Plant Science 181 (2011) 696–700

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Analysis of expressed sequence tags from biodiesel plant *Jatropha curcas* embryos at different developmental stages

Mao-Sheng Chen[a], Gui-Juan Wang[a], Ru-Ling Wang[a], Jun Wang[c], Song-Quan Song[b,*], Zeng-Fu Xu[a,**]

[a] *Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, 88 Xuefu Road, Kunming 650223, Yunnan, China*
[b] *Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China*
[c] *Beijing Genomics Institute at Shenzhen, Shenzhen 518000, China*

ARTICLE INFO

Article history:
Received 1 September 2010
Received in revised form 7 August 2011
Accepted 10 August 2011
Available online 31 August 2011

Keywords:
Bioenergy plant
Embryo
Expressed sequence tags
Lipid metabolism
Physic nut
Seed development

ABSTRACT

Jatropha curcas is considered a potential biodiesel feedstock plant whose seeds contain up to 40% oil. However, little is currently known about the seed biology of *Jatropha*. Therefore, it would be valuable to understand the mechanisms of development and lipid metabolism in *Jatropha* seeds. In the present study, three cDNA libraries were constructed with mRNA from *Jatropha* embryos at different stages of seed development. A total of 9844 expressed sequence tags (ESTs) were produced from these libraries, from which 1070 contigs and 3595 singletons were obtained. One hundred and seven unigenes were found to be differentially expressed in the three cDNA libraries of *Jatropha* embryos, indicating that these genes may play key roles in seed development. We have identified 59 and 61 unigenes that might be involved in the development and lipid metabolism in *Jatropha* seeds, respectively. Some of these genes may also play important roles in embryogenesis, morphogenesis, defense response and adaptive mechanisms in plants.

1. Introduction

Jatropha curcas L., belonging to the family Euphorbiaceae, is a drought-resistant oil plant that is widely distributed in tropical and subtropical areas [1,2]. *Jatropha* has been traditionally used as a living fence against domestic animals and for fertilizer, medicine production and control of soil erosion [1,3,4]. Unlike crops such as oil-seed rape, maize and soybean, *Jatropha* does not occupy farmland, and can produce non-edible seeds with high oil content (up to 40%) [5]. In recent years, the use of *Jatropha* as a biodiesel feedstock plant has greatly interested a large number of researchers. China, India and several African countries have initiated large-scale plantations of *Jatropha* [5,6]. *Jatropha* has small genome size (about 416 Mb) and few chromosome number ($2n = 22$) [7], and is amenable to genetic transformation [8–10]. Recently, *Jatropha* whole genome sequence was made publicly available [11], which make it suitable as a model plant for biodiesel feedstock research. As a wild plant, however, seed yield of *Jatropha* is poor and insufficient for the biodiesel industry [6,12,13]. Moreover, little is currently known about the genetic information and molecular biology of *Jatropha*. Therefore, it would be valuable to understand the mechanism of seed development and lipid metabolism in *Jatropha*, which would be helpful in using genetic engineering to develop new *Jatropha* cultivars.

Expressed sequence tag (EST) analysis can provide a convenient and efficient method for identification of genes expressed in specific tissues and cells, as well as allow the characterization of transcript expression levels [14]. Combined with breakthroughs in highly parallel designs for gene expression analysis, EST analysis can also provide valuable information for understanding the molecular basis of important agricultural traits in plants [15]. In this study, we sequenced the 5′-ends of about 10,000 cDNA clones randomly selected from three cDNA libraries derived from different developmental stages of *Jatropha* seeds. A large number of putative genes associated with signal transduction, synthesis of stored reserves, and metabolism and transport of amino acids, lipids, and proteins, have been identified. Therefore, this study provides a basis for understanding the mechanism of seed development and lipid metabolism in *Jatropha*. Moreover, the study can also provide a valuable resource for the cloning of new genes, annotation of genomic sequences and the development of molecular markers for gene mapping, polymorphism and marker-assisted selection breeding of *Jatropha*, which are prerequisites for the application of genetic engineering in *Jatropha* breeding [16].

* Corresponding author. Tel.: +86 10 6283 6484; fax: +86 10 6259 0833.
** Corresponding author. Tel.: +86 871 514 4423; fax: +86 871 516 0916.
E-mail addresses: sqsong@ibcas.ac.cn (S.-Q. Song), zfxu@xtbg.ac.cn, zengfu.xu@gmail.com (Z.-F. Xu).

0168-9452/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.plantsci.2011.08.003

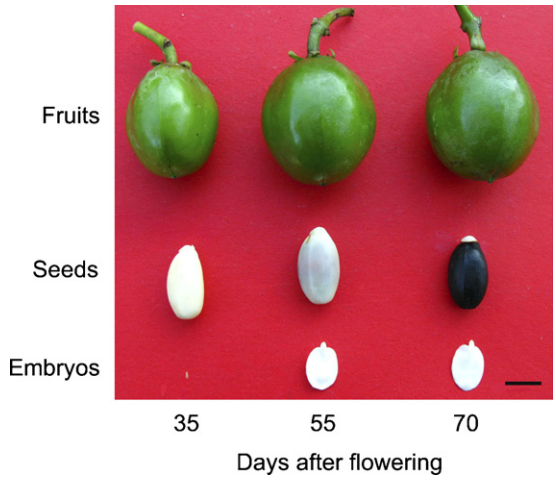

Fig. 1. *Jatropha curcas* fruits, seeds and embryos at different developmental stages. Seeds were harvested at 35–55, 56–70 and 71–95 days after flowering (DAF). Total RNA was extracted from embryos to construct cDNA libraries. Scale bar = 1.0 cm.

2. Materials and methods

2.1. cDNA library construction and sequencing

Jatropha seeds were harvested at 35–55, 56–70 and 71–95 days after flowering (DAF), which represent the early, middle and late stage of seed development, respectively (Fig. 1). Embryos were excised from seeds at the three development stages. mRNAs were purified with Oligotex mRNA Isolation Kits (Qiagen, Valencia, CA) from total RNA extracted from embryos. cDNAs were synthesized with the SuperScript II-RT system (Invitrogen). Three cDNA libraries, early stage library (35–55 DAF, library I), middle stage library (56–70 DAF, library II) and late stage library (71–95 DAF, library III), were generated with the plasmid pBluescript II SK (+). cDNA colonies in each library were picked randomly, and sequenced once from the 5′-end of each clone using an automated DNA sequencer (GE MegaBase 1000 sequencers) at the Beijing Genomics Institute.

2.2. Sequence assembly and analysis

After removing ribosomal RNA, poly(A), vector and low-quality sequences, high-quality ESTs (length of sequence ≥ 100 bp and phred quality value ≥ 20) were assembled into clusters using the PHRED/PHRAP/CONSED software package [17,18]. The 4665 unigenes were searched against the NCBI non-redundant nucleotide databases (NT) using the blastn program with an *E*-value ≤ 1.0E−05, and the NCBI non-redundant protein database (NR) and the SWISS-PROT database using the blastx program with an *E*-value ≤ 1.0E−05, and Kyoto Encyclopedia of Genes and Genomes database (KEGG, http://www.genome.jp/kegg/) using the blastx program with an *E*-value ≤ 1.0E−10 [19,20]. Functional classification of the ESTs was further examined according to the NCBI Clusters of Orthologous Groups of Proteins (COGs) database using the blastx program with an *E*-value ≤ 1.0E−10 [21].

2.3. EST expression profiling

Data analysis on gene expression profiles in three cDNA libraries was performed using R statistics with Bonferonni correction at the significance threshold of 1.85E−05 using the web tool IDEG6 [22,23]

Table 1
Summary of the number of unigenes in three cDNA libraries from *Jatropha curcas* embryos.

Library	Unigenes	Redundancy (%)	Mean length (bp)
I. Early stage (35–55 DAF)	2295	32.5	444.4
II. Middle stage (56–70 DAF)	1646	55.3	476.2
III. Late stage (71–95 DAF)	1512	46.7	536.4
Total	4665	52.6	496.2

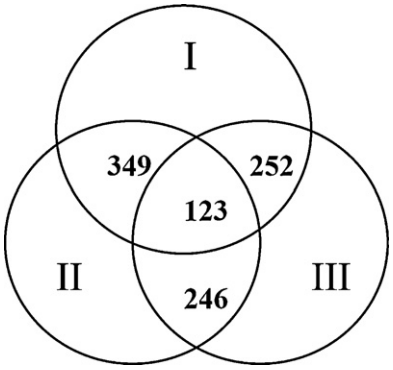

Fig. 2. Overlapping of unigenes detected in three libraries. A total of 2295, 1646, and 1512 unigenes were present in libraries I, II and III, respectively. Among them, 123 unigenes were common in the three libraries. Libraries I and II shared 349 unigenes, libraries II and III shared 252, and libraries III and I shared 246.

3. Results and discussion

3.1. Sequencing and analysis of Jatropha ESTs

Three cDNA libraries were constructed from *Jatropha* embryos at different developmental stages of seeds (Fig. 1). From each cDNA library, 3000–4000 clones were sequenced, from which a total of 10,913 5′-end sequences were generated. After trimming low-quality and vector sequences and removing ribosomal RNA sequences, 9844 high-quality ESTs with a minimum of ≥100 bp, phred quality value ≥ 20 and average length of 496 bp were obtained. All sequences have been deposited in DDBJ/EMBL/GenBank (FM887038–FM896881). The 9844 ESTs were assembled into 4665 unigenes containing 1070 contigs and 3595 singletons using PHRED/PHRAP/CONSED software, which represent putative transcripts that vary during seed development in *Jatropha* (Additional file 1). Libraries I, II and III contains 2295, 1646 and 1512 unigenes, respectively (Table 1). There was an 8.86% overlap between libraries I and II, a 6.62% overlap between I and III, and a 7.79% overlap between II and III, respectively (Table 1; Fig. 2). Surprisingly, there was a 2.64% overlap among three libraries and only 123 unigenes were common (Fig. 2). These results show that the development of *Jatropha* embryo is a complex process with progressive changes in physiology and morphology, involving a large number of genes that are specifically expressed during different developmental stages. In addition, because only 3000–4000 clones have been sequenced from each library, some common genes expressed in the three libraries have not been identified.

3.2. Annotation and functional classification of Jatropha unigenes

A homology search by using the blastx program revealed that 65%, 40.8% and 59.5% of the 4665 *Jatropha* unigenes had significant matches with sequences in the NCBI non-redundant protein database (NR), SWISS-PROT protein database and the Kyoto Encyclopedia of Genes and Genomes database (KEGG), respectively. By using the blastn program, about 97.9% of these unigenes were