

## 生物技术与方法

# 基于高通量测序的慈竹笋转录组分析与基因功能注释

陈宇鹏, 曹颖, 胡尚连, 黄艳, 卢学琴, 徐刚, 龙治坚

西南科技大学 生命科学与工程学院 四川省生物质资源利用与改性工程技术研究中心, 四川 绵阳 621010

陈宇鹏, 曹颖, 胡尚连, 等. 基于高通量测序的慈竹笋转录组分析与基因功能注释. 生物工程学报, 2016, 32(11): 1610–1623.

Chen YP, Cao Y, Hu SL, et al. Transcriptome analysis and gene function annotation of *Bambusa emeiensis* shoots based on high-throughput sequencing technology. Chin J Biotech, 2016, 32(11): 1610–1623.

**摘要:** 慈竹是我国四川当地的优势丛生竹种之一, 其纤维长度和质量较优异, 是造纸、纺织等工业的良好原料。本文利用 Illumina HiSeq™ 2000 平台, 对 10、50、100 和 150 cm 高的慈竹笋进行转录组分析, 共得到 69.28 M 条读长 (Reads), 经从头拼接、组装和聚类后得到 111 137 条非重复序列基因 Unigene, 其中共有 63 094 条注释到 COG、GO、KEGG、Swiss-Prot 和 Nr 数据库中。这些 Unigene 不仅具有一般的功能, 如转录和信号转导等, 还涉及到蔗糖转运与代谢、次级代谢产物及细胞壁的生物合成等方面。不同高度慈竹笋的纤维素合成酶基因存在差异表达, 发现了可能调控慈竹生长发育以及纤维素和木质素生物合成的相关基因, 为慈竹品种改良提供一定的理论基础。

**关键词:** 慈竹, 转录组, 高通量测序, 功能注释

**Received:** March 9, 2016; **Accepted:** July 20, 2016

**Supported by:** National Natural Science Foundation of China (Nos. 31400257, 31400333), Breeding Program Fund Project by the 13th Five-Year Plan of Sichuan Province, Fund of Engineering Research Center for Biomass Resource Utilization and Modification of Sichuan Province (Nos. 12zxsk07, 13zxsk01, 14tdgc05), Major Project of Education Department in Sichuan Province (No. 16ZA0145), Postgraduate Innovation Fund Project by Southwest University of Science and Technology (No. 15ycx089).

**Corresponding author:** Shanglian Hu. Tel/Fax: +86-138-81194095; E-mail: hushanglian@126.com

国家自然科学基金 (Nos. 31400257, 31400333), 四川省“十三五”育种公关项目, 四川省生物质资源利用与改性工程技术研究中心基金 (Nos. 12zxsk07, 13zxsk01, 14tdgc05), 四川省教育厅科研项目 (No. 16ZA0145), 西南科技大学研究生创新基金 (No. 15ycx089) 资助。

# Transcriptome analysis and gene function annotation of *Bambusa emeiensis* shoots based on high-throughput sequencing technology

Yupeng Chen, Ying Cao, Shanglian Hu, Yan Huang, Xueqin Lu, Gang Xu, and Zhijian Long

Engineering Research Center for Biomass Resource Utilization and Modification of Sichuan Province, School of Life Science and Engineering, Southwest University of Science and Technology, Mianyang 621010, Sichuan, China

**Abstract:** *Bambusa emeiensis* is one of the preponderant species of sympodial bamboos in Sichuan province of China, and has excellent fiber length and quality as raw materials for papermaking, textile and other industries. In this study, with the application of Illumina HiSeq™ 2000 platform, we analyzed transcriptome in *B. emeiensis* with different heights of 10, 50, 100 and 150 cm. A total of 69.28 M reads were obtained, and a sum up of 111 137 bands of Unigenes were acquired following *de novo* stitching, assembly and clustering, among which there were 63 094 bands that had been integrated in the COG, GO, KEGG, Swiss-Prot and Nr databases using annotated methods. These Unigenes not only had general functions, such as transcription and signal transduction, but were also involved in sucrose transport and metabolism, secondary metabolites and cell wall biosynthesis. There was significant difference regarding the expression of cellulose synthase gene in *B. emeiensis* at different heights, relevant genes were found that might be responsible for the regulation of the growth and development of *B. emeiensis* as well as the biosynthesis of cellulose and lignin. Our findings could provide some elementary theories for breed improvement of *B. emeiensis*.

**Keywords:** *Bambusa emeiensis*, transcriptome, high-throughput sequencing, function annotation

竹子是重要的非木材可再生资源,以生长速度快、产量高、易加工、用途广泛、经济价值高等特点著称。近年来随着分子生物学研究手段的发展,尤其是基因组和转录组测序技术的广泛应用,毛竹、麻竹等竹种的基因组和(或)转录组数据已经被公布,一些与竹笋快速生长以及竹材纤维形成的相关基因被报道。如 Peng 等<sup>[1-2]</sup>通过基因组和转录组测序发现了 19 个毛竹纤维素合酶基因,和大量与植物激素、细胞周期调控、细胞壁合成、细胞形态等相关的候选基因,为毛竹竹材特性的改良提供了依据。Liu 等<sup>[3]</sup>分析了麻竹的转录组,筛选出 105 个木质素生物合成途径相关关键酶基因,并与毛竹的相关基因进行了比对,确定了潜在的与生长和发育相关的候选基因。

慈竹 *Bambusa emeiensis*, 是四川本土大型丛生竹之一,其秆径较小,秆壁较薄,纤维长度和质量均优于其他竹种,是造纸、纺织等行业的优良原料之一,具有比较重要的应用前景<sup>[4]</sup>。目前,对于慈竹的研究主要集中于生长状况<sup>[5]</sup>、基因克隆<sup>[6]</sup>、遗传多样性<sup>[7]</sup>、生理生化指标<sup>[8]</sup>等方面。与毛竹、麻竹等竹种相比,慈竹的分子生物学研究相对滞后。有研究表明慈竹可能有 70 条染色体,为典型的多倍体复合体<sup>[9]</sup>,其基因组测序的难度较大,而且目前慈竹转录组分析以及纤维素、木质素生物合成相关的分子机理等也未见报道。

本文以不同高度的慈竹笋为材料,比较了不同高度慈竹笋的纤维素含量,并采用 Illumina HiSeq™ 2000 平台对其进行了转录组测序。将

得到的 Unigene 进行 Cluster of Orthologous Groups of proteins (COG)、Gene Ontology (GO)、Kyoto Encyclopedia of Genes and Genomes (KEGG)、Swiss-Prot 和 Nonredundant protein (Nr) 注释, 并与其他相似物种进行比对, 旨在发现调控慈竹生长发育以及与纤维素、木质素生物合成相关基因, 挖掘慈竹中的珍贵基因资源。同时, 本文也进一步分析了 4 个样品中与纤维素合成相关的差异表达基因, 为慈竹的分子遗传育种以及品种改良等提供一定的理论基础。

## 1 材料与方法

### 1.1 材料

供试材料采自西南科技大学生命科学与工程学院资源圃 (年平均气温 17.2 °C, 年平均降雨量 793.5 mm)。采样时, 以露出地面 10、50、100 和 150 cm 的笋为对象, 每个高度取 3 株独立的笋 (测量误差小于 2 cm), 并取其基部第 2 个节间的部分, 分别切碎、混匀, 其中一部分储存于 -80 °C 冰箱中备用, 另一部分置于烘箱中烘干用于纤维素含量测定。

### 1.2 不同高度慈竹笋纤维素含量测定

纤维素含量的测定参照 FOSS 公司 Fibertec™ M6 1020/1021 型纤维素测定仪的操作手册进行。通过酸性洗涤纤维 (ADF)<sup>[10]</sup>法, 得到纤维素含量, 每个样品生物重复 3 次, 每组样品运行 1 个空白对照。

### 1.3 转录组测序

#### 1.3.1 总 RNA 琼脂糖凝胶电泳检测及测序

RNA 提取采用 OMEGA BIO-TEK 公司 Plant RNA Kit 试剂盒中的试剂及提取方法。总 RNA 样品采用 Nanodrop 检测和 Agilent Technologies 2100 Bioanalyzer 检测。检测合格

后用 NEBNext Poly (A) mRNA Magnetic Isolation Module (NEB, E7490) 富集 mRNA, 将 mRNA 片段化处理, 以 mRNA 片段为模板, 采用随机引物法, 用 NEBNext mRNA Library Prep Master Mix Set for Illumina (NEB, E6110) 和 NEBNext Multiplex Oligos for Illumina (NEB, E7500) 构建上机文库。制备好的文库用 1.8% 琼脂糖凝胶电泳进行检测文库插入片段大小, 然后用 Library Quantification Kit-Illumina GA Universal (Kapa, KK4824) 进行 QPCR 定量, 再在 cDNA 片段两端加上接头。检测合格的文库在 Illumina cbot 上进行簇的生成, 最后用 Illumina HiSeq™ 2000 进行测序。

#### 1.3.2 数据分析

转录组测序得到的原始数据经过过去除杂质和冗余处理后, 利用 Trinity 软件<sup>[11]</sup>对经过过滤后的高质量数据进行从头拼接, 获得重叠群 Contig。根据其结果利用双末端信息将来自同一转录本的不同 Contig 连在一起, 做进一步的序列拼接, 得到转录本 Transcripts。在 Transcripts 聚类单元中选取最主要的转录本作为 Unigene 序列, 并对 Unigene 数据进行聚类分析和进一步去冗余处理, 最终得到非冗余 Unigene 库。对于多样品的组装, 由于后续的 Open reading frame (ORF) 预测、Simple sequence repeats (SSR) 分析、表达丰度分析等都建立在同一套参考基因的基础上, 因此对各样品得到的 Unigene 作进一步的聚类分析, 最后得到慈竹的 Unigene 数据库。

#### 1.3.3 功能注释

使用 Basic Local Alignment Search Tool (Blast) 将组装后的 Unigene 序列与 COG、GO、KEGG、Swiss-Prot 和 Nr 数据库比对, 得到最高序列相似性的 Unigene。

### 1.3.4 基因的差异表达分析

在 NR、Swiss-Prot 注释结果中进行关键字检索,记录“GeneID”以及相对应的注释结果。根据“GeneID”记录相应基因的相对表达丰度值,并将值导入 MeV 4.9.0 分析软件中,制作热图并进行聚类分析。

### 1.4 差异表达基因 qRT-PCR 验证

利用 primer premier 5.0 对已测序正确的序列和慈竹内参基因 *Tublin* 进行 qRT-PCR 引物的设计。在 iQ Multicolor Real-Time PCR 自动扩增仪进行反应,每个样品重复 3 次,反应体系 20  $\mu$ L。以内参基因 *Tublin* 为对照,利用公式  $2^{-\Delta\Delta CT}$  计算其相对表达量。

## 2 结果与分析

### 2.1 测序数据结果统计

利用 Illumina HiSeq<sup>TM</sup> 2000 平台,对供试

材料进行转录组测序,共获得 13.98 Gb 的原始数据,得到 69.28 M 条的读长 (reads) (表 1)。其中 10、50 和 150 cm 慈竹笋总碱基数均大于 4 Gb, reads 数大于 20 M, GC 含量分别为 52.78%、53.12%和 52.21%;而 100 cm 慈竹笋因测序深度较低,仅获得 7 829 716 条 reads,GC 含量为 49.49%。同时,供试材料的不确定碱基数 (N%) 所占比例均 $\leq$ 0.03%,测序质量值 $\geq$ 30 的碱基 (Q30%) 所占比例均 $>$ 80%,这表明测序质量可靠。

通过进一步分析后得到的非冗余 Unigene 库,含有 111 137 条慈竹笋的 Unigene。其中 200 bp–300 bp 之间的 Unigene,为 41 814 条,占 37.62%;大于 1 000 bp 的 Unigene 为 21 518 条,占 19.35% (图 1A, 1B)。

### 2.2 基因功能注释

#### 2.2.1 基因功能注释

通过数据库比对,得到最高序列相似性的

表 1 测序数据统计

Table 1 Statistics of sequencing data

Samples	Total reads	Total nucleotides (bp)	GC (%)	N (%)	Q20 (%)	Q30 (%)
10 cm	20 125 331	4 062 680 957	52.78	0.01	89.12	80.27
50 cm	20 823 597	4 200 352 710	53.12	0.02	89.08	80.35
100 cm	7 829 716	1 579 745 023	49.49	0.03	92.01	87.12
150 cm	20 504 408	4 137 822 934	52.21	0.01	89.12	80.35

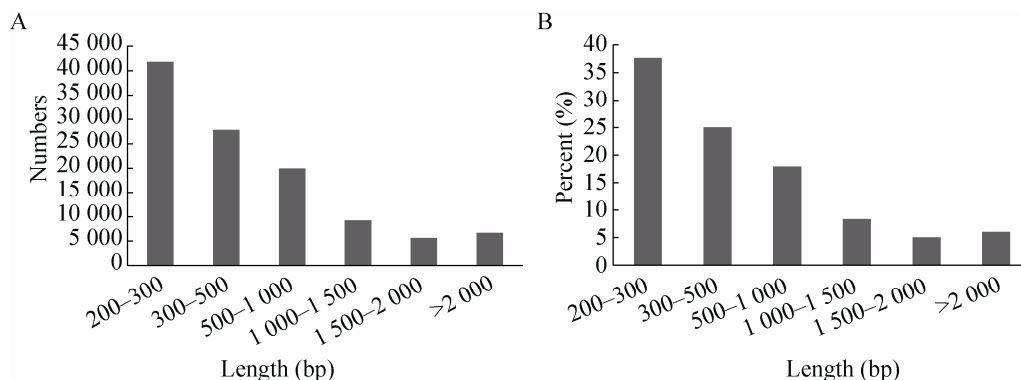


图 1 慈竹功能基因长度分布和所占比例

Fig. 1 Distribution and percentage of Unigene length from *Bambusa emeiensis*.

Unigene, 其中有 16 151 条 Unigene 被匹配到 COG 数据库中; 52 872 条 Unigene 被匹配到 GO 数据库中; 11 032 条 Unigene 被匹配到 KEGG 数据库中; 44 552 条 Unigene 被匹配到 Swiss-Port 数据库中; 而 62 795 条 Unigene 被匹配到 Nr 数据库中 (表 2)。

### 2.2.2 功能基因的 COG 注释

COG 的功能注释与基因产物有直接的联系<sup>[12]</sup>。在慈竹笋 Unigene 数据库的 COG 功能注释中, 有 16 151 条 Unigene 具有蛋白质功能定义。在 COG 功能分类中包含复制、重组、修复、

信号转导、新陈代谢、分子过程等 25 个分类。其中一般功能基因代表最大的一类, 有 4 221 条 Unigene; 其次是复制、重组和修复, 有 2 701 条。此外, 还涉及到多个植物生长发育相关以及纤维素、木质素生物合成相关的生理生化过程, 如碳水化合物转运与代谢、氨基酸转运与代谢、辅酶转运与代谢、次级代谢产物生物合成、运输和分解代谢、细胞壁与细胞膜的生物合成、信号转导机制等过程 (图 2)。

### 2.2.3 功能基因的 GO 分类

GO 数据库可对基因和蛋白质进行分类和

表 2 慈竹转录组功能注释

Table 2 Transcriptome function annotation of *Bambusa emeiensis*

Database	Annotated number	3 00	Length<1 000	Length 1 000
COG_Annotation	16 151		5 991	8 640
GO_Annotation	52 872		23 297	19 361
KEGG_Annotation	11 032		4 315	5 066
Swissprot_Annotation	44 552		19 210	18 551
Nr_Annotation	62 795		28 348	20 587
All_Annotated	63 094		28 491	20 599

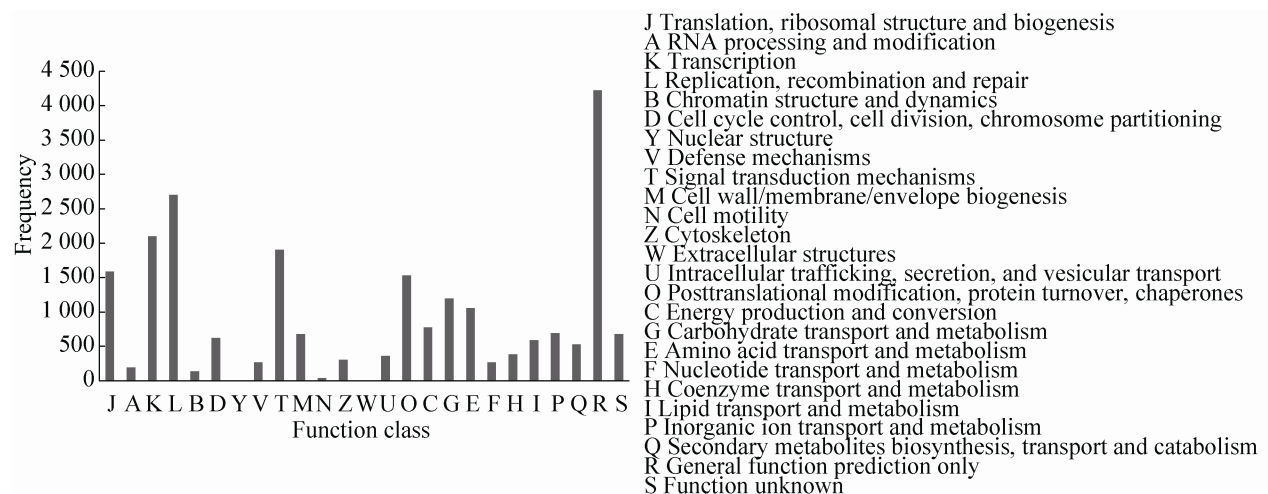


图 2 Unigene 的 COG 功能分类

Fig. 2 Unigene of COG function categories.

注释,其对很多物种都是适用的<sup>[13]</sup>。GO 数据库包含 3 大类,即细胞组分、分子功能和生物学过程。在慈竹笋 Unigene 数据库中,有 52 872 条基因注释到 GO 数据库中。其中第一大类中的细胞部分、细胞和细胞器所占比例较高,分别达到 85.71%、84.85%和 80.39%;第二大类中的结合和催化活性所占比例较高,分别达到 68.90%和 50.82%;而第三大类中的细胞过程和代谢过程所占比例较高,分别达到 75.86%和 73.31% (图 3)。

### 2.2.4 功能基因的 KEGG 注释

KEGG 数据库是有关基因产物代谢通路的

主要数据库,基于代谢通路的分析对于进一步分析解读基因的功能非常有帮助<sup>[14]</sup>。在慈竹笋 Unigene 数据库中,有 11 032 条 Unigene 注释到 KEGG 数据库中。其中注释序列较多的 3 条代谢通路为 RNA 转运、植物激素信号转导与核糖体。而与植物纤维素生物合成相关的代谢通路,如淀粉和蔗糖代谢、光合生物的固碳作用和光合作用分别有 257 条、189 条和 105 条注释到 KEGG 数据库中;与木质素生物合成相关的代谢途径,如苯丙烷类的生物合成和苯丙氨酸代谢分别有 118 条和 110 条在 KEGG 代谢通路中得到注释 (表 3)。

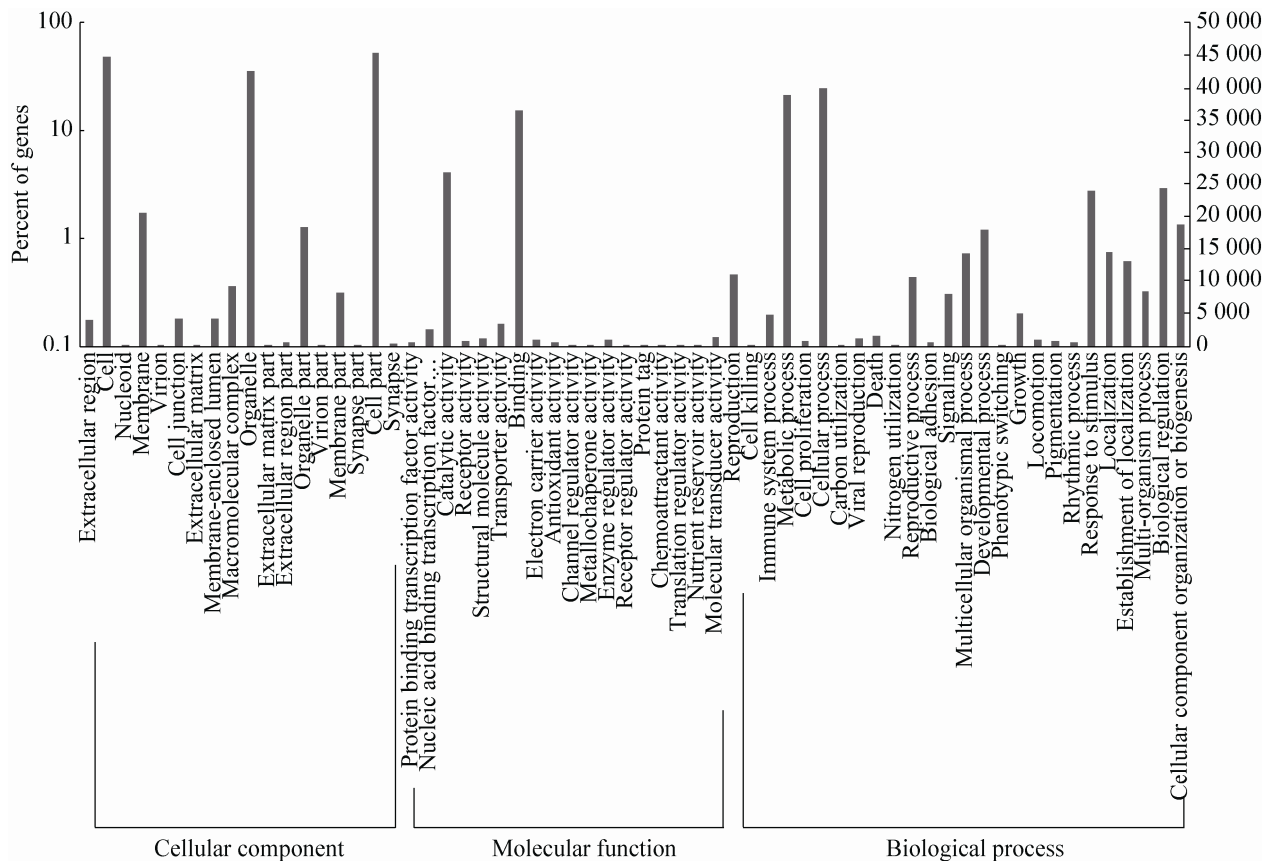


图 3 Unigene 的 GO 功能分类

Fig. 3 Unigene of GO function categories.

表 3 Unigene 数目最多的 5 个代谢通路及与纤维素、木质素生物合成相关的代谢通路

Table 3 Top five metabolic pathways and the metabolic pathways of cellulose and lignin biosynthesis involving *Bambusa emeiensis* Unigene

Name of metabolic pathway	Number of Unigene	Percent (%)
RNA transport	408	5.47
Plant hormone signal transduction	407	5.46
Ribosome	404	5.42
Spliceosome	371	4.98
Protein processing in endoplasmic reticulum	345	4.63
Starch and sucrose metabolism	257	3.45
Carbon fixation in photosynthetic organisms	189	2.54
Phenylpropanoid biosynthesis	118	1.58
Phenylalanine metabolism	110	1.48
Photosynthesis	105	1.41

### 2.3 慈竹 Unigene 与其他物种的比较分析

首先将慈竹的 Unigene 与水稻 *Oryza sativa*、毛竹 *Phyllostachys edulis*、短柄草 *Brachypodium*、玉米 *Zea mays* 和高粱 *Sorghum* 的 Unigene 进行功能比对。所得结果如表 4 所示,慈竹 Unigene 与毛竹匹配度最高,达到 52.46%,可能是因为慈竹与毛竹同属竹类植物的原因。其次是与水稻相匹配达到 50.41%,可能是因为水稻为模式植物,其基因序列信息较其他几种禾本科更全面。

表 4 慈竹 Unigene 与其他物种的比较分析

Table 4 Analysis on Unigene between *Bambusa emeiensis* and other species

	Number of Unigene	Percent (%)
With hits to <i>Oryza sativa</i>	28 666	50.41
With hits to <i>Phyllostachys edulis</i>	29 831	52.46
With hits to <i>Brachypodium</i>	27 106	47.67
With hits to <i>Zea mays</i>	25 662	45.13
With hits to <i>Sorghum</i>	27 046	47.56
With hits to all above	21 524	37.85
With no hits to all above	23 062	40.56

根据表 4 的结果,再将慈竹的 Unigene 与水稻转录本序列和毛竹编码序列数据库进行 Blast 比对,选择  $E\text{-value} < 1e-10$ 。63 094 个慈竹 Unigene 中 34 127 个 (54.09%) Unigene 与毛竹转录组相匹配,33 685 个 (53.39%) Unigene 与水稻相匹配。在这些配对的序列中 29 591 个 (46.90%) Unigene 与毛竹和水稻都相匹配,4 536 个 (7.19%) Unigene 只与水稻相匹配,4 094 个 (6.49%) Unigene 只与毛竹相匹配。还有许多预测的 Unigene 与水稻和毛竹都不匹配,它们编码的蛋白质主要与生理过程、结合活性和催化功能有关,其中有大量预测的 Unigene 可能是慈竹特有的 (图 4)。

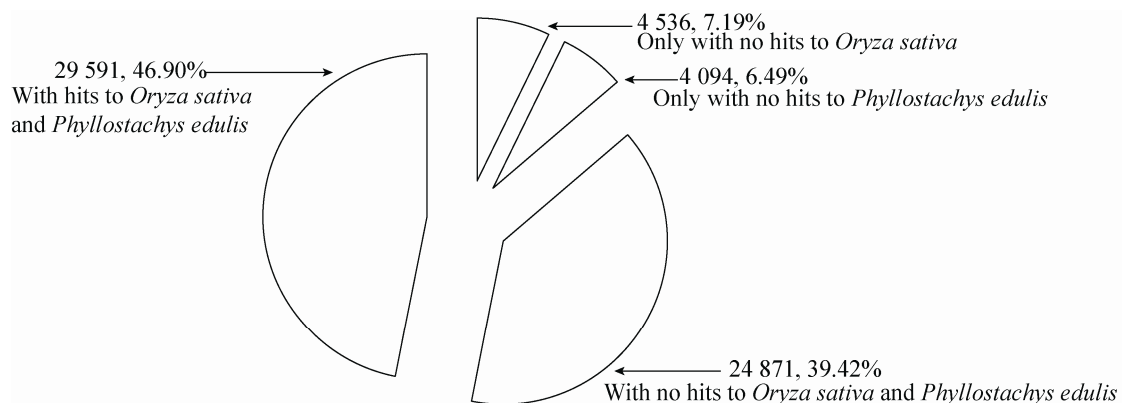


图 4 慈竹 Unigene 与水稻和毛竹的比较分析

Fig. 4 Analysis on Unigene between *Bambusa emeiensis*, *Oryza sativa* and *Phyllostachys edulis*.

## 2.4 纤维素、木质素生物合成相关基因与其他物种比对分析

### 2.4.1 纤维素生物合成相关酶基因与其他物种比对分析

将鉴定出的慈竹纤维素生物合成途径中的关键酶基因与水稻、短柄草、玉米、高粱和毛竹进行比对分析,毛竹的相关数据引用自 Peng 等<sup>[1]</sup>。由表 5 可知,在 Swiss-Prot 数据库中,慈竹纤维素合酶 Cellulose synthase (*CesA*) 的数量与其他几个物种相比差异较大,*CesA* 仅有 9 条。纤维素合酶相似蛋白 Cellulose synthase-like protein (*Csl*) 则是毛竹的最多有 76 条,慈竹仅有 11 条,与毛竹相比差异比较明显。而短柄草中最少,只在 Swiss-Prot 数据库中鉴定到 7 条 *Csl* 基因。蔗糖合酶 Sucrose synthase (*SuSy*) 数

量最多的是水稻 (21),其次是短柄草 (11),慈竹只在 Swiss-Prot 数据库中鉴定出 7 条。

### 2.4.2 木质素生物合成相关酶基因与其他物种比对分析

经 Swiss-Prot 数据库功能注释,筛选出 188 个木质素生物合成途径相关关键酶基因。与其他禾本科植物相比,慈竹苯丙氨酸解氨酶 Phenylalanine ammonia-lyase (*PAL*)、4-香豆酸辅酶 A 连接酶 4-coumarate-CoA ligase (*4CL*)、肉桂酰辅酶 A 还原酶 Cinnamoyl-CoA reductase (*CCR*)、肉桂醇脱氢酶 Cinnamoyl alcohol dehydrogenase (*CAD*) 等 7 个关键酶基因均略少于水稻 (表 6)。在 Swiss-Prot 数据库中,慈竹、短柄草、玉米和高粱的漆酶 Laccase (*LAC*) 基因是数量最丰富的,分别有 17、32、20 和 20 条,

表 5 慈竹纤维素合成的功能基因与水稻、短柄草、玉米、高粱和毛竹的比较

**Table 5** *Bambusa emeiensis* functional Unigene of cellulose biosynthesis comparison with *Oryza sativa*, *Brachypodium*, *Zea mays*, *Sorghum* and *Phyllostachys edulis*

Gene	<i>Oryza sativa</i>	<i>Brachypodium</i>	<i>Zea mays</i>	<i>Sorghum</i>	<i>Bambusa emeiensis</i>	<i>Phyllostachys edulis</i> <sup>a</sup>
<i>CesA</i>	21	10	21	17	9	19
<i>SuSy</i>	21	11	7	8	7	—
<i>Csl</i>	18	7	10	13	11	76

<sup>a</sup> The results were cited from Peng et al. <sup>[1]</sup>.

表 6 慈竹木质素合成的功能基因与水稻、短柄草、玉米、高粱和毛竹的比较

**Table 6** *Bambusa emeiensis* functional Unigene of lignin biosynthesis comparison with *Oryza sativa*, *Brachypodium*, *Zea mays* and *Sorghum* and *Phyllostachys edulis*

Gene	<i>Oryza sativa</i>	<i>Brachypodium</i>	<i>Zea mays</i>	<i>Sorghum</i>	<i>Bambusa emeiensis</i>	<i>Phyllostachys edulis</i> <sup>a</sup>
<i>PAL</i>	7	4	5	3	5	8
<i>4CL</i>	26	0	0	0	13	6
<i>CCR</i>	45	24	0	0	2	3
<i>CAD</i>	18	10	10	21	10	1
<i>LAC</i>	23	32	20	20	17	—
<i>CCoAOMT</i>	10	0	0	0	3	2
<i>COMT</i>	10	0	1	4	2	1

<sup>a</sup> The results were cited from Peng et al. <sup>[1]</sup>.



毛竹则未检测到。水稻的 *CCR* 咖啡酰辅酶 A-O-甲基转移酶 Caffeoyl-CoA O-methyltransferase (*CCoAOMT*) 和咖啡酸/5-羟基阿魏酸-O-甲基转移酶 Caffeic acid-O-methyltransferase (*COMT*) 基因数量较其他物种丰富, 分别为 45 条、10 条和 10 条。在慈竹中数量较少的是 *COMT*, 为 2 条。在毛竹中 *PAL* 最多, 为 8 条, 其余则较少, 与慈竹相比差异较大。毛竹的相关数据引用自 Peng 等<sup>[1]</sup>。

#### 2.4.3 转录因子相关基因与其他物种比对分析

转录因子是由核基因编码的一类蛋白质, 能选择性地激活或抑制某些基因的转录与表达<sup>[15]</sup>。从 Swiss-Prot 数据库中得到的转录因子如表 7 所示。慈竹 *MYB* 和 *bHLH* 转录因子数量较水稻、短柄草、玉米和高粱多, 且转录因子整体数量上与毛竹相比差异较明显。

## 2.5 差异表达基因分析

### 2.5.1 差异表达基因的筛选与统计

为了解 4 个不同高度慈竹笋的差异表达基因, 采用 DESeq 软件, 设定  $FDR < 0.01$ 、Fold Change (FC)  $\geq 2$  作为筛选差异表达基因的关键指标。将 10、50、100 和 150 cm 四个样品进行两两比对后, 得到以下结果, 10 vs 50 cm 组有 1 828 条差异表达基因; 10 vs 100 cm 组有 1 542 条差异表达基因; 10 vs 150 cm 组有 2 601 条差异表达基因; 50 vs 100 cm 组有 1 224 条差异表达基因; 50 vs 150 cm 组有 1 745 条差异表达基因; 100 vs 150 cm 组有 1 231 条差异表达基因。将同一组合在不同的数据库间进行比较, 提取并统计筛选得到的差异表达基因的注释信息 (表 8)。

表 7 慈竹转录因子相关基因与水稻、短柄草、玉米、高粱和毛竹的比较

Table 7 Transcription factor of *Bambusa emeiensis* comparison with *Oryza sativa*, *Brachypodium*, *Zea mays* and *Sorghum* and *Phyllostachys edulis*

Gene	<i>Oryza sativa</i>	<i>Brachypodium</i>	<i>Zea mays</i>	<i>Sorghum</i>	<i>Bambusa emeiensis</i>	<i>Phyllostachys edulis</i> <sup>b</sup>
<i>MYB</i>	4	1	0	0	24	116
<i>WRKY</i>	133	82	119	100	47	116
<i>NAC</i>	120	71	101	87	23	125
<i>bZIP</i>	6	7	6	8	3	107
<i>bHLH</i>	1	1	1	1	54	168
<i>HD-ZIP</i>	0	0	0	0	0	51
<i>MADS-box</i>	97	87	79	78	23	16

<sup>b</sup>The results were cited from plant transcription factor database.

表 8 注释的差异表达基因数目统计

Table 8 Summary of differential expression gene annotated

Combination types	COG	GO	KEGG	Swiss-port	Nr
10 vs 50 cm	311	1 171	192	111	1 347
10 vs 100 cm	271	1 089	253	972	1 254
10 vs 150 cm	587	1 881	457	1 779	2 086
50 vs 100 cm	222	807	213	725	961
50 vs 150 cm	400	1 188	348	1 112	1 332
100 vs 150 cm	268	799	255	720	919

### 2.5.2 纤维素生物合成相关差异表达基因分析

慈竹纤维素含量直接影响着造纸效率。通过酸性洗涤纤维 (ADF) 法分析了 4 个不同高度的慈竹笋基部的纤维素含量。结果表明,在笋的早期生长阶段,随着笋的伸长,慈竹笋基部的纤维素含量逐渐增加 (图 5)。通过进一步的转录组数据分析,可以从分子机理上了解慈竹笋在发育过程中纤维素含量的变化。

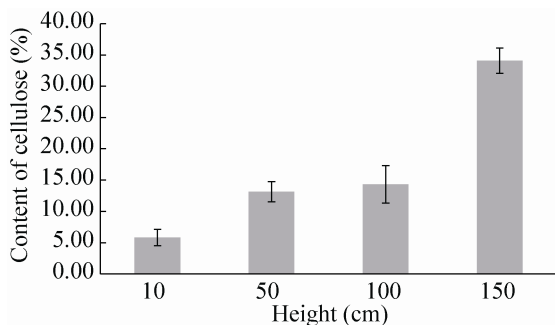


图 5 不同高度慈竹笋纤维素含量  
Fig. 5 The content of cellulose in different height of *Bambusa emeiensis*.

*CesA* 在纤维素生物合成过程中起着关键的调节作用。转录组数据分析表明, *CesA* 基因的 RPKM 值在 10 和 50 cm 的样品中聚在一簇;而 100 和 150 cm 的样品聚为另一簇,表明 *CesA* 基因随慈竹笋高度的增长呈现不同的表达模式 (图 6A, 6B)。其中, *CesA1*、*CesA5*、*CesA6* 和 *CesA8* 同源的 Unigene 有着较高的相对表达量,但却随着笋的伸长,呈现出降低的趋势 (图 6A, 6B)。在 4 个不同生长高度的慈竹笋中,与 *CesA4*、*CesA7*、*CesA9* 同源的 10 条 Unigene,虽然呈现着较低的表达,却随着笋的伸长,呈现出增高的趋势 (图 6A, 6B)。这些 Unigene 在 GO 数据库的功能注释中都涉及多个 GO 功能条目,包括细胞壁、初生细胞壁生物合成、次生细胞壁生物合成、纤维素合酶活性、纤维素生物合成过程、木质部发育、细胞壁增厚等,而在 KEGG 注释中,10 条 Unigene 都参与的 KEGG 通路是 K10999 (纤维素合酶) (表 9)。

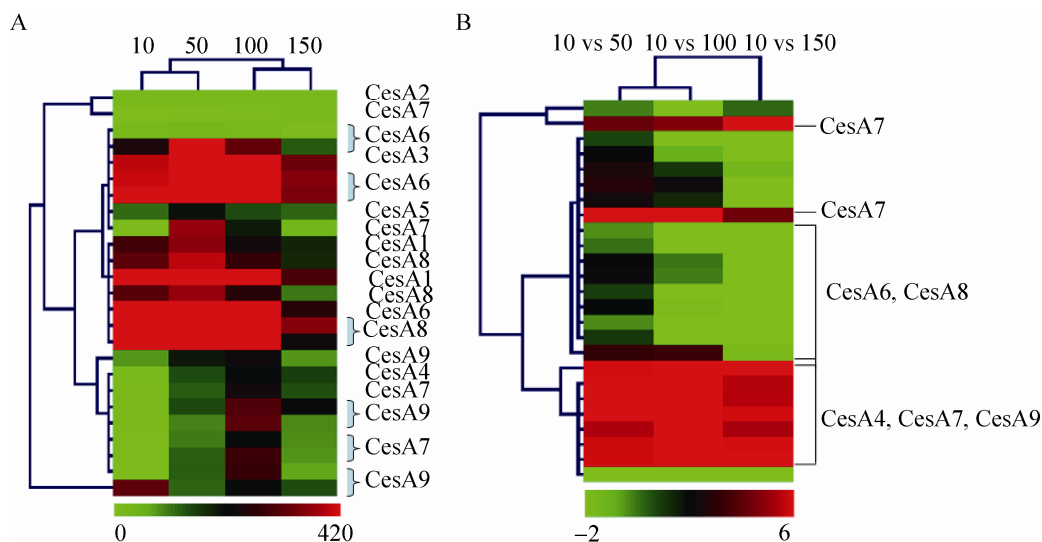


图 6 不同生长高度的慈竹笋中 *CesA* 的相对表达水平 (RPKM 值) 和差异表达水平 (Log<sub>2</sub> 值)  
Fig. 6 The relative expression level (RPKM value) and the differential expression level (Log<sub>2</sub> value) of *CesA* in the shoots of different height from *Bambusa emeiensis*.

表 9 纤维素合酶基因的功能注释

Table 9 Annotation of cellulose synthase genes

Unigene	KEGG pathway	Annotation
CL10Contig1	K10999	Cellulose synthase A catalytic subunit 9
T2Unigene23059	K10999	Cellulose synthase A catalytic subunit 7
CL1669Contig1	K10999	Cellulose synthase A catalytic subunit 7
T2Unigene28744	K10999	Cellulose synthase A catalytic subunit 9
T1Unigene23998	K10999	Cellulose synthase A catalytic subunit 7
T1Unigene22002	K10999	Cellulose synthase A catalytic subunit 9
T4Unigene34723	K10999	Cellulose synthase A catalytic subunit 9
CL1525Contig2	K10999	Cellulose synthase A catalytic subunit 4
T2Unigene23060	K10999	Cellulose synthase A catalytic subunit 7
T3Unigene29112	K10999	Cellulose synthase A catalytic subunit 7

## 2.6 差异表达基因 *qRT-PCR* 验证

根据纤维素生物合成相关差异表达基因热图分析,筛选出 *CesA9* (T1\_Unigene\_BMK.22002) 进行 *qRT-PCR* 验证。如图 7 所示,在 100 cm 的慈竹笋中, *CesA9* 的相对表达量大约是 10 cm 笋的 9.5 倍,表明其表达水平可能直接影响着笋的纤维素含量。

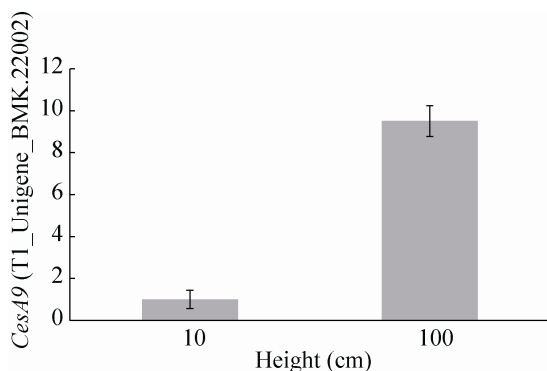


图 7 *CesA9* (T1\_Unigene\_BMK.22002) 基因 *qRT-PCR* 验证

Fig. 7 Quantitative RT-PCR verification of *CesA9* (T1\_Unigene\_BMK.22002) gene.

## 3 讨论与结论

测序技术日趋成熟,水稻<sup>[16]</sup>和小麦<sup>[17]</sup>等禾本科比较有代表性植物的基因组测序已经完成,并且已经转向实际应用方面的研究。在竹子中毛竹<sup>[18]</sup>基因组测序也已经完成,为竹类植物的分子生物学研究奠定了良好基础。转录组测序已经不再仅限于拟南芥等模式植物,越来越多非模式植物的转录组也开始得到研究,其对应的生长发育状态与机理也逐渐被人们所了解。本文利用转录组测序技术,对处于早期生长的 4 个不同高度的慈竹笋进行转录组从头测序,得到了非冗余的 111 137 条慈竹笋的 Unigene。COG 和 KEGG 功能注释表明,除了基本的功能外,这些 Unigene 涉及到蔗糖转运与代谢、次级代谢产物(如与木质素相关的苯丙烷类的生物合成)及细胞壁的生物合成等方面(图 2,表 3)。

值得一提的是竹类植物 *MYB* 类转录因子数量明显多于另外 4 个物种。有研究指出 *MYB* 转录因子家族能通过调控参与苯丙烷代谢途径的

相关基因表达,进而影响木质部细胞的形成<sup>[19]</sup>。此外,MYB类转录因子也会调控部分纤维素生物合成相关的基因表达,如MYB46可直接调控与次生壁相关的部分纤维素合酶基因,从而影响纤维素的合成<sup>[20]</sup>。因此,进一步探讨MYB类转录因子在慈竹中的功能,对慈竹分子遗传改良具有重要的意义。

慈竹纤维素含量丰富,是较好的造纸原料,其与硬头黄<sup>[21]</sup>等竹种在四川已形成了以其为龙头的竹浆造纸工业产业链。竹材中的纤维素含量直接影响竹浆造纸的效率。10、50、100和150 cm高度的笋中,纤维素含量具有明显差异。因此本文进一步深入分析了4个样品中与纤维素合成相关的差异表达基因。纤维素的生物合成是一个高度复杂的生物过程,有多个酶参与其生物合成,其中包括CesA、Csl、SuSy等<sup>[22]</sup>。CesA是目前研究最多的纤维素生物合成基因,它编码的纤维素合酶不仅是纤维素生物合成的场所,而且在纤维素生物合成过程中起着关键的调节作用。纤维素合酶本身具有多基因现象,在纤维素生物合成过程中,多个CesA基因彼此相关,协同作用<sup>[23]</sup>。在慈竹Unigene数据库中,分别与Swiss-port和Nr数据库比对时,与CesA基因同源的Unigene分别有49条和33条;与Csl基因同源的Unigene分别有36条和27条;与SuSy基因同源的Unigene分别有14条和7条。这些基因序列与毛竹、水稻、玉米等物种的CesA、Csl和SuSy基因具有很高的同源性。进一步分析表明,不同高度慈竹笋的纤维素合酶CesA4、CesA7和CesA9基因存在差异表达且随着笋的伸长,总体呈上升的趋势(图6)。有研究表明,竹类植物笋期的高速生长与笋的居间分生组织中细胞的快速生长与伸长有关,初

生壁的发育确保了细胞的正常生长,而次生壁的发育所起到的支撑作用也确保了笋的快速生长<sup>[24-25]</sup>。另有研究表明,不同的CesA基因分别在细胞初生壁和次生壁合成时期起不同的作用<sup>[23]</sup>。水稻的OsCesA4、OsCesA7和OsCesA9是细胞次生壁合成所必需的,在木质部的细胞中高度表达。OsCesA4、OsCesA7或OsCesA9突变后会导致水稻茎秆变脆且纤维素含量降低,进而影响水稻的正常生长<sup>[26]</sup>。因此推测慈竹的纤维素合酶基因CesA4、CesA7和CesA9与细胞次生壁的合成有关且可能影响慈竹笋的生长发育,这有待进一步研究。

## REFERENCES

- [1] Peng ZH, Lu Y, Li LB, et al. The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). *Nat Genet*, 2013, 45(4): 456–463.
- [2] Peng ZH, Zhang CL, Zhang Y, et al. Transcriptome sequencing and analysis of the fast growing shoots of moso bamboo (*Phyllostachys edulis*). *PLoS ONE*, 2013, 8(11): e78944.
- [3] Liu MY, Qiao GR, Jiang J, et al. Transcriptome sequencing and *De novo* analysis for ma bamboo (*Dendrocalamus latiflorus Munro*) using the illumina platform. *PLoS ONE*, 2012, 7(10): e46766.
- [4] Qi JQ, Chi B, Xie JL, et al. Study on variations of fiber morphology and tissue proportion of *Neosinocalamus affinis* culm. *Trans China Pulp Paper*, 2013, 28(3): 1–4 (in Chinese).  
齐锦秋, 池冰, 谢九龙, 等. 慈竹纤维形态及组织比量的研究. *中国造纸学报*, 2013, 28(3): 1–4.
- [5] Liu Q, He H, Shen ZP. Study on the growth status and correlative environment factors of bamboo, *Neosinocalamus affinis* from Chengdu region. *Sichuan Environ*, 2001, 20(4): 43–46 (in Chinese).  
刘庆, 何海, 沈昭萍. 成都地区慈竹生长状况及其

- 与环境因子关系的初步分析. 四川环境, 2001, 20(4): 43-46.
- [6] Hu SL, Cao Y, Huang SX, et al. Cloning and bioinformation analysis of 4CL gene in *Neosinocalamus affinis*. J Northwest A&F Univ: Nat Sci Ed, 2009, 37(8): 204-210 (in Chinese).  
胡尚连, 曹颖, 黄胜雄, 等. 慈竹 4CL 基因的克隆及其生物信息学分析. 西北农林科技大学学报: 自然科学版, 2009, 37(8): 204-210.
- [7] Chen QB, Jiang Y, Lu XQ, et al. Assessment of genetic diversity in *Neosinocalamus affinis* from different regions of Sichuan province. J Northwest A&F Univ: Nat Sci Ed, 2009, 37(6): 187-193 (in Chinese).  
陈其兵, 蒋瑶, 卢学琴, 等. 四川不同地区慈竹的遗传多样性研究. 西北农林科技大学学报: 自然科学版, 2009, 37(6): 187-193.
- [8] Hu SL, Jiang Y, Chen QB, et al. Physical and chemical properties of 2 species in bamboos from the different regions in Sichuan province. Bull Botan Res, 2010, 30(6): 708-712 (in Chinese).  
胡尚连, 蒋瑶, 陈其兵, 等. 四川 2 种丛生竹理化特性及纤维形态研究. 植物研究, 2010, 30(6): 708-712.
- [9] Li XL, Lin RS, Fung HL, et al. Chromosome numbers of some caespitose bamboos native in or introduced to China. Acta Phytotaxonom Sin, 2001, 39(5): 433-442 (in Chinese).  
李秀兰, 林汝顺, 冯学琳, 等. 中国部分丛生竹类染色体数目报道. 植物分类学报, 2001, 39(5): 433-442.
- [10] Tang WW, Wu XL, Zhou J. Comparison of neutral detergent fiber, acid detergent fiber and acid detergent lignin contents of Dongxiang wild rice and cultivated rice. Agric Sci Technol, 2010, 11(3): 11-14.
- [11] Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol, 2011, 29(7): 644-652.
- [12] Tatusov RL, Galperin MY, Natale DA, et al. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res, 2000, 28(1): 33-36.
- [13] Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. Nat Genet, 2000, 25(1): 25-29.
- [14] Kanehisa M, Goto S, Kawashima S, et al. The KEGG resource for deciphering the genome. Nucleic Acids Res, 2004, 32(Suppl 1): D277-D280.
- [15] Zhang CY, Long Y, Feng J, et al. Transcriptional regulation of plant genes and its significance in biology. Hereditas (Beijing), 2007, 29(7): 793-799 (in Chinese).  
张椿雨, 龙艳, 冯吉, 等. 植物基因在转录水平上的调控及其生物学意义. 遗传, 2007, 29(7): 793-799.
- [16] Lin CY, Trinh NN, Lin CW, et al. Transcriptome analysis of phytohormone, transporters and signaling pathways in response to vanadium stress in rice roots. Plant Physiol Biochem, 2013, 66: 98-104.
- [17] Zhang J, Liu W, Han H, et al. *De novo* transcriptome sequencing of *Agropyron cristatum* to identify available gene resources for the enhancement of wheat. Genomics, 2015, 106(2): 129-136.
- [18] Zhao GZ, Sun HY, Zhao HS, et al. A study of genome sequencing of *Phyllostachys edulis* and its data applications. World Bamb Ratt, 2015, 13(3): 8-13 (in Chinese).  
赵广枝, 孙化雨, 赵韩生, 等. 毛竹基因组测序及数据应用研究现状. 世界竹藤通讯, 2015, 13(3): 8-13.
- [19] Zhao CS, Craig JC, Petzold HE, et al. The xylem and phloem transcriptomes from secondary tissues of the *Arabidopsis* root-hypocotyl. Plant Physiol, 2005, 138(2): 803-818.
- [20] Kim WC, Ko JH, Kim JY, et al. *MYB46* directly regulates the gene expression of secondary wall associated cellulose synthases in *Arabidopsis*. Plant J, 2013, 73(1): 26-36.
- [21] Chen QB, Gao SP, Liu L. The selection of paper-pulp bamboo species and the development of bamboo paper sector in Sichuan. J Bamb Res, 2002, 21(4): 47-51 (in

- Chinese).
- 陈其兵, 高素萍, 刘丽. 四川省优良纸浆竹种选择与竹纸产业化发展. 竹子研究汇刊, 2002, 21(4): 47–51.
- [22] Cheng X, Hao HQ, Peng L. Recent progresses on cellulose synthesis in cell wall of plants. *J Trop Subtr Bot*, 2011, 19(3): 283–290 (in Chinese).
- 程曦, 郝怀庆, 彭励. 植物细胞壁中纤维素合成的研究进展. 热带亚热带植物学报, 2011, 19(3): 283–290.
- [23] McFarlane HE, Döring A, Persson S. The cell biology of cellulose synthesis. *Annu Rev Plant Biol*, 2014, 65: 69–94.
- [24] Gan XH, Ding YL. Investigation on the variation of fiber wall in *Phyllostachys edulis* culms. *For Res*, 2006, 19(4): 457–462 (in Chinese).
- 甘小洪, 丁雨龙. 毛竹茎秆纤维发育过程中细胞壁的变化规律研究. 林业科学研究, 2006, 19(4): 457–462.
- [25] Cui K, He CY, Zhang JG, et al. Characteristics of temporal and spatial tissue development during the rapidly growing stage of moso bamboo culms. *For Res*, 2012, 25(4): 425–431 (in Chinese).
- 崔凯, 何彩云, 张建国, 等. 毛竹茎秆组织速生的时空发育特征. 林业科学研究, 2012, 25(4): 425–431.
- [26] Tanaka K, Murata K, Yamazaki M, et al. Three distinct rice cellulose synthase catalytic subunit genes required for cellulose synthesis in the secondary wall. *Plant Physiol*, 2003, 133(1): 73–83.

(本文责编 陈宏宇)