

钩苞大丁草高通量转录组测序及差异表达分析

陈 菁¹, 郑 伟², 王谈笑¹, 王 炜², 徐晓丹^{3*}

(1 昆明理工大学 现代农业工程学院, 昆明 650500; 2 昆明理工大学 建筑与城市规划学院, 昆明 650500; 3 昆明理工大学 艺术与传媒学院, 昆明 650500)

摘 要: 该研究采用 Illumina Hi-Seq2500 高通量测序技术, 对叶背有纤维和无纤维发育的两组钩苞大丁草(*Gerbera delavayi* Franch.) 叶片样品的 cDNA 进行转录组测序, 分析其叶背毡毛纤维发育机理。测序结果得到了 108 694 条单基因序列, 进一步筛选得到了 1 605 条差异表达基因, 838 条差异表达基因在 GO 数据库具有功能定义, 512 条差异表达基因在 COG 分类体系中具有详细的蛋白功能释义, 315 条差异表达基因注释到了 KEGG 数据库中。其中, 氨基糖和核苷酸糖代谢途径中控制纤维素合成的相关基因, 苯丙烷类生物合成途径中控制木质素合成的相关基因, 以及激素信号转导途径中控制生长素信号转导的相关基因表达量下调; 激素信号转导途径中控制细胞分裂素、脱落酸信号转导的相关基因表达量上调。研究结果在一定程度上丰富了钩苞大丁草的基因信息, 并为后续的遗传改良提供了基础数据。

关键词: 钩苞大丁草; 转录组; 高通量测序; 差异表达基因; 纤维发生

中图分类号: Q786; Q789

文献标志码: A

Transcriptome Analysis of *Gerbera delavayi* Based on High-throughput Sequencing Technology and Differential Expression Analysis

CHEN Jing¹, ZHENG Wei², WANG Tanxia¹, WANG Wei², XU Xiaodan^{3*}

(1 Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology, Kunming 650500, China; 2 Faculty of Architecture and City Planning, Kunming University of Science and Technology, Kunming 650500, China; 3 Faculty of Art and Communication, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: In the present study, two samples (abaxial leaves with and without penniform fiber)'s cDNA were sequenced based on Illumina Hi-Seq2500 to analyze the fiber development mechanism. The results indicated that 108 694 unigenes were obtained. Then, 1 605 differential expressed genes (DEGs) were selected. 830 DEGs were divided into 39 GO terms, 512 DEGs were assigned to 25 categories with COG database. Function annotation against KEGG database obtained 315 DEGs. 10 significantly reliable enrichment pathways were selected by enrichment analysis of 79 KEGG pathways. DEGs in "amino sugar and nucleotide sugar metabolism" which control cellulose biosynthesis, DEGs in "phenylpropanoid biosynthesis" which control lignin biosynthesis and DEGs in "plant hormone signal transduction" which control auxin's signal transduction were down-regulated. While those in "plant hormone signal transduction" which control cytokinin and abscisic acid signal transduction were up-regulated. These results greatly en-

收稿日期: 2016-12-09; 修改稿收到日期: 2017-01-19

基金项目: 国家自然科学基金(31560086)

作者简介: 陈 菁(1992—), 女, 在读硕士研究生, 主要从事植物种质资源研究。E-mail: 936570200@qq.com

* 通信作者: 徐晓丹, 硕士生导师, 主要从事观赏园艺与种质资源研究。E-mail: 125128370@qq.com

riched genetic information of *G. delavayi* and provided basic data for function verification and genetic improvement of fiber traits in the future.

Key words:*Gerbera delavayi*; transcriptome; Illumina Hi-seq 2500; DEGs; fiber development

钩苞大丁草(*Gerbera delavayi* Franch.)为菊科大丁草属多年生草本植物,因其叶片干枯后具有良好的助燃性,又名“火草”。钩苞大丁草叶背的毡毛纤维是中国西南少数民族数编织火草服饰的原料。火草衣具有冬暖夏凉、越洗越白的特点^[1]。火草纺织技艺已有 500 多年历史,现被列为云南省和四川省的非物质文化遗产。钩苞大丁草毡毛纤维的主要化学成分是纤维素,其含量占到 87%,与棉纤维中纤维素的含量(90%左右)接近,拉伸率、回潮率也和棉纤维相差不大,因此钩苞大丁草的叶背毡毛纤维具有较大的利用潜力。然而,其纤维的长度、细度、断裂强度均低于棉纤维^[2],可见,钩苞大丁草的纤维品质还有较大的提升空间。

研究表明,对纤维发育相关基因的深入研究有助于纤维的遗传改良,提升纤维品质。例如,棉花中 *GhPRP5* 调控因子的表达可使棉纤维变长^[3],吡哆乙酸合成基因 *iaaM* 的超量表达可提高棉纤维品质^[4]。为了进一步提升钩苞大丁草的纤维品质,有必要筛选与钩苞大丁草叶片毡毛纤维发育相关的基因,进而可进一步通过基因克隆与功能验证,并最终实现钩苞大丁草叶背毡毛纤维的遗传改良。

前期实验中,通过在基本培养基添加 6-BA 及 NAA,诱导培养出了叶背无纤维特征的钩苞大丁草组培植株,为纤维的发育研究提供了良好的对比材料。本研究用 Illumina Hi-seq 2500 高通量测序平台进行测序,基于转录组数据分析与钩苞大丁草叶背毡毛纤维发育相关的功能基因,代谢途径及关键基因,同时也积累钩苞大丁草的基因信息,为后续关键基因的克隆、功能验证及遗传改良提供基础数据。

1 材料和方法

1.1 试验材料

选取 MS 基本培养基中培养 20 d 的带芽短缩茎的钩苞大丁草样品 A (图 1, A, 叶片有纤维发育),在基本培养基 MS+2 mg · L⁻¹6-BA+0.1 mg · L⁻¹NAA 上增殖培养 20 d 的丛生芽钩苞大丁草样品 B (图 1, B, 叶片无纤维发育),分别取幼嫩的叶片 50~100 mg。

1.2 方 法

1.2.1 高通量测序及数据组装 采用 Trizol 法提取 2 个钩苞大丁草样品的总 RNA。检测 RNA 样品的纯度、浓度和完整性等后,构建 cDNA 文库,并对文库质量进行检测,质控合格后,用 Illumina Hi-Seq2500 进行高通量测序,测序读长为 PE125。得到原始数据后,截除 Reads 中的测序引物、接头等人工序列并过滤低质量值数据后得到有效数据(clean data),用 Trinity 对 clean data 进行合并组装^[5-6]。

1.2.2 基因表达量估计及差异表达基因筛选 采用 Bowtie 将各样品测序得到的 Reads 比对到单基因序列库^[7],得到 Mapped reads,根据比对结果,进行表达量水平估计,利用 FPKM (Fragments per kilobase of transcript per million mapped reads),即每百万个 Reads 中来自比对到某一基因每千碱基长度的 Fragments 数目,表示对应的单基因序列表达丰度。使用 EBSseq 进行差异表达分析,采用 Benjamini-Hochberg 方法对原有假设检验得到的显著性 *P* 值进行校正,并最终采用校正后的 *P* 值,即 FDR (False discovery rate)作为差异表达基因(Differen-

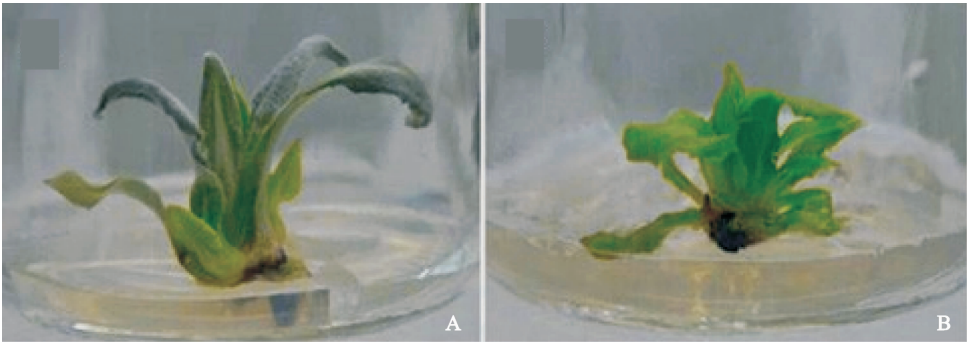


图 1 钩苞大丁草转录组检测样品
Fig. 1 Transcriptome detection samples

tially expressed genes, DEGs)筛选的关键指标^[8]。在筛选过程中,将 $FDR < 0.01$ 且差异倍数 FC (Fold change) ≥ 2 作为筛选标准,其中, FC 表示两样品(组)间 FPKM 值的比值。

1.2.3 差异表达基因的功能注释和富集分析 使用 Blast^[9] 软件将差异表达基因的 Unigene 序列与 NR、Swiss-prot、KEGG、COG 和 GO 数据库进行比对,得到其功能注释。为了进一步系统分析基因产物在细胞中的代谢途径以及这些基因产物的功能,将差异表达基因比对到 KEGG 数据库,得到差异表达基因的 Pathway 后,结合富集通路筛选出与纤维发生相关的关键差异表达基因。

2 结果与分析

2.1 测序数据产出及组装结果统计

表 1 显示,经过测序质量控制,共得到 11.13 Gb 有效数据,各样品 Q30 碱基百分比均不小于88.40%,表明测序数据可靠,可用于接下来的分析。

对组装结果进行统计,得到序列重叠群(Contig),转录本序列(Transcript)和单基因序列(Unigene)长度分布表(表 2),在序列重叠群中,所占比例最大的是 0~300 bp,有 4 161 013 条,占 98.18%;

长度大于 2 000 bp 的有 5 478 条,占 0.13%,平均长 67.86 bp,N50 长 61 bp。由序列重叠群组装出 204 352条转录本序列,平均长 918.79 bp,N50 长 1 660 bp。由转录本序列组装出 108 694 条单基因序列,平均长 593.90 bp,N50 长 912 bp,单基因长度主要分布在 200~2 000 bp,占 94.95%,长度大于 2 000 bp 的有 5 496 条,占 5.06%。

2.2 功能注释

使用 Blast 软件将单基因序列序列与 NR、Swiss-prot、KEGG、COG 和 GO 比较,得到其注释信息(表 3),结果显示,在 108 694 条单基因序列中,有

表 3 单基因序列注释统计表

Table 3 Unigene functions annotated

生物信息数据库 Bioinformation database	注释单基因数 Unigene annotated	百分比 Percentage/%
COG	11 369	10.46
GO	21 378	19.67
KEGG	8 078	7.43
Swiss-Prot	25 905	23.83
NR	39 559	36.39
总注释单基因序列数 All unigenes annotated	40 915	37.64

表 1 样品有效数据评估

Table 1 Statistics of clean data

样本 Sample	编号 Number	读数 Read	碱基 Base	GC 含量 GC content/%	Q30/%
有纤维叶片(Y) Abaxial leaves with penniform fiber	A	24 356 692	6 136 154 059	44.82	88.57
无纤维叶片(N) Abaxial leaves without penniform fiber	B	19 825 172	4 994 320 337	45.10	88.40

表 2 序列重叠群,转录本序列和单基因序列分布数量及比例

Table 2 Number and percentage of length distribution of contig, transcript and unigene

长度区间 Length Range/bp	序列重叠群数量 Contig number	转录本序列数量 Transcript number	单基因序列数量 Unigene number
200~300	4 161 013 (98.18%) *	61 631 (30.16%)	48 114 (44.27%)
300~500	38 003 (0.90%)	42 862 (20.97%)	28 386 (26.12%)
500~1 000	21 760 (0.51%)	38 189 (18.69%)	16 578 (15.25%)
1 000~2 000	11 783 (0.28%)	36 714 (17.97%)	10 120 (9.31%)
≤2 000	5 478 (0.13%)	24 956 (12.21%)	5 496 (5.06%)
总数 Total number	4 238 037	204 352	108 694
总长度 Total length/bp	287 572 789	187 756 722	64 553 876
N50 长度 N50 length/bp	61	1 660	912
平均长度 Mean length/bp	67.86	918.79	593.90

注: * 表示统计的是长度在区间 0~300 内序列重叠群的数量和百分比
Note: * means contig number and percentage of 0—300 length range

表 4 差异表达基因 GO 功能注释

Table 4 DEGs GO annotation

	项目 Item	单基因序列数量 Number of unigenes
细胞组分 Cellular component	细胞部分 Cell part	302
	细胞 Cell	301
	细胞器 Organelle	206
	胞膜 Membrane	202
	胞膜部分 Membrane part	109
	细胞器部分 Organelle part	82
	大分子复合物 Macromolecular complex	57
	胞外区 Extracellular region	49
	细胞功能 Cell junction	26
	细胞外基质 Extracellular matrix	1
	细胞外基质部分 Extracellular matrix part	1
分子功能 Molecular function	酶催化活性 Catalytic activity	497
	蛋白结合 Binding	368
	转运活性 Transporter activity	73
	电子载体活性 Electron carrier activity	31
	抗氧化活性 Antioxidant activity	15
	结构分子活性 Structural molecule activity	13
	核苷酸结合转录因子活性 Nucleic acid binding transcription factor activity	13
	酶调节器活性 Enzyme regulator activity	9
	分子转导活性 Molecular transducer activity	7
	营养库活性 Nutrient reservoir activity	3
	蛋白结合转录因子活性 Protein binding transcription factor activity	3
生物学过程 Biological process	受体活性 Receptor activity	1
	代谢过程 Metabolic process	585
	细胞过程 Cellular process	447
	单一有机体过程 Single-organism process	437
	刺激反应 Response to stimulus	192
	生物调节 Biological regulation	158
	定位 Localization	143
	细胞组分组织或生物形成 Cellular component organization or biogenesis	103
	发育过程 Developmental process	72
	多细胞有机体过程 Multicellular organismal process	48
	繁殖过程 Reproductive process	39
	多有机体过程 Multi-organism process	30
	信号 Signaling	28
	生长 Growth	15
	繁殖 Reproduction	8
	免疫系统过程 Immune system process	6
	生物黏附 Biological adhesion	3

40 915 条单基因序列获得注释,占单基因序列总数的 37.64%,另有 67 779 条未被注释,占单基因序列总数的 62.36%,认为可能是新基因。单基因序列与 NR 数据库比对,得到 39 559 条同源序列,占单基因序列总数的36.39%。与 Swiss-Prot 数据库比对,

有 25 905 条单基因序列找到了同源序列,占单基因序列总数的 23.83%。

2.3 差异表达基因的筛选与功能注释

通过对差异表达基因的筛选,共获得 1 605 个差异表达基因,其中 755 个基因在样品 A (叶片有

纤维发育)中上调,850 个基因在样品 B (叶片无纤维发育)中上调。将 DEGs 的单基因序列与 NR、Swiss-prot、KEGG、COG 和 GO 数据库比较,一共有 1 450 条单基因序列得到注释,注释到 GO 数据库的有 838 条,注释到 COG 数据库的有 512 条,有 315 条基因注释到了 KEGG 数据库中。

2.3.1 差异表达基因的 GO 功能注释 GO 分类统计显示,差异表达基因在细胞组分、分子功能和生物学过程 3 大类中包含 39 个功能小类(表 4)。其中,细胞组分这一类中,差异表达基因在“细胞部分”和“细胞”这 2 个功能小类中所占的比例最高;分子功能这一类中,差异表达基因在“酶催化途径”和“蛋白结合”这 2 个功能小类中所占的比例最高;生物学过程这一类中,差异表达基因在“代谢过程”、“细胞过程”和“单一有机体过程”这 3 个功能小类中所占的比例最高(表 4),这与李锡花等^[10]对棉花发育早期 RNA 序列转录组的分析结果基本一致。

上述基因在钩苞大丁草的转录本中显著富集,是进行纤维发生功能基因深入分析的首要对象。此外,注释到生物学过程中细胞过程这一功能小类的

差异表达基因较多,可能与 6-BA 有关。

2.3.2 差异表达基因的 COG 功能注释 将注释到 COG 数据库的 512 条差异表达基因进行直系同源分类,结果显示共分为 23 个类别(表 5),其中 R (一般功能预测)包含的基因数目最多,其次是 G (碳水化合物转运和代谢)、L (复制、重组、修复)、K (转录)、Q (次级代谢物生物合成、运输和分解)以及 E (氨基酸转运和代谢)、I (脂类转运和代谢)、T (信号转导机制)。信号转导机制这一功能类富集可能是由于本研究中对对照组激素的添加造成的(表 5)。

2.4 差异表达基因的 KEGG 注释

2.4.1 差异表达基因 KEGG 代谢通路分析 为了鉴定在代谢或者信号通路中显著富集的基因,将差异表达基因映射到 KEGG 数据库,得到了 315 个功能定义,被注释到 79 条代谢通路中。差异表达基因注释序列最多的 10 条代谢通路为:淀粉和蔗糖代谢、植物激素信号转导、苯丙烷类生物合成、氨基糖和核苷酸糖代谢等(表 6)。

本研究选择与纤维发生相关的且差异表达基因注释序列较多的氨基糖和核苷糖代谢途径、苯丙烷

表 5 差异表达基因 COG 功能注释
Table 5 DEGs COG annotation

COG	功能分类 Function classification	单基因序列数量 Number of unigenes
A	RNA 加工和修饰 RNA processing and modification	14
B	染色质结构和动力学 Chromatin structure and dynamics	16
C	能量产生和转换 Energy production and conversion	35
D	细胞周期调控、细胞分裂、染色体分配 Cell cycle control, cell division, chromosome partitioning	25
E	氨基酸转运和代谢 Amino acid transport and metabolism	44
F	核苷酸转运和代谢 Nucleotide transport and metabolism	5
G	碳水化合物转运和代谢 Carbohydrate transport and metabolism	66
H	辅酶转运和代谢 Coenzyme transport and metabolism	9
I	脂类转运和代谢 Lipid transport and metabolism	43
J	翻译、核糖体结构和生源 Translation, ribosomal structure and biogenesis	2
K	转录 Transcription	51
L	复制、重组和修复 Replication, recombination and repair	58
M	细胞壁/膜/被膜生源 Cell wall/membrane/envelope biogenesis	23
N	细胞运动 Cell motility	1
O	翻译后修饰、蛋白质转换、伴侣蛋白 Posttranslational modification, protein turnover, chaperones	31
P	无机离子转运和代谢 Inorganic ion transport and metabolism	38
Q	次级代谢物生物合成、运输和分解 Secondary metabolites biosynthesis, transport and catabolism	48
R	一般功能预测 General function prediction only	113
S	未知功能 Function unknown	18
T	信号转导机制 Signal transduction mechanisms	42
U	细胞内运输、分泌和囊泡转运 Intracellular trafficking, secretion, and vesicular transport	4
V	防御机制 Defense mechanisms	12
Z	细胞骨架 Cytoskeleton	19

表 6 差异表达基因 KEGG 功能注释

Table 6 Differential expression genes KEGG function annotation

KEGG 功能注释 KEGG function annotation	单基因序列数目 Unigene numbers	比例 Percentage/%
淀粉和蔗糖代谢 Starch and sucrose metabolism	17	5.04
植物激素信号转导 Plant hormone signal transduction	17	5.04
不饱和脂肪酸生物合成 Biosynthesis of unsaturated fatty acids	16	5.08
苯丙烷类生物合成 Phenylpropanoid biosynthesis	13	4.13
氧化磷酸化 Oxidative phosphorylation	12	3.18
苯丙氨酸代谢 Phenylalanine metabolism	11	3.49
半胱氨酸和甲硫氨酸代谢 Cysteine and methionine metabolism	11	3.49
戊糖和葡萄糖醛酸互变 Pentose and glucuronate interconversions	10	3.17
氨基糖和核苷酸糖代谢 Amino sugar and nucleotide sugar metabolism	9	2.86
类黄酮生物合成 Flavonoid biosynthesis	8	2.54

表 7 关键基因的差异表达情况

Table 7 Expression of the key genes

酶 Enzyme	名称 Name	基因表达 Genes expression	基因 ID Gene ID	log ₂ FC
2. 4. 2. 24	类似纤维素合酶 Cellulose synthase-like (Csl)	下调 Down	c61335, graph_c0	−1. 51
2. 4. 1. 12	纤维素合酶 Cellulose synthase(Ces)	下调 Down	c20932, graph_c0 c47446, graph_c0 c47929, graph_c0 c25197, graph_c0	−1. 02 −1. 04 −1. 03 −1. 21
1. 14. 13. 11	肉桂酸-4-羟化酶 Cinnamate 4-hydroxylase (C4H)	下调 Down	c24824, graph_c0 c63002, graph_c0 c63078, graph_c0	−2. 27 −2. 40 −2. 27
AUX/IAA	生长素诱导蛋白 Anxin/ Indole acetic acid	上调 Up	c39986, graph_c0	1. 69
ARF	生长素影响因子 Auxin response factor	下调 Down	c37813, graph_c0 c50728, graph_c0 c61957, graph_c0	−3. 07 −2. 77 −2. 24
A-ARR	A 型反应调节因子 A- Arabidopsis response regulator	下调 Down	c53892, graph_c1	−1. 25
PP2C	2C 型蛋白磷酸酶 Protein phosphatase 2C	下调 Down	c51077, graph_c0	−1. 45

类生物合成途径、植物激素信号转导途径作为切入点进行分析,寻找激素处理后钩苞大丁草中发生差异表达的关键基因。

2.4.2 关键基因的差异表达分析 由表 7 可知,氨基糖和核苷酸糖代谢途径中,类似纤维素合酶(Csl)的基因表达量下调,另外,纤维素合酶(Ces)的基因表达量也下调,该结果说明在激素诱导条件下,钩苞大丁草叶片中纤维素的合成量减少。苯丙烷类代谢途径中,肉桂酸-4-羟化酶(C4H)的基因表达量下调,该结果说明在激素诱导条件下钩苞大丁草叶片中木质素的合成量减少。植物激素信号转导途径中,生长素诱导蛋白(AUX/IAA)复合体通过与生长素响应因子(ARF)形成异源二聚体而起生长素响应抑制子作用,从而抑制了生长素响应基因的激活^[11-12]。AUX/IAA 的基因表达量上调,生长素响应因子(ARF)的基因表达量下调,抑制了下游生长素反应;A 型反应调节因子(A-ARR)作为细胞分

裂素的负调控因子可以抑制 B 型 ARR 的活性,从而抑制下游的细胞分裂素反应^[13-14],其基因表达量下调,相对促进了下游细胞分裂素反应;由于 2C 型蛋白磷酸酶(PP2C)类蛋白对 ABA 信号的负调控作用,其基因表达量的下调,促进了 ABA 反应。

3 讨 论

3.1 钩苞大丁草叶背毡毛纤维发生的关键基因
3.1.1 氨基糖和核苷酸糖代谢途径及苯丙烷类合成途径中的关键基因 纤维素在植物中由 36 条 1, 4-β-D-葡聚糖链相互以氢键结合形成拟晶体,在一些特殊细胞,如棉纤维中,纤维素的含量占到 98%^[15]。有报道表明,纤维素的生物合成依赖 Ces 家族,在拟南芥和水稻的基因组中均包含 10 个 CesA 基因,大麦中有 9 个 CesA 基因^[16-18]。此外,Cls 基因家族也参与纤维素的生物合成,其基因序列与 Ces 部分同源,拟南芥中发现 9 个 Cls 基因^[19],本研

究中,在氨基糖和核苷酸糖代谢途径中发现一个 *ClS* 基因,其基因表达量发生下调,此外,发现 4 个 *CesA* 基因,其基因表达量也发生下调,这些基因表达量的下调不利于钩苞大丁草纤维素的合成。张高阳等用含有黄麻 *CcCesA1* 基因反义载体的质粒转化拟南芥,发现转基因拟南芥中,纤维素含量有不同程度地降低,其植株矮小,且茎部易倒伏^[20]。庞景等^[21]发现在正常杉木木质部中,*CesA1* 和 *CesA2* 基因的表达量是矮生杉木的 2~12 倍,并推测这 2 个基因在杉木木材形成过程中具有极其重要的作用。研究钩苞大丁草纤维素合酶基因的表达规律,并鉴定出特异性表达或高表达的纤维素合酶基因,将为钩苞大丁草纤维的改良提供一定的条件。

此外,钩苞大丁草差异表达基因的苯丙烷类代谢途径中,*C4H* 基因发生下调,抑制了肉桂酸生成对香豆酸,而对香豆酸是木质素合成的重要前体物质。有研究表明,*C4H* 基因的转录丰度直接影响植物中木质素的生物合成^[22-24]。因此,有必要对 *C4H* 基因在钩苞大丁草中的调控机制做更进一步的研究,从而达到纤维成分改良的目的。

3.1.2 激素信号转导途径中的关键基因 在本研究中,NAA 和 6-BA 的添加不仅造成了丛生芽的产生,同时导致了叶片毡毛纤维发生受阻。植物激素对纤维发育的调控是一个非常复杂的过程,通常情况下是由多种激素共同调控纤维的起始和伸长。生长素、赤霉素、乙烯、油菜内酯素等可能对纤维的发育起促进作用,而细胞分裂素和脱落酸可能对纤维的发育起抑制作用^[25-27]。有研究表明,由人工合成的生长激素-萘乙酸(NAA)会抑制棉纤维次生细胞壁纤维素的合成^[28]。本研究中,*AUX/IAA* 表达量

的上调以及 *ARF* 表达量的下调,抑制了下游生长素反应;*A-ARR* 表达量的下调,抑制了下游细胞分裂素反应;*PP2C* 表达量的下调,促进了 ABA 反应,可能由于这些激素的共同作用抑制了钩苞大丁草叶背纤维的发生。

梁海泳等将源于紫穗槐的尿苷二磷酸葡萄糖焦磷酸化酶(*UGPase*)的基因、反义 4-香豆酸辅酶 A 连接酶(*4CL*)的基因以及两者的双价基因分别转移至烟草中,增强 *UGPase* 基因的表达,抑制 *4CL* 基因的表达,显著提高了转基因植株的纤维素含量,同时降低转基因植株的木质素含量^[29]。在后续研究中,通过对钩苞大丁草中纤维素发生的关键基因的表达调控,可能会实现对钩苞大丁草纤维发育的调控,提高纤维品质和产量。

3.2 转录组测序丰富了钩苞大丁草的序列信息

本次研究基于 SBS 技术使用 Illumina Hi-seq 2500 高通量测序平台进行测序,共得到 108 694 条单基因序列,将单基因序列比对到 NR 等数据库,共有 40 915 条单基因序列得到注释,占 37.64%,在一定程度上丰富了钩苞大丁草的序列资源,也为后续研究提供了基础数据。但由于钩苞大丁草生物学信息研究的空白以及数据库的信息的不完整,所以尚有 62.36%的单基因序列未得到注释,有待于进一步丰富数据库信息。

本研究为钩苞大丁草叶片毡毛纤维调控基因的挖掘、克隆及功能验证等后续研究积累了资料,为最终实现钩苞大丁草纤维性状的优化改良,提升钩苞大丁草的经济价值,保护火草纺织技艺这一非物质文化遗产提供了基础研究数据。

参考文献:

[1] 欧 丽. 彝族“罗噜颇”的火草麻布纺织[J]. 毕节学院学报, 2009,11(27): 42-46.
OU L. The fire weed gunny cloth spinning and weaving of Yi nationality Luolupo[J]. *Bijie University Journal*, 2009,11(27): 42-46.

[2] 徐晓丹. 钩苞大丁草的纤维特征及遗传多样性研究[D]. 武汉: 华中农业大学, 2016.

[3] XU W L, ZHANG D J, WU Y F, *et al.* Cotton *PRP5* gene encoding a proline-rich protein is involved in fiber development [J]. *Plant Molecular Biology*, 2013,82(4-5): 353-365.

[4] ZHANG M, ZHENG X L, SONG S Q, *et al.* Spatiotemporal manipulation of auxin biosynthesis in cotton ovule epidermal

cells enhances fiber yield and quality[J]. *Nature Biotechnology*, 2011,29: 453-458.

[5] HASS 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): 1 494-1 512.

[6] GRABHERR M G, HAAS B J, YASSOUR M, *et al.* Full length transcriptome assembly from RNA Seq data without a reference genome[J]. *Nature Biotechnology*, 2011,(29): 644-652.

[7] LANGMEAD B, TRAPNELL C, POP M, *et al.* Ultrafast and memory-efficient alignment of short DNA sequences to the human genome[J]. *Genome Biology*. 2009,10(3): R25.

- [8] FINN R D, BATEMAN A, CLEMENTS J, *et al.* Pfam: the protein families database [J]. *Nucleic Acids Research*, 2013; gkt1223.
- [9] AITSCHUL S F, MADDEN T L, SCHAFER AA, *et al.* Gapped BLAST and PSI BLAST: A New Generation of Protein Database Search Programs[J]. *Nucleic Acids Research*, 1997, **25**(17): 3 389-3 402.
- [10] 李锡花, 吴 嫚, 于霁雯, 等. 棉花纤维发育早期 RNA-Seq 转录组分析[J]. 棉花学报, 2013, **25**(3): 189-196.
LI X H, WU M, YU J W, *et al.* Transcriptome analysis of early developing cotton fiber by RNA-Seq[J]. *Cotton Science*, 2013, **25**(3): 189-196.
- [11] MOCKAITS K, ESTELLE M. Auxin receptors and plant development: a new signaling paradigm[J]. *Annual Review of Cell and Developmental Biology*, 2008, (24): 55-80.
- [12] BENJAMINS R, SCHERES B. Auxin: the looping star in plant development[J]. *Annual Review of Cell and Developmental Biology*, 2008, (59): 443-465.
- [13] TO J P, HABERER G, FERREIRA F J, *et al.* Type-A Arabidopsis response regulator are partially redundant negative regulators of cytokinin signaling[J]. *Plant Cell*, 2004, (16): 658-671.
- [14] TO J P, KIEBER J J. Cytokinin signaling: two-components and more[J]. *Trends Plant Science*, 2008, (13): 85-92.
- [15] WU Y R, MACHADO A C, WHITE R G, *et al.* Expression profiling identifies genes expressed early during lint fiber initiation in cotton[J]. *Plant Cell Physiology*, 2006, **47**(1): 107-127.
- [16] BURTON R A, SHIRLEY N J, KING B J, *et al.* The *CesA* gene family of barley. Quantitative analysis of transcripts reveals two groups of coexpressed genes[J]. *Plant Physiology*, 2004, 134: 224-236.
- [17] TANAKA K, MURATA K, YAMAZAKI M, *et al.* Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall[J]. *Plant Physiology*, 2003, 133: 73-83.
- [18] HOLLAND N, HOLLAND D, Helentjaris T, *et al.* A comparative analysis of the plant cellulose synthase (*CesA*) gene family[J]. *Plant Physiology*, 2000, **123**(4): 1 313-1 324.
- [19] SOMERVILLE C. Cellulose synthesis in higher plants[J]. *Annual Review of Cell and Developmental Biology*, 2006, 22: 53-78.
- [20] 张高阳, 祁建民, 徐建堂, 等. 圆果黄麻纤维素合成酶基因 *CcCesA1* 的克隆及利用反义载体转化拟南芥[J]. 作物学报, 2014, **40**(5): 816-822.
ZHANG G Y, QI J M, XU J T, *et al.* Cloning cellulose synthetase Gene *CcCesA1* from jute (*Corchorus capsularis* L.) and transformation of *Arabidopsis* via antisense vector[J]. *Acta Agronomica*, 2014, **40**(5): 816-822.
- [21] 庞 景, 童再康, 黄华宏, 等. 杉木纤维素合成酶基因 *CesA* 的克隆及表达分析[J]. 浙江农林大学学报, 2015, **32**(1): 40-46.
- [22] PANG J, TONG Z K, HUANG H H, *et al.* Isolation and expression analysis of cellulose synthase genes in Chinese fir (*Cunninghamia lanceolata*) [J]. *Journal of Zhejiang A&F University*, 2015, **32**(1): 40-46.
- [23] LU S, ZHOU Y, LI L, *et al.* Distinct roles of cinnamate 4-hydroxylase genes in *Populus* [J]. *Plant Cell Physiology*, 2006, **47**(7): 905-914.
- [24] MILLAR D J, LONG M, DONOVAN G, *et al.* Introduction of sense constructs of cinnamate 4-hydroxylase (CYP73A24) in transgenic tomato plants shows opposite effects on flux into stem lignin and fruit flavonoids[J]. *Phytochemistry*, 2007, **68**(11): 1 497-1 509.
- [25] 谭国飞, 王 枫, 王广龙, 等. 鸭儿芹肉桂酸 4-羟化酶基因的克隆与不同温度下的表达分析[J]. 西北植物学报, 2014, **34**(7): 1 298-1 304.
TAN G F, WANG F, WANG G L, *et al.* Isolation and expression analysis of cinnamic acid 4-hydroxylase gene under different temperatures in *Cryptotaenia japonica* Hassk [J]. *Acta Botanica Boreali-Occidentalia Sinica*, 2014, **34**(7): 1 298-1 304.
- [26] GOU J Y, WANG L J, CHEN S P, *et al.* Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis[J]. *Cell Research*, 2007, 17: 422-434.
- [27] 张国华, 张艳洁, 丛日晨, 等. 赤霉素作用机制研究进展 [J]. 西北植物学报, 2009, **29**(2): 412-419.
ZHANG G H, ZHANG Y J, CONG R C, Advances on gibberellins mechanism[J]. *Acta Botanica Boreali-Occidentalia Sinica*, 2009, **29**(2): 412-419.
- [28] 白玉林, 韩春丽, 勾 玲, 等. 棉花纤维发育过程中内源激素含量和纤维品质变化及其关系研究 [J]. 西北植物学报, 2008, **28**(11): 2 278-2 284.
BAI Y L, HAN C L, GOU L, *et al.* Relationship between endogenous hormone contents and fiber quality during the cotton fiber development[J]. *Acta Botanica Boreali-Occidentalia Sinica*, 2008, **28**(11): 2 278-2 284.
- [29] SINGH B, CHEEK H D, HAIGLER C H. A synthetic auxin (NAA) suppresses secondary all cellulose synthesis and enhances elongation in cultured cotton fiber[J]. *Plant Cell Reports*, 2009, 28: 1 023-1 032.
- [30] 梁海泳, 夏秀英, 冯雪松. *UGPase* 和反义 *4CL* 基因对转基因烟草纤维素和木质素合成的调控 [J]. 植物生理学通讯, 2006, **42**(6): 1 067-1 072.
LIANG H Y, XIA X Y, FENG X S. *UGPase* and anti-sense *4CL* and their regulation of synthesis of lignin and cellulose in transgenic tobacco (*Nicotiana tabacum* L.) [J]. *Plant Physiology Communication*, 2006, **42**(6): 1 067-1 072.

(编辑:宋亚珍)