica sinensis*

## 2 方法

**2.1 全长转录本文库测序与校正** 选取下机数据中 full passes 数目 $\geq 1$ 的序列开展环型一致性序列 (CCS)分析,去除接头和 polyA、嵌合体序列后得到全长非嵌合序列 (FLNC),即为当归原始全长转录本序列。然后进行 Reads 聚类和校正,利用 Quiver 算法对 UC isoforms 进一步校正,根据输出的序列准确度 ( $\geq 0.99$ ) 预测高质量序列。

**2.2 全长转录本的基本注释** 对样本的全长转录组进行 NR (NCBI 非冗余蛋白序列数据库)、京都基因和基因组百科全书 (KEGG)、蛋白质真核直系同源簇数据库 (KOG) 及基因功能 (GO) 分类数据库注释等。首先,通过 blastx<sup>[18]</sup> 将 isoform 序列比对到蛋白数据库 NR、蛋白质序列数据库 (SwissProt) 和 KEGG,得到与给定 isoform 具有最高序列相似性的蛋白,在此基础上注释信息;再根据 NR 注释信息,使用 Blast2GO 软件得到 Isoform 的 GO 注释信息;最后统计所有 isoform 在 NR、SwissProt、KEGG 和 KOG 四大数据库中的比对 E 值 (isoform 与数据库中匹配序列为同源序列的假阳性概率)。将比对 E 值分为 5 个范围,统计每个范围内的基因个数,利用 blastx 将每个 isoform 在 NR 库中 E 值最低的序列作为对应同源序列,如果有并列序列则取第一条序列,最后比对确定到所属物种的同源序列数量。

**2.3 二代转录组文库构建与测序** 以被打断的 mRNA 为模版,加入随机寡核苷酸引物后在 M-MuLV 逆转录酶体系中合成 cDNA 第一条链,然后加入 RNaseH 将 RNA 链降解,以 dNTPs 为原料在 DNA polymerase I 体系下合成 cDNA 第二条链。末端修复纯化后的双链 cDNA,加 A 尾连接测序接头

后用筛选 200 bp 左右的 cDNA,进行聚合酶链式反应 (PCR) 扩增后再次进行 AMPure XP beads 纯化,完成测序样本文库构建。检测合格后的样本利用 Illumina HiSeq 4000 进行测序分析。

**2.4 差异表达基因分析** 以岷归 1 号和岷归 2 号样本  $|\log_2(\text{fold-change})| > 1$  和错误发现率 (FDR)  $< 0.05$  为标准,筛选出 DEGs,将 DEGs 在 SwissProt 数据库中进行比对,去除未鉴定到的和重复的 DEGs 后,再利用 Uniprot 数据库对鉴定的和去重后 DEGs 的生物学和分子功能进行检索,并将其精细分类。

## 3 结果与分析

**3.1 全长转录本原始数据统计与校正** 经 SMRT 测序和分析的岷归 1 号和岷归 2 号混合样品的原始数据,此转录本总碱基数为 49 491 865 361 bp, subreads 数量为 29 448 198 条, subreads 平均长度为 1 680 bp, N50 为 2 100 bp。经过 CCS 分析及 Quiver 计算校正得到了 45 026 条高质量序列 (准确度  $> 99\%$ ), 803 条低质量序列 (准确度  $< 99\%$ )。使用 LoRDEC 软件将得到的低质量序列与二代转录组测序得到的短 reads 进行校正,对校正后的转录本序列进一步进行去冗余,得到的当归样本的最终全长转录组序列数量 34 528 条,具体信息参数见表 1。由表 1 可知当归转录本的测序结果良好,测序数据质量较高,可用于进行下一步分析研究。

表 1 当归全长转录组序列信息 (校正后)

Table 1 Full-length transcriptome sequence information of *A. sinensis* (after correction) bp

转录本	SMRT 结果
总碱基数	72 410 837
最长序列碱基数	11 835
最短序列碱基数	53
序列平均碱基数	2 097
subreads 平均长度	1 680
N50	2 461

**3.2 比对率统计** 使用 RSEM<sup>[19]</sup> 软件对三代组装出来的 isoform 进行定量。所有样品未比对上参考基因的 reads 数占有效 reads 比例在 10% 左右,唯一比对上参考基因的 reads 数占有效 reads 比例均在 16% 以上,多处比对参考基因的 reads 数占有效 reads 比例在 72% 以上,全部的可以定位到基因上的 reads 数量占有效 reads 比例均在 88% 以上,比对参考统计见表 2,由此可知,当归转录组数据样本数据质量较高,可进行当归全长转录本的功能注释。

表2 当归比对参考统计

Table 2 *A. sinensis* comparison reference statistical table

样本	读长/bp	未比对上参考基因的读长数(比例)	唯一比对上参考基因的读长数(比例)	多重比对上参考基因的读长数(比例)	全部可以定位到基因上的读长数(比例)
M1-1	39 109 128	4 640 123(11.86)	6 270 466(16.03)	28 198 539(72.10)	34 469 005(88.14)
M1-2	39 694 294	4 710 866(11.87)	6 363 166(16.03)	28 620 262(72.10)	34 983 428(88.13)
M1-3	37 140 284	4 205 973(11.32)	5 992 902(16.14)	26 941 409(72.54)	32 934 311(88.68)
M2-1	36 548 970	3 828 383(10.47)	5 916 631(16.19)	26 803 956(73.34)	32 720 587(89.53)
M2-2	40 000 590	4 323 555(10.81)	6 431 430(16.08)	29 245 605(73.11)	35 677 035(89.19)
M2-3	39 651 314	4 207 661(10.61)	6 402 314(16.15)	29 041 339(73.24)	35 443 653(89.39)

3.3 全长转录本的基本注释 将当归全长转录本的34 528个序列在四大数据库NR、SwissProt、KEGG和KOG中进行比对和功能注释,最终有注释的全长转录本为34 003个,没有注释的为525个。由表3可知能注释到蛋白质数据库NR的全长转录本为33 947个(99.8%)、KEGG为33 241个(97.8%)、SwissProt数据库为29 150个(85.7%)、KOG为22 601个(66.5%)。

利用NR数据库比对发现在33 947个unigenes中被注释到表4所示前10个植物物种的同源序列有29 092个(85.7%),其余的4 855个(14.3%)unigenes被注释到其他物种中。

通过KOG数据库比对,被注释的22 601个unigenes的KOG功能可分为25个功能类别,见图2,其中一般功能预测基因(General function prediction only),翻译后的修饰、蛋白周转、分子伴侣

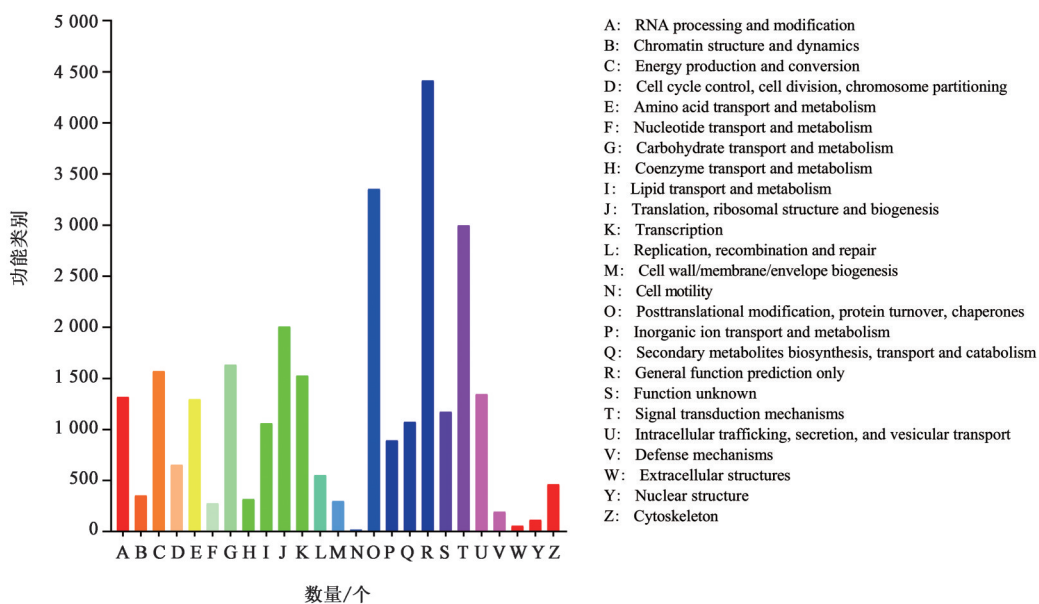
表3 当归全长转录组测序获得的转录本注释

Table 3 Transcripts annotated by full-length transcriptome sequencing of *A. sinensis*

公共数据库	注释数目/个	百分比/%
NR数据库	33 947	99.8
KEGG数据库	33 241	97.8
SwissProt数据库	29 150	85.7
KOG数据库	22 601	66.5
总数	34 003	100

(Posttranslational modification, proteinturnover, chaperones)和信号转导机制(Signal transduction mechanisms)3个功能的unigenes分布数量较多。

利用GO数据库比对分析后,发现有247 149个unigenes得到GO功能注释,其中分布在生物过程的GO terms有124 044个、分布在细胞组分的有89 160个、分布在分子功能的有33 945个,3个大类功能又被



注:横坐标表示KOG功能类别;纵坐标为长度在一定范围内的unigenes数量

图2 当归全长转录本KOG功能分类

Fig. 2 KOG functional classification of full-length transcripts of *A. sinensis*

表4 NR数据库注释前10个植物物种数量

Table 4 Number of top 10 plant species in NR database notes

物种	注释数目/个	百分比/%
胡萝卜( <i>Daucus carota</i> )	25 235	86.74
猕猴桃( <i>Actinidia chinensis</i> )	2 055	7.06
扁桃( <i>Prunus dulcis</i> )	347	1.19
茶树( <i>Camellia sinensis</i> )	311	1.07
番木瓜( <i>Carica papaya</i> )	236	0.81
当归( <i>A. sinensis</i> )	205	0.70
欧芹( <i>Petroselinum crispum</i> )	196	0.67
葡萄( <i>Vitis vinifera</i> )	175	0.60
欧洲油菜( <i>Brassica napus</i> )	171	0.60
黄花蒿( <i>Artemisia annua</i> )	161	0.55
总量	29 092	100

分为24、17、12个共53个小类功能,见图3。

**3.4 二代转录组测序分析** 为了进一步对岷归1号和岷归2号的差异表达基因进行分析,对M1和M2进行转录测序,并比较分析两组转录本。将约3%的低质量数据从原始数据中过滤后,其碱基的组成及质量分布较平衡,见表5,说明过滤后得到的转录本质量较高,可保证后续分析的准确性。

**3.5 岷归1号和岷归2号差异表达基因分析** 通过M1/M2转录比较,从34 528个unigenes中共获得2 210个DGEs,其中1 110个上调表达基因(UR),1 100个下调表达基因(DR);通过差异基因聚类热图可看出M1和M2的差异基因分组明显,说明统计得到的差异表达基因数目较准确。见图3。

进一步对表达基因进行精细分类,即从2 210个差异基因中去除1 099个重复基因,有1 111个基因在M1/M2中差异表达,其中有368个基因在SwissProt数据库中未鉴定。在743个已鉴定的DEGs中,38个无生物学特性,其余705个具有已知功能(267UR,438DR)。利用Uniprot数据库检索

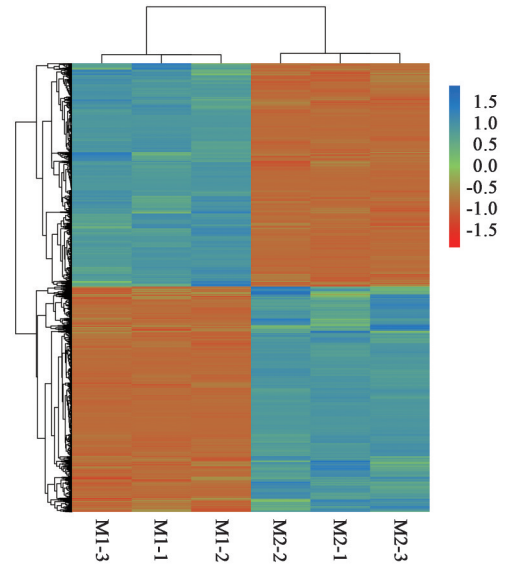


图3 岷归1号和岷归2号差异基因聚类热

Fig. 3 Mingui 1 and Mingui 2 differential gene clustering heat map

后,具有生物学功能和分子功能的705个DGEs可分为11类,其中72个为光合作用(Photosynthesis)和呼吸作用(Respiration)、126个为初级代谢(Primary metabolism)、81个为次级代谢(Secondary metabolism)、11个为激素生物合成(Hormone biosynthesis)、48个为细胞形态发生(Cell morphogenesis)、72个为生物信号(Bio-signaling)、24个为多核苷酸生物合成(Polynucleotide biosynthesis)、29个为转录因子(Transcription factors)、59个具有翻译(Translation)功能、81个具有转运(Transport)功能及105个为逆境响应(Stressresponse)。

KO功能富集结果显示,岷归1号和岷归2号颜色调控基因参与的最主要的生化代谢途径和信号转导途径富集于103个KO条目,主要包括氧化磷酸化(Oxidative phosphorylation)、代谢途径(Metabolic pathways)、亚麻酸代谢(Linolenic acid metabolism)、

表5 当归二代数据处理

Table 5 Data processing table of *A. sinensis* second generation

样本	高质量有效数据量数目 (比例)	含有 adapter 的有效数据量 数目(比例)	低质量有效数据量 数目(比例)	polyA 有效数据量 数目(比例)	含N比率高于过滤有效 数据量数目(比例)
M1-1	39 109 128(98.38)	272 560(0.69)	372 340(0.94)	0(0.00)	0(0.00)
M1-2	39 694 294(98.38)	95 140(0.24)	556 534(1.38)	0(0.00)	0(0.00)
M1-3	37 140 284(98.27)	196 342(0.52)	458 044(1.21)	0(0.00)	0(0.00)
M2-1	36 548 970(97.76)	334 484(0.89)	504 052(1.35)	0(0.00)	0(0.00)
M2-2	40 000 590(97.88)	401 702(0.98)	466 350(1.14)	0(0.00)	0(0.00)
M2-3	39 651 314(98.58)	188 304(0.47)	383 874(0.95)	0(0.00)	2(0.00)

个(%)

ABC转运蛋白(ABC transporters)、丙酮酸代谢(Pyruvate metabolism)等生化代谢途径或信号转导途径,前20个条目见增强出版附加材料。

对次生代谢功能中的DEGs进行比对,发现8个DEGs与类黄酮生物合成有关,其中6个DEGs表达下调,2个DEGs表达上调。见表6。

表6 次级代谢中参与类黄酮生物合成的DEGs及表达量

Table 6 DEGs and their expression levels involved in flavonoid biosynthesis in secondary metabolism

序号	基因名称	蛋白质名称	SwissProt数据库中序列名	差异倍数(岷归1号/岷归2号)
1	<i>CHS1</i>	chalcone synthase 1	Q9ZS41	-8.634
2	<i>CHI3</i>	probable chalcone-flavonone isomerase 3	Q8VZW3	-1.056
3	<i>FHT</i>	flavanone 3-dioxygenase	Q7XZQ7	-1.973
4	<i>DFR</i>	dihydroflavonol 4-reductase	P51105	-6.504
5	<i>F3GT1</i>	anthocyanidin 3-O-galactosyltransferase F3GT1	A0A2R6Q8R5	-1.174
6	<i>UGT73C6</i>	UDP-glycosyltransferase 73C6	Q9ZQ95	1.525
7	<i>ANS</i>	leucoanthocyanidin dioxygenase	P51091	-7.506
8	<i>3MAT</i>	malonyl-coenzyme A: anthocyanin 3-O-glucoside-6"-O-malonyltransferase	Q8GSN8	1.280

## 4 讨论

**4.1 当归转录组学与测序分析** 与拟南芥、玉米、小麦、高粱等<sup>[20-21]</sup>模式植物和农作物相比,药用植物的全长转录组研究相对较缓慢,仅在药用模式植物丹参的研究上相对较深入<sup>[22]</sup>。由于绝大多数非模式植物的全基因组数据尚未获得,因此全长转录本的获取可有力地推进非模式植物功能基因组的基础研究水平<sup>[23]</sup>。与二代RNA-seq技术相比,三代Iso-Seq测序技术可以构建质量更高的参考序列文库,如果将二代RNA-seq测序结果与三代Iso-Seq测序结果联合分析,获得的差异表达基因和数据分析结果将具有更高的可信度和准确性<sup>[17]</sup>。胡豪等<sup>[24]</sup>利用二代测序和SMRT全长转录组测序数据,通过靶基因的功能注释和富集分析,揭示lncRNA在红苞凤梨叶片细胞失绿白化及生长发育过程中的作用,为其表观遗传调控机制研究提供参考。王丰青等<sup>[25]</sup>基于二代转录组测序获得地黄转录组数据,利用Pacific Biosciences RS II平台进行测序获得三代全长转录组,推导并鉴定出参与苯乙醇苷成分生物合成的催化酶基因及其序列,为进一步研究地黄苯乙醇苷类成分生物合成的分子机制提供了参考。因此,本研究采用二代转录组和三代无参转录组联合分析,获得的当归三代全长转录本序列数量34 528条,总碱基数72 410 837 bp,最长序列碱基数11 835 bp,最短序列碱基数53 bp,序列平均碱基数2 097 bp,N50为2 461 bp,GC含量41.02%。

**4.2 当归全长转录本功能注释** 利用单分子实时

测序技术对岷归1号和岷归2号的茎叶进行无参全长转录本分析,通过进行当归全长转录组比对率统计分析,当归转录组数据样本数据质量较高,可用于进行当归全长转录本的功能注释。将获得的全长转录本通过蛋白质功能注释、KOG功能注释、GO功能注释等,可以进一步挖掘基因的功能以及在细胞中的代谢途径,约90%以上的转录本在数据库中得到功能注释,注释信息基本符合当归遗传信息的整体表达特征。通过以上数据库的注释,确定了与当归同源序列排在前十位的物种,明确了当归全长转录本的KOG功能类别,系统分析了当归基因产物在细胞中的代谢途径以及这些基因的功能。

**4.3 岷归1号和岷归2号差异表达基因分析** 当归全长转录本作参考,应用二代RNA-seq技术对岷归1号和岷归2号进行转录比较,共获得2 210个DGEs(1 110个上调,1 100个下调),差异基因聚类热图分析显示得到的差异表达基因数目较准确;通过对表达基因的精细分类,共有705个(267个上调,438个下调)具有生物学功能和分子功能,两种颜色当归的DGEs主要富集在初级代谢(17.87%)、逆境响应(14.47%)以及次级代谢(11.49%)等功能上;对次生代谢功能中的DEGs进行比对,发现*CHS1*、*CHI3*、*FHT*、*DFR*、*F3GT1*、*UGT73C6*、*ANS*和*3MAT* 8个基因都与类黄酮的生物合成有关,由此推测,岷归1号和岷归2号茎叶中的次生代谢物差异性主要表现在类黄酮的生物合成上,这也可能是导致2种当归茎叶色差异的主要原因,可为当归中调控类黄酮关键

基因的功能验证以及阿魏酸、绿原酸等药效成分的调控机制研究奠定基础。

[利益冲突] 本文不存在任何利益冲突。

[参考文献]

- [1] 国家药典委员会. 中华人民共和国药典:一部[M]. 北京:中国医药科技出版社,2020:139.
- [2] 中国科学院《中国植物志》编辑委员会. 中国植物志(第五十五卷第三分册)[M]. 北京:科学出版社,1992:41-43.
- [3] ZHU T T, ZHANG M H, SU H Y, et al. *Angelica sinensis* integrated metabolomic and transcriptomic analysis reveals differential mechanism of flavonoid biosynthesis in two cultivars[J]. *Molecules*, 2022, 27(1): 306-323
- [4] DIXON R A, ACHNINE L, KOTA P, et al. The phenylpropanoid pathway and plant defence-a genomics perspective[J]. *Mol Plant Pathol*, 2002, 3(5):371-390.
- [5] FORKMANN G, MARTENS S. Metabolic engineering and applications of flavonoids [J]. *Curr Opin Biotechnol*, 2001, 12(2):155-160.
- [6] 贾婕楠. 当归新品系 90-01 选育报告[J]. *中药材*, 2005, 28(8):649-650.
- [7] 颜红梅,刘效瑞,荆彦明,等. 甘肃当归 90-02 选育研究[J]. *中国现代中药*, 2007, 9(3):37-38, 42.
- [8] 荔淑楠,王引权,王富胜,等. 基于 UPLC-Q-TOF-MS 技术的不同品种当归代谢组学分析[J]. *中国实验方剂学杂志*, 2020, 26(8):138-147.
- [9] 王明伟,李硕,李敏,等. 基于熵权 TOPSIS 模型对当归不同栽培品种(品系)药材质量的综合评价[J]. *中国实验方剂学杂志*, 2017, 23(5):63-68.
- [10] 徐小琼,陈博,张小波,等. 甘肃各主产区当归 UPLC 多成分含量测定及评析[J]. *天然产物研究与开发*, 2020, 32(6):1014-1022.
- [11] 王星,罗双霞,于萍,等. 茄科蔬菜苯丙烷类代谢及相关酶基因研究进展[J]. *园艺学报*, 2017, 44(9):1738-1748.
- [12] 崔凯,吴伟伟,刁其玉. 转录组测序技术的研究和应用进展[J]. *生物技术通报*, 2019, 35(7):1-9.
- [13] BRUNO V M, WANG Z, MARJANI S L, et al. Comprehensive annotation of the transcriptome of the human fungal pathogen *Candida albicans* using RNA-seq[J]. *Genome Res*, 2010, 20(10):1451-1458.
- [14] CAMARENA L, BRUNO V, EUSKIRCHEN G, et al. Molecular mechanisms of ethanol-induced pathogenesis revealed by RNA-sequencing [J]. *PLoS Pathog*, 2010, 6(4):e1000834.
- [15] ZENONI S, FERRARINI A, GIACOMELLI E, et al. Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq [J]. *Plant Physiol*, 2010, 152(4):1787-1795.
- [16] 祁云霞,刘永斌,荣威恒. 转录组研究新技术:RNA-Seq 及其应用[J]. *遗传*, 2011, 33(11):1191-1202.
- [17] 赵陆滢,曹绍玉,龙云树,等. 全长转录组测序在植物中的应用研究进展[J]. *植物遗传资源学报*, 2019, 20(6):1390-1398.
- [18] LUO M, LIU X. Regulatory networks of flowering genes in *Angelica sinensis* during vernalization [J]. *Plants (Basel, Switzerland)*, 2022, 11(10):1355-1374.
- [19] DEWEY C N, LI B. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome[J]. *BMC Bioinformatics*, 2011, 12(1):323.
- [20] ZHU F Y, CHEN M X, YE N H, et al. Proteogenomic analysis reveals alternative splicing and translation as part of the abscisic acid response in *Arabidopsis* seedlings[J]. *Plant J*, 2017, 91(3):518-533.
- [21] ABDEL-GHANY S E, HAMILTON M, JACOBI J L, et al. A survey of the sorghum transcriptome using single-molecule long reads[J]. *Nat Commun*, 2016, 7:11706.
- [22] XU Z, PETERS R J, WEIRATHER J, et al. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis [J]. *Plant J*, 2015, 82(6):951-961.
- [23] 王瑞娴,李川. 全长转录组测序技术在非模式植物转录组学研究中的应用[J]. *分子植物育种*, 2019, 17(2):502-508.
- [24] 胡豪,蔺珍,薛彦斌,等. 基于联合测序分析技术挖掘红苞凤梨 lncRNA 信息[J]. *广西植物*, 2021, 41(8):1237-1250.
- [25] 王丰青,杨旭,左鑫,等. 地黄全长转录组测序及苯乙醇苷合成途径催化酶基因鉴定[J]. *药学学报*, 2022, 57(3):831-838.

[责任编辑 顾雪竹]