

## 低氮胁迫灯盏花全植株的转录组文库构建及其测序

田宝强<sup>1,2</sup>, 李玥<sup>1,3</sup>, 赖思晨<sup>1</sup>, 木佳<sup>1,3</sup>, 应宇翔<sup>1,2</sup>, 严胜柒<sup>1,3</sup>, 张云峰<sup>1,2,3</sup>

1. 云南师范大学生命科学学院, 云南 昆明 650500

2. 生物能源持续开发利用教育部工程研究中心, 云南 昆明 650500

3. 云南师范大学-台湾新竹清华遗传-环境互作联合实验室, 云南 昆明 650500

**摘要:** 目的 为探寻药用植物灯盏花 *Erigeron breviscapus* 的遗传背景, 利用低氮胁迫的灯盏花全植株构建了转录组文库, 并利用新一代测序技术进行测序。方法 采用改良异硫氰酸胍-CTAB 法, 提取低氮胁迫灯盏花植株及其对照植株的总 RNA, 经富集 mRNA、打断、构建测序用 cDNA 文库。结果 通过测序, 低氮和正常样本分别获得 3 587 万条和 2 582 万条测序读长 (raw reads), 总数据量超过 6 G, 碱基错误率低于 1% (Q20) 的数据分别为 98.37% 和 98.67%。经 *de novo* 组装, 总共得到 101、156 条 Unigene, 平均读长 768 bp, N50 为 1 290 bp, 其中 44.39% 长度超过 500 bp。101、156 条 Unigene 中, 58.86% 在公共数据库中比对到相似序列, 89.08% Unigene 在 Nr 数据库比对到相似序列。得到灯盏花黄酮类合成途径中包括苯丙氨酸解氨酶 (PAL)、肉桂酰-4-羟化酶 (C4H)、查耳酮合成酶 (CHS)、查耳酮异构酶 (CHI)、黄酮 3'-羟化酶 (F3H)、类黄酮 3'-羟化酶 (F3'H)、花色素还原酶 (ANS) 等序列。结论 灯盏花的转录组信息得到较好的保存, 为下一步灯盏花遗传环境互作研究及分子辅助育种奠定基础。

**关键词:** 灯盏花; 低氮胁迫; 转录组文库; 测序; *de novo* 组装; Unigene

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2015)21-3235-07

DOI: 10.7501/j.issn.0253-2670.2015.21.020

## Transcriptome library construction in whole plant of *Erigeron breviscapus* stressed by low nitrogen and its sequencing

TIAN Bao-qiang<sup>1, 2</sup>, LI Yue<sup>1, 3</sup>, LAI Si-chen<sup>1</sup>, MU Jia<sup>1, 3</sup>, YING Yu-xiang<sup>1, 2</sup>, YAN Sheng-Qi<sup>1, 3</sup>, ZHANG Yun-feng<sup>1, 2, 3</sup>

1. School of Life Sciences, Yunnan Normal University, Kunming 650500, China

2. Engineering Research Center of Biological Energy Development and Utilization, Ministry of Education, Kunming 650500, China

3. Join-Laboratory of YNNU-NTHU for Genetic & Environment Interaction, Kunming 650500, China

**Abstract: Objective** To explore the genetic background of *Erigeron breviscapus*, a very important herb, we used the plantlets under low nitrogen and normal condition cultured *in vitro* as material to construct the transcriptome library, and to sequence the library via next generation sequencing technique. **Methods** Modified guanidinium isothiocyanate-CTAB method was used to isolate the total RNA from low nitrogen and normal condition cultured plantlets. The mRNA was enriched from the total RNA and broken into short fragments, and then the cDNA library was established for RNA-Seq. **Results** In total, 35.87 million and 25.82 million raw reads were generated from LD and CK libraries via next generation sequencing, respectively. The overall sequencing outputs were over 6 Gb. Among all of the raw reads, more than 98.37% and 98.67% had Phred-like quality scores at Q20 level (an error probability of 1%), respectively. After filtered to remove low quality reads, the high quality sequencing sequence was used for *de novo* assembling. Unigenes of 101 and 156 pieces with the average length of 768 bp (N50 1 290 bp) were obtained, and the length of 44 908 pieces (about 44.39%) is more than 500 bp. Among 101 and 156 Unigenes, 59 538 (58.86%) showed the significant BLAST hits in the public databases. Many sequences concerning flavanoids bio-synthesis which included PAL, C4H, CHS, CHI, F3H, F3'H, and ANS were

收稿日期: 2015-03-19

基金项目: 国家自然科学基金资助项目 (30660075, 31360263); 云南科技厅社会发展科技计划基础研究面上项目 (2009CD052); 云南省应用基础研究计划 (2011FA016)

作者简介: 田宝强 (1989—), 男, 硕士在读, 研究方向为植物资源与植物遗传。Tel: (0871)65941599 Fax: (0871)65941599

E-mail: tianbaoqiang\_0330@163.com

\*通信作者 张云峰, 男, 教授, 硕士生导师, 主要从事植物资源与植物遗传研究。

Tel: (0871)65941599 Fax: (0871)65941599 E-mail: zhyunfeng001@aliyun.com

obtained from the experiment. **Conclusion** Transcriptome information of *E. breviscapus* has been better preserved, which provides the foundation for the further analysis in genetic-environment interaction and molecular assistant breeding.

**Key words:** *Erigeron breviscapus* (Vant.) Hand. -Mazz.; low nitrogen stress; transcriptome library; sequencing; *de novo* assembling; Unigene

灯盏花又名灯盏细辛，为菊科（Compositae）飞蓬属 *Erigeron* L. 植物短葶飞蓬 *Erigeron breviscapus* (Vant.) Hand. -Mazz.，广泛分布于我国西南的高山、亚高山地区，尤以云南较多，资源量占全国 95%以上<sup>[1]</sup>。云南的苗、壮、白、彝、傣等民族均有利用该植物的历史，以全草入药，用于感冒头痛、牙痛、胃痛、风湿疼痛、脑血管性头痛、瘫痪、骨髓炎、慢性气管炎等<sup>[2]</sup>。灯盏花临床上主要用于治疗高血压、脑栓塞、冠心病、心绞痛、心肌缺血等<sup>[3]</sup>，被列为我国治疗心血管疾病临床用药的基本药物。近年来，由于生境恶化，外加野生资源的过度采挖，导致野生资源面临枯竭，虽人工栽培已解决灯盏花的资源供应瓶颈。但鉴于对灯盏花遗传学背景了解的匮乏，导致栽培灯盏花在育种方面滞后，使得栽培灯盏花在有效组分量、产量、抗性上极不稳定。对其遗传背景的了解，不仅有助于阐述灯盏花药材道地性的形成机制，也有助于资源的评价及分子辅助育种。基于二代测序技术的转录组测序，由于具高通量、精确性高、低成本、高敏感度等优点，可不需了解其全基因组背景下，通过测序结果与模式植物的数据库比对，掌握非模式生物基因组的基本信息及其表达状况，目前已成为转录本结构、转录本结构变异、基因表达水平、非编码区域功能、低丰度全新转录本，甚至是基因调控网络构建的重要手段<sup>[4-9]</sup>。在模式生物酿酒酵母 *Saccharomyces cerevisiae*<sup>[10]</sup>、裂殖酵母 *Fission yeast*<sup>[11]</sup>、拟南芥 *Arabidopsis thaliana* L.<sup>[12]</sup>、小鼠<sup>[13]</sup>、人<sup>[14]</sup>中广泛运用，在非模式生物的药用植物中，目前在银杏 *Ginkgo biloba* L.<sup>[15]</sup>、虎杖 *Polygonum cuspidatum* Sieb. et Zucc.<sup>[16]</sup>、罗汉果 *Siraitia grosvenorii* (Swingle) C. Jeffrey ex A. M. Lu & Zhi Y. Zhang<sup>[17]</sup>、铁皮石斛 *Dendrobium officinale* Kimura et Migo 和人参 *Panax ginseng* C. A. Meyer<sup>[18]</sup>等也用于有效成分生物合成途径的研究。灯盏花的黄酮类及咖啡酸酯量与土壤中 N、P、K 量具有明显的相关性，与全 N 量呈显著的负相关<sup>[19-20]</sup>。本研究在对低氮胁迫的灯盏花全植转录组测序文库构建的基础上，利用 Illumina HiSeqTM 2000 测序平台株进行转录组文库测序，以期获得灯盏花的低氮胁迫响

应机制，发掘与黄酮类化合物合成途径相关的基因，为下一步灯盏花分子遗传图的构建及分子辅助育种奠定基础。

## 1 材料与方法

### 1.1 材料及处理

材料为云南师范大学-台湾新竹-清华遗传环境互作联合实验室培养的无菌苗，经笔者鉴定为短葶飞蓬 *Erigeron breviscapus* (Vant.) Hand. -Mazz. 的幼苗（原始材料采集于云南丘北，实验室编号为 E6）。以 MS 培养基为基本培养基，培养条件：25 °C、2 000 lx、光周期 16 h 光照/8 h 黑暗。实验中，首先将无菌丛生苗转入生根培养基进行培养，18~20 d 后，当丛生苗长出约 1 cm 的根时，选取长势相近的灯盏花生根苗，分别转入正常对照组（CK）和低氮处理组（LN）培养基中进行培养，培养条件与无菌苗培养相同。CK 培养基 N 素量为 60 mmol/L，LN 培养基的 N 量降为 CK 的 75%（45 mmol/L），其他条件相同。培养 14 d 后，分别随机从 CK 和 LN 培养瓶中选取植株各 15 株，鲜质量约 500 mg，液氮速冻后于 -80 °C 冰箱保存备用。

### 1.2 灯盏花总 RNA 的提取

采用 Piotr Chomczynski 和 Nicoletta Sacchi 改良的异硫氰酸胍 CTAB 法提取灯盏花样品总 RNA<sup>[21]</sup>。Agilent 2100 生物分析仪对总 RNA 完整性计数（RIN）值及 28 S/18 S 值进行检测，验证 RNA 的完整性。

### 1.3 cDNA 文库的制备

检验合格的 RNA 用于构建 cDNA 文库，使用 Illumina TruSeq RNA sample pre Kit 试剂盒及其推荐的配套试剂。具体流程如下：（1）mRNA 的富集纯化：用带有寡核苷酸 Oligo-（dT）的磁珠富集 mRNA。磁珠与 mRNA 3'端 poly（A）在结合缓冲液中相互结合，用加热的 10mmol/L Tris-HCl Buffer 洗脱，得到纯化的 mRNA。（2）mRNA 的片段化：纯化的 mRNA 加入片段化缓冲液，94 °C 处理 5 min 可将 mRNA 打断成短片段，然后用乙醇沉淀法回收产物。（3）cDNA 第 1 链的合成：以短片段 mRNA 为模板，加入第 1 链合成反应体系，即缓冲液 2 μL、RNase 抑制剂 0.25 μL、反转录酶 1 μL、1 μL dNTPs（10mmol/L）和 1 μL 六碱基随机引物（12 μmol/L）。混合均匀后室温静置 10 min，42 °C 水浴反应 1 h，

反应结束后置于冰上 2 min, 即合成 cDNA 的第 1 链。(4) cDNA 第 2 链的合成: cDNA 第 1 链的合成反应结束后, 在其体系中再加入第 2 链合成体系, 即缓冲液 5  $\mu\text{L}$ 、dNTPs (10mmol/L) 2  $\mu\text{L}$ 、DNA 聚合酶 I 100U、RNase H 50U 和 DNA 连接酶 60U。混合均匀后 16  $^{\circ}\text{C}$  反应 3 h, 纯化合成产物。(5) 末端修复: 用 T4 DNA Polymerase、Klenow DNA Polymerase 和 T4 PNK 对合成的 cDNA 3' 末端补平、5' 端进行磷酸化修饰。之后对产物进行纯化。(6) cDNA 3' 加 dA: 用 Klenow Exo-给经过末端修饰的 cDNA 3' 末端加入 1 个腺苷酸, 以便与接头连接。(7) 接头连接: 用 T4 DNA Ligase 将经过磷酸化和腺苷酸化的 cDNA 片段与接头进行连接。纯化产物, 并用琼脂糖凝胶电泳和胶回收, 收集 (200 $\pm$ 25) bp 的 cDNA 片段。所用接头序列信息如下: TruSeq Universal Adapter (5' 端接头): 5'-3'AATGATACGG-CGACCACCGAGATCTACACTCTTCCCTACAC-GACGCTCTTCCGA TCT; TruSeq Adapter (3' 端接头, 带下划线的 6 bp 碱基为随机序列 Index, 共有 24 种 Index): 5'-3'GATCGGAAGAGCACACGTCT-GAACTCCAGTCACNNNNNNATCTCGTATGCCGTCTTCTGCTTG。PCR 扩增: 用与接头相对应的引物对回收的 (200 $\pm$ 25) bp cDNA 片段进行 PCR 扩增, 回收产物并纯化, cDNA 文库制备完成。构建好的文库用 Illumina HiSeq<sup>TM</sup> 2000 测序平台进行测序。

#### 1.4 原始数据过滤

测序仪测序得到的原始图像数据经碱基读出转化为序列数据, 并以 Fasta 格式记录序列及其碱基质量信息, 每个读出碱基的测序质量值 ( $Q_{\text{phred}}$ ) 范围为 2~35。 $Q_{\text{phred}}$  与测序错误率 ( $e$ ) 有如下关系:  $Q_{\text{phred}} = -10 \lg e$ , 即测序质量值 10、20、30 分别对应着测序错误率 10%、1% 和 0.1%。原始数据中的测序读长 (raw reads) 中包含有很多带接头、低质量的 [ $Q_{\text{phred}} \leq 5$  的碱基数占整个测序读长 (reads) 超过 50%] 和含有 N(不确定碱基) 的比例超过 10% 的 reads。为了保证信息分析质量, 必须对测序读长过滤, 过滤之后的序列称为有效读长 (clean reads), 后续分析都基于 clean reads。

#### 1.5 转录组 *de novo* 组装

灯盏花及其近属植物目前都还没有基因组测序信息, 所以本研究采用无参考基因组的转录组分析方法。使用短序列拼接软件 Trinity (v2012-10-05) 对测序数据集 (reads set) 进行 *de novo* 组装<sup>[22]</sup>。Trinity 的

拼接是基于 reads 之间具有一定长度的重叠区域, reads 的初步拼接产物称为 Contig。相同的原理, 将 Contig 通过重叠区域进一步组装得到更长的序列, 直至所得序列两端无法再延长。Trinity 组装的结果称为转录本 (transcript), 一个转录本即可看作是一个广泛意义上的基因。组装过程中, 可能会产生具有主要路径相同、局部分支不同的结果, 转录本序列号中用 comp\_\* 为具有相同主干的一组序列编号, 用 seq1、seq2 等表示不同的分支序列。挑选每组中最长的转录本作为该组基因的拼接基因 (Unigene) (图 1)。

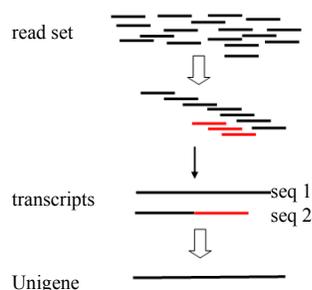


图 1 *de novo* 组装流程图

Fig. 1 Flow chart of *de novo* assembling

#### 1.6 Unigene 基因注释

以  $e$  值  $< 1 \times 10^{-5}$  为阈值, 将 Unigene 序列与 Nr (NCBI non-redundant protein sequences)、Nt (NCBI nucleotide sequences)、Swiss-Prot、KOG (euKaryotic Ortholog Groups, 真核生物直系同源蛋白聚簇)、KO (KEGG Ortholog database, Kyoto Encyclopedia of Genes and Genomes, 京都基因与基因组百科全书) 等数据库做 Blast 比对, 用 Hmmscan 在 Pfam (Protein family) 数据库和用 Blast2GO 在 GO (Gene Ontology) 数据库中也做同样的比对, 得到该 Unigene 在相应数据库的蛋白功能及结构注释信息。如果不同库之间的比对结果有矛盾, 按 Nr、Swiss-Prot、KO 及 COG 的优先级顺序确定 Unigene 的比对结果。跟以上库皆比对不上的 Unigene 用软件 ESTScan 预测其编码区并确定序列的方向。

## 2 结果

### 2.1 灯盏花植株总 RNA 的获得

使用改良异硫氰酸胍-CTAB 法提取灯盏花植株总 RNA, 经 Agilent 2100 生物分析仪检测, CK 及 LN 样品总 RNA 分别为 505.67、672.42 ng/ $\mu\text{L}$ , 而其相应 RIN 值分别为 7.3、6.8, 28 S/18 S 值分别为 1.3、1.2, 28 S/18 S 值均大于 1.0, RNA 总量远大于 20  $\mu\text{g}$ , 满足转录组测序需求, 可用于下一步测序文库的构建。

### 2.2 测序数据输出

通过 Illumina HiSeq™ 2000 进行高通量测序，结果见表 1。从灯盏花 LN 和 CK 样本中分别获得 35、869、305 个和 25、823、354 个 raw reads，平均读长 100 bp。碱基质量分值达到 20 分的碱基比例均超过 98%，分值达到 30 分的碱基均超过 94%。灯盏花 LN 和 CK 样本 GC 量分别为 43.42% 和 43.59%。

表 1 灯盏花转录组 Illumina 测序数据统计

Table 1 Sequencing data of transcriptome via Illumina sequencing in *E. breviscapus*

指标	LN	CK
测序读长数	35 869 305	25 823 354
平均读长/bp	100	100
Q20/%	98.17	98.29
Q30/%	94.00	94.17
GC/%	43.42	43.59
N 读数	63 545 (0.18%)	32 709 (0.13%)
低质量读长比例	546 553 (1.52%)	294 377 (1.14%)
接头比例	393 493 (1.10%)	271 933 (1.05%)
高质量读长	34 865 714 (97.20%)	25 224 335 (97.68%)
核苷酸总数/G	3.49	2.52

Q20-测序质量值达到 20 ( $e < 1\%$ ) 的碱基占总体碱基的百分比；  
Q30-测序质量值达到 30 ( $e < 0.1\%$ ) 的碱基占总体碱基的百分比；  
GC-碱基 G 和 C 的数量占总的碱基数量的百分比；N-含有 N (不确定碱基) 的比例大于 10% 的 reads 数及其占相应样本总 raw reads 数的百分比 (下同)

Percentage of base with Q20-sequencing up to 20 ( $e < 1\%$ ) in total base; Percentage of base with Q30-sequencing quality value up to 30 ( $e < 0.1\%$ ) in total base; Percentage of content-GC base in total base; Percentage of reads containing N (uncertain base) in total raw reads over 10% (same as below)

### 2.3 转录组 de novo 组装结果

用组装软件 Trinity 对全部 clean reads 做 de novo 组装，共得到 226 401 条转录本，平均读长 1 058 bp，N50 (将转录本按长度从大到小排列，累加转录本长度，当总长达到总碱基数的 50% 时的转录本长度) 为 1 644 bp，有 142 919 条 (63.13%) 转录本读长超过 500 bp，有 88 914 条 (39.27%) 读长超过 1 000 bp。挑选每组 (具有相同 comp 序列号) 转录本中最长的转录本作为该组基因的 Unigene，总共得到 101 156 条 Unigene，平均读长 768 bp，N50 为 1 290 bp。其中有 44 908 条 (44.39%) Unigene 读长超过 500 bp，有 24 185 条 (23.91%) Unigene 读长超过 1 000 bp。转录本和 Unigene 的长度频数分布见图 2。

将用于组装的 clean reads 定位到组装结果转录本上 (图 3)，发现 reads 在转录本上的覆盖度分布特点为距离转录本的 5' 端和 3' 端越近，平均测序深度越低，总体分布较为均匀，总测序数据的组装效率较高。

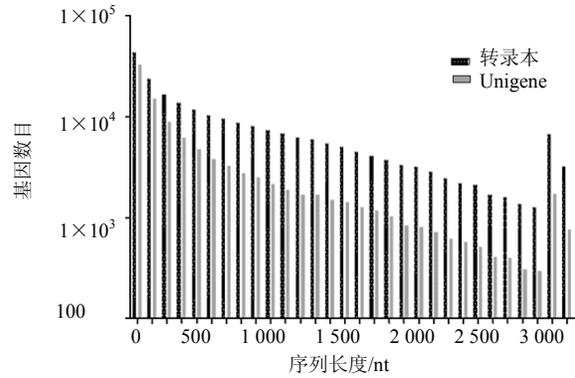


图 2 组装所得转录本和 Unigene 长度频数分布  
Fig. 2 Length frequency distribution of transcript and Unigene after assembling in *E. breviscapus*

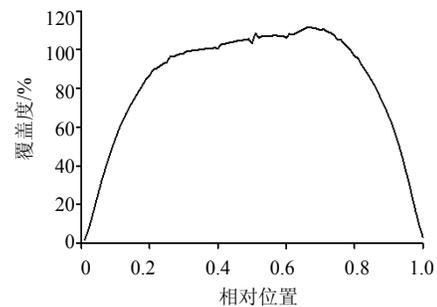


图 3 reads 在组装转录本上覆盖度的分布  
Fig. 3 Overlap distribution of reads in assembled transcript in *E. breviscapus*

### 2.4 Unigene 功能注释

通过将组装获得的 Unigene 与各公共数据库进行比对，结果见表 2。将组装和筛选得到的 Unigene 序列与 Nr、Nt、Swiss-Prot、COG 和 KO 等数据库进行同源性比对，以  $e$  值  $< 1 \times 10^{-5}$  为阈值，得到与 Nr、Nt、Swiss-Prot 的比对结果，以及 GO、COG、KEGG Pathway 的功能注释。另外，通过 Hmmscan 搜索 Pfam 得到 Unigene 的功能域及蛋白家族注释信息。经过序列比对，101 156 条 Unigene 中总共有 59 538 条 (58.86%) 在公共数据库寻得相似序列。其中，Nr、Nt 与 Swiss-Prot 中大多数 Unigene 有多于 1 个的比对结果 (最多 10 个)。Nr 中总共得到 512 010 条相似序列，与所比对的 Unigene 同源性  $\geq 80\%$  的有 81 567 条 (15.94%)，同源性在 80%~60% 的有 221 326 条 (43.23%)。Swiss-Prot 中总共得到 306 015 条相似序列，其中，与所比对的 Unigene 同源性  $\geq$

表2 灯盏花转录组中 Unigene 在公共数据库的功能注释  
Table 2 Functional annotation of Unigene in *E. breviscapus* transcriptome in public protein databases

数据库	Unigene 在相应数据库中的匹配数	匹配比例/%
Nr	53 039	89.08
Nt	18 032	30.29
Swiss-Prot	38 660	64.93
COG	22 462	37.73
KO	22 391	37.61
GO	43 392	72.88
Pfam	41 045	68.94

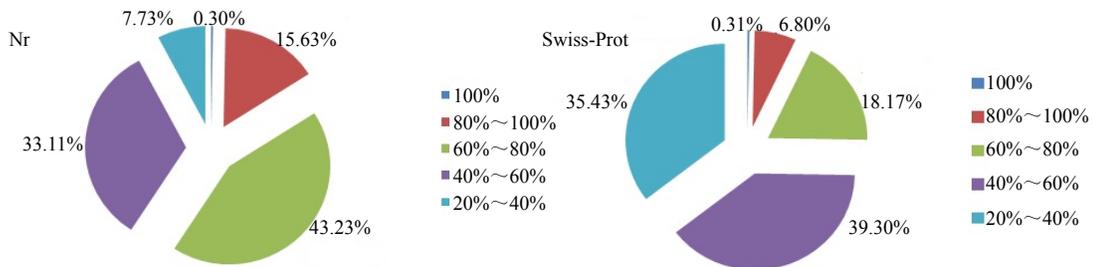


图4 在 Nr 和 Swiss-Prot 中的同源性比对结果

Fig. 4 Homologous comparison between Nr and Swiss-Prot database

表3 Nr 注释中的物种分布

Table 3 Species distribution of Unigene annotated from Nr database

品种	注释 Unigene 数	所占比例/%
葡萄	65 536	12.80
大豆	65 536	12.80
三角叶扬	61 983	12.11
拟南芥	47 561	9.29
蓖麻	39 372	7.69
蕨藜苜蓿	34 740	6.79
琴叶拟南芥	23 975	4.68
水稻	21 824	4.26
其他	151 493	29.59

## 2.5 灯盏花类黄酮生物合成相关基因挖掘

通过 Nr、Swiss-Prot、GO、COG、KO 等数据库的注释信息,在本研究构建的灯盏花转录组数据库中获得了参与黄酮类生物合成途径的多个酶,包括 PAL、C4H、CHS、CHI、F3H、F3'H、ANS 等(表4)。通过与公共数据库比对获得的注释信息,从中获得了类黄酮生物合成途径中的主要基因,包括 PAL、C4H、CHS、CHI、F3H、F3'H、ANS 等(表4)。

## 3 讨论

### 3.1 对测序数据和转录组 *de novo* 组装结果质量评估

GC 量是衡量测序质量可信度的一个指标,本实

80%的有 21 738 条(7.11%),80%~60%的有 55 588 条(18.17%)。对 Nt 因为是直接进行核酸序列比对,同源性要求较高,得到的比对结果同源性全部大于 69%,在 90 776 条相似序列中,同源性 $\geq$ 90%的有 9 065 条(9.99%),80%~90%的有 34 665 条(38.19%)(图4)。Nr 数据库的注释结果中,比对到葡萄 *Vitisv vinifera* L. 和大豆 *Glycine max* Merrill. 最多,同为 65 536 条(12.80%)。其次为三角叶扬 *Populus trichocarpa* Torr. 和拟南芥 *Arabidopsis thaliana* (L.) Heynbh, 分别有 61 983 条(12.11%)和 47 561 条(9.29%)。种属的注释见表3。

验 GC 量在(50 $\pm$ 10)%内说明样本测序质量较好。比较转录本和筛选之后的 Unigene 的序列读长分布,发现被筛选掉的序列数量与序列读长呈正比,这可能是因为组装过程中较长的序列是经过较为复杂的组装路径,这其中产生了较多的分支,而最终筛选 Unigene 只挑选最长的序列,故而其他分支的结果就都被略去了。本研究中对灯盏花转录组测序序列组装的结果质量较高,获得了读长较长的转录本和 Unigene。序列长度通常与比对结果呈正相关,研究获得了较长的序列组装结果,为之后的数据库比对奠定了良好的基础。

### 3.2 影响灯盏花黄酮类生物合成相关基因

植物体内合成类黄酮的种类和数量除了与黄酮类合成途径、黄酮和黄酮醇合成途径、异黄酮合成途径直接相关外,还与其上游的苯丙烷合成途径及下游的类黄酮化合物在植株体内的合成种类和数量有关,受 PAL、F3H、F3'H 等酶相对活性的调控<sup>[23-24]</sup>。通过转录组文库测序可以发掘到大量涉及类黄酮合成的相关基因,如 PAL、C4H、CHS、CHI、F3H、F3'H、ANS 等,可为下一步研究灯盏花的低氮胁迫、功能基因发掘及分子辅助育种奠定基础。

本研究以无菌培养的灯盏花组培苗为材料,

表 4 灯盏花类黄酮合成相关酶基因

Table 4 Related enzyme gene to flavonoid biosynthesis in *E. breviscapus*

酶蛋白	基因	Unigenes/条	相关途径
苯丙氨酸解氨酶 (phenylalanine ammonialyase)	PAL	18	P
4-香豆酸辅酶 A 连接酶 (4-coumarate-CoA ligase)	4CL	12	P、F
查耳酮合成酶 (chalcone synthase)	CHS	4	F
查耳酮异构酶 (chalcone isomerase)	CHI	2	F
黄烷酮 3-羟化酶 (flavanone 3-hydroxylase)	F3H	2	F
肉桂酸 4-羟化酶 (trans-cinnamate 4-hydroxylase)	C4H	7	P、F
黄酮合成酶 I/II (flavone synthase I/II)	FNS I/II	11	F
黄酮醇合成酶 (flavonol synthase)	FLS	41	F
类黄酮 3'-羟化酶 (flavonoid 3'-monooxygenase)	F3'H	2	F
黄酮醇 3-氧甲基转移酶 (flavonol 3-O-methyltransferase)	FMT	10	F
二氢黄酮醇-4-还原酶 (dihydroflavonol-4-reductase)	DFR	14	F
尿苷二磷酸葡萄糖: 黄芩素 7-氧葡萄糖醛酸基转移酶 (UDP-glucuronate: baicalein 7-O-glucuronosyltransferase)	UBGAT	9	F
黄酮 7-氧糖基 2"-鼠李糖基转移酶 (flavanone 7-O-glucoside 2"-O-β-L-rhamnosyltransferase)	7GRT	7	F
花色苷还原酶 (anthocyanidin reductase)	ANR	1	F、A
花色苷 3-氧糖苷转移酶 (anthocyanidin 3-O-glucosyltransferase)	AN3GT	26	A
花色苷 5,3-氧糖苷转移酶 (anthocyanidin 5,3-O-glucosyltransferase)	5,3GT	13	A
花色苷 3'-氧糖苷转移酶 (anthocyanidin 3'-O-glucosyltransferase)	3'GT	3	A
花色苷 6'-氧糖苷转移酶 (anthocyanidin 6'-O-glucosyltransferase)	6'GT	1	A
花色苷 5-芳香酰基转移酶 (anthocyanin 5-aromatic acyltransferase)	5AAT	8	A
白花色素双加氧酶 (leucoanthocyanidin dioxygenase)	LANH	11	A
花色苷 3-鼠李糖基转移酶 (anthocyanidin-3-glucoside rhamnosyltransferase)	3GRT	1	A
阿魏酸-5-羟化酶 (ferulate-5-hydroxylase)	F5H	9	P
肉桂醇脱氢酶 (cinnamyl-alcohol dehydrogenase)	CADH、REF1	12	P
香豆酰莽草酸 3'-羟化酶 (coumaroylshikimate 3'-monooxygenase)	C3'H	4	P
肉桂酰辅酶 A 还原酶 (cinnamoyl-CoA reductase)	CCR	2	P
松柏醛脱氢酶 (coniferyl-aldehyde dehydrogenase)	CALDH	5	P
过氧化物酶体 (peroxidase)	POD	78	P
咖啡酰辅酶 A 氧甲基转移酶 (caffeoyl-CoA-O-methyltransferase)	COMT	2	P、F
莽草酸氧羟基肉桂酰转移酶 (shikimate-O-hydroxycinnamoyltransferase)	HCT	14	P
松柏醇葡萄糖基转移酶 (coniferyl-alcohol glucosyltransferase)	CALGT	7	P

P-苯丙烷类生物合成途径 F-类黄酮生物合成途径 A-花色苷生物合成途径

P-phenylpropanoid biosynthesis pathway F-flavonoid biosynthesis pathway A-anthocyanin biosynthesis pathway

经 75%低氮胁迫处理和正常对照处理, 提取总 RNA, 构建文库, 并通过 Illumina HiSeq™ 2000 测序平台进行转录组高通量测序, 对数据进行质量评估, 进而挖掘到影响灯盏花类黄酮生物合成的主要相关基因。为下一步进行灯盏花转录组及其与氮元素的互作规律研究奠定了良好的基础, 同时为下一步灯盏花遗传环境互作研究及分子辅

助育种奠定基础。以期能够通过这些基础, 在转基因作物行不通的情况下, 培育出高产、含药用成分高的种质资源。

#### 参考文献

- [1] 杨生超, 文国松, 刘雪玲, 等. 灯盏花种质资源遗传关系的 ISSR 分析 [J]. 中草药, 2010, 41(9): 1523-1527.
- [2] 刘宏, 杨祥良, 徐辉碧. 灯盏花的研究进展 [J]. 中

- 草药, 2002, 33(6): 89-91.
- [3] 林雄, 褚克丹. 灯盏花素的药理学研究进展 [J]. 海峡药学, 2006, 17(6): 5-8.
- [4] Mortazavi A, Williams B A, McCue K, *et al.* Mapping and quantifying mammalian transcriptomes by RNA-Seq [J]. *Nat Methods*, 2008, 5(7): 621-628.
- [5] Zhang G J, Guo G W, Hu X D, *et al.* Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome [J]. *Genome Res*, 2010, 20(5): 646-654.
- [6] Lu T T, Lu G J, Fan D L, *et al.* Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq [J]. *Genome Res*, 2010, 20(9): 1238-1249.
- [7] Filichkin S A, Priest H D, Givan S A, *et al.* Genome-wide mapping of alternative splicing in *Arabidopsis thaliana* [J]. *Genome Res*, 2010, 20(1): 45-58.
- [8] Shah S P, Morin R D, Khattra J, *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution [J]. *Nature*, 2009, 461(7265): 809-813.
- [9] Ponting C P, Oliver P L, Reik W. Evolution and functions of long noncoding [J]. *RNA Cell*, 2009, 136(4): 629-641.
- [10] Yao Y Y, Guo G G, Ni Z F, *et al.* Cloning and characterization of microRNAs from wheat (*Triticum aestivum* L.) [J]. *Genome Biol*, 2007, 8(6): R96.
- [11] Nagalakshmi U, Wang Z, Waern K, *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing [J]. *Science*, 2008, 320(5881): 1344-1349.
- [12] Wilhelm B T, Marguerat S, Watt S, *et al.* Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution [J]. *Nature*, 2008, 453(7199): 1239-1243.
- [13] Lister R, O'Malley R C, Tonti-Filippini J, *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis* [J]. *Cell*, 2008, 133(3): 523-536.
- [14] Mortazavi A, Williams BA, McCue K, *et al.* Mapping and quantifying mammalian transcriptomes by RNA-Seq [J]. *Nat Methods*, 2008, 5(7): 621-628.
- [15] Wang E T, Sandberg R, Luo S, *et al.* Alternative isoform regulation in human tissue transcriptomes [J]. *Nature*, 2008, 456(7221): 470-476.
- [16] 林小涵. 银杏和木兰属植物的转录组; 叶绿体基因组及其相关研究 [D]. 北京: 中国医学科学院北京协和医学院, 2011.
- [17] 郝大程, 马培, 穆军, 等. 中药植物虎杖根的高通量转录组测序及转录组特性分析 [J]. 中国科学: 生命科学, 2012, 42(5): 398-412.
- [18] 唐其. 罗汉果转录组, 表达谱的高通量测序及甜苷 V 生物合成关键酶的克隆 [D]. 北京: 北京协和医学院, 2010.
- [19] 郭淑. 基于转录组测序的石斛生物碱和人参皂苷生物合成相关基因的发掘, 克隆及鉴定 [D]. 北京: 北京协和医学院, 2013.
- [20] 苏文华, 张光飞, 周鸿, 等. 氮素对短葶飞蓬生长和次生代谢产物积累的影响 [J]. 云南植物研究, 2010, 32(1): 41-46.
- [21] 苏文华, 张光飞, 周鸿, 等. 短葶飞蓬黄酮及咖啡酸酯的含量与土壤氮供应量的关系 [J]. 植物生态学报, 2009, 33(5): 885-892.
- [22] Piotr C, Nicoletta S. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol chloroform extraction: twenty-something years on nature protocols [J]. *Nat Protoc*, 2006, 1(2): 581-584.
- [23] Grabherr M G, Haas B J, Yassour M, *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome [J]. *Nat Biotechnol*, 2011, 29(7): 644-652.
- [24] 刘春霞, 王玥, 张云峰, 等. 氮, 磷营养限制及光照, 蔗糖浓度对黄酮代谢相关基因表达的影响 [J]. 江苏农业科学, 2010(1): 4-8.