Pak. J. Bot., 50(3): 989-998, 2018.

CLONING AND EXPRESSION ANALYSIS OF *JcAACT*, *JcMDC* AND *JcFPS*, INVOLVED IN TERPENOID BIOSYNTHESIS IN *JATROPHA CURCAS* L.

YAOYAO HUANG[1], JINFEN WEN[2*] AND MINGHUA DENG[3]

[1]*Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology, Kunming 650500*
[2]*Faculty of Architecture and City Planning，Kunming University of Science and Technology, Kunming 650500*
[3]*College of Landscape and Horticulture, Yunan Agricultural University, Kunming 650224*
[*]*Corresponding author's: wenjf888@163.com*

Abstract

To better understand the functions of key genes involved in terpenoid biosynthesis in *Jatropha curcas*, we cloned and characterized three genes, namely acetyl CoA acyltransferase (*JcAACT*), diphosphate mevalonate decarboxylase (*JcMDC*) and farnesyl pyrophosphate synthase (*JcFPS*). The opening reading frames (ORFs) of *JcAACT*, *JcMDC* and *JcFPS* were 1239 bp,1248 bp and 1029 bp, respectively, encoding a 412-amino acid, 415-amino acid and 342-amino acid polypeptide, respectively. Results of homology analysis showed that *JcAACT*, *JcMDC* and *JcFPS* encoded proteins that all had the highest identity and closest relationship with the corresponding genes in *Hevea brasiliensis*, with identities of 89%, 92% and 93%, respectively. *JcAACT*, *JcMDC* and *JcFPS* were expressed in all organs tested of *J. curcas*; the highest expression level for each gene occurred in seeds. In the early growth stage of seeds, the expression level of each of these three genes increased with time, with *JcAACT* and *JcMDC* expression level reaching a peak at the late stage of seed development (50 d), while *JcFPS* expression level reached a peak at the mid-late stage (40 d). Following the peak, the expression of each gene then declined. The expression level of *JcAACT* was the highest of the three genes, regardless of the organ or the stage of seed growth, indicating its important role in *J. curcas*. This study lays the foundation for a better understanding of the important role of the *JcAACT*, *JcMDC* and *JcFPS* genesin the terpenoid biosynthesis pathway of *J. curcas*.

Key words: Terpenoid biosynthesis pathway; *JcAACT; JcMDC; JcFPS;* Clone; Expression analysis.

Introduction

Terpenoids, based on the isoprene unit, are the largest class of plant secondary metabolites. Terpenoid compounds play not only an important role in maintaining plant life and regulating the relationship between plants and the environment, such as enhancing the ability of plants to resist disease or to tolerate abiotic stresses, but also are widely used in industrial products, such as biodiesel or in pharmaceuticals, which are of significant commercial value. The terpenoid biosynthesis pathway is also called the isoprenoid pathway, and it is one of the most important metabolic pathways in plants. Terpenoid biosynthesis involves two routes, the mevalonate (MVA) pathway operating in the cytoplasm of plant cells, and the 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway in plastids (Laule *et al*., 2003), with the MVA pathway being recognized as the main pathway for the synthesis of terpenoid compounds. Although the MVA and MEP pathways exist in different compartments of the plant cell, their separation is not absolute (Kasahara *et al*., 2002). Both pathways use isopentenyl pyrophosphate (IPP) as the main intermediate product. After IPP has polymerized, the C_5 chain continues to extend, finally forming a variety of compounds (Haralampidis *et al*., 2002; Kim *et al*., 2010).

Acetyl coenzyme A transferase (*AACT*) is the first enzyme in the terpenoid biosynthesis pathway and is specific for the thiolysis of acetoacetyl-CoA. It catalyzes the formation of acetoacetyl-CoA by transferring an acetyl group from one acetyl-CoA molecule to another (Jin *et al*., 2012). Because this is the first step of many biosynthetic pathways, AACT plays a fundamental role in the carbon skeleton assembly patterns in many biological systems, including the synthesis of steroid hormones, cholesterol and ketone bodies.

AACT has been cloned from *Salvia miltiorrhiza* (Cui *et al*., 2010), *Camellia oleifera* (Zhang *et al*., 2011), *Ganoderma lucidum* (Xing *et al*., 2013), *Houttuynia cordata* (Yao *et al*., 2015) and *Tripterygium wilfordii* (Zhao *et al*., 2015). By analysis of the *AACT* semi-quantitative RT-PCR expression profiles in *Ganoderma lucidum*, Xing *et al*. (2013) found that, during the development process of *Ganoderma lucidum*, *AACT* expression and triterpene content were positively correlated, indicating that *AACT* plays a key role in triterpene biosynthesis. Cui *et al*., (2010) found that, in the presence of yeast extracts and Ag^+, the expression level of *AACT* was associated with the content of ketones.

Mevalonate pyrophosphate decarboxylase (*MDC*) is the last rate-limiting enzyme before IPP synthesis on the MVA pathway. *MDC* catalyzes mevalonate diphosphate and ATP to form IPP (Byres *et al*., 2007; Shi *et al*., 2012; Simkin *et al*., 2011. It is the least-studied member of the GHMP kinase superfamily. *MDC* has been cloned from *Arabidopsis thaliana* (Cordier *et al*., 1999), *Catharanthus roseus* (Simkin *et al*., 2011), *Panax notoginseng* (Shi *et al*., 2012), *Acanthopanax senticosus* (Xing *et al*., 2012) and *Gentiana rigescens* (Zhang *et al*., 2015). In *G. lucidum*, the expression of *GlMDC* was positively correlated with the content of triterpenes (Shi *et al*., 2012). Over-expression of *GlMDC* in *Panax ginseng* can increase triterpene content by 17%-101.4% (Shi *et al*., 2012), while, in the hairy roots of ginseng, over-expression of *PgMDC* resulted in a 4.4 times increase in the content of the triterpenoid compound stigmasterol (Kim *et al*., 2014).

Farnesyl pyrophosphate synthase (FPS) catalyzes 1'-4 condensation of the 5-carbon isoprenoid compound IPP and the 10-carbon geranyl diphosphate (GPP) to form the 15-carbon product farnesyl diphosphate (FPP) (Closa *et*

al., 2010). FPP is used as a precursor for the synthesis of many kinds of terpenoid derivatives, such as steroids, saponins, triterpenes, sesquiterpenes, and rubber (Szkopifiska & Plochocka, 2005). These substances play an important role in the growth and development of plants and in the process of disease resistance. FPS is one of the most extensively researched prenyl transferases. Due to the important role of this enzyme in terpenoid biosynthesis, *FPS* has been cloned from *Arabidopsis thaliana* (Cunillera *et al*.,1996), *Panax notoginseng* (Chen *et al*., 2006), rice (Sanmiya *et al*., 1997), cotton (Liu *et al*., 1998), maize (Li &Larkins , 1996) and dozens of other species. Overexpression of the *FPS* gene in transgenic *Artemisia annua* resulted in an increase in artemisinin content of almost four times (Han *et al*., 2006; Banyai & Kirdmanee, 2010). The *FPS* gene of mint was transferred into tobacco, resulting in a significant increase in its resistance to brown spot (Cui *et al*., 2006). The expression level of the *FPS* gene of *Alisma orientale* was positively correlated with the content of the active component alisol B 23-acetate, demonstrating that *FPS* plays an important role in the process of the synthesis of terpenoid compounds (Gu *et al*., 2011).

Jatropha curcas L. belongs to the Euphorbiaceae, and the high quantity and quality of the seed oil makes it a potential candidate for biodiesel production, which could not only protect the environment but also solve the problem of global fossil fuel depletion (Zeng *et al*., 2006). Terpenoid compounds are among the main chemical constituents of *J. curcas*. Modern pharmacological research has shown that terpenoid compounds are a useful raw material for drug chemistry research and development. Diterpenoid and triterpenoid compounds isolated from the roots, stems, leaves and seeds of *J. curcas* have antibacterial, antitumor and anti-HIV effects (Quan *et al*., 2009).After oil extraction, the remaining seed cake is one of the largest by-products of *Jatropha* biodiesel production; with its high protein content and appropriate amino acid balance, the seed cake can provide a highly-nutritious animal protein feed,but, because it contains a lot of toxic substances, it cannot be fed directly. One tetracyclic diterpenoid-phorbol ester is the main toxic component, being toxic to many microorganisms and animals, as well as being the most difficult to remove. As a consequence, the phorbol ester has become the main reason for restricting the use of the *Jatropha* seed cake as an animal feed resource (Bao *et al*., 2011). Of the various *Jatropha* organs, seeds have the highest content of the phorbol ester (Martinez *et al*., 2006; Adolf *et al*., 1984; Makkar *et al*., 1997), and, while several methods have been tested to remove the ester from the seed cake, such as alkali treatment or methanol extraction, so far, no ideal method has been developed. An alternative approach is to achieve removal of the phorbol ester at the molecular level. As part of this strategy, this study cloned three genes involved in terpenoid biosynthesis, *JcAACT*, *JcMDC* and *JcFPS*, and analyzed the bioinformatics and expression profiles of these genes, to provide a theoretical basis for the molecular regulation of terpenoid synthesis in *J. curcas*.

Materials and Methods

Plant material: The plants used in this study were grown at Kunming University of Science and Technology, Yunnan Province, China. Roots, stems, leaves and seeds of *J. curcas* were harvested and snap-frozen by immediate immersion in liquid nitrogen. At the same time, seeds were also collected at different development stages after pollination (10, 20, 30, 40, 50 and 60 d) and immediately snap-frozen in liquid nitrogen. All these *Jatropha* tissues were stored at -80ºC before use. Each tissue was sampled from ten plants of *J. curcas.* Tissue samples from three trees were pooled to represent a replicate tissue sample, and repeated four times.

RNA isolation and cDNA cloning: RNA was isolated using the TIANGEN RNAprep pure Kit (DP432; TIANGEN, Beijing, China), treated with RNase-Free DNaseI (Takara, Dalian, China) and subsequently used as the template for TaKaRa Prime Script™ II 1st Strand cDNA Synthesis Kit reverse transcription (RT) (TaKaRa, Dalian, China). The cDNA was stored at -20ºC before being used. For PCR amplification, TaKaRa Ex Taq Hot Start Version (TaKaRa), 5×PCR buffer and gene-specific primers (Table 1) were used.

Table 1. Oligonucleotide primers used in the study.

Name	Sequence
JcAACT-F	5’-GCTTAAATTCAAAATCCATCG-3’
JcAACT-R	5’-ATTCAAGTAACGGAACATGCA-3’
JcMDC-F	5’-CCCAGACTATGACATCTCCCTAC-3’
JcMDC-R	5’-CAAGACAATAGACGTGCAAGAAA-3’
JcFPS-F	5’-TCTCCTCACTACTGCCCTCCCT-3’
JcFPS-R	5’-CAATCATTGACTGTGCTTCTGC-3’
JcAACT-qF	5’-GTCTTCTTTGGCAATGTTCTTAG-3’
JcAACT-qR	5’-ATCCCACCAACCACAACAATA-3’
JcMDC-qF	5’-TCGCAGTTTGTTTGGTGGTT-3’
JcMDC-qR	5’-TTCAACAGTCTCACGCATTCC-3’
JcFPS-qF	5’-CAATGTGCCTGGAGGGAAG-3’
JcFPS-qR	5’-CTTGGAGCCATTCAATACACC-3’
β-actin-F	5’-GCAGGCATCCACGAGACTACT-3’
β-actin-R	5’-GTCAGCAATACCAGGGAACATAG-3’

Bioinformatic and phylogenetic analyses: Sequence analysis was performed using software from NCBI (http://www.ncbi.nlm.nih.gov) and ExPaSy (http://www.expasy.org). The blastp program (http://www.ncbi.nlm.nih.gov/Blast) and the Conserved Domain Architecture Retrieval Tool were used to search for similar proteins and conserved domains, respectively. The alignment of the nucleotide sequences and the deduced amino acid sequences were computed using DNAMAN, and the phylogenetic trees were computed using the Mega 7.0 software, with standard parameters. The theoretical isoelectric point (pI) and molecular weight (Mw) were predicted using the Compute pI/Mw Tool (http://us.expasy.org/tools/pi_tool.html). The putative signal peptide was predicted, using the SignalP 3.0 server (http://www.cbs.dtu.dk/services/SignalP/). The putative protein subcellular localization was predicted using PSort II (http://psort.hgc.jp/). Transmembrane topology prediction was performed using the TMHMM-2.0 server (http://www.cbs.dtu.dk/services/TMHMM-2.0/). Secondary structures of deduced amino acid sequences were predicted with SOPMA (http://npsa-pbil.ibcp.fr/). The tertiary structures were predicted based on the existing primary structures, using the amino acid homology modeling on the SWISS-MODEL server (http://swissmodel.expasy.org/).

Expression analysis: RNA was isolated and cDNA synthesized as described in subsection 2.2. qRT-PCR was carried out using SYBR Premix Ex Taq^{TM}II (Tli RNaseH Plus; manufacturer, Dalian, China), according to the manufacturer's recommendations; the experiment was carried out on the Roche Light Cycler 480 system. Analysis of gene expression data was based on the $2^{-\Delta\Delta CT}$ method reported by Livak (Livak & Schmittgen, 2001).In the tissue expression analysis, the lowest expression value was used as the reference to obtain "Relative Expression". In the expression analysis during seed development, the sample collected at the first development stage (10 d) was used as the reference to obtain "Relative Expression".

Results

Cloning and characterization: *JcAACT, JcMDC* and *JcFPS* genes were cloned by RT-PCR. After agarose gel electrophoresis, we found three products that were more than 1000 bp in length,consistent with the expected product lengths. After sequencing, we found that the full-length sequence of *JcAACT* cDNA contained a 1239 bp ORF, encoding a 412-amino acid protein. The full-length sequence of *JcMDC* cDNA contained a 1248 bp ORF, encoding a 415-amino acid protein, while the full-length sequence of *JcFPS* cDNA contained a 1029 bp ORF, encoding a 342-amino acid protein (Figs. 1-4).

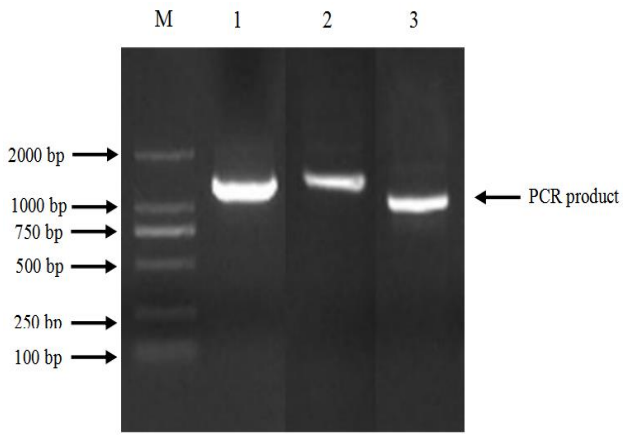

Fig. 1. PCR products of *JcAACT*, *JcMDC* and *JcFPS*.
M: DNA marker; 1: PCR product of *JcAACT*; 2: PCR product of *JcMDC* 3: PCR product of *JcFPS*

Protein primary structure analysis: ProtParam was used to analyze the protein primary structure of the products of these three genes, with results being shown in Table 2.

Protein secondary structure prediction: The prediction of secondary structure by SOPMA indicated that the deduced *JcAACT* protein contained 170 alpha helices, 17 extended strands, 59 beta turns and 112 random coils. The *JcMDC* protein contained 142 alpha helices, 98 extended strands, 35 beta turns and 140 random coils, while the *JcFPS* protein contained 198 alpha helices, 37 extended strands, 28 beta turns and 79 random coils. The number of domains and active sites in each protein was determined, using prosite (http://expasy.org/prosite/) software. Six types of site were found in *JcAACT*, six in *JcMDC*, and four in *JcFPS*, as shown in Table 3.

Protein tertiary structure prediction: The SWISS-MODEL site (http://swissmodel.expasy.org/ MODELING) was used to predict protein tertiary structure of *JcAACT, JcMDC* and *JcFPS* (Fig. 5).

Prediction and analysis of protein transmembrane structure and signal peptide: Cell location prediction revealed that the proteins encoded by *JcAACT, JcFPS* and *JcMDC* were most probably located in the cytoplasm. Using a hidden Markov model algorithm, transmembrane topology predictions made by the TMHMM program showed that JcAACT, JcFPS and JcMDD were not potential membrane proteins, while the same proteins were predicted not to contain a signal peptide.

Homology analysis

Sequence analysis of the deduced *JcAACT* amino acid sequence: Multiple alignments of the *JcAACT* deduced amino acid sequence showed strong similarities to other *AACT* genes from various plant species, including *Hevea brasiliensis* (AFJ74323) with an identity of 89%, *Ricinus communis*(XP 015577103) with an identity of 88%, *Nelumbo nucifera* (XP 010267979) with an identity of 87%, *Populus euphratica* (XP 011002351)and *Cicer arietinum* (XP 004507588)with identities of 86%, *Euphorbia helioscopia* (ALC76524)with an identity of 85%, and *Prunus mume* (XP 008224638), *Morus alba* (ALD84318), *Nicotiana tabacum* (XP 016475988) and *Solanum tuberosum* (XP 006350251) with identities of 84%. As shown by InterProScan, the deduced *JcAACT* protein contained an acetyl-CoA acetyltransferase family distinctive domain, belonging to the cond enzyme superfamily, and two thiolase active sites, one of which was characteristic of thiolase 2 (348-364, NvhGGaVSlGHPlGcSG) and one characteristic of thiolase 3 (383-396, GVAAICNGgGgAsA) (Fig. 6).

Sequence analysis of the deduced *JcMDC* amino acid sequence: The multiple alignment of the *JcMDC* deduced amino acid sequence showed strong similarities to other MDC proteins from various plant species, including *Hevea brasiliensis*(AFJ74330) with an identity of 92%, *Ricinus communis* (XP_002521172) and *Populus trichocarpa* (XP_002311015,) with identities of 90%, *Populus euphratica* (XP_011032862) with an identity of 89%, *Euphorbia helioscopia* (ALC76525) with an identity of 87%, *Nicotiana tabacum* (XP_016498510), *Morus alba* (ALD84324) and *Theobroma cacao* (XP_007010314) with identities of 85%, and *Capsicum annuum* (XP_016551358) and *Astragalus membranaceus* (AID51442) with identities of 84%. InterProScan indicated that the *JcMDC* protein contained three conserved domains, namely the ribosomal protein S5 domain 2 (6-196), the GHMP kinase N-domain (113-170) and the GHMP kinase C-domain (199-413). The signature sequence of the GHMP kinase superfamily is 110–322 amino acid residues (Fig. 7).

```
1       ATGGCGGCTTCTTCGGATTCTATCAATCCTCGAGATGTTTGTGTCGTGGGTGTTGCACGCACACCAATGGGTGGT
1        M  A  A  S  S  D  S  I  N  P  R  D  V  C  V  V  G  V  A  R  T  P  M  G  G
76      TTTCTTGGTTCACTTTCATCTCTTTCAGCTACAAAGCTCGGCTCTATAGCTATTCAGTCTGCTCTTAAAAGAGCA
26       F  L  G  S  L  S  S  L  S  A  T  K  L  G  S  I  A  I  Q  S  A  L  K  R  A
151     AATGTTGATCCATCACTCGTGCAAGAGGTCTTCTTTGGCAATGTTCTTAGTGCTAATTTAGGACAAGCTCCTGCA
51       N  V  D  P  S  L  V  Q  E  V  F  F  G  N  V  L  S  A  N  L  G  Q  A  P  A
226     AGGCAGGCTGCTTTAGGTGCGGGTATTCCTAATTCAGTGAATTGCACCACAATTAATAAAGTTTGTTCTTCGGGG
76       R  Q  A  A  L  G  A  G  I  P  N  S  V  N  C  T  T  I  N  K  V  C  S  S  G
301     ATGAAAGCAACTATGATTGCTGCACTAAGTATCCAAGCGGGTATTAATGATATTGTTGTGGTTGGTGGGATGGAA
101      M  K  A  T  M  I  A  A  L  S  I  Q  A  G  I  N  D  I  V  V  V  G  G  M  E
376     AGCATGTCCAAAGCACCTAAGTATCTTGCAGAAGCAAGAAAGGGTTCTCGACTAGGACATGATACCATCATTGAC
126      S  M  S  K  A  P  K  Y  L  A  E  A  R  K  G  S  R  L  G  H  D  T  I  I  D
451     GGCATGCTCAAAGATGGTCTGTGGGATGTATATAATGACTTTGGAATGGGAGTTTGTGCAGAAATTTGTGCTGAC
151      G  M  L  K  D  G  L  W  D  V  Y  N  D  F  G  M  G  V  C  A  E  I  C  A  D
526     CAACATAAAATTACAAGAGAAGAGCAGGATTCTTATGCTGTACGGAGCTTTGAGCGTGGAATTTCTGCACAAAAT
176      Q  H  K  I  T  R  E  E  Q  D  S  Y  A  V  R  S  F  E  R  G  I  S  A  Q  N
601     GGTGGTTTTTTTTCGTGGGAAATAGTTCCGGTTGAAGTTCCTGGGGGAAGAGGGAAACCTGCCACTATCATTAAT
201      G  G  F  F  S  W  E  I  V  P  V  E  V  P  G  G  R  G  K  P  A  T  I  I  N
676     AAGGATGAAGGTTTAGGAACGTTTGATGCTGCAAAATTGAGGAAGCTTAGACCAAGTTTCAAGGAGAATGGTTCT
226      K  D  E  G  L  G  T  F  D  A  A  K  L  R  K  L  R  P  S  F  K  E  N  G  S
751     GTTACAGCTGGAAATGCGTCTATCATAAGTGATGGTGCAGCTGCATTAGTGCTGGTGAGTGGGGAAAAAGCCATT
251      V  T  A  G  N  A  S  I  I  S  D  G  A  A  A  L  V  L  V  S  G  E  K  A  I
826     AAGCTTGGTTTGCAAGTGATTGCTAGGATAAGAGGATATGCTGATGCTGCCCAGGCCCCTGAGTTGTTTCCAACT
276      K  L  G  L  Q  V  I  A  R  I  R  G  Y  A  D  A  A  Q  A  P  E  L  F  P  T
901     GCACCAGCCCTTGCAATACCAAAAGCTATTTCAAATGCTGGTTTGAAGACTTCCCAGATTGATTACTACGAAATA
301      A  P  A  L  A  I  P  K  A  I  S  N  A  G  L  K  T  S  Q  I  D  Y  Y  E  I
976     AATGAAGCATTTTCTGTCGTGGCTCTTGCCAATCAAAAGCTTCTTAATCTTAAACCAGAAAAAGTAAATGTTCAT
326      N  E  A  F  S  V  V  A  L  A  N  Q  K  L  L  N  L  K  P  E  K  V  N  V  H
1051    GGTGGAGCAGTATCTTTGGGACATCCATTAGGATGCAGTGGGGCACGCATCTTGGTCACATTGTTAGGGGTGCTT
351      G  G  A  V  S  L  G  H  P  L  G  C  S  G  A  R  I  L  V  T  L  L  G  V  L
1126    AGACATAAAAATGGTAAGTATGGGGTTGCTGCCATCTGCAATGGAGGAGGGGGGGCATCTGCCCTTGTTCTTGAG
376      R  H  K  N  G  K  Y  G  V  A  A  I  C  N  G  G  G  G  A  S  A  L  V  L  E
1201    CTCATGTCAAATCCGACGGTGCGACGGTCTTCGTTATGA
401      L  M  S  N  P  T  V  R  R  S  S  L  *
```

Fig. 2. Nucleotide and amino acid sequences of *JcAACT*.

```
1       ATGGCGGATCTGAAATCAACATTCTTGGAGGTCTATTCTGTTCTCAAGAAGGAGCTACTTCAAGACCCGGCTTTC
1        M  A  D  L  K  S  T  F  L  E  V  Y  S  V  L  K  K  E  L  L  Q  D  P  A  F
76      GAATGGACACCAGATTCTCGTGAATGGGTCGAAAGGATGCTGGACTACAATGTGCCTGGAGGGAAGCTGAATAGG
26       E  W  T  P  D  S  R  E  W  V  E  R  M  L  D  Y  N  V  P  G  G  K  L  N  R
151     GGGCTCTCTGTGATTGACAGCTACAAATTGTTGAAAGATGGACAGGAATTAACAGAAGAAGAAATCTTTCTCGCA
51       G  L  S  V  I  D  S  Y  K  L  L  K  D  G  Q  E  L  T  E  E  E  I  F  L  A
226     AGCGCTCTTGGTTGGTGTATTGAATGGCTCCAAGCCTATTTCCTTGTCCTTGATGATATTATGGATAGCTCTCAT
76       S  A  L  G  W  C  I  E  W  L  Q  A  Y  F  L  V  L  D  D  I  M  D  S  S  H
301     ACACGACGTGGTCAACCATGTTGGTTTATGGTGCCCAAGGTTGGTCTTATTGCAGCAAATGATGGGATTTTGCTT
101      T  R  R  G  Q  P  C  W  F  M  V  P  K  V  G  L  I  A  A  N  D  G  I  L  L
376     CGAAATCACATTCCCAGGATTCTTAAAAAGCACTTCCGAGGGAAAGCATACTATGTAGATCTTCTAGATTTATTT
126      R  N  H  I  P  R  I  L  K  K  H  F  R  G  K  A  Y  Y  V  D  L  L  D  L  F
451     AATGAGGTGGAGTTTCAAACAGCCTCAGGACAGATGATAGATCTGATTACAACACTTGAAGGAGAAAAGGATTTA
151      N  E  V  E  F  Q  T  A  S  G  Q  M  I  D  L  I  T  T  L  E  G  E  K  D  L
526     TCGAAGTACAATTTATCGCTTCACCGGCGAATTGTTCAGTACAAAACTGCCTACTACTCATTTTACCTTCCTGTT
176      S  K  Y  N  L  S  L  H  R  R  I  V  Q  Y  K  T  A  Y  Y  S  F  Y  L  P  V
601     GCTTGTGCATTGCTCATGGCTGGTGAGAATCTGGACAGCCATATTGATGTACAGAATATTCTTGTCCAGATGGGA
201      A  C  A  L  L  M  A  G  E  N  L  D  S  H  I  D  V  Q  N  I  L  V  Q  M  G
676     ATCTACTTCCAAGTACAGGATGATTATTTGGATTGCTTTGGTGATCCCAAGACAATTGGCAAGATAGGGACAGAT
226      I  Y  F  Q  V  Q  D  D  Y  L  D  C  F  G  D  P  K  T  I  G  K  I  G  T  D
751     ATTGAAGATTTTAAGTGCTCTTGGTTGGTCGTGAAGGCTTTGGAGCGATGCAATGAAGAACAAAAGAAAGTTCTA
251      I  E  D  F  K  C  S  W  L  V  V  K  A  L  E  R  C  N  E  E  Q  K  K  V  L
826     CATGAGCATTATGGGAAACCTGACCCAGCCAGTGTGTCAAAGGTGAAAGTCCTCTATGATGAGCTGGACCTTCAG
276      H  E  H  Y  G  K  P  D  P  A  S  V  S  K  V  K  V  L  Y  D  E  L  D  L  Q
901     GGGGTATTTATGGAGTATGAGAACCAAAGCTATGATAAACTAGTAACCTCCATTGAGGCTCACCCTAGCAAGGCA
301      G  V  F  M  E  Y  E  N  Q  S  Y  D  K  L  V  T  S  I  E  A  H  P  S  K  A
976     GTGCAAGCAGTGTTGAAGTCTTTCCTTGCCAAAATTTACAAGAGACAGAAATAA
326      V  Q  A  V  L  K  S  F  L  A  K  I  Y  K  R  Q  K  *
```

Fig. 3. Nucleotide and amino acid sequences of *JcMDC*.

```
1          ATGGCAGAAAAATGGGTGAGGATGGTTACTGCACAGACGCCAACAAATATAGCGGTGATTAAGTATTGGGGAAAG
1           M  A  E  K  W  V  R  M  V  T  A  Q  T  P  T  N  I  A  V  I  K  Y  W  G  K
76         AGAGATGAGACCCTTATTTTGCCTGTTAATGATAGTATAAGTGTTACATTAGATCCTTCCCATCTTTGTACTACT
26          R  D  E  T  L  I  L  P  V  N  D  S  I  S  V  T  L  D  P  S  H  L  C  T  T
151        ACAACTGTTGCTGTTAGTCCTACTTTTGATCAGGATCGTATGTGGCTTAATGGAAAGGAGATTTCCCTTTCTGGA
51          T  T  V  A  V  S  P  T  F  D  Q  D  R  M  W  L  N  G  K  E  I  S  L  S  G
226        GGCAGGTTCCAGAGTTGTTTAAGAGAAATTCGTGCTCGAGCCTGTGATGTTGAGGATAAAGAAAAGGGTATCAAG
76          G  R  F  Q  S  C  L  R  E  I  R  A  R  A  C  D  V  E  D  K  E  K  G  I  K
301        ATTGTAAAGAAGGATTGGGATAAATTACACGTGCATATGGCATCATATAACAATTTCCCTACTGCTGCTGGACTG
101         I  V  K  K  D  W  D  K  L  H  V  H  M  A  S  Y  N  N  F  P  T  A  A  G  L
376        GCTTCTTCAGCAGCTGGTTTTGCTTGTCTTGTGTTTGCCCTAGCAAAGCTTATGAATGCTAAAGAAGATAATGGT
126         A  S  S  A  A  G  F  A  C  L  V  F  A  L  A  K  L  M  N  A  K  E  D  N  G
451        GAGCTCTCTGCTATTGCAAGGCAAGGTTCAGGCAGTGCTTGTCGCAGTTTGTTTGGTGGTTTTGTGAAATGGAAC
151         E  L  S  A  I  A  R  Q  G  S  G  S  A  C  R  S  L  F  G  G  F  V  K  W  N
526        ATGGGCAAAGTTGAAGATGGAAGTGACAGTCGTGCTGTTCAAGTTGTTGATGACAAGCACTGGGATGATCTTGTT
176         M  G  K  V  E  D  G  S  D  S  R  A  V  Q  V  V  D  D  K  H  W  D  D  L  V
601        ATTATTATTGCTGTGGTAAGTTCACGGCAGAAAGAAACAAGTAGTACCACAGGAATGCGTGAGACTGTTGAAACT
201         I  I  I  A  V  V  S  S  R  Q  K  E  T  S  S  T  T  G  M  R  E  T  V  E  T
676        AGCCTGCTTTTGCAACACAGAGCTAAGGAGGTTGTACCAAAACGCATTATAAAAATGGAAGAGGCCATAAAGAAC
226         S  L  L  L  Q  H  R  A  K  E  V  V  P  K  R  I  I  K  M  E  E  A  I  K  N
751        CGTGATTTTGCATCTTTTGCACAATTAACCTGTGCTGATAGTAATCAGTTCCACGCTGTCTGCTTAGATACATCC
251         R  D  F  A  S  F  A  Q  L  T  C  A  D  S  N  Q  F  H  A  V  C  L  D  T  S
826        CCCCCTATTTTCTACATGAATGATACATCCCACAGGATAATAAGCTGCATTGAGAAATGGAATTGCTGTGAGGGA
276         P  P  I  F  Y  M  N  D  T  S  H  R  I  I  S  C  I  E  K  W  N  C  C  E  G
901        ACACCTCAGGTGGCATATACATTTGATGCTGGGCCTAATGCTGTTCTAATTGCACAAAATAGAAAGACTGCTGCC
301         T  P  Q  V  A  Y  T  F  D  A  G  P  N  A  V  L  I  A  Q  N  R  K  T  A  A
976        CAGTTGCTGCAGAAGTTGCTTTTCTTTTTCCCTCCAAATTCTGATACTGATTTAAACAGTTATGTTATTGGTGAT
326         Q  L  L  Q  K  L  L  F  F  F  P  P  N  S  D  T  D  L  N  S  Y  V  I  G  D
1051       AAGTCAATACTAAAAGATGCTGGGATTCAAGAGATAAAGGATGTGGAAGCATTGCCACCACCTCCAGAAATTAAG
351         K  S  I  L  K  D  A  G  I  Q  E  I  K  D  V  E  A  L  P  P  P  P  E  I  K
1126       GATGCCTCAAGATACAAAGGAGATGTTAGTTATTTCATCTGCACAAGACCTGGCAGGGGTCCTGTTTTGCTCTCC
376         D  A  S  R  Y  K  G  D  V  S  Y  F  I  C  T  R  P  G  R  G  P  V  L  L  S
1201       GACGAAAGTCATGCTCTTCTCAATCCCGAAACTGGTCTGCCTAAGTAA
401         D  E  S  H  A  L  L  N  P  E  T  G  L  P  K  *
```

Fig. 4. Nucleotide and amino acid sequences of *JcFPS*.

Table 2. Physicochemical properties of *JcAACT*, *JcMDC* and *JcFPS*.

Protein	***JcAACT***	***JcMDC***	***JcFPS***
Formula	$C_{1877}H_{3066}N_{534}O_{569}S_{16}$	$C_{2023}H_{3216}N_{558}O_{614}S_{20}$	$C_{1793}H_{2784}N_{454}O_{518}S_{14}$
Total number of atoms	6062	6431	5563
Molecular weight (kDa)	42.73	45.82	39.44
Theoretical isoelectric point (pI)	9.08	6.47	5.39
Negatively charged residues	33	51	50
Positively charged residues	41	49	40
Instability index	31.21	38.31	41.60
Aliphatic index	98.33	97.46	84.63
Grand average of hydropathicity	0.134	-0.218	-0.231

Table 3. Protein secondary structure prediction.

Protein	***JcAACT***	***JcMDC***	***JcFPS***
N-myristoylation sites	18	3	3
Protein kinase C phosphorylation sites	2	2	2
N-glycosylation sites	3	2	2
Casein kinase II phosphorylation sites	4	3	7
cAMP- and cGMP-dependent protein kinase phosphorylation sites	2	0	0
Thiolase active sites	1	0	0
Amidation sites	0	1	0
Tyrosine kinase phosphorylation sites	0	1	0

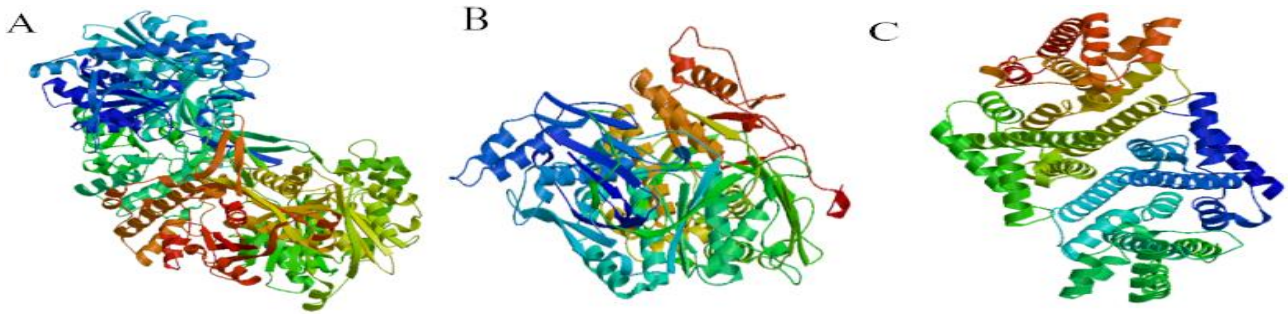

Fig. 5. Tertiary structure of A. *JcAACT*, B. *JcMDC and* C. *JcFPS*.

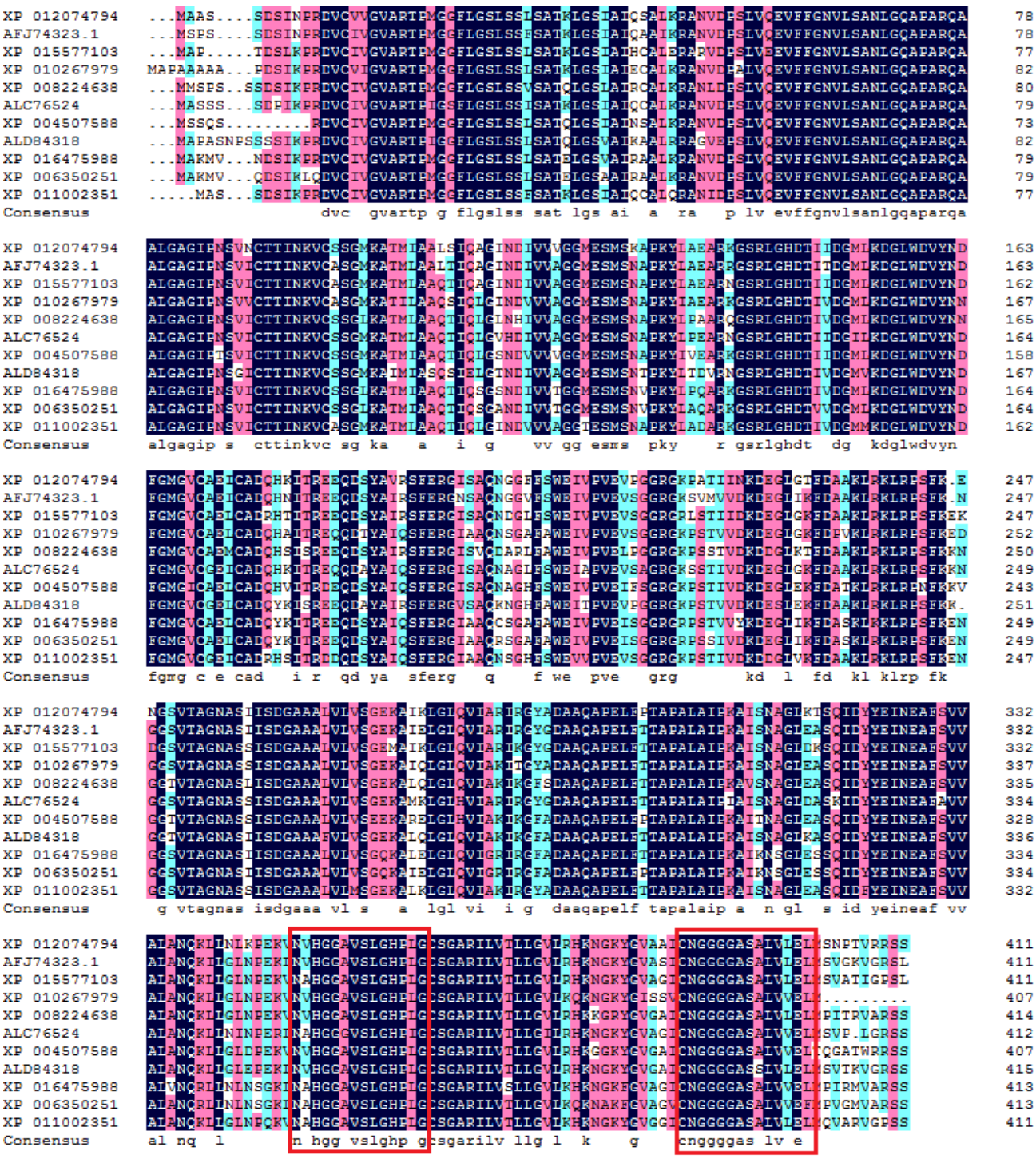

Fig. 6 Multiple alignment of JcAACT protein with the 12 most homologous AACT proteins from other species.
Two thiolase active sites are shown in the two red boxes

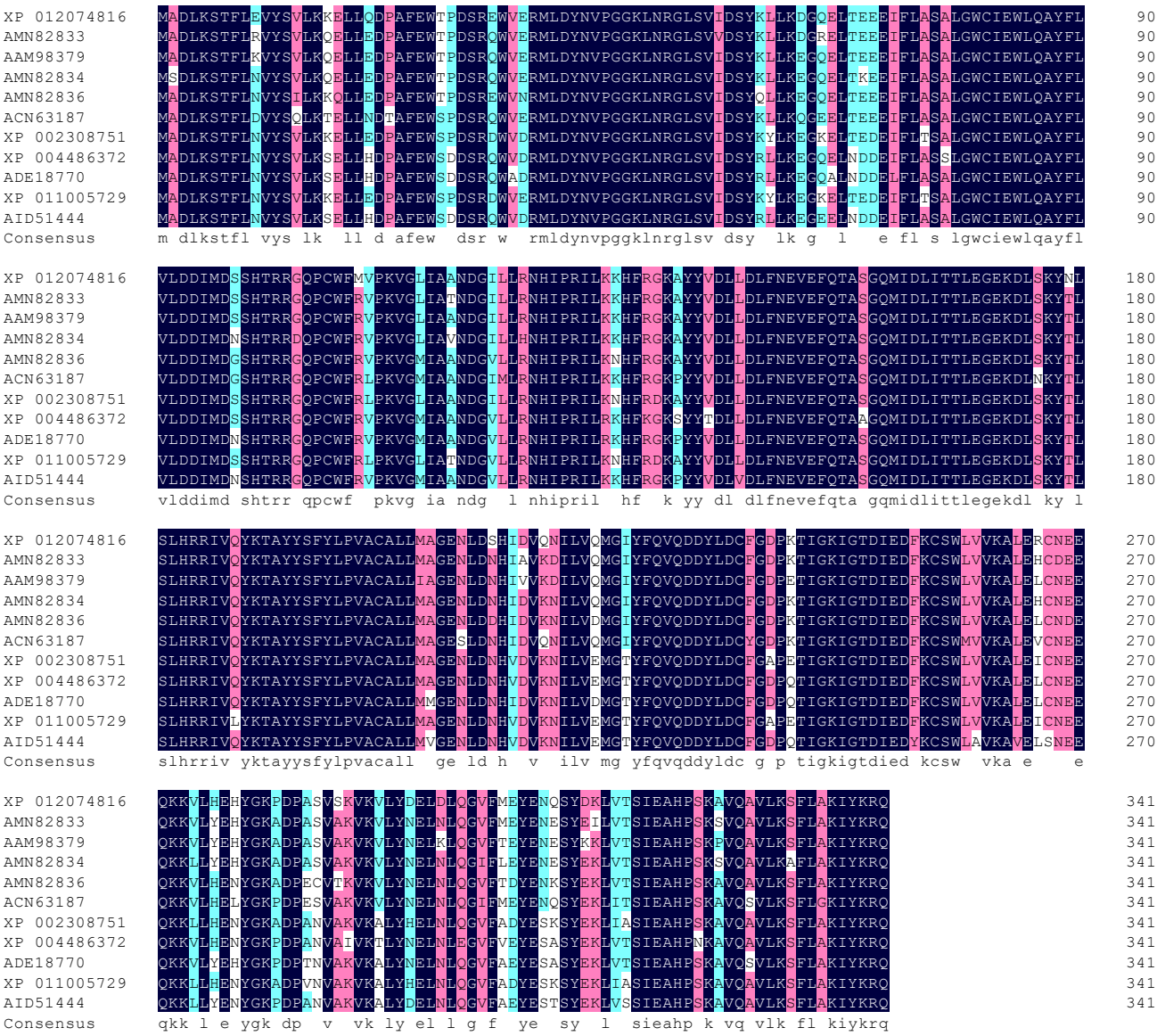

Fig.7. Multiple alignments of JcMDC with the ten most homologous MDC proteins from other species.

Sequence analysis of the deduced *JcFPS* amino acid sequence: The multiple alignment of the *JcFPS* deduced amino acid sequence showed strong similarities to otherFPS from various plant species, including *Hevea brasiliensis* (AAM98379) and *Manihot esculenta* FPS1(AMN82833) with identities of 93%, *Manihot esculenta* FPS2(AMN82834) and *Ricinus communis* (AMN82836) with identities of 92%, *Euphorbia pekinensis* (ACN63187) with an identity of 91%, *Populus trichocarpa* (XP 002308751) with an identity of 89%, *Cicer arietinum* FPS1 (XP 004486372), *Glycyrrhiza uralensis* (ADE18770) and *Populus euphratica* (XP_011005729) with identities of 88%, and *Astragalus membranaceus* (AID51444)with an identity of 87%.The functional domain amino acid composition of JcFPS is consistent with FPS proteins from other plants, all having sites which are necessary for *FPS* activity, including five conserved functional domains (I-V), in which II and V riched of Asp(DDXXD) (Fig. 8).

Phylogenetic analysis: MEGA7.0 software was used to construct the phylogenetic tree of JcAACT, JcMDC and JcFPS. The results indicated that the evolution of *JcAACT*, *JcDMC* and *JcFPS* was consistent with plant taxonomy, and the tree has obvious family characteristics, such as members of the Euphorbiaceae, Salicaceae and Solanaceae being clustered together. The clustering of JcAACT, JcDMC and JcFPS with the corresponding proteins of *H. brasiliensis* demonstrate that the *Jatropha* proteins were most closely related to their counterparts in *H. brasiliensis* (Fig. 9).

Expression analysis: Results of qRT-PCR analysis of *JcAACT, JcMDC* and *JcFPS* showed that each was expressed in roots, stems, leaves and seeds, the highest expression level of all three genes occurring in the seeds (Fig. 10). During the growth and development of the seeds, the trends of changes in expression level with seed development for the three genes were the same, with expression initially increasing, then, after reaching the highest value, decreasing. *JcAACT* and *JcMDC* expression reached the peak value at the late stage (50 d) of seed growth, while *JcFPS* expression reached its peak value at the mid-late stage (40 d) of seed growth (Fig. 11). The expression level of *JcAACT* was the highest of the three genes, regardless of the organ or the seed growth stage.

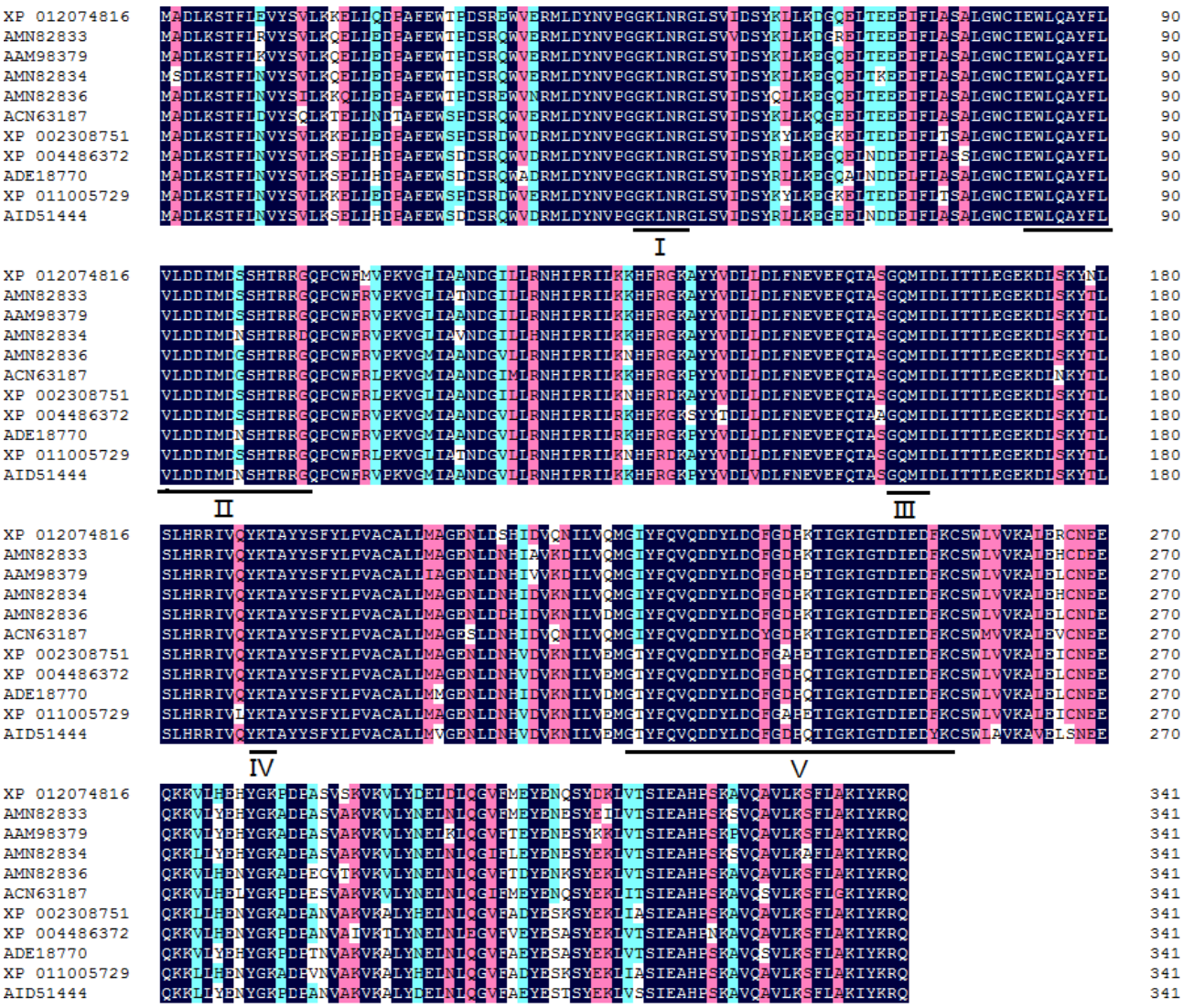

Fig. 8. Multiple alignments of JcFPS with the ten most homologous FPS proteins from other species.
The five conserved domains are numbered and indicated by the upline

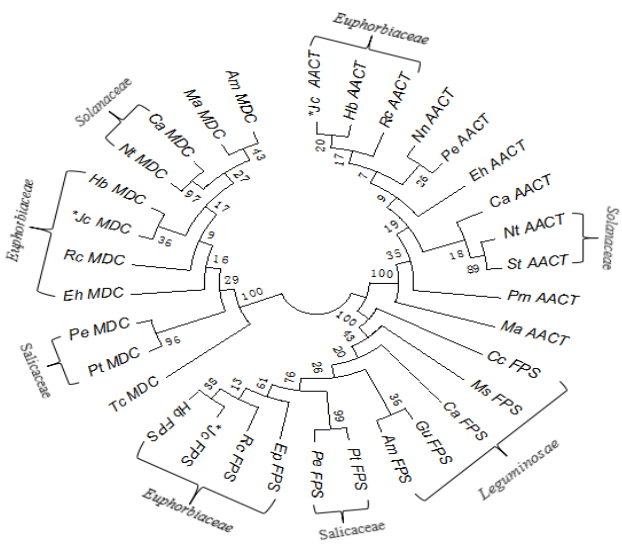

Fig. 9. The phylogenetic tree of amino acid sequences of JcAACT, JcMDC and JcFPS and the corresponding proteins from other plants.
Numbers represent the bootstrap percentage values calculated from 1000 replicates

Discussion

In this study, we cloned three key genes which are involved in the terpenoid biosynthesis pathway of *J. curcas*. The opening reading frames (ORFs) of *JcAACT*, *JcMDC* and *JcFPS*were 1239 bp, 1239 bp and 1029 bp, respectively, encoding 412-amino acid, 415-amino acid and 342-amino acid proteins, respectively. The molecular weights ofJcAACT, JcMDC and *JcFPS* were 42.73, 45.82 and 39.44 kDa, respectively, and the theoretical isoelectric points were 9.08, 6.47 and 5.36, respectively. Cell location prediction revealed that the JcAACT, JcFPS and JcMDC proteins were all most probably located in the cytoplasm, while transmembrane topology prediction showed that JcAACT, JcFPS and JcMDC were not potential membrane proteins. Signal peptide prediction showed that none of these three genes encoded a signal peptide. The results of Homology Analysis showed that the JcAACT, JcMDC and JcFPS proteins all exhibited highest identities and closest relationships with the corresponding proteins from *H. brasiliensis*, with identities of 89%, 92% and 93%, respectively.

JcAACT, *JcMDC* and *JcFPS* were expressed in root, stem, leaf and seed, the highest expression level occurring in seeds. In the early stages of seed growth, the expression level of all three genes increased, with the expression of *JcAACT* and *JcMDC* both peaking at the late stage (50 d), while expression of *JcFPS* reached its highest value at the mid-late stage (40 d); after reaching peak levels, expression of all three genes declined. The time difference in peak expression among the genes was probably because JcAACT and JcMDC belong to the class I enzymes which act before IPP synthesis, whereas JcFPS belongs to the class II enzymes, which act after IPP synthesis (Chen & Zhao, 2004).

The expression level of *JcAACT* was the highest, regardless of the organ concerned or the stage of seed growth at which samples were collected, indicating its important role in the terpenoid biosynthesis pathway; because AACT is the first enzyme of this pathway, it directly affects the rate of production of terpenoid compounds. The expression level of *JcMDC* was lowest, agreeing with the research results of Pang (2005), probably because it is the rate-limiting enzyme of the pathway before IPP synthesis. The similar trend of expression change as the seeds aged in all three genes was basically in accordance with seed growth, with *JcAACT*, *JcMDC* and *JcFPS* being most highly expressed during seed maturation (40-50 d), so it also confirmed that their production of terpenoid compounds has an important role in seed growth.

This study lays the foundation for research into the regulation of terpenoid biosynthesis at the molecular level in *J. curcas*. Because of their high oil content, the seeds are the most valuable part of the plant, producing excellent biodiesel. After oil extraction, the remaining seed cake could also be used as a high-protein animal feed, once removal of toxic components, such as toxic proteins and terpenoids, primarily a tetracyclic diterpenoid-phorbol ester, have been removed. The medicinal value of the seed also should not be ignored; many terpenoids, which have been extracted from seeds, have a medicinal use, while diterpenoid andtriterpenoid compounds also have obvious antitumor activity. This experiment studied three key genes involved in the terpenoid biosynthesis pathway of *J. curcas*, providing a theoretical basis for regulation of terpenoid biosynthesis, and laying the foundation for promoting the development of industries producing biodiesel, biopharmaceuticals and animal feed from *J. curcas*.

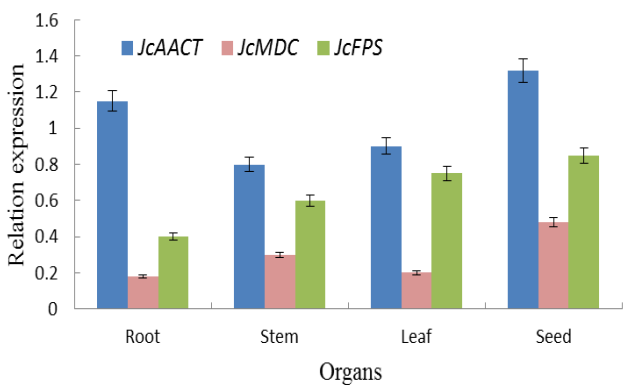

Fig. 10. Expression levels of *JcAACT*, *JcMDC* and *JcFPS* genes in different organs of *J. curcas*.

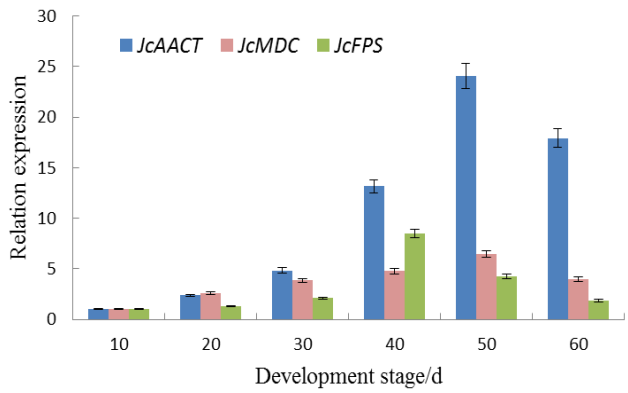

Fig. 11. Expression levels of *JcAACT*, *JcMDC* and *JcFPS* gene in seeds at different development stages of *J. curcas*.

Acknowledgements

This study was supported by National Natural Science Foundation of China (Grant No. 31260064, 31460355, 31460059).

References

Adolf, W., H.J Opferkuch and E. Hecker. 1984. Irritant phorbol derivatives from four *Jatropha* species. *Phytochem.*, 23(1):129-132.

Banyai, W. and C. Kirdmanee. 2010. Overexpression of Farnesyl pyrophosphate synthase (FPPS) gene affected artemisinin content and growth of *Artemisia annua* L. *Plant Cell Tiss. Organ Cult.*, 103(2): 255-265.

Bao, Z.H., L.L. Fu and X.H. Wang. 2011. Developing the Seed Cake of *Jatropha curcas* as Animal Feed is an Important Measure far the Industrailazation of *Jatropha Biodiesel. Net. Prod. Res. Dev.*, 23:288-293. (in Chinese with English abstract)

Byres, E., M.S. Alphey and T.K. Smith. 2007. Crystal structures of *Trypanosoma brucei* and Staphylococcus aureus mevalonate diphosphate decarboxylase inform on the determinants of specificity and reactivity. *J. Mol. Bio.*, 371(2): 540-553.

Chen, J. and D.G. Zhao. 2004. Research advances on the enzymes and their coding gene involved in plant *Terpene. Biosynthesis. Mol. Plant Breeding*, 2(6): 757-764. (in Chinese with English abstract)

Chen, L., X.W. Lan and H. Zhu. 2006. Genetic cloning and sequence analysis of farnesyl pyrophosphate synthase in *Panax notoginseng* [J].*Chin. Tra. and Her. Dru.*, 37(7): 1080-1083. (in Chinese with English abstract)

Closa, M., E. Vranova and C. Bortolotti. 2010. The Arabidopsis thaliana FPP synthase isozymes have overlapping and specific functions in isoprenoid biosynthesis, and complete loss of FPP synthase activity causes early developmental arrest. *Plant J.*, 63: 512-525.

Cordier, H., F. Karst and T. Berges. 1999. Heterologous expression in *Saccharomyces cerevisiae* of an *Arabidopsis thaliana* cDNA encoding mevalonate diphosphate decarboxylase. *Plant Mol. Bio.*, 39(5): 953-967.

Cui, H., H.Q. Liu and X.J. Li. 2006. Expression of foreign farnesyl diphosphate synthase gene in transgenic tobacco enhances disease resistance to *Alternaria alternata In vitro. Acta. Agron. Sin.*, 32(6): 817-820.(in Chinese with English abstract)

Cui, H.G., X.Y. Wang and H. Feng .2010. Molecular cloning and SNP analysis of a acetyl-CoA C-acetyltransferase gene

(*SmAACT*) from *Salvia miltiorrhiza*.*Acta. Pharm. Sin.*, 45(6): 785-790. (in Chinese with English abstract)

Cunillera, N. and D. Delourme. 1996. Arabidopsis thaliana contains 2 differentially expressed farnesyl-diphosphate synthase genes. *J. Biol. Chem.*, 271: 7774-7 780.

Gu, W., Q.N. Wu and J.G. Chao. 2011. Molecular cloning of farnesyl pyrophosphate synthase from *Alisma orientale* (Sam.) Juzep. and its distribution pattern and bioinformatics analysis. *Acta. Pharm. Sin.*, 46(5): 605-612.

Han, J.L, B.Y. Liu and H.C. Ye. 2006. Effects of overexpression of the endogenouse farnesyl diphosphate synthase on the artemisinin content in *Artemisia annua* L. *J. Integr. Plant Biol.*, 48: 482-487.

Haralampidis, K., M. Trojanowska and A. Osboun. 2002. Biosynthesis of triterpenoid saponins in plants. *Adv. Biochem. Eng. Biotech.*, 75: 31.

Jin, H., Z. Sony and B.J. Nikolau. 2012. Reverse genetic characterization of two paralogousacetoacetyl CoA thiolase genes in Arabidopsis reveals their importance in growth and development. *Plant J.*, 70(6): 1015.

Kasahara, H., A. Hanada and T. Kuzuyama. 2002. Contribution of the mevalonate and methylerythritol phosphate pathways to the biosynthesis of gibberellins in Arabidopsis. *J. Biol. Chem.*, 277: 45188-45194.

Kim, O., K. Bang and S. Jung. 2010. Molecular characterization of ginseng farnesyl diphosphate synthase gene and its up-regulation by methyl jasmonate. *Biol. Planta*, 54(1): 47-50.

Kim, Y.K., Y.B. Kim and M.R. Uddin. 2014. Enhanced triterpene accumulation in Panax ginseng hairy roots overexpressing mevalonate-5-pyrophosphate decarboxylase and farnesyl pyrophosphate synthase. *ACS. Syn. Biol.*, 3(10): 773-779.

Laule, O., A. Fürholz and H.S. Chang. 2003. Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. *PNAS*, 100(11): 6866-6871.

Li, C.P. and B.A. Larkins. 1996. Identification of a maize endosperm-specific cDNA encoding farnesyl pyrophosphate synthetase. *Gene*, 171: 193-196.

Liu, C.J., Y.L. Meng and G.S. Hou. 1998. Cloning and sequencing of a cDNA encoding from *Grossypium arbreum* and its expression pattern in developing seeds of *Grossypium hirsutum* cv "Sumian-6". *Acta. Bot. Sin.*, 40(8): 703-710.

Livak, K.J. and T.D. Schmittgen. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, 25: 402-408.

Makkar, H.P.S, K. Becker and F. Sporer. 1997. Studies on nutritive potential and toxic constituents of different provenances of *Jatropha curcas*. *J. Agri. Food Chem.*, 45(8): 3152-3157.

Martinez-Herrera, J., P. Siddhuraju and G. Francis. 2006. Chemical composition, toxic/antimetabolic constituents, and effects of different treatments on their levels, in four Provenances of *Jatropha curcas* from Mexico. *Food Chem.*, 96(1): 80-89.

Pang, Y.Z. 2005. Molecular cloning and characterization of important genes involved in the biosynthetic pathways of flavonoids and terpenoids from *Ginkgo biloba* L. Shanghai: Fudan University.

Quan, Z.G., J.F. Zhao and X.D. Yang. 2009. Research progress on chemical constituents and pharmacological effects of *Jatropha curcas* L. *Yunnan Chem. Tech.*, 36(2): 39-43.

Sanmiya, K., T. Iwasaki and M. Matsuoka. 1997. Cloning of a cDNA that encodes farnesyl diphosphate synthase and the blue-light-induced expression of the corresponding gene in the leaves of rice plants. *Biochim. Biophys. Acta.*, 1350: 240-246.

Shi, L., L. Qin and Y. Xu. 2012. Molecular cloning, characterization and function analysis of a mevalonate pyrophosphate decarboxylase gene from *Ganoderma lucidum*. *Mol. Biol. Rep.*, 39(5): 6149-6159.

Simkin, A.J., G. Guirimand and N. Papon. 2011. Peroxisomal localisation of the final steps of the mevalonic acid pathway in planta. *Planta*, 234(5): 903-914.

Szkopifiska, A. and D. Plochocka. 2005. Farnesyl diphosphate synthase: Regulation of product specificity. *Acta. Biochem. Physiol.*, 52: 45-55.

Xing, F., S. Liang and R. Ang. 2013.The cloning, characterization and functional analysis of a gene encoding an acetyl-CoA acetyltransferase involved in triterpene biosynthesis. *Mycosci*., 54(1): 100-105.

Xing, Z.B., Y.H. Long and S. He. 2012. Cloning and expression analysis of mevalonate diphosphate decarboxylasa gene in *Eleutherococcus senticosus*. *Acta. Bot. Boreal. Occident Sin.*, 32(10): 1950-1956.

Yao, Y.Z., X.Y. Li and L. Wei. 2015. Cloning, expression, and bioinformatics analysis of acetyl-CoA C-acetyltransferase gene in *Houttuynia cordata*. *Chin. Tra. & Her. Dru.*, 46(1): 107-111.

Zeng, J.M. 2006. Biomass energy plant with potential development: *Jatropha curcas* L. *Yunnan For.*, 27(2): 21-22.

Zhang, L., X.F. Tan and J. Hu. 2011. Cloning and sequence characterization of cDNA encoding aeetyl-CoA C-acetyltransferase in *Camellia oleifera*. *J. Cen. Sou. Univ. For & Tec.*, 31(8): 108-112.

Zhang, X.D., C.X. Li and Y.Z. Wang. 2015. Cloning and expression analysis of mevalonate diphosphate decarboxylase gene in *Gentiana rigescens*. *Guizhou Agri. Sci.*, 43(12):164-169.

Zhao, Y.J., M. Zhang and Y.J. Liu. 2015. Cloning and expression analysis of a acetyl-CoA C-acetyltransferase gene (*TwAACT*) from *Tripterygium wilfordii*.*Chin. Tra. & Her. Dru.*, 40(5): 847-851.

(Received for publication 28 December 2016)