Front. Agric. China 2007, 1(1): 17–23
DOI 10.1007/s11703-007-0003-1

RESEARCH ARTICLE

CAIYIN Qinggele, LI Mingchun, WEI Dongsheng, CAI Yi, XING Laijun

Isolation and sequencing analysis on the seed-specific promoter from soybean

Abstract The low level of foreign genes' expression in transgenic plants is a key factor that limits plant genetic engineering. Because of the critical regulatory activity of the promoters on gene transcription, they are studied extensively to improve the efficiency of the plant transgenic system. The constitutive promoters, such as CaMV 35S promoter, are usually used in plant genetic engineering. But those constitutive promoters continuously express their downstream genes during the whole life span in all the tissues of the host plants. This is not only wasteful to host plant's energy, but also harmful to host plants and usually affects their agronomic characteristics. In contrast, the seed-specific promoter only expresses its downstream genes from mid to late stage of seed maturation, and there is no expression or much lower expression in other tissues. So the seed-specific promoters are distinguished for their improvement and what they have brought to plant quality engineering. The aim of this article is to characterize a new seed-specific promoter and improve grain quality. The promoter region of *β*-conglycinin α-subunit gene was isolated from the genomic DNA of soybean Jilin 43 by PCR method, and successfully extended this fragment by TAIL PCR method and obtained the promoter fragment BCSP666. Sequencing analysis showed that the cloned fragment BCSP666 contained all of the motifs, such as RY repeat element, AG/CCCCA motif, TACACAT motif, ACGT motif, A/T rich motif and E-box etc., which constituted the seed-specific promoter activity. Based on this sequencing analysis, the seed-specific promoter activity of the fragment BCSP666 was predicted. And then the seed-specific expression vector pBI121-666, which contained GUS reporter gene, was constructed with the fragment BCSP666. Transformation of *Arabidopsis thaliana* plants by *Agrobacterium*-mediated floral-dip method with the recombined vector pBI121-666 was conducted. The transgenic plants were selected on the kanamycin-resistant MS medium, and confirmed by Southern Blot Analysis. Fluorometric and histochemical analysis of GUS enzyme activity of the transgenic *Arabidopsis thaliana* plants support our original suggestions. Therefore, the seed-specific promoter BCSP666 is obtained and characterized.

Keywords soybean, seed-specific promoter, motif, TAIL PCR

1 Introduction

Soybeans contain two major kinds of seed-storage proteins, *β*-conglycinin (7S proteins) and glycinins (11S proteins), of which the *β*-conglycinin accumulates up to 30% of the total seed proteins (Beachy et al., 1981). The *β*-conglycinin consists of three subunits, a-, α′- and *β*-subunits. Analysis of the promoter activities of these subunits indicated that the promoter of α′-subunit possesses more seed-specific activities than the others and it contains all of the motifs required for seed-specific expression in the upstream region of this α′-subunit gene promoter (Beachy et al., 1985). Cahoon, et al. (1999, 2000, 2001) successfully expressed the fatty acid desaturase gene in soybean somatic embryos by means of this α′-subunit gene promoter. Therefore, it is a feasible way to control the interested genes in the transgenic plants temporally and specially with seed-specific promoter, and it is an important breakthrough and improvement in plant genetic engineering technology.

Except for the report by Yoshino, et al. (2001), in which the 489 bp fragment of the α-subunit gene promoter was merely cloned, there is no report about the function of α-subunit gene promoter nowadays. In order to further explore the function of the α-subunit gene promoter of the soybean *β*-conglycinin and develop a novel seed-specific promoter, the 489 bp fragment of the α-subunit gene promoter was firstly cloned by PCR method and successfully extended by Thermal Asymmetric Interlaced PCR (TAIL PCR) to gain

Translated from *Acta Agronomoca Sinica*, 2005, 31(1): 11–17 [译自: 作物学报]

CAIYIN Qinggele, LI Mingchun, WEI Dongsheng, CAI Yi,
XING Laijun (✉)
Tianjin Key Laboratory of Microbial Functional Genomics, Department of Microbiology, Nankai University, Tianjin300071, China
E-mail: xinglaij@eyou.com

666 bp fragment, and then the sequence, structure and function of this fragment were analyzed in this study.

2 Materials and methods

2.1 Materials

2.1.1 Experimental materials

The seeds of the Soybean Jilin 43S were kindly provided by the Academy of Agricultural Sciences in Jilin province. The seeds of *Arabidopsis thaliana* were kindly provided by Prof. Wang Ningning from Nankai University. *Escherichia coli* DH5α, *Agrobacterium* LBA4404, and plant expression vector pBI121 were available in our labs. The sequencing vector pGEM-T was purchased from Promega Corporation.

2.1.2 Experimental reagents

DNA Restriction enzymes, T_4 NDA ligase, Taq DNA polymerase, and DNA gel extraction kit were purchased from Takara Bio Inc, where the DNA sequencing was also accomplished. Other chemicals were commercially available at home and abroad. The PCR primers were synthesized in Shanghai Sangon Biotechnology Co. Ltd. PCR reactions were performed at Tgradient Thermoblock (Biometra). Fluorometrical analysis was accomplished with a Spectrofluorophotometer RF-540 (Shimadzu Corporation).

2.2 Methods

2.2.1 Extraction of soybean total DNA

Total DNA of the seeds of the Soybean Jilin 43 was extracted by SDS-CTAB method with some improvement (Liu et al., 1997). Its purity and concentration of 5 μL of the DNA samples were analyzed by gel electrophoresis.

2.2.2 PCR amplification

(1) Amplification of BCSP489 fragment The primer P1 and P2 were designed and synthesized according to the sequence of the α-subunit gene promoter (Yoshino et al., 2001), and the restriction enzyme site *Hin*dIII and *Xba*I were added to 5′-end of Primer P1 and P2 for further cloning, Primer P1: 5′-GCG AAG CTT (*Hin*dIII) AAG CAA CCA TAT CAG CAT ATC-3′, Primer P2: 5′-GAA TCT AGA (*Xba*I) AAC CGC GCT CTC ATC ATA GTA TAT-3′. And then the total DNA of Soybean Jilin 43 was used as template, PCR reaction was conducted in the following conditions, 97°C – 5 min for pre-denature, followed by 35 cycles of repeating 94°C – 30 sec, 52°C – 1 min, 72°C – 2 min, and then 72°C – 10 min for post-extension. The PCR products were extracted by a DNA gel extraction kit and ligated to pGEM-T, to obtain pT-BCSP489 and sequence it.

(2) TAIL PCR extension Three nested primers SP1, SP2, SP3 (SP1: 5′-GTT GCG CAT GCA TGA TCC AAG AGA-3′; SP2: 5′-CTT GGA CAT TGC TTT CGA AAG GAT A-3′; SP3: 5′-TGA AGT GGG GTG AGG TTG CAT T-3′) and three arbitrary degenerate primers AD1, AD2, and AD3 (AD1: 5′-NTCGA(G/C)T(A/T)T(G/C)G(A/T) GTT-3′; AD2: 5′-NGTCGA(G/C)(A/T)GANA(A/T)GAA-3′; AD3: 5′-(A/T)GTGNAG(A/T)ANCA NAGA-3′) were synthesized respectively according to the sequence of BCSP489 obtained by Liu et al. (1995). TAIL PCR was conducted with the above primers and PCR conditions were referred to the report of Liu et al. (1995) with some alterations. The composition of TAIL PCR reaction is shown in Table 1 and the procedure of TAIL PCR is shown in Table 2. TAIL PCR products (named TAIL1, approximately 200 bp) were obtained and sequenced.

(3) Amplification of BCSP666 fragment The upstream primer P3, 5′-GCG AAG CTT (*Hin*dIII) CAA AAA CGC AAT CAC ACA CA-3′ was synthesized according to the sequence of TAIL1, and then PCR products BCSP666 were obtained with the primer P3, P2 and the total DNA of Soybean Jilin 43 was used as template. The PCR conditions were as follows, 97°C – 5 min for pre-denature, followed by 35 cycles of repeating 94°C –30 secs, 52°C – 1 min, 72°C – 2 min, then 72°C – 10 min for post-extension. The PCR products BCSP666 were extracted by DNA gel extraction kit, ligated to sequencing vector pGEM-T to generate pT-BCSP666, and then sequence it.

2.2.3 Construction of the seed-specific expression vector with BCSP666 and transformation of *Arabidopsis thaliana* plant

The plasmid pT-BCSP666 and plant expression vector pBI121 were double-digested with *Hin*dIII and *Xba*I, respectively and ligated with each other so as to construct vector pBI121-666 (Fig. 6). Then the vector pBI121-666 was transformed into

Table 1 The compose of TAIL PCR reaction mixture

Reagent	Amount in primary reaction mixture/μL	Amount in secondary reaction mixture/μL	Amount in tertiary reaction mixture/μL
10 × PCR buffer	2	2	5
2 mmol/L dNTPs	2	2	5
10 mmol/L AD	4	4	10
1 mmol/L GSP	4	4	10
ddH_2O	6.5	6.5	17
Taq Pol (2 U/μL)	0.5	0.5	2
Template	1(100 nmol)	1(1/50 primary PCR product)	1(1/50 secondary PCR product)

Table 2 TAIL PCR procedure

Reaction	File No.	Thermal cycling condition	Cycle No.
	1	95°C 2 min, 7°C 1 min	1
	2	95°C 15s, 65°C 15s, 72°C 30s	5
	3	95°C 15s, 5°C 3 min, ramping to 72°C over 3 min, 72°C 2 min	1
Primary	4	95°C 15s, 44°C 15s, 72°C 30s 95°C 10s, 65°C 15s, 72°C 30s	3
	5	95°C 10s, 65°C 15s, 72°C 30s 95°C 10s, 44°C 15s, 72°C 30s	12
	6	72°C 7 min 95°C 10s, 61°C 15s, 72°C 30s	1
	7	95°C 10s, 61°C 15s, 72°C 30s	15
Secondary		95°C 10s, 48°C 15s, 72°C 30s	
	8	72°C 7 min 95°C 10s, 65°C 15s, 72°C 30s	1
	9	95°C 10s, 65°C 15s, 72°C 30s	15
Tertiary		95°C 10s, 48°C 15s, 72°C 30s	
	10	72°C 7 min	1

Agrobaterium LBA4404 by means of the freeze-and-thaw method (Hofen and Willmitzer, 1988). The positive recombined *Agrobaterium* LBA4404 was selected to transform *Arabidopssi thaliana* by Agrobacterium-mediated floral-dip method (Clough and Bent, 1998) and the positive plant strains were selected on the 1/2 MS medium containing 50 mg/L kanamycin. The positive *Arabidopsis thaliana* plant strains were further confirmed by Southern Blot Analysis with the help of PCR products of the GUS gene as a probe. The amplification of GUS gene primers were P_{gus1}, 5′-GAA GAG GAT CCC CGG GTG GT-3′, and P_{gus2}, 5′-ACA GAG CTC GAT GGT GCG CCA GGA GAG TTG-3′.

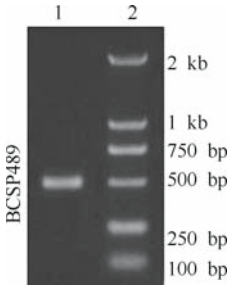
2.2.4 Fluorescence analysis of the GUS protein in transgenic *Arabidopsis thaliana* plants

Fluorescence and histochemistry analyses of the GUS protein were referred to the work of Jefferson (1987). For the fluorescence analysis, GUS proteins were extracted from 100 mg leaves and seeds of the transgenic *Arabidopsis thaliana* plants, and same amount samples of the wild-type *Arabidopsis thaliana* strains as the negative controls respectively. Fluorescence activities of these samples were analyzed at Spectrofluorophotometer RF-540. For the histochemical analysis, the seeds of the transgenic *Arabidopsis thaliana* plants and wild-type plants were harvested respectively, and dipped in X-Gluc solution [0.1 mol/L phosphate buffer containing 1 mmol/L X-Gluc (5-bromo-4-chloro-3-indolyl glucuronide)] at 37°C for a whole night. Then the samples were destained with 75% ethanol at 37°C for 4–6 h, and then observed under optical microscopes.

3 Results

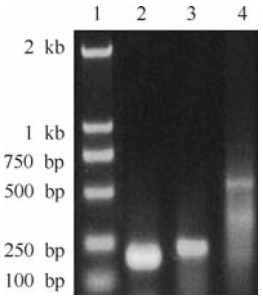
3.1 Cloning of seed-specific promoter fragment BCSP666

The promoter fragment BCSP489 was obtained by PCR amplification via primers P1 and P2, and the result is shown in Fig. 1. This fragment was ligated to pGEM-T to obtain pT-BCSP489 and sequence it.



1: PCR product BCSP489
2: DNA Marker
Fig. 1 Cloning of BCSP489

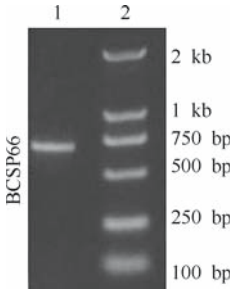
According to the sequence of the BCSP489, the primers, required for TAIL PCR, were synthesized and the BCSP489 was extended by TAIL PCR, so that the fragment TAIL1 was obtained and then sequenced, as shown in Fig. 2.



1: DNA Marker
2: Primary TAIL PCR product
3: Secondary TAIL PCR product
4:Tertiary TAIL PCR product
Fig. 2 TAIL PCR product

The upstream primer P3 was synthesized according to the sequence of TAIL1 (217 bp), and the PCR fragment

BCSP666 was obtained by primer P2, P3 via the total DNA of Soybean Jilin 43 as the PCR template. The result is shown in Fig. 3. Then the PCR fragment BCSP666 was ligated to sequencing vector pGEM-T, to get pT-BCSP666 obtained and sequenced.

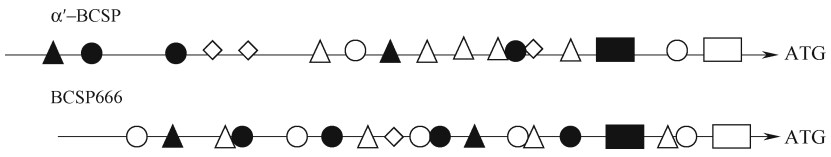


1: PCR product BCSP666
2: DNA Marker
Fig. 3 Cloning of BCSP666

3.2 Sequencing analysis of the seed-specific promoter fragment BCSP666

Sequencing analysis of the seed-specific promoter fragment BCSP666 indicated that this fragment contains 666 pairs of bases and several kinds of seed-specific promoter motifs as shown in Fig. 4 and Fig. 5. The motifs are as follows: (1) A/T rich motif, which is an important region for binding of transcriptional factors and required for seed-specific expression (Stalberg et al., 1996); (2) RY repeat element (or Legeumin box), which is also a sequence motif required for seed-specific expression (Chamberland et al., 1992); (3) AGCCCA motif, which is a sequence motif required for transcriptional regulation and seed-specific expression (Chen et al., 1986); (4) TACACAT motif, which allows one base to mismatch and is responsible for activating seed-storage protein expression (Josefsson et al., 1987); (5) ACGT motif, which is also a sequence motif required for seed-specific expression (Vincentz et al., 1997); (6) E-box, which is cooperated with other seed-specific promoter motifs to activate the seed-specific

A: RY repeat motif; B: ACGT motif; C: AGCCCA motif; D: TACACAT motif; E: E-box; F: CAAT box; G: TATA box
Fig. 4 Sequence analysis of soybean seed-specific promoter BCSP666



□: TATA box; ■: CAAT box; ○: RY repeat elements; △: AG/CCCCA motifs; ●: TACACAT motifs; ▲: ACGT motifs; ◇: E-box
Fig. 5 Schematic diagram summarizing distribution of seed-specific promoter related motifs on nucleotide sequences of the a′-subunit gene promoter and BCSP666

promoter for heterologous expression (Kawagoe and Murai, 1992). Although the sequence alignment between BCSP666 and β-conglycinin α'-subunit gene promoter was very low (40%), both of them exhibited similar sequence structure and similar amount of seed-specific promoter motifs, such as multi-copy of A/T rich motif, RY repeat element, AGCCCA motif and TACACAT motif as shown in Fig. 4 and Fig. 5. Based on the above analysis, it indicates that both of the promoter fragment BCSP666 and β-conglycinin α'-subunit gene promoter possess similar promoter activity. Therefore, the promoter fragment BCSP666 may promote the seed-specific expression of heterologous gene.

3.3 Construction of the seed-specific expression vector and transformation of *Arabidopsis thaliana* plants

The procedure to construct the seed-specific expression vector is shown in Fig. 6. It proved that the seed-specific expression vector pBI121-666 was confirmed by double digestion via *Hin*dIII and *Xba*I, and that the promoter fragment BCSP666 inserted in front of the GUS gene of the plant expression vector pBI121 and substituted the original CaMV 35S promoter, as shown in Fig. 7. The vector pBI121-666 was transformed into *Agrobacterium* LBA4404 by the freeze-and-thaw method, further transformed into *Arabidopsis thaliana* plants via *Agrobacterium*-mediated fluoral-dip method and approximately 2 000 seeds of *Arabidopsis thaliana* were harvested. The seeds were planted on the MS medium containing kanamycin (50 mg/L) and cultivated for about 15 days, and 20 transgenic plant strains were harvested. Then they were transplanted into a flowerpot to harvest their seeds. The Southern Blot Analysis of the total DNA from the transgenic *Arabidopsis thaliana* plants indicated (Fig. 8) that the GUS gene was inserted as a single copy into the genomic DNA of the transgenic *Arabidopsis thaliana* plants.

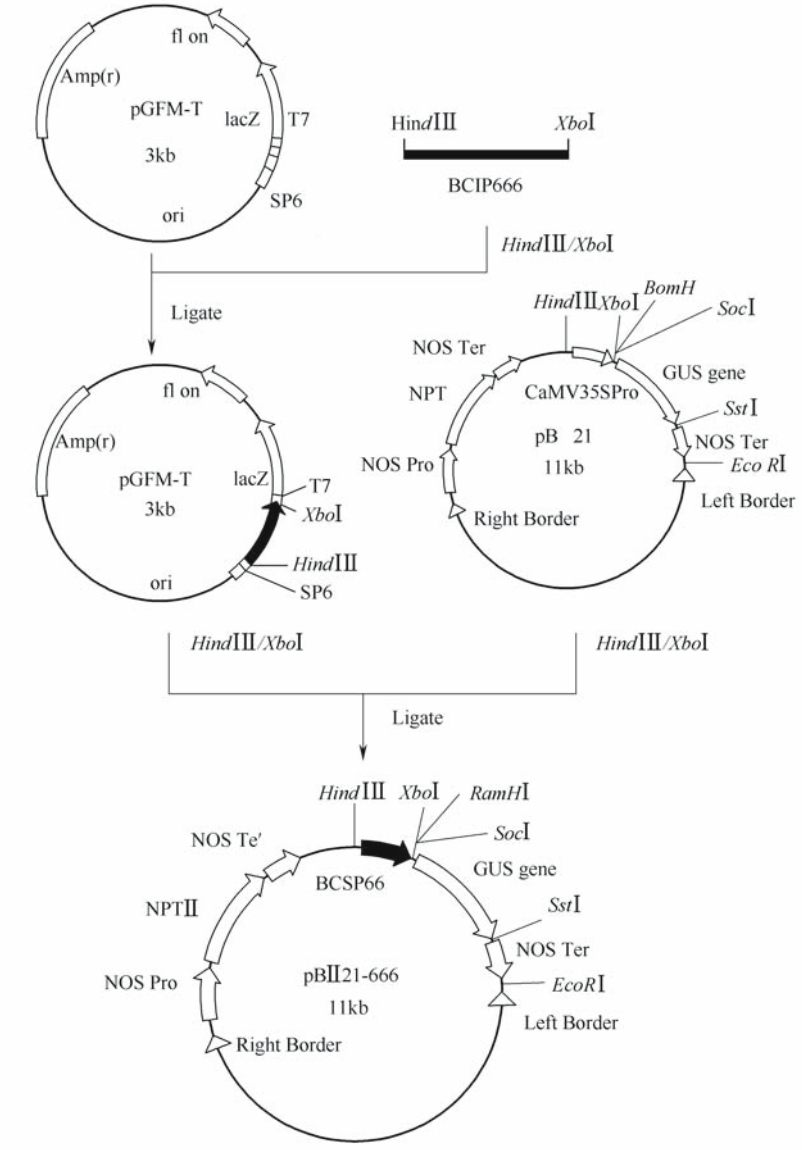
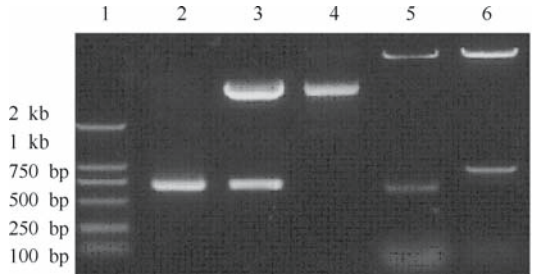
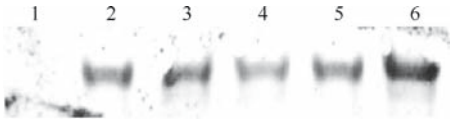


Fig. 6 Construction of seed-specific expression vector pBI121-666



1:The DNA Marker; 2:The PCR product BCSP489; 3:Double digestion product pT-BCSP666/*Hin*dIII/*Xba*I; 4:Double digestion product pGEM-T/*Hin*dIII/*Xba*I; 5:Double digestion product pBI121-666/*Hin*dIII/*Xba*I; 6:Double digestion product pBI121/*Hin*dIII/*Xba*I

Fig. 7 Construction and identification of the seed-specific vector pBI121-666

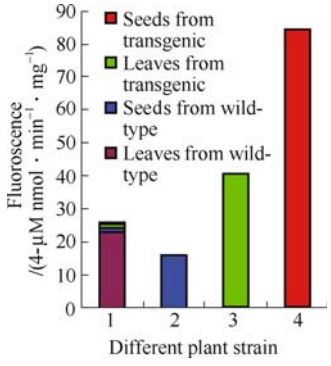


1: Wild-type plants as a negative control; 2–5: Transgenic plant strains; 6: PCR products of GUS gene as a positive control

Fig. 8 Southern Blot Analysis of the transgenic plants

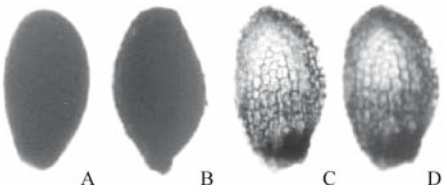
3.4 Seed-specific expression of GUS gene in the transgenic *Arabidopsis thaliana* plants

The fluorescence analysis of leaves and seeds from the transgenic *Arabidopsis thaliana* plants and wild-type strains indicated that fluorescence activities of the samples from the transgenic *Arabidopsis thaliana* plants were higher than those of the wild-type strains (Fig. 9), while the histochemistry analysis of the seeds from the transgenic *Arabidopsis thaliana* plants and wild-type strains further confirmed the above mentioned result, as shown in Fig. 10. Therefore, it is concluded that the GUS gene in the transgenic *Arabidopsis thaliana* plants was expressed in a seed-specific manner under the control of the seed-specific promoter fragment BCSP666.



1: Leaves of wild type plants; 2: Seeds of wild type plants; 3: Leaves of transgenic plants; 4: Seeds of transgenic plants

Fig. 9 Fluorescence analysis of transgenic plants and wild-type plants



A and B: Transgenic plant strains; C and D: Wild-type plants as negative control

Fig. 10 Histochemical analysis of the transgenic plants

4 Discussion

The β-conglycinin α-subunit gene promoter BCSP666 was cloned from the total DNA of Soybean Jilin 43 in this study. Sequencing analysis indicated that the fragment BCSP666 contains several kinds of multi-copy of seed-specific promoter motifs, such as 5 copies of RY repeat motifs, 4 copies of AGCCCA motifs and TACACAT motifs, two copies of ACGT motifs and E-box, and one copy of A/T rich motif. These kinds of multi-copy of seed-specific promoter motifs in fragment BCSP666 may be responsible for its seed-specific promoter activity. It is interesting to find that there are two copies of TATA box and CAAT box in the promoter fragment BCSP666, as shown in Fig. 4. And they seldom happen in other promoter sequences and their functions are still under study.

TAIL PCR protocol was successfully applied to extend the promoter fragment BCSP489 in this study. Although the extended fragment was not long enough, it provided us with a simple and convenient way to clone the unknown sequence adjacent to a known sequence fragment. And it is especially useful for cloning the unknown promoter fragment because it is laborious and consumptive to clone promoters via genomic library construction, or linker PCR (Siebert et al., 1995), or inverse PCR (Ochman et al., 1988). For TAIL PCR, it just requires to synthesize three gene-specific primers and 3 to 10 arbitrary degenerated primers, and the final results can be achieved in 6–8 h. Therefore, TAIL PCR possesses the nice cloning characters of convenience, quickness, efficiency, and specificity. The extended fragment of this study was only 217 bp. The possible reason was the quantity of the arbitrary degenerate primers (only three primers), and it was improved by increasing the quantity of the arbitrary degenerate primers. According to some related reports (Liu et al., 1998), the products of the TAIL PCR can be reached to 0.2–2 kb. Therefore, it comes to the conclusion that TAIL PCR is an effective and convenient protocol to clone unknown promoter fragments.

Acknowledgements The authors wish to acknowledge the financial support provided by the National Natural Science Foundation of China (No. 39870020) and Key Foundation of Tianjin (No. 013802511).

References

Beachy R N, Chen Z L, Horsch R B, Rogers S G, Hoffmann N J, Fraley R T (1985). Accumulation and assembly of soybean β-conglycinin in seeds of transformed petunia plants. The EMBO Journal, 4(12): 3047–3053

Beachy R N, Jarvis N P, Barton K A (1981). Biosynthesis of subunits of the soybean 7S storage protein. J Mol Appl Genet, 1(1): 19–27

Cahoon E B, Carlson T J, Ripp K G, Schweiger B J, Cook G A, Hall S E, Kinney A J (1999). Biosynthetic origin of conjugated double bonds: production of fatty acid components of high-value drying oils in transgenic soybean embryos. Proc Natl Acad Sci USA, 96(22): 12935–12940

Cahoon E B, Marillia E F, Stecca K L, Hall S E, Taylor D C, Kinney A J (2000). Production of fatty acid components of Meadow foam oil in somatic soybean embryos. Plant Physiol, 124: 243–251

Cahoon E B, Ripp K G, Hall S E, Kinney A J (2001). Formation of conjugated Δ^8,Δ^{10}-double bonds by Δ^{12}-oleic-acid desaturase-related enzymes. The J Bio Chem, 276(4): 2637–2643

Chamberland S, Daile N, Bernier F (1992). The legumin boxes and the 3′ part of a soybean β-conglycinin promoter are involved in seed gene expression in transgenic tobacco plants. Plant Molecular Biology, 19: 937–949

Chen Z L, Schuler M A, Beachy R N (1986). Functional analysis of regulatory elements in a plant embryo-specific gene. Proc Natl Acad Sci USA, 83: 8560–8564

Clough S J, Bent A F (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. The Plant Journal, 16(6): 735–743

Hofen R, Willmitzer L (1988). Storage of competent cells for *Agrobacterium* transformation. Nucleic Acids Research, 16: 9877

Jefferson R A (1987). Assaying chimeric genes in plants: The GUS gene fusion system. Plant Molecular Biology Reporter, 5(4): 387–405

Josefsson L G, Lenman M, Ericson M L, Rask L (1987). Structure of a gene encoding the 1.7S storage protein, Napin from *Brassica napus*. The Journal of Biological Chemistry, 262: 1196–1208

Kawagoe Y, Murai N (1992). Four distinct nuclear proteins recognize *in vitro* the proximal promoter of the bean seed storage protein P-phaseolin gene conferring spatial and temporal control. Plant J, 2: 927–936

Liu X Y, Tian S Z, Qin G F, Shen R Y (1997). An improved method for extracting DNA from plants and microorganisms using SDS-CTAB. Journal of Beijing Forestry University, 19(3):101–104 (in Chinese)

Liu Y G, Huang N (1998). Efficient amplification of insert end sequences from Bacterial Artificial Chromosome clones by thermal asymmertric interlaced PCR. Plant Molecular Biology Reporter, 16: 175–181

Liu Y G, Mitsukawa N, Oosumi T, Whittier R F (1995). Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junction by thermal asymmertric interlaced PCR. The Plant J, 8(3): 457–463

Ochman H, Gerber A S, Hartl D L (1998). Genetic applications of an inverse polymerase chain reaction. Genetics, 120: 621–623

Siebert P D, Chenchik A, Kellogg D E, Lukyanov K A, Lukyanov S A (1995). An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Research, 23: 1087–1088

Stalberg K, Ellerstrom M, Ezcurra I, Ablov S, Rask L (1996). Disruption of an overlapping E-box/ABRE motif abolished high transcription of the nap A storage protein promoter in transgenic *Brassica napus* seeds. Planta, 199: 515–519

Vincentz M, Leite A, Neshich G, Vriend G, Mattar C, Barros L, Weinberg D, de Almeida E R, de Carvalho M P, Aragao F, Gander E S (1997). ACGT and vicilin core sequences in a promoter domain required for seed-specific expression of a 2S storage protein gene are recognized by the opaque-2 regulatory protein. Plant Molecular Biology, 34: 879–889

Yoshino M, Kanazawa A, Tsutsumi K I, Nakamura I, Shimamoto Y (2001). Structure and characterization of the gene encoding alpha subunit of soybean beta-conglycinin. Genes Genet Syst, 76(2): 99–105