Plant Gene 9 (2017) 34–44

Contents lists available at ScienceDirect

Plant Gene

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

Isolation and characterization of curcin genes with distinct expression patterns in leaves and seeds of *Jatropha curcas* L.

Lifang Wu [a,1], Mei Ling Goh [a], Dongsheng Tian [a], Keyu Gu [a], Yan Hong [a,b,c], Zhongchao Yin [a,d,*]

[a] Temasek Life Sciences Laboratory, National University of Singapore, 1 Research Link, Singapore 117604, Republic of Singapore
[b] JOil (S) Pte Ltd, 1 Research Link, Singapore 117604, Republic of Singapore
[c] School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, Republic of Singapore
[d] Department of Biological Sciences, National University of Singapore, 14 Science Drive, Singapore 117543, Republic of Singapore

ARTICLE INFO

Article history:
Received 7 September 2016
Received in revised form 25 November 2016
Accepted 19 December 2016
Available online 21 December 2016

Keywords:
Curcin
Jatropha curcas L.
Ribosome inactivating protein
Tissue-specific gene expression
Endosperm-specific promoter

ABSTRACT

Jatropha (*Jatropha curcas* L.) is an emerging biofuel plant, of which its seed oil is suitable for biodiesel production. Curcin, a type 1 ribosome inactivating protein, is the major toxic protein found in jatropha seeds. Here we report the cloning and characterization of three curcin genes from *J. curcas* MD44. The *Curcin 1* (*C1*) gene encodes a Type I curcin precursor containing 293 amino acid residues. The *C1* gene is specifically expressed in the endosperm of jatropha seeds at late developmental stage. The *Curcin 2A* (*C2A*) gene encodes a Type II curcin precursor containing 309 amino acid residues and is mainly expressed in young leaves of *J. curcas*. The *Curcin 2B* (*C2B*) gene is another Type II curcin gene that is physically linked to the *C1* gene. However, the expression of the *C2B* gene was not detected in leaves or seeds of jatropha plants growing under normal growth condition. A conserved 227-bp or 228-bp intron was identified or predicted in the 5′ untranslated region of the three curcin genes. A *C1* promoter (−2888 bp to +293 bp) was identified to harbor all *cis*-elements that are required for full promoter activity in jatropha endosperm. The isolation of the three curcin genes and the endosperm-specific *C1* promoter provide useful information and research materials for further functional study of curcin proteins and genetic engineering of *J. curcas*.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Jatropha (*Jatropha curcas* L.) is a tropical shrub yielding high quality non-edible oil. In recent years, jatropha has become a potential oilseed crop for the production of renewable bioenergy (Maghuly and Laimer, 2013). The kernel of its seed has high crude protein (22–28%) and oil (54–58%) contents (Devappa et al., 2010). Production of jatropha oil could generate similar amount of seed cake. However, jatropha seeds contain toxic and anti-nutritive compounds, such as phorbol esters and curcins (Sabandar et al., 2013). The presence of these compounds in jatropha has resulted in seed cake being unsuitable to be used as animal feed and hence raised safety and environment concerns on jatropha plantation and processing (He et al., 2011; Pradhan et al., 2012).

Ribosome-inactivating proteins (RIPs) are toxic N-glycosidases that depurinate the universally conserved α-sarcin loop of large rRNAs and inactivate the ribosome by blocking its further participation in protein synthesis (Nielsen and Boston, 2001; Sikriwal and Batra, 2010). Based on their overall structure, RIPs are classified into two major groups. Enzymes that consist exclusively of a single polynucleotide adenine glycosylase (PAG) domain are referred to as type 1 RIPs whereas type 2 RIPs are chimeric proteins with PAG domain linked to a C-terminal lectin domain. Curcin is a type 1 RIP that is commonly present in the members of Euphorbiaceae family. Lin et al. (2003a) isolated the first curcin gene that encodes a seed-specific curcin precursor. Previous genome sequencing analysis suggested that *J. curcas* has three curcin genes and two additional curcin-like genes (Sato et al., 2011). However, the sequences of the DNA contigs that contain the curcin or curcin-like genes are not available in public domains or on the website provided (http://www.kazusa.or.jp/jatropha/). Recent gene annotation work identified one Type I curcin (XP012074335), two Type II curcins (XP012074346 and XP012074358) and three curcin-like proteins (XP012074106, XP012074368 and XP012074421) in *J. curcas* with their genomic sequences available in GenBank (Zhang et al., 2014). So

Abbreviation: BAC, bacterial artificial chromosome; *C1*, *Curcin 1*; *C2A*, *Curcin 2A*; *C2B*, *Curcin 2B*; ERE, ethylene responsive element; GUS, β-glucuronidase; *JcActin1*, jatropha actin gene 1; kDa, kilo-Dalton; PAG, polynucleotide adenine glycosylase; qRT-PCR, quantitative reverse transcription polymerase chain reaction; RACE, rapid amplification of cDNA ends; RIP, ribosome-inactivating protein; SD, standard deviation; SNP, single nucleotide polymorphism; TAIL-PCR, thermal asymmetric interlaced polymerase chain reaction; *Ubi1*, maize ubiquitin gene 1; X-Gluc, 5-bromo-4-chloro-3-indolyl glucuronide; 5′UTR, 5′ untranslated region; 3′UTR, 3′ untranslated region.

* Corresponding author at: Temasek Life Sciences Laboratory, National University of Singapore, 1 Research Link, Singapore 117604, Republic of Singapore.
E-mail addresses: lfwu@ipp.ac.cn (L. Wu), bleumoon_mei@yahoo.com (M.L. Goh), tiands@tll.org.sg (D. Tian), keyu@tll.org.sg (K. Gu), yhong@ntu.edu.sg (Y. Hong), yinzc@tll.org.sg (Z. Yin).
[1] Present address: Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, Anhui, China.

http://dx.doi.org/10.1016/j.plgene.2016.12.005
2352-4073/© 2016 Elsevier B.V. All rights reserved.

far, more than ten curcin genes and several curcin-like genes have been identified from different jatropha accessions and the predicted amino acid sequences of curcin precursors are available in GenBank. These curcin precursors share at least 86% identity at amino acid level and are classified into two types. Type I curcin precursors have 293 amino acid residues, whereas Type II curcin precursors contain 309 amino acid residues. Each curcin precursor carries a conserved signal peptide containing 42 amino acid residues, which is removed during curcin maturation (Lin et al., 2003a). The mature Type I and Type II curcins are 28 kilo-Dalton (kDa) and 30 kDa, respectively (Lin et al., 2003a; Wei et al., 2005; Qin et al., 2010). Type I curcins were only identified in jatropha seeds (Lin et al., 2010; He et al., 2011; King et al., 2011), whereas Type II curcins were mainly found in jatropha leaves and induced by biotic and abiotic stress (Wei et al., 2005; Qin et al., 2010). The biochemical function of curcins in jatropha is not known. Type I curcins in seeds may serve as storage proteins to provide nutrition during seed germination. The toxic proteins may also prevent animals from consuming jatropha seeds (Stirpe, 2004). Type II curcins in leaves may play a role in defense against biotic and abiotic stress (Wei et al., 2005; Huang et al., 2008; Qin et al., 2010). Besides, curcins were found to demonstrate antitumor activity, and hence they have promising potential in cancer therapy (Lin et al., 2003b; Luo et al., 2006; Mohamed et al., 2014a; Mohamed et al., 2014b; Jaramillo-Quintero et al., 2015). In addition, the promoters of Type I and Type II curcin genes have been isolated by PCR and characterized in transgenic tobacco (Qin et al., 2009a; Qin et al., 2009b). Here we report the isolation of one Type I curcin gene, *Curcin 1* (*C1*), and two Type II curcin genes, *Curcin 2A* (*C2A*) and *Curcin 2B* (*C2B*), from *J. curcas* and the characterization of their distinct expression patterns in leaves and seeds of *Jatropha curcas* L.

2. Material and methods

2.1. Plant materials and growth condition

J. curcas MD44 was used for the experiments. Plants were grown in experimental field in Singapore with temperatures ranging from 32 °C during the day to 26 °C at night with 85% relative humidity and photoperiod of 12 to 13 h.

2.2. Constructs

To make a construct that carries *GFP* reporter gene for promoter analysis, a synthetic *GFP* gene was amplified from pSSZ41 (Kolesnik et al., 2004) and cloned into binary vector pC1300 to generate pCGFP. A 216-bp *C1* terminator was cloned into pCGFP at sites between *Kpn*I and *Eco*RI to generate pCGFPT1. *C1* promoters of different length were amplified by PCR and cloned into pCGFPT1 at the sites between *Hind*III and *Pst*I to generate transcriptional fusion constructs. *C1* promoters were also used to replace the CaMV 35S promoter at the upstream of the intron-containing β-glucuronidase gene (*GUSPlus*) in CAMBIA vector pC1305.1 to create promoter-*GUSPlus* fusion genes.

2.3. Rapid amplification of cDNA ends (RACE) and thermal asymmetric interlaced polymerase chain reaction (TAIL-PCR)

5′ RACE and 3′ RACE were carried out according to the manufacturer's instruction (Invitrogen, Carlsbad, CA, USA). The primers for 5′ RACE of *C1* were primer pair GSP1 and AAP, and primer pair GSP1A and AUAP (Table 1). The primers for 3′ RACE of *C1* were primer pair C2RTF2 and AUAP, and primer pair 3′RACEF4 and AUAP (Table 1). The primers for 5′ RACE of *C2A* were primer pair C2RTR2 and AAP, and primer pair 5′RACEG4 and AUAP (Table 1). The primers for 3′ RACE of *C2A* were primer pair CTRTF2 and AUAP, and primer pair 3′ RACEF1 and AUAP (Table 1). TAIL-PCR was performed according to the procedures published previously (Liu and Whittier, 1995). The nested primers were C2AFAR1, C2AFAR2 and C2AFAR3 (Table 1). The arbitrary degenerate primer was AD2 (Table 1).

Table 1
DNA oligo primers used in this study.

Primer	Nucleotide sequence (5′ to 3′)
GSP1	GTCTGCTGGCTTTGAACTTT
GSP1A	CACCCCTAAATCCACATCCT
AAP	GGCCACGCGTCGACTAGTACGGGGGGGGGGGGGGGG
AUAP	GGCCACGCGTCGACTAGTAC
C2RTF2	GGGCATCGGCTAGGGAAATA
3′RACEF4	TCTCATCAAACCAAAACTACAAA
C2RTR2	AGAGGTCTCCCCAGTTGTTC
5′RACEG4	TGAATCTTGCTGCCTCTG
CTRTF2	GGGCATCGGCTAGGGAAATA
3′RACEF1	GAGAGGATGTGGATTTAGG
C2AFAR1	CTAAACTCAAACTCAAATCATC
C2AFAR2	CTAACGCTCATCTTAAGCTTAGTA
C2AFAR3	CCATAACCAATGTATGATTTGGTA
AD2	NGTCGASWGANAWGAA
Cantigen F1	GCTGGTTCCACTCCAACTTTAAC
Cantigen R1	TCAGACTTTGTATTTGACTGCATTC
Cantigen R2	CCCAAGCTTAATTATCTAATGCCTGCACCCC
CF2	GATATTTGTGTTTCTTCAT
CR2	TTTGTTGTCCTTTATTTATGCT
C2PF4	CTTCAAGACACTAGTTCAAAAA
C1CDSR	TGATGAGAATGGACAAACTAT
C2ACDSF	CCCAAGTAAAAGGTCTCA
C2TR2EcoRI	CGGAATTCGTATTATTTGGATGGTAGAAAATT
C2AP3PmlF	CGGCACGTGAAAGTTAGAGTTAGGGTATAA
C2APstR	AACTGCAGATTGATTTCACCTGTCCAGTTGTA
CF	TTTACTTCCCCGTTTGCTCA
C1SF	TGCAGTCAATTACAAAGTCTG
C2ASF	AGTAAAAGGTCTCATGGGAGTC
C1SR	TATAGCATTAGCAACAATAATAAT
C2ASR	CATTAGCAACATGTTTGGACAAC
JcActin1-F	TAATGGTCCCTCTGGATGTG
JcActin1-R	AGAAAAGAAAAGAAAAAAGCAGC

2.4. Quantitative reverse transcription polymerase chain reaction (qRT-PCR)

qRT-PCR was conducted in accordance with the procedures described previously (Gu et al., 2011). Total RNA extracted from endosperm of developing seeds, leaves at different developmental stages or leaves infected with aphids (*Aphis craccivora*) were used for DNase I treatment and 1st strand cDNA synthesis according to the manufacturer's instructions (Bio-Rad). A standard reaction mixture (15 μl) consisted of 2 μl 1st strand cDNA, 1× SsoFast EvaGreen supermix (Bio-Rad) and 500 nM forward and reverse primers. The PCR reaction consisted of 95 °C for 30 s, followed by 40 cycles of 95 °C for 5 s, 60 °C for 10 s. A melting-curve reaction was subsequently performed for 10 s, starting at 65 °C with 0.5 °C increments. The expression of jatropha actin gene 1 (*JcActin1*) was used as the internal control. The average expression levels of *C1* in 6-week-old jatropha endosperm and *C2A* in young leaves were set as "1", respectively. The qRT-PCR experiments were performed in triplicate, and the data are presented as means ± standard deviation (SD). The specific primer pairs for the *C1* and *C2A* genes were C1SF/C1SR and C2ASF/C2ASR, respectively (Table 1). The specific primer pair for the *JcActin1* gene was JcActin1-F/JcActin1-R (Table 1).

2.5. Southern blot analysis

Southern blot analysis was carried out according to the standard procedures (Sambrook and Russell, 2001). Plant genomic DNA was isolated from leaves as described previously (Dellaporta et al., 1983). About 2–5 μg of DNA was digested with restriction enzymes by *Bam*HI, *Eco*RI, *Hind*III or *Xba*I, separated on 0.8% agarose gel and then blotted onto Hybond™-N + nylon membrane (Amersham Biosciences,

Little Chalfont, Buchinghamshire, UK). The DNA probe for Southern blot analysis of curcin genes in jatropha genome was a 756-bp fragment amplified from the 3′ coding region of the *C1* gene using primers Cantigen F1 and Cantigen R1 (Table 1). DNA probes were labelled with [^{32}P]-dCTP by Rediprime II random prime labelling system (Amersham Biosciences, Piscataway, NJ, USA).

2.6. Generation of anti-curcin antibody

Polyclonal antibodies were raised against N-terminal region of C1 (C1N, 44–184 amino acid residues). The cDNA fragment that encodes C1N was amplified from a *C1* cDNA clone using primers Cantigen F1 and Cantigen R2 (Table 1). The PCR products were cloned into vector pQE30 (Qiagen) to generate pQCF1R2. The construct was then introduced into *E. coli* BL21 strain for antigen expression. The antigen proteins were extracted and purified using a polyhistidine-tagged protein purification kit (MACHEREY-NAGEL, Duren, Germany). The anti-C1N antibodies were raised using standard procedure for polyclonal antibody production in rabbits.

2.7. Western blot analysis

Western blot analysis was performed with 20 μg total proteins from different tissues of *J. curcas* MD44. Proteins were separated through sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE, 12% or 15%) followed by blotting onto nitrocellulose membrane. Curcin proteins were detected with anti-C1N antibodies and a horseradish peroxidase-coupled secondary antibody (Bio-Rad, Hercules, CA 94547, USA) according to the product manual.

2.8. Particle bombardment

One micro-meter gold particles were coated with plasmid DNA according to the manufacturer's instructions. Endosperm from six-week-old immature seeds or young leaves of *J. curcas* were manually cut into thin slices of about 0.5-mm thickness or leaf disks of 2 × 2 cm^2. For each bombardment, five to six slices of endosperm or one piece of leaf disk were bombarded at 1100 psi with a flight distance of 15 cm using a PDS-1000/He system (Bio-Rad). The bombarded samples were cultured on MS_0 medium (Murashige and Skoog, 1962) for 1–2 days at 22–25 °C. The maize ubiquitin gene 1 (*Ubi1*) promoter-driving GFP gene in pSSZ41 (Kolesnik et al., 2004) or 35S-driving intron-containing *GUSPlus* gene in pC1305.1 were used as positive controls, respectively. The empty vector pC1300 was used as the negative control.

2.9. Fluorescence and confocal microscopy

The transient expression of GFP in bombarded endosperm tissues was examined in fluorescence and confocal microscopy. Briefly, endosperm slices at 1–2 days after bombardment were screened with a fluorescence stereomicroscope (Model SZX12, Olympus, Japan) for GFP expression and then imaged on LSM510 META inverted confocal microscope (Zeiss, Jena, Germany) with the excitation wavelength at 488 nm and emission wavelength at 505–530 nm.

2.10. Histochemical assay of β-glucuronidase (GUS) activity

GUS activity was detected by X-Gluc (5-bromo-4-chloro-3-indolyl glucuronide) staining according to the method as described previously with slight modification (Hiei et al., 1994). Leaf disks at 2 days after bombardment were incubated in phosphate buffer (50 mM $NaPO_4$, pH 6.8) that contained 1% Triton X-100 at 37 °C for 1 h. The buffer was removed and fresh phosphate buffer containing 1.0 mM X-Gluc and 20% methanol was added to the leaf disks. The reaction mixture was placed under a mild vacuum for 5 min and incubated overnight at 37 °C. Samples were washed twice in 99% ethanol for 2 h each time before microscopy examination.

3. Results

3.1. Isolation of cDNA clones of C1 and C2A

To isolate the curcin gene that is expressed in jatropha seeds, we constructed and sequenced a normalized cDNA library derived from developing endosperm of *J. curcas* MD44 (Gu et al., 2011). A consensus cDNA was obtained from four independent clones identified from the cDNA library. The open reading frame of the cDNA encodes a Type I curcin precursor containing 293 amino acid residues (accession no.: ADN39429), which was designated as Curcin 1 (C1). C1 shows 99% identity to another Type I curcin precursor (accession no.: AAL58089), which is specifically expressed in jatropha seeds (Lin et al., 2003a). The 5′ and 3′ ends of the cDNA of the *C1* gene were isolated by 5′ or 3′ RACE. The full-length *C1* cDNA consists of 1161 bp, including a 66-bp 5′ untranslated region (5′UTR), an 882-bp open reading frame and a 213-bp 3′UTR.

To isolate the curcin gene that is expressed in jatropha leaves, we conducted RT-PCR and 5′ and 3′ RACE using primers derived from the *C1* cDNA. An 1172-bp consensus cDNA was obtained from leaf tissues of *J. curcas*. The 1172-bp cDNA was derived from a Type II curcin gene, which was designated as *Curcin 2A* (*C2A*) in this study. The *C2A* gene encodes a Type II curcin precursor containing 309 amino acid residues (accession no.: ADN39428), which is identical to the Type II curcin precursor ABZ04128 in GenBank.

3.2. Isolation of genomic clones of C1, C2A and C2B

Using a DNA probe derived from the *C1* gene, two major hybridization bands were detected in Southern blot analysis when the jatropha genomic DNA was digested with restriction enzymes *Bam*HI, *Eco*RI or *Hind*III (Fig. 1A). Among which, two curcin gene-containing *Hind*III fragments had the molecular size at about 4.0 kb and 5.4 kb, respectively (Fig. 1A). Previously, a bacterial artificial chromosome (BAC) library with 7 times coverage of jatropha genome was constructed and was used for isolation of genes from *J. curcas* (Gu et al., 2011). Four BAC clones (121E10, 127F22, 147I10 and 176I18) were identified from the BAC library using the *C1* probe. These BAC clones shared some common DNA bands when they were digested by *Hind*III and separated on a 0.8% agarose gel, indicating that the DNA inserts in the four BAC clones were overlapped (Fig. 1B). In addition, they all harbored the curcin gene-containing 4.0-kb and 5.4-kb *Hind*III fragments, which were detected by the *C1* probe in DNA hybridization to the DNA fingerprinting blot (Fig. 1B). BAC clone 121E10 was sequenced using shotgun DNA sequencing

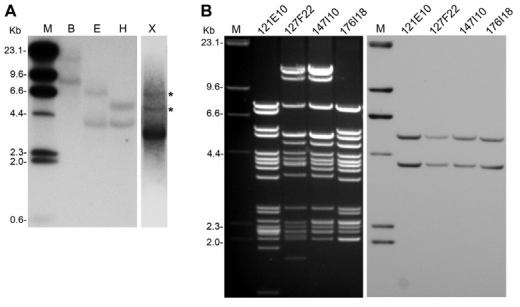

Fig. 1. Detection of curcin genes in jatropha genome. A) Southern blot analysis of genomic DNA of *J. curcas* using the *C1* probe. The two bands in the last lane marked with asterisks may contain the two additional curcin or curcin-like genes in the jatropha genome, which were not isolated in this study. B) Identification of BAC clones that contain curcin genes. Four BAC clones, which harbor *C1* and *C2B* genes, were digested with *Hind*III and separated on a 0.8% agarose gel (left panel), and Southern blot was hybridized with the *C1* probe (right panel). M, λ DNA digested with *Hind*III; B, *Bam*HI; E, *Eco*RI; H, *Hind*III; X, *Xba*I.

approach. The BAC insert in 121E10 is 65,841 bp (accession no.: GQ925454). One of the two curcin genes in 121E10 was identified to be the *C1* gene, while the other curcin gene encodes a Type II curcin precursor containing 309 amino acid residues, which was designated as *Curcin 2B* (*C2B*) gene in this study. *C1* and *C2B* genes are closely linked and arranged in tandem manner with the *C1* gene locating at the 3′ end of *C2B*. The physical distance of the open reading frames of the two genes is 8131 bp. By comparing the *C1* cDNA with its genomic sequence, a previously unidentified 227-bp intron was found in the genomic sequence of the 5′ UTR of the *C1* gene (Fig. 2).

The genomic sequence of *C2A* was isolated from *J. curcas* using PCR and TAIL-PCR. Firstly, a 1018-bp genomic fragment that comprises *C2A* coding region was amplified by PCR using primer pair CF2 and CR2 (Table 1). Secondly, a 1216-bp fragment, including a 1105-bp 5′ regulation region at the upstream of the *C2A* coding region, was obtained by PCR using primer C1CDSR from *C2A* and degenerated primer C2PF4 derived from *C2B* promoter (Table 1). Thirdly, another 979-bp fragment at further upstream of the 1105-bp 5′ regulation region of *C2A* was isolated by TAIL-PCR. Finally, an 838-bp fragment, including a 734-bp 3′ regulation region at downstream of the *C2A* coding region, was obtained by PCR using primer C2ACDSF from *C2A* coding region and degenerated primer C2TR2EcoRI derived from the putative 3′ UTR of the *C2B* gene (Table 1). In addition, the contiguous 2084-bp *C2A* promoter was further cloned by PCR amplification using primers C2AP3PmlF and C2APstR (Table 1), followed by DNA sequencing. In total, we have obtained a 3748-bp genomic clone of the *C2A* gene, which includes a 2084-bp *C2A* promoter, a 930-bp coding region and a 734-bp 3′ UTR. Similar to the *C1* gene, a 228-bp intron was identified in the 5′ UTR of the *C2A* gene after comparing the *C2A* cDNA to its genomic sequence (Fig. 3).

The attempt to isolate *C2B* cDNA from endosperm or leaf tissues of *J. curcas* was not successful, indicating that the *C2B* gene might not be expressed in normal growth condition. Nevertheless, after comparison of genomic sequences among *C1*, *C2A* and *C2B*, a similar gene organization could be predicted for the *C2B* gene, including a putative 227-bp intron in the predicted 5′ UTR and a 930-bp open reading frame encoding a putative Curcin 2B precursor (C2B) (accession no.: AJW31110) (Fig. 4).

Compared to other curcin genes whose genomic sequences are available in GenBank, the *C1*, *C2A* and *C2B* genes show 99% identity in promoter, coding and terminator regions to the genes of XP012074335, XP012074358 and XP012074346, respectively. Interestingly, compared to conserved coding regions among curcin genes, the Type I and Type II curcin genes show great polymorphism in the promoter regions. Only short homologous regions with high identity

```
-180 TGGATTTTCAATAATTCTTTTTCATTGGGATATGTTGTTGTTTGTCTTTATTTTACATCT
-120 TATGAAACTGTCGTTTGGTTGCTAATAATAATAATAATAATAATAAAGAAAAAAAATTGA
 -60 CAAAATAAAGGGCCGGGAAGGCAGTTTCCCTATAAAAGCAGGTGATGGGGGAAGGCAAAA
   1 GACCATCTCTTGCTCTCTTCTTCTTTACTTCCCCGTTTGCTCAGTTGCTTTCTTTgtaag
  61 taatatagaagcctctgcccttctttgttgttgacaaattccatttttttgttttactatt
 121 agcatgttaatttctagcttctggaaatgagtttattatccttttatatgataaacttgtg
 181 accattctatctctttttaattatttttataatttttatgcaattctattaaaataatcgt
 241 attcgtataatgatatttgtgtttcttcatacaactggacagGTGAAATCAATATGAAAG
                                                       M  K  G
 301 GTGGAAAGATGAATCTCTCCATTATGGTGGCTGCATGGTTTTGCTGGAGTAGTATTATAT
       G  K  M  N  L  S  I  M  V  A  A  W  F  C  W  S  S  I  I  F
 361 TCGGATGGGCATCGGCTAGGGAAATAGTTTGTCCATTCTCATCAAACCAAAACTACAAAG
       G  W  A  S  A  R  E  I  V  C  P  F  S  S  N  Q  N  Y  K  A
 421 CTGGTTCCCCTCCAACTTTAACCATTACTTATGACGCTACTACTGATAAGAAAAACTACG
       G  S  P  P  T  L  T  I  T  Y  D  A  T  T  D  K  K  N  Y  A
 481 CCCAGTTCATTAAAGATCTAAGAGAAGCATTTGGCTTCAGTTATTCAAGCCATGAAATAC
       Q  F  I  K  D  L  R  E  A  F  G  F  S  Y  S  S  H  E  I  P
 541 CAGTCTTACGGGCCACAGTTGCTCCAAATCAGAAATTTATTGTAGCCAAAGTCATAAATG
       V  L  R  A  T  V  A  P  N  Q  K  F  I  V  A  K  V  I  N  V
 601 TAGCGAATTTAGAAGTATCATTAGGATTAAACGTCGTTAATGCGTATTTAGTGGGTTATA
       A  N  L  E  V  S  L  G  L  N  V  V  N  A  Y  L  V  G  Y  K
 661 AGGTAGGAGGTACTTCCTATTTCTTTAACGATCCGGAATCTTTGGCTGATGCAAAAACAT
       V  G  G  T  S  Y  F  F  N  D  P  E  S  L  A  D  A  K  T  Y
 721 ATCTTTTCACAGACACAAAGCAACAAACGCTATCATTTACTGGTAGCTATGCAGATTTTC
       L  F  T  D  T  K  Q  Q  T  L  S  F  T  G  S  Y  A  D  F  L
 781 TATCTAGGGCAAACGTACACAGAGAGGATGTGGATTTAGGGGTGCAGGCATTAGATAATT
       S  R  A  N  V  H  R  E  D  V  D  L  G  V  Q  A  L  D  N  Y
 841 ACATATATACACTTGAAAAAAGTTCAAAGCCAGCAGACATTGCTAAACCTCTAGTTGGTT
       I  Y  T  L  E  K  S  S  K  P  A  D  I  A  K  P  L  V  G  F
 901 TTATCGAAATGGTTCCAGAGGCAGCAAGATTCAAATATATTGAGAAAAAAGTATTAAGTC
       I  E  M  V  P  E  A  A  R  F  K  Y  I  E  K  K  V  L  S  Q
 961 AAATTAGCAAAACCTTTAGGCCGGGTGGTGACATAATTAGCCTTGAGAACAACTGGGGAG
       I  S  K  T  F  R  P  G  G  D  I  I  S  L  E  N  N  W  G  D
1021 ACCTCTCTTATCAAATACAGAAATGTGTAAATGGTGTATTTCTGAAGCCAGTTCAATTAC
       L  S  Y  Q  I  Q  K  C  V  N  G  V  F  L  K  P  V  Q  L  Q
1081 AACGTGAAAACTATACCAATATCCTAGTGAACAATGTCACCCAAGTAGCAGGTGTCATGG
       R  E  N  Y  T  N  I  L  V  N  N  V  T  Q  V  A  G  V  M  G
1141 GAGTCTTGTTGAATGCAGTCAATTACAAAGTCTGAATGGAAGAAATTATTTTCAACTACC
       V  L  L  N  A  V  N  Y  K  V  *
1201 AAAAGTGGCTGCCATGGCTTTAATCCTACTTTTGCTCTATATATAGAGTAGCATAAATAA
1261 AGGACAACAAATTTATTATTATTGTTGCTAATGCTATATGCTATTTCCCTGTAATATCCT
1321 CATCTTTCCAATGTATGAATATGATGATGAATTATATATGACAAATAAAGTTTCTACTAG
1381 TTCTTAAT
```

Fig. 2. Nucleotide sequence of the *C1* gene and the deduced amino acid sequence of the C1 protein. The transcription initiation site of the *C1* gene is marked as the position +1. The nucleotides of the intron in the 5′UTR of the *C1* gene are shown in lower-case letters. The two *Sau3A*I sites (GATC) in the coding region of the *C1* gene and the polyadenylation signal (AATAAA) are shown in bold letters. The annotation of the *C1* gene was made based on the experimental data obtained in this study.

```
-180 TTCTTTTTTTTTTTTTTTTTGACAGTTGTTATTGTTTGTCTTTATTTTTACATCTTATGA
-120 AACTGTCGTTTGGTAGCTAATAATCAAAATAATAATAAAAAATAATAATAATAAAGAAAA
 -60 AAAAATTGACAAAATAAAGGGCAGTTTCCCTATAAAAGCAGGTGATGGGGGAAGGCAAAA
   1 GACCATCTCTCGCTCTCTTCTTCTTTACTTCCCCGTTTGCTCAGTTGCTTTCTTTgtaag
  61 taatattgaagcctctgcccttctttttttgttgacaaattccatttttttgttttactaa
 121 tagcatgttaatttctagcttctggaaatgagtttattatactttatatgataaacttgt
 181 gaccattctatctcttttaatcatttttataattttatgcaaatctattataataatcg
 241 tattcgtataatgatatttgtgtttcttcatacaactggacagGTGAAATCAATATGAAA
                                                           M  K
 301 GGTGGAAAGATGAACCTCTCCATTATGGTGGCTGCCTGGTTTTGCTGGAGTAGTATTATA
     G  G  K  M  N  L  S  I  M  V  A  A  W  F  C  W  S  S  I  I
 361 TTCGGATGGGCATCGGCTAGGGAAATAGTTTGTCCATTCTCATCAAACCAAAACTACAAA
     F  G  W  A  S  A  R  E  I  V  C  P  F  S  S  N  Q  N  Y  K
 421 GCTGGTTCCACTCCAACTTTAGCCATTACTTATGACGCTACTACTGATAAGAAAAACTAC
     A  G  S  T  P  T  L  A  I  T  Y  D  A  T  T  D  K  K  N  Y
 481 GCCCAGTTCATTGAAGATCTAAGAGAAGCATTTGACTTCAGTTATTTAAGCCATAAAATA
     A  Q  F  I  E  D  L  R  E  A  F  D  F  S  Y  L  S  H  K  I
 541 CCAGTCTTACGGGCCACGGTTGCTGCAAATCAGAAATTTATTGTAGCCAAAGTCATAAAT
     P  V  L  R  A  T  V  A  A  N  Q  K  F  I  V  A  K  V  I  N
 601 TCTGGGGACATAGAAGTATCAGTAGGATTAAACGTCATTAATGCATATCTAGTGGCTTAT
     S  G  D  I  E  V  S  V  G  L  N  V  I  N  A  Y  L  V  A  Y
 661 AAGGTAGGAAGTAATTCCTATTTCTTTAACGATTCGGAATCTTTGGCTGATGCAAAAAAA
     K  V  G  S  N  S  Y  F  F  N  D  S  E  S  L  A  D  A  K  K
 721 AATCTTTTCACAGACACAAACCAACAAACACTAGCATTTACTGGTAGCTATGCAGATTTT
     N  L  F  T  D  T  N  Q  Q  T  L  A  F  T  G  S  Y  A  D  F
 781 GAATCTAGGGCAAAGTTACATAGAGAGGAAGTGGATTTAGGAGTGGTGGCATTGGATAAT
     E  S  R  A  K  L  H  R  E  E  V  D  L  G  V  V  A  L  D  N
 841 TACGTATATACACTTGAAAAAAGTTCTCAGCCAGCAGACATTGCTAAACCTCTAGTTGGT
     Y  V  Y  T  L  E  K  S  S  Q  P  A  D  I  A  K  P  L  V  G
 901 TTTATCGAAATGGTTCCAGAGGCAGCAAGATTCAAATATATTGAGAAAAAAATATCAACT
     F  I  E  M  V  P  E  A  A  R  F  K  Y  I  E  K  K  I  S  T
 961 CAAATTAGCAAAACCTTTAGGCCGCGTGGTGACATAATTAGCCTTGAGAACAACTGGGGA
     Q  I  S  K  T  F  R  P  R  G  D  I  I  S  L  E  N  N  W  G
1021 GACCTCTCTTATCAAATACAGAAATCTGTTGATGATGTATTTCTGAAGCCAGTTCAATTG
     D  L  S  Y  Q  I  Q  K  S  V  D  D  V  F  L  K  P  V  Q  L
1081 CAACGTGAAAACTATACCAATATCCTAGTGAACAATGTCACCCAAGTAAAAGGTCTCATG
     Q  R  E  N  Y  T  N  I  L  V  N  N  V  T  Q  V  K  G  L  M
1141 GGAGTCTTGTTGAATGCAGTCAATTACAAAGTCTCAATGGAAGAAATTATTTTCAACGAC
     G  V  L  L  N  A  V  N  Y  K  V  S  M  E  E  I  I  F  N  D
1201 CAAAAGTGGCTGCCATGGCTTTAATCCTACTTTTGCTCTATATATAGTAGCATAAATAAA
     Q  K  W  L  P  W  L  *
1261 GGACAACAAATTTAGTATTATTGTTGTTGTCCAAACATGTTGCTAATGATATATGCTCTT
1321 TCCCTGTAATATCCTCGTCTTTCCAATGTATGAATATGATGATGAATTATATATGACAAA
1381 TAAAGTTTCTACTAGTTCTTAAT
```

Fig. 3. The nucleotide sequence of the *C2A* gene and the deduced amino acid sequence of the C2A protein. The transcription initiation site of the *C2A* gene is marked as the position +1. The nucleotides of the intron in the 5′UTR of the *C2A* gene are shown in lower-case letters. The *Sau3A*I site (GATC) in the coding region of the *C2A* gene and the polyadenylation signal (AATAAA) are shown in bold letters. The annotation of the *C2A* gene was made based on the experimental data obtained in this study.

could be identified between the region from −1 to −283 in the *C1* promoter (the transcription initiation site = +1) and the region from −1 to −348 in the *C2A* promoter (identity = 84–86%) or between the region from −1 to −286 in the *C1* promoter and the region from −1 to −275 in the *C2B* promoter (identity = 89%) (the predicted transcription initiation site = +1). Between the two Type II genes, *C2A* and *C2B* promoters share 88% to 95% identity between the region from −1 to −1345 in the *C2A* promoter and the region from −1 to −1229 in the *C2B* promoter. In addition, the *C2A* and *C2B* terminators also share 97% identity to each other, while both of them show great polymorphism when compared to the *C1* terminator, respectively. The results demonstrate that the *C1* gene has quickly evolved or acquired a new promoter and a new terminator for tissue- and organ-specific gene expression after the duplication from a Type II curcin gene.

To investigate if *J. curcas* harbors more curcin or curcin-like genes, the genomic DNA of *J. curcas* was digested with restriction enzyme *Xba*I and subjected to Southern blot analysis using the *C1* probe. The expected *Xba*I fragments of *C1*, *C2A* and *C2B* are 2604 bp, 2763 bp and 2729 bp, respectively, which are difficult to be separated on a 0.8% agarose gel. As expected, a strong hybridization band with molecular size at about 2.7 kb was observed in the Southern blot analysis (Fig. 1A). In addition to this 2.7-kb band, two other hybridization bands with one at 4.7 kb and another one at 6.0 kb were detected by the *C1* probe (Fig. 1A). The results indicated that at least two additional curcin or curcin-like genes might be present in the jatropha genome, which remain to be isolated.

3.3. Analysis of curcin proteins

The three curcin precursors identified in this study show high identity to other thirteen non-redundant curcin or curcin-like precursors with amino acid sequences available in GenBank (Fig. 5). The sixteen curcin precursors can be grouped to five Type I curcins, eight Type II curcins, and three curcin-like proteins (Fig. 5). Interestingly, the Type I curcin precursor C1 shows higher identity to the Type II curcin precursor C2B than that of the Type II curcin precursor C2A (Fig. 5). Phylogenetic tree analysis also confirmed that C1 is more closely related to C2B than C2A (Fig. 6). The results are consistent with the fact that *C1* and *C2B* share a higher identity in the coding regions than that of *C1* and *C2A* or *C2A* and *C2B*. There are 35 single nucleotide polymorphisms (SNP) between *C1* and *C2B*, 57 SNPs between *C1* and *C2A* and 59 SNPs

```
-180 CCTTGCATTTTCAATAATTCTTTTTCATTGGGATTTGTTGTTGTTTGTCTTTATTTTACA
-120 TCTTATGAAACTGTCGTTTGGTAGCTAATAATAATAATAATAATAATAATAAAGAAA
 -60 AAAAATTGACAAAATAAAGGGCAGTTTCCCTATAAAAGCAGGTGATGGGGGAAGGCAAAA
   1 GACCATCTCTCGCTTTCTTTgtaagtaatagtgaagcctctgcccttctttttttgttgac
  61 aaattccatttttttgttttactaatagcatgttaatttctagcttctggaaatgagttta
 121 ttatcctttatatgataaacttgtgaacattctatctctttttaattatttttataattt
 181 tatgcaaagctattaaaataatcgtattcgtataatgatatttgtgtttcttcatacaac
 241 tggacagGTGAAATCAATATGAAAGGTGGCAAGATGAACCTCTCTATTATGGTGGCTGCA
                      M  K  G  G  K  M  N  L  S  I  M  V  A  A
 301 TGGTTTTGCTGGAGTTGTATTATATTCGGATGGGCATCGGCTAGGGAAATAGTTTGTCCA
     W  F  C  W  S  C  I  I  F  G  W  A  S  A  R  E  I  V  C  P
 361 TTCTCATCAAACCAAAACTACAAAGCTGGTTCCACTCCAACTTTAACCATTACTTATGAC
     F  S  S  N  Q  N  Y  K  A  G  S  T  P  T  L  T  I  T  Y  D
 421 GCTGCTGCTGATAAGAAAAACTACGCCAACTTCATTAGAGATCTAAGAGAAGCATTTGGC
     A  A  A  D  K  K  N  Y  A  N  F  I  R  D  L  R  E  A  F  G
 481 TTCAGTTATTCAAGCCATGAAATACCAGTCCTACGGGCCACGGTTGCTGCAAATCAGAAA
     F  S  Y  S  S  H  E  I  P  V  L  R  A  T  V  A  A  N  Q  K
 541 TTTATTGTAGCCAAAGTCATAAATGTAGCGAATTTAGAAGTATCATTAGGATTAAACGTC
     F  I  V  A  K  V  I  N  V  A  N  L  E  V  S  L  G  L  N  V
 600 GTTAATGCATATTTAGTGGCTTATAAGGCAGGAGGTACATCCTATTTCTTTAACGATCCC
     V  N  A  Y  L  V  A  Y  K  A  G  G  T  S  Y  F  F  N  D  P
 661 GAATCTTTGGCTGATGCAAAAAAATATCTTTTCACAGACACAAAGCAACAAACGCTATCA
     E  S  L  A  D  A  K  K  Y  L  F  T  D  T  K  Q  Q  T  L  S
 721 TTTACTGGTAGCTATGCAGATTTTCTATCTAGGGCAAACGTACACAGAGAGGATGTGGAT
     F  T  G  S  Y  A  D  F  L  S  R  A  N  V  H  R  E  D  V  D
 781 TTAGGGGTGCTGGCATTAGATAATTACATATATATACTTCACAAAAGTTCTCAACCAGCA
     L  G  V  L  A  L  D  N  Y  I  Y  I  L  H  K  S  S  Q  P  A
 841 GACATTGCTAAACCTCTAGTTGGTTTTATCGAAATGGTTCCAGAGGCAGCAAGATTCAAA
     D  I  A  K  P  L  V  G  F  I  E  M  V  P  E  A  A  R  F  K
 901 TATATTGAGAAAAAAGTATTAACTCAAATTAGCGAAACCTTTAGGCCGCGTGGTGTCATA
     Y  I  E  K  K  V  L  T  Q  I  S  E  T  F  R  P  R  G  V  I
 961 ATTAGCCTTGAGAACAACTGGGGAGACCTCTCTTATCAAATACAGAAATCTGTAAATGGT
     I  S  L  E  N  N  W  G  D  L  S  Y  Q  I  Q  K  S  V  N  G
1021 ATATTTCTGAAGCCAGTTCAATTGCAACGTGAAAACTATACCAATATCCTAGTGAACAAT
     I  F  L  K  P  V  Q  L  Q  R  E  N  Y  T  N  I  L  V  N  N
1081 GTCACCCAAGTAACAGGTCTCATGGGAGTCTTGTTGAATGCAGTCAATTACAAAGTCTCA
     V  T  Q  V  T  G  L  M  G  V  L  L  N  A  V  N  Y  K  V  S
1141 ATGGAAGAAATTATTTTCAACTACCAAAAGTGGCTGCCATGGCTTTAATCCTACTTTTGC
     M  E  E  I  I  F  N  Y  Q  K  W  L  P  W  L  *
1201 TCTATATATAGTAGCATAAATAAAGGACAACAAATTTAGTATTATTGTTGTTGTCCAAAC
1261 ATGTTGCCAATGATATATGCTATTTCCCTGTAATATCCTCATCTTTCCAATGTATGAATA
1321 TGATGATGTATTATATATGACAAATAAAGTTTCTACTAGTTCTTAAT
```

Fig. 4. The nucleotide sequence of the *C2B* gene and the deduced amino acid sequence of the C2B protein. The predicted transcription initiation site of the *C2B* gene is marked as the position +1. The nucleotides of the predicted intron in the 5′UTR of the *C2B* gene are shown in lower-case letters. The two *Sau3A*I sites (GATC) in the coding region of the *C2B* gene and the putative polyadenylation signal (AATAAA) are shown in bold letters. The annotation of the *C2B* gene was predicted based on the gene organizations of the *C1* and *C2A* genes.

between *C2A* and *C2B* in the coding regions. The results collectively demonstrate that the duplication of *C2A* from *C2B*, or vice versa, was earlier than that of *C1* from *C2B* during evolution. In addition, *C1* might be a result of *C2B* duplication, followed by the mutation of the nucleotide "C" in the triplet codon "TCA", which encodes the serine residue at position 294 of C2B, to the nucleotide "G", which encodes the stop codon "TGA" in *C1*. This observation suggests that the emergence of Type I curcin gene might arise after the duplication of the Type II curcin genes followed by the point mutation from "C" in "TCA" to "G" in the stop codon "TGA".

3.4. Expression patterns of C1 and C2A in J. curcas

The mRNA transcripts of the *C1* and *C2A* genes in jatropha seeds and leaves were detected by RT-PCR using a common forward primer CF in the exon 1 of the *C1* or *C2A* gene and either one of the two specific reverse primers, C1SR from *C1* 3′ UTR and C2ASR from *C2A* 3′ UTR (Table 1). The *C1* mRNA transcripts were only detected in developing endosperm, whereas the *C2A* mRNA transcripts were only detected in young leaves (Fig. 7A). The RT-PCR products of *C1* and *C2A* were further verified by *Sau3A*I digestion. The CF-C1SR products of *C1* (1048 bp) produced three bands with size at 195 bp, 244 bp and 609 bp after *Sau3A*I digestion. On the other hand, the CF-C2ASR products of *C2A* (1057 bp) produced two bands with size at 244 bp and 813 bp (Fig. 7B). The temporal expression of *C1* in developing endosperm was determined by qRT-PCR. The *C1* transcripts were highly detected at 6 weeks after pollination, but were undetectable at 2, 4 or 8 weeks after pollination (Fig. 7C). Similarly, the *C2A* transcripts were only detected in young leaves, but not in fully expanded leaves, old leaves or leaves infected with aphids (*Aphis craccivora*) (Fig. 7D).

The presence of C1 and C2A proteins in different tissues and organs of *J. curcas* was detected by western blot analysis using in-house anti-C1N antibody (Fig. 8). The expected mature C1 proteins (~28 kDa) were detected in jatropha seeds, while the expected mature C2A proteins (~30 kDa) were detected in young leaves of jatropha (Fig. 8). Neither C1 nor C2A was detected in roots or calli (Fig. 8). It was noted that another protein with molecular size at about 17 kDa was specifically detected in proteins from jatropha seeds (Fig. 8). The results are consistent with the specific and differential expression of mRNA transcripts of the two curcin genes in these tissues and organs.

3.5. Characterization of C1 promoter

The *C1* promoter has specific promoter activity in endosperm cells of jatropha developing seeds. Several *cis*-acting DNA elements that are putatively involved in endosperm-specific or storage protein gene

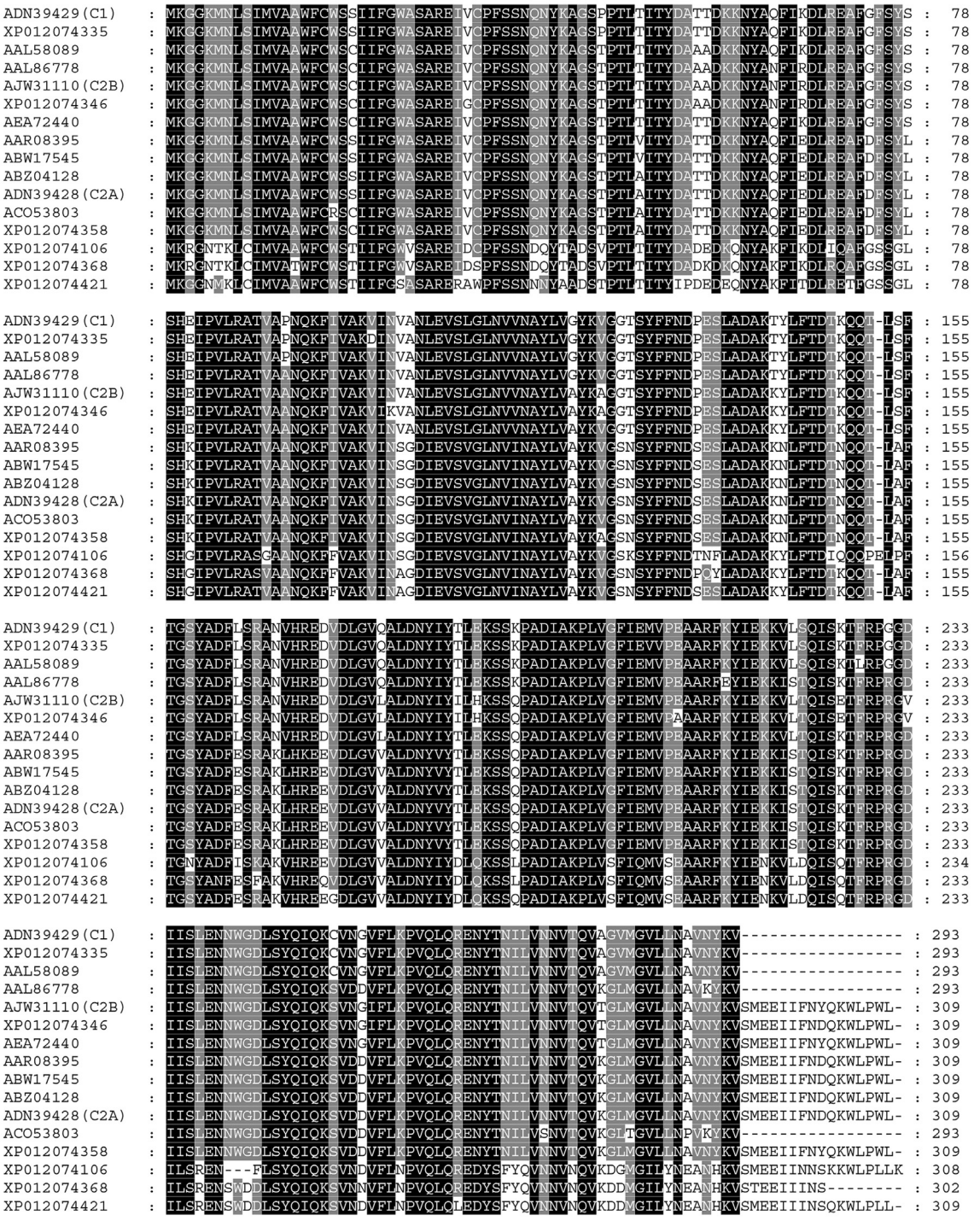

Fig. 5. Alignment of curcins and curcin-like proteins in *J. curcas*. The amino acid sequences were aligned using the program Clustal W and shaded with GENEDOC software. The identical and conserved amino acids are highlighted in black or grey background, respectively. Of the sixteen proteins, ADN39429 (C1), ADN39428 (C2A) and AJW31110 (C2B) were identified in this study and the other curcins or curcin-like proteins were downloaded from GenBank.

expression were identified in *C1* promoter through the survey for plant *cis*-acting regulatory DNA elements (PLACE) (http://www.dna.affrc.go.jp/PLACE/signalscan.html). Among which, nine E-box (CANNTG) elements were identified in the *C1* promoter (−2888 to +1) at positions −22, −823, −1041, −1331, −2273, −2316, −2335, −2596 and −2649. E-box element is required for high activity of the *napA* promoter in transgenic *Brassica napus* seeds (Stålberg et al., 1996). Another DNA element called "RY repeat (CATGCAY)" or legumin box was identified at position −1388. The legumin box was found in the promoters of seed-storage protein genes in legume, such as soybean (*Glycine* max) (Fujiwara and Beachy, 1994). The last DNA element termed "AACA element" was identified at position −1534. The core of AACA element was found in rice (*Oryza sativa*) glutelin genes and involved in controlling endosperm-specific expression (Wu et al., 2000).

To define the DNA elements in the *C1* promoter that are required for endosperm-specific gene expression, we performed promoter deletion assay. *C1* promoters with different deletion from 5′ region were transcriptionally fused with the *GFP* or *GUSPlus* genes. The promoter activity was determined by transient expression of GFP in endosperm cells or GUS activity in leaf cells of *J. curcas*. *C1* promoters P_{C1P3} (−2888 to

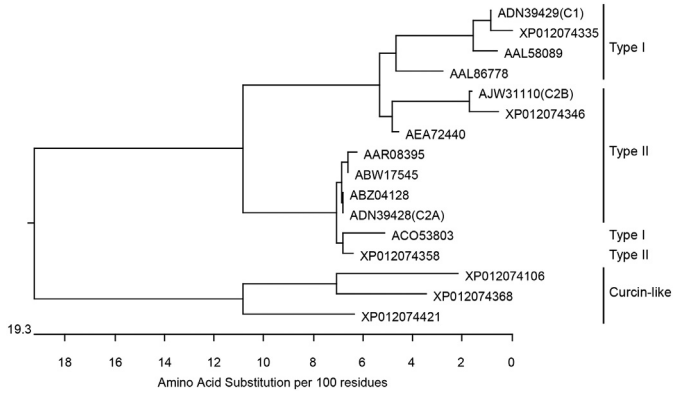

Fig. 6. Phylogenetic tree of curcins and curcin-like proteins. The protein alignment and the generation of un-rooted phylogenetic tree were performed with software Clustal W under MegAlign Pro™. The curcins and curcin-like proteins are the same as described in Fig. 5. Type I, Type I curcin; Type II, Type II curcin; Curcin-like, curcin-like proteins.

+293) or P_{C1P4} (−4956 to +293), which contain necessary *cis*-elements for endosperm-specific or storage protein gene expression, such as the AACA element, the RY repeat and the E-box elements, strongly drove GFP expression in endosperm cells (Table 2 and Fig. 9A). Deletion of the E-box elements at the upstream of −1746 in the *C1* promoter (P_{C1P2}) greatly attenuated *C1* promoter activity in endosperm cells (Table 2). Further deletion of the downstream E-box elements, the legumin box and the AACA element in the *C1* promoter (P_{C1P1}) abolished its activity in endosperm cells (Table 2). As the control, the maize ubiquitin gene promoter displayed strong activity in jatropha endosperm cells (Fig. 9A). In another experiment, none of the *C1* promoters showed any promoter activity in driving *GUSPlus* gene expression in jatropha leaf cells (Table 2 and Fig. 9B). The deletion of the 227-bp intron in the 5′UTR of the *C1* gene did not significantly affect *C1* promoter activity and its specific expression in endosperm cells of *J. curcas* (data not shown).

4. Discussion

We have isolated three curcin genes, *C1*, *C2A* and *C2B*, from *J. curcas* MD44. Genomic sequence analysis indicates that *C1*, *C2A* and *C2B* are allelic to the three curcin genes that encode XP012074335, XP012074358 and XP012074346. The *C1* gene encodes a Type I curcin that is specifically expressed in seed endosperm. The *C2A* gene encodes a Type II curcin that is specifically expressed in your leaves. The *C2A* gene encodes another Type II curcin whose expression remains to be characterized. Previous studies indicate that curcin-like genes are also presented in jatropha genome (Sato et al., 2011; Zhang et al., 2014). Southern blot analysis with the *C1* probe in this study indicates that there are at least two additional curcin or curcin-like genes in *J. curcas* MD44, which remain to be identified. The presence of multiple copies of curcin genes in jatropha genome with distinct promoters enables the curcin genes to be either developmentally regulated or responsive to an environmental stimulus. For instance, RIPs with distinct expression patterns to meet different defense-related function were observed in maize (Bass et al., 2004).

A 227- or 228-bp intron was identified in the 5′ UTR of the *C1* and *C2A* genes or predicted in the putative 5′UTR of the *C2B* gene. These

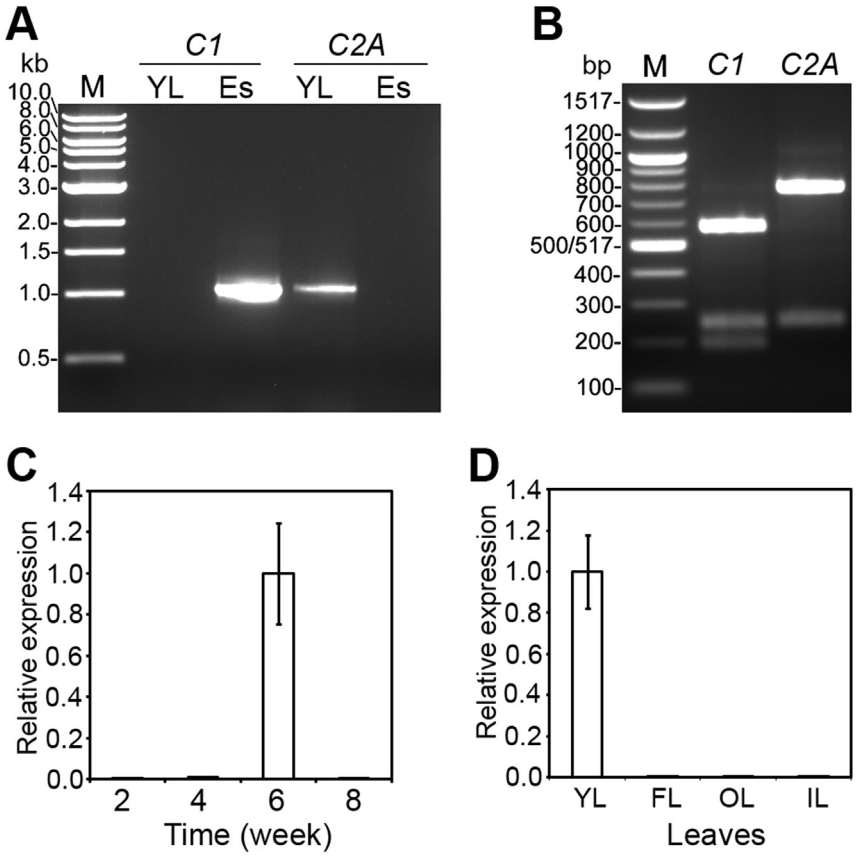

Fig. 7. Detection of curcin gene expression in *J. curcas*. A) RT-PCR analysis of *C1* and *C2A* transcripts in young leaves (YL) and endosperm (Es) of *J. curcas*. B) Verification of *C1* and *C2A* transcripts by digesting their RT-PCR products with restriction enzyme *Sau*3AI. C) Relative gene expression of *C1* in jatropha endosperm at different developmental stages detected by qRT-PCR. D) Relative gene expression of *C2A* in jatropha leaves at different developmental stages and jatropha leaves infected with aphids (*Aphis craccivora*) detected by qRT-PCR. For qRT-PCR, the average expression level of *C1* in 6-week-old jatropha endosperm (C) and *C2A* in young leaves (D) was set as "1", respectively. The qRT-PCR experiments were performed in triplicate, and the data are presented as means ± SD. M, DNA markers. YL, young leaves; FL, fully expanded leaves; OL, old leaves; IL, leaves infected with *A. craccivora*.

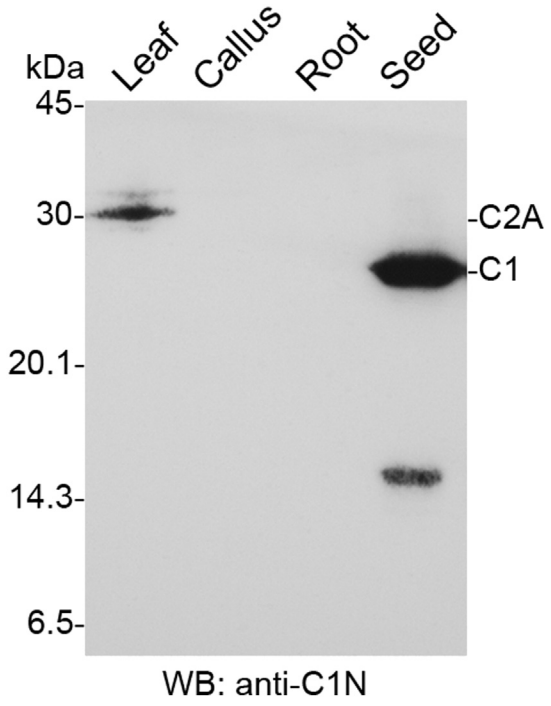

Fig. 8. Detection of curcin proteins in *J. curcas*. Total proteins isolated from young leaves, calli, roots and seeds of *J. curcas* were separated on 12% SDS-PAGE gel, blotted onto nitrocellulose membrane and detected by anti-C1N polyclonal antibodies. Standard protein markers are shown in kilo-Dalton (kDa). The bands corresponding to the matured C1 and C2A proteins are indicated.

common introns in the 5′ UTR of curcin genes were not found in the previous reports (Lin et al., 2003a; Wei et al., 2005). Previous studies demonstrated that the presence of an intron in the 5′'UTR is essential or enhances gene expression in plants (McElroy et al., 1990; Chung et al., 2006). However, our experimental results demonstrated that deletion of the 5′UTR intron from the *C1* promoter did not significantly affect its activity and specific expression in jatropha endosperm (data not shown). Nevertheless, considering the curcin genes encoding toxic RIP precursors, the presence of an intron in the 5′UTR of the curcin genes may play an important role in regulating transcription, mRNA export, and translation (Bicknell et al., 2012).

The *C1*, *C2A* and *C2B* genes are relatively conserved in their coding regions but show great polymorphism in their promoters. Qin et al. (2009a) isolated a 0.6-kb curcin gene promoter from *J. curcas* and claimed that the Type I curcin gene promoter was sufficient to drive the GUS gene expression in tobacco seeds. By comparing the five DNA primers listed in the paper with the three curcin gene promoters isolated in this study, we found that the reported 0.6-kb (620-bp) promoter and the 5 DNA primers matched perfectly to a 620-bp region of the same position in the *C2B* promoter, but not to that of the *C1* promoter, to which the DNA primer P1F cannot be aligned (Qin et al., 2009a) (data not shown). However, in this study we found that *C2B* is not a seed-specific-expressing curcin gene as no *C2B* transcripts were detected in leaves or seeds of *J. curcas* under normal growth condition. Qin et al. (2009b) characterized a 654-bp curcin-L gene promoter in transgenic tobacco and concluded that this gene promoter was leaf specific and able to drive the expression of the GUS reporter gene under stress conditions. After comparing with the DNA primers listed in the report with *C1*, *C2A* or *C2B* promoters, we found that the 654-bp curcin-L gene promoter was identical to the *C2A* promoter in this study. Furthermore, *C2A* was found to be expressed in young leaves of *J. curcas* when the plants were grown under normal growth condition, but was not induced by aphid infection (Fig. 7D). In both reports, the reverse DNA primers, P2R (Qin et al., 2009a) or CP2R (Qin et al., 2009b), were designed based on the DNA sequences in the introns in the 5′ UTR of *C2B* or *C2A* genes. Therefore, the 654-bp curcin-L gene promoter in previous report (Qin et al., 2009b) may be a truncated *C2A* promoter that disrupts normal intron splicing in the 5′ UTR during mRNA maturation.

The *C1* promoters P_{C1P3} and P_{C1P4} harbor all *cis*-elements that are required for high and specific promoter activity in endosperm cells of *J. curcas*. These findings make *C1* promoter an ideal promoter for driving gene expression specifically in jatropha endosperm cells. Recently, we used a 3765-bp *C1* promoter to make an RNAi construct for endosperm-specific gene silencing of the *C1* gene in jatropha and the transgenic RNAi plants produced curcin-deficient jatropha seeds (Gu et al., 2015). The C1 protein is a seed storage protein. In addition to the predicted 28-kDa C1 protein, a 17-kDa protein from endosperm was detected by the anti-C1N antibody in this study (Fig. 8). Interestingly, this 17-kDa protein was knocked down in curcin-deficient jatropha seeds from the *C1* RNAi plants (Gu et al., 2015), suggesting that it might be derived from the *C1* gene or other curcin-like genes that have not been isolated in this study. It has been reported that the type 1 RIP polypeptide could undergo proteolytic cleavage to yield smaller protein fragments. For instance, the maize RIP b32 (or ZmR1) is synthesized in the cytoplasm as an inactive precursor and subsequently becomes an active form after proteolytic cleavage of an N-terminal, a C-terminal and an internal sequence (Walsh et al., 1991). Further experiments needs to be done to verify the two possibilities.

Table 2
Characterization of *C1* promoters.

Promoter[a]	Size of promoter[b]	Reporter gene	Total number of tissue sections bombarded[c]	Positive cells or spots[d]
Empty vector	N.A.	N.A.	80, endosperm section	0
P_{Ubi1}	N.A.	*GFP*	81, endosperm section	960
P_{C1P1}	−200 to +293	*GFP*	80, endosperm section	0
P_{C1P2}	−1746 to +293	*GFP*	80, endosperm section	10
P_{C1P3}	−2888 to +293	*GFP*	80, endosperm section	336
P_{C1P4}	−4956 to +293	*GFP*	75, endosperm section	328
Empty vector	N.A.	N.A.	5, leaf disk	0
P_{35S}	N.A.	*GUSPlus*	5, leaf disk	65
P_{C1P1}	−200 to +293	*GUSPlus*	5, leaf disk	0
P_{C1P2}	−1746 to +293	*GUSPlus*	5, leaf disk	0
P_{C1P3}	−2888 to +293	*GUSPlus*	5, leaf disk	0
P_{C1P4}	−4956 to +293	*GUSPlus*	5, leaf disk	0

N.A., not applicable.
[a] Empty vector pC1300 was used as the negative control. Maize ubiquitin gene promoter (P_{Ubi1})-driving *GFP* gene in pSSZ41 (Kolesnik et al., 2004) and CaMV 35S promoter (P_{35S})-driving *GUSPlus* gene in pC1305.1 were used as the positive controls.
[b] The transcription initiation site of the *C1* promoter is marked as +1.
[c] The endosperm slices of *J. curcas* were used for transient GFP assay while leaf disks of *J. curcas* were used for transient GUS assay.
[d] For transient GFP assay, individual cells that showed GFP expression were counted; for transient GUS activity assay, blue spots stained by X-Gluc solution, which showed GUS activity, were scored.

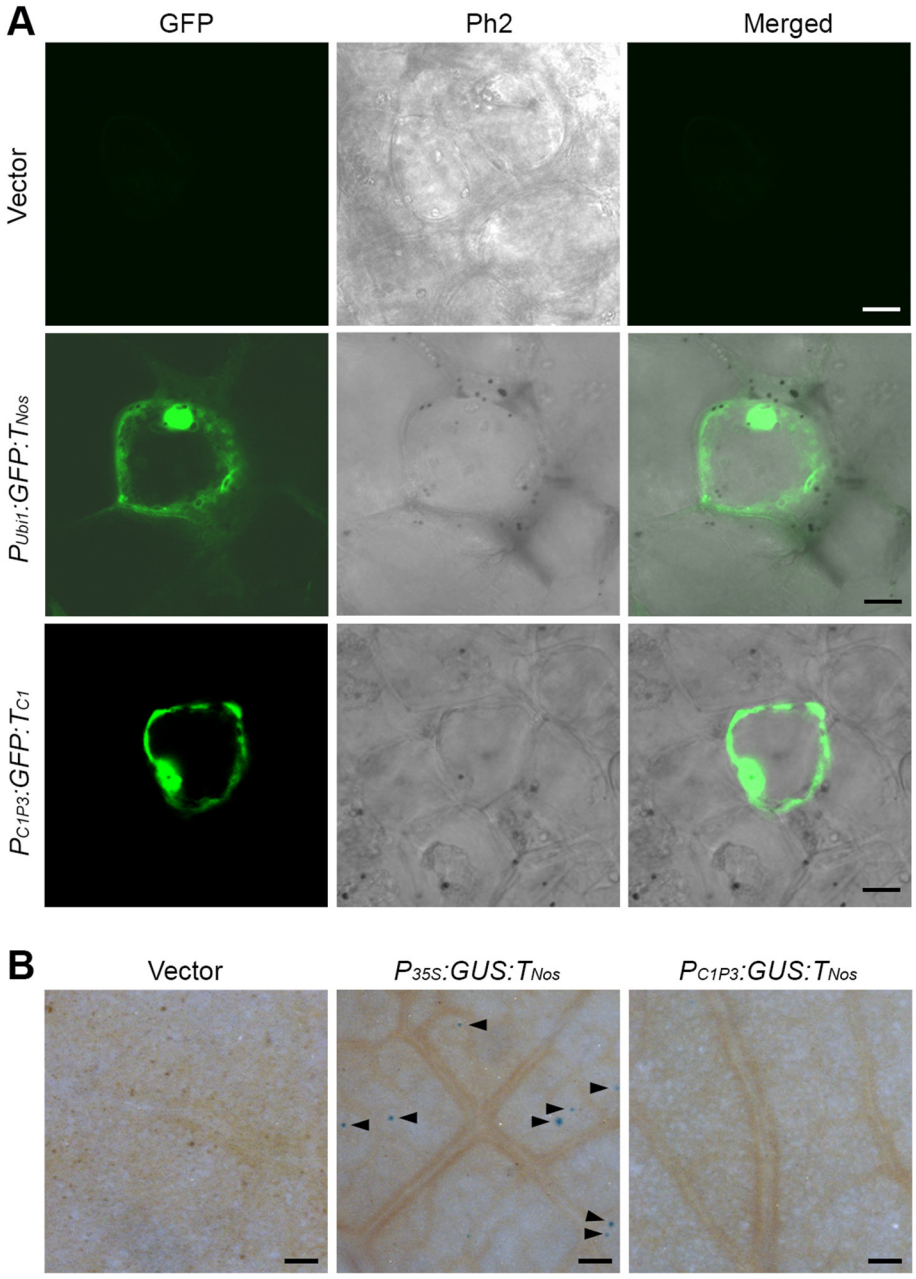

Fig. 9. Detection of endosperm-specific activity of *C1* promoter in *J. curcas*. A) Detection of *C1* promoter-driven GFP expression in developing endosperm cells of *J. curcas*. Transient expression of GFP in endosperm cells was imaged by confocal microscopy at 24 h after bombardment with empty vector pC1300 or constructs carrying P_{Ubi1}:*GFP*:T_{Nos} or P_{C1P3}:*GFP*:T_{C1} genes. GFP, fluorescence channel; Ph2, phase contrast channel; Merged, merged channels. Bar = 10 μm. B) Detection of *C1* promoter-driven transient GUS expression in leaf cells of *J. curcas*. GUS activity was detected by X-Gluc staining of leaf disks after bombardment with empty vector pC1300 or constructs carrying P_{35S}:*GUS*:T_{Nos} or P_{C1P3}:*GUS*:T_{C1} genes. Arrows indicate leaf cells with GUS activity, which were stained in blue color. Bar = 2 mm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

No *C2B* transcripts were detected in leaves or seeds of *J. curcas* in this study. Previously, the Type II curcin genes were found to be induced by biotic and abiotic stress (Wei et al., 2005; Qin et al., 2010). PLACE analysis with *C2B* promoter identified several putative *cis*-elements, including light-regulated and stress-inducible elements that were also identified in *C2A* promoter. Two of the putative *cis*-elements, which are unique to *C2B* promoter, are T/G-box at position −269 and ethylene responsive element (ERE) at position −782. T/G-boxes, found in promoters of tomato proteinase inhibitor II (pin2) and leucine aminopeptidase genes, are involved in jasmonate (JA) induction of these genes (Boter et al., 2004). ERE elements were found to mediate ethylene-induced activation of the *U3* promoter in the 5′-LTR region of TLC1.1 retrotransposon family in *Lycopersicon chilense* (Tapia et al., 2005). The presence of T/G-box and ERE element in the *C2B* promoter suggests that the curcin gene might be activated by JA upon attack by insects (Howe and Jander, 2008) and/or ethylene during defense reaction or leaf senescence (Van Loon et al., 2006). One possible future study would be to investigate whether *C2B* could be induced by any biotic and/or abiotic stress.

5. Conclusion

Three curcin genes, *C1*, *C2A* and *C2B*, have been isolated from *J. curcas* in this study. *C1* encodes a Type I curcin precursor and is specifically expressed in the endosperm of developing seeds. *C2A* encodes a Type II curcin precursor and is specifically expressed in young leaves of *J.*

curcas. *C2B* is another Type II curcin gene, whose expression was not detected under normal growth condition. A *C1* promoter (P_{C1P3}, −2888 to +293) was identified to carry all *cis*-elements that are required for full endosperm-specific promoter activity. The isolation and characterization of the three curcin genes and the identification of the endosperm-specific *C1* promoter provide useful information and research materials for further functional study of curcin proteins and genetic engineering of *J. curcas*.

Conflict of interest

A patent titled "*Jatropha curcas* curcin gene, tissue-specific promoters and generation of curcin-deficient transgenic jatropha plants" has been filed by Temasek Life Sciences Laboratory with Z. Yin, L. Wu, H. Mao and C. Qiu as inventors.

Acknowledgements

The authors thank Richard Jefferson of the Center for the Application of Molecular Biology to International Agriculture (CAMBIA) for binary vectors pC1300 and pC1305.1 and Kar Hui Ong for critical reading and comments on the manuscript. This work was supported by intramural research funds from the Temasek Life Sciences Laboratory TLL-3030 and JOil-Temasek Life Sciences Laboratory TLL-JOil-9141 Jatropha Biotechnology Program. LW and ZY designed the experiments and analyzed the experimental data. LW, MLG, DT and KG conducted the experiments. YH collected germplasm, prepared samples for cDNA library construction and supervised field work. ZY wrote the manuscript and MLG, DT, KG and YH participated in manuscript revision. All authors read and approved the final manuscript.

References

Bass, H.W., Krawetz, J.E., OBrian, G.R., Zinselmeier, C., Habben, J.E., Boston, R.S., 2004. Maize ribosome-inactivating proteins (RIPs) with distinct expression patterns have similar requirements for proenzyme activation. J. Exp. Bot. 55, 2219–2233.

Bicknell, A.A., Cenik, C., Chua, H.N., Roth, F.P., Moore, M.J., 2012. Introns in UTRs: why we should stop ignoring them. BioEssays 34, 1025–1034.

Boter, M., Ruíz-Rivero, O., Abdeen, A., Prat, S., 2004. Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and *Arabidopsis*. Genes Dev. 18, 1577–1591.

Chung, B.Y., Simons, C., Firth, A.E., Brown, C.M., Hellens, R.P., 2006. Effect of 5′UTR introns on gene expression in *Arabidopsis thaliana*. BMC Genomics 7, 120.

Dellaporta, S.L., Wood, J., Hicks, J.B., 1983. A plant DNA minipreparation: version II. Plant Mol. Biol. Report. 1, 19–21.

Devappa, R.K., Makkar, H.P., Becker, K., 2010. Nutritional, biochemical, and pharmaceutical potential of proteins and peptides from *Jatropha*: review. J. Agricult. Food. Chem. 58, 6543–6555.

Fujiwara, T., Beachy, R.N., 1994. Tissue-specific and temporal regulation of a β-conglycinin gene: roles of the RY repeat and other *cis*-acting elements. Plant Mol. Biol. 24, 261–272.

Gu, K., Chiam, H., Tian, D., Yin, Z., 2011. Molecular cloning and expression of heteromeric ACCase subunit genes from *Jatropha curcas*. Plant Sci. 180, 642–649.

Gu, K., Tian, D., Mao, H., Wu, L., Yin, Z., 2015. Development of marker-free transgenic *Jatropha curcas* producing curcin-deficient seeds through endosperm-specific RNAi-mediated gene silencing. BMC Plant Biol. 15, 242.

He, W., King, A.J., Khan, M.A., Cuevas, J.A., Ramiaramanana, D., Graham, I.A., 2011. Analysis of seed phorbol-ester and curcin content together with genetic diversity in multiple provenances of *Jatropha curcas* L. from Madagascar and Mexico. Plant Physiol. Biochem. 49, 1183–1190.

Hiei, Y., Ohta, S., Komari, T., Kumashiro, T., 1994. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. Plant J. 6, 271–282.

Howe, G.A., Jander, G., 2008. Plant immunity to insect herbivores. Annu. Rev. Plant Biol. 59, 41–66.

Huang, M.-X., Hou, P., Wei, Q., Xu, Y., Chen, F., 2008. A ribosome-inactivating protein (curcin 2) induced from *Jatropha curcas* can reduce viral and fungal infection in transgenic tobacco. Plant Growth Regul. 54, 115–123.

Jaramillo-Quintero, L.P., Contis Montes de Oca, A., Romero Rojas, A., Rojas-Hernández, S., Campos-Rodríguez, R., Martínez-Ayala, A.L., 2015. Cytotoxic effect of the immunotoxin constructed of the ribosome-inactivating protein curcin and the monoclonal antibody against Her2 receptor on tumor cells. Biosci. Biotechnol. Biochem. 1–11.

King, A.J., Li, Y., Graham, I.A., 2011. Profiling the developing *Jatropha curcas* L. seed transcriptome by pyrosequencing. Bioenerg. Res. 4, 211–221.

Kolesnik, T., Szeverenyi, I., Bachmann, D., Kumar, C.S., Jiang, S., Ramamoorthy, R., Cai, M., Ma, Z.G., Sundaresan, V., Ramachandran, S., 2004. Establishing an efficient *Ac/Ds* tagging system in rice: large-scale analysis of *Ds* flanking sequences. Plant J. 37, 301–314.

Lin, J., Chen, Y., Xu, Y., Yan, F., Tang, L., Chen, F., 2003a. Cloning and expression of curcin, a ribosome-inactivating protein from the seeds of *Jatropha curcas*. Acta Bot. Sin. 45, 858–863.

Lin, J., Yan, F., Tang, L., Chen, F., 2003b. Antitumor effects of curcin from seeds of *Jatropha curcas*. Acta Pharmacol. Sin. 24, 241–246.

Lin, J., Zhou, X., Wang, J., Jiang, P., Tang, K., 2010. Purification and characterization of curcin, a toxic lectin from the seed of *Jatropha curcas*. Prep Biochem Biotechnol 40, 107–118.

Liu, Y.-G., Whittier, R.F., 1995. Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics 25, 674–681.

Luo, M.J., Yang, X.Y., Liu, W.X., Xu, Y., Huang, P., Yan, F., Chen, F., 2006. Expression, purification and anti-tumor activity of curcin. Acta Biochim. Biophys. Sin. 38, 663–668.

Maghuly, F., Laimer, M., 2013. *Jatropha curcas*, a biofuel crop: functional genomics for understanding metabolic pathways and genetic improvement. Biotechnol. J. 8, 1172–1182.

McElroy, D., Zhang, W., Cao, J., Wu, R., 1990. Isolation of an efficient actin promoter for use in rice transformation. Plant Cell 2, 163–171.

Mohamed, M.S., Veeranarayanan, S., Baliyan, A., Poulose, A.C., Nagaoka, Y., Minegishi, H., Iwai, S., Shimane, Y., Yoshida, Y., Maekawa, T., Kumar, D.S., 2014a. Structurally distinct hybrid polymer/lipid nanoconstructs harboring a type-I ribotoxin as cellular imaging and glioblastoma-directed therapeutic vectors. Macromol. Biosci. 14, 1696–1711.

Mohamed, M.S., Veeranarayanan, S., Poulose, A.C., Nagaoka, Y., Minegishi, H., Yoshida, Y., Maekawa, T., Kumar, D.S., 2014b. Type 1 ribotoxin-curcin conjugated biogenic gold nanoparticles for a multimodal therapeutic approach towards brain cancer. Biochim. Biophys. Acta 1840, 1657–1669.

Murashige, T., Skoog, F., 1962. A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol. Plant. 15, 473–497.

Nielsen, K., Boston, R.S., 2001. Ribosome-inactivating proteins: a plant perspective. Annu. Rev. Plant Biol. 52, 785–816.

Pradhan, S., Naik, S., Khan, M., Sahoo, P., 2012. Experimental assessment of toxic phytochemicals in *Jatropha curcas*: oil, cake, bio-diesel and glycerol. J. Sci. Food Agric. 92, 511–519.

Qin, X., Shao, C., Hou, P., Gao, J., Lei, N., Jiang, L., Ye, S., Gou, C., Luo, S., Zheng, X., Gua, X., Zhua, X., Xua, Y., Chen, F., 2010. Different functions and expression profiles of curcin and curcin-L in *Jatropha curcas* L. Z. Naturforsch. C 65, 355.

Qin, X., Zhang, J., Shao, C., Lin, S., Jiang, L., Zhang, S., Xu, Y., Chen, F., 2009a. Isolation and characterization of a curcin promoter from *Jatropha curcas* L. and its regulation of gene expression in transgenic tobacco plants. Plant Mol. Biol. Report. 27, 275–281.

Qin, X., Zheng, X., Shao, C., Gao, J., Jiang, L., Zhu, X., Yan, F., Tang, L., Xu, Y., Chen, F., 2009b. Stress-induced curcin-L promoter in leaves of *Jatropha curcas* L. and characterization in transgenic tobacco. Planta 230, 387–395.

Sabandar, C.W., Ahmat, N., Jaafar, F.M., Sahidin, I., 2013. Medicinal property, phytochemistry and pharmacology of several *Jatropha* species (Euphorbiaceae): a review. Phytochemistry 85, 7–29.

Sambrook, J., Russell, D.W., 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sato, S., Hirakawa, H., Isobe, S., Fukai, E., Watanabe, A., Kato, M., Kawashima, K., Minami, C., Muraki, A., Nakazaki, N., 2011. Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. DNA Res. 18, 65–76.

Sikriwal, D., Batra, J.K., 2010. Ribosome inactivating proteins and apoptosis. Toxic Plant Proteins. Springer, pp. 167–189.

Stålberg, K., Ellerström, M., Ezcurra, I., Ablov, S., Rask, L., 1996. Disruption of an overlapping *E*-box/ABRE motif abolished high transcription of the napA storage-protein promoter in transgenic *Brassica napus* seeds. Planta 199, 515–519.

Stirpe, F., 2004. Ribosome-inactivating proteins. Toxicon 44, 371–383.

Tapia, G., Verdugo, I., Yañez, M., Ahumada, I., Theoduloz, C., Cordero, C., Poblete, F., González, E., Ruiz-Lara, S., 2005. Involvement of ethylene in stress-induced expression of the *TLC1. 1* retrotransposon from *Lycopersicon chilense* Dun. Plant Physiol. 138, 2075–2086.

Van Loon, L.C., Rep, M., Pieterse, C., 2006. Significance of inducible defense-related proteins in infected plants. Annu. Rev. Phytopathol. 44, 135–162.

Walsh, T.A., Morgan, A., Hey, T.D., 1991. Characterization and molecular cloning of a proenzyme form of a ribosome-inactivating protein from maize. Novel mechanism of proenzyme activation by proteolytic removal of a 2.8-kilodalton internal peptide segment. J. Biol. Chem. 266, 23422–23427.

Wei, Q., Huang, M.-X., Xu, Y., Zhang, X.-S., Chen, F., 2005. Expression of a ribosome inactivating protein (curcin 2) in *Jatropha curcas* is induced by stress. J. Biosci. 30, 351–357.

Wu, C.Y., Washida, H., Onodera, Y., Harada, K., Takaiwa, F., 2000. Quantitative nature of the prolamin-box, ACGT and AACA motifs in a rice glutelin gene promoter: minimal cis-element requirements for endosperm-specific gene expression. Plant J. 23, 415–421.

Zhang, L., Zhang, C., Wu, P., Chen, Y., Li, M., Jiang, H., Wu, G., 2014. Global analysis of gene expression profiles in physic nut (*Jatropha curcas* L.) seedlings exposed to salt stress. PloS One 9, e97878.