Gene 553 (2014) 17–23

Contents lists available at ScienceDirect

Gene

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

Seed-specific expression of seven *Arabidopsis* promoters

Hee-Jeong Jeong [a,b], Jun Young Choi [a], Hyun Young Shin [a], Jung-Myung Bae [a], Jeong Sheop Shin [a,*]

[a] Division of Life Sciences, Korea University, Seoul 136-701, Republic of Korea
[b] Department of Plant Molecular Systems Biotechnology & Crop Biotech Institute, Kyung Hee University, Yongin 446-701, Republic of Korea

ARTICLE INFO

Article history:
Received 25 August 2014
Received in revised form 16 September 2014
Accepted 23 September 2014
Available online 26 September 2014

Keywords:
ATH1 microarray
Seed tissue-specific
Promoters
Arabidopsis thaliana
Gene expression

ABSTRACT

Seeds contain storage compounds, from various carbohydrates to proteins and lipids, which are synthesized during seed development. For the purposes of many plant researches or commercial applications, developing promoter systems expressing specifically in seeds or in particular constituents or tissues/compartments of seeds are indispensable. To screen genes dominantly or specifically expressed in seed tissues, we analyzed *Arabidopsis* ATH1 microarray data open to the public. Thirty-two candidate genes were selected and their expressions in seed tissues were confirmed by RT-PCR. Finally, seven genes were selected for promoter analysis. The promoters of seven genes were cloned into pBI101 vector and transformed into *Arabidopsis* to assay histochemical β-glucuronidase (GUS) activity. We found that *Pro-at3g03230* promoter drove GUS expression in a chalazal endosperm, *Pro-at4g27530:GUS* expressed in both chalazal endosperm and embryo, *Pro-at4g31830* accelerated GUS expression both in radicle and procambium, *Pro-at5g10120* and *Pro-at5g16460* drove GUS expression uniquely in embryo, *Pro-at5g53100:GUS* expressed only in endosperm, and *Pro-at5g54000* promoted GUS expression in both embryo and inner integument. These promoters can be used for expressing any genes in specific seed tissues for practical application.

1. Introduction

Plant molecular technology makes it possible to finely regulate a specific gene, while traditional breeding-based technology can be used to control this gene in the genomic unit. The molecular approach to the manipulation of genes will therefore become the leading-edge strategy in the next generation-agricultural industry. In plant transformation, the expression of a transgene in an expression vector is regulated by a specific promoter that has been cloned into that vector at a location preceding the gene. This tissue- or developmental stage-specific promoter can maximize the effect of transgenic expression without affecting other tissues during growth and development.

Plant seeds are multifunctional in terms of the benefits they provide the plant which produces them, providing nourishment of the embryo, dispersal to a new location, and dormancy under unfavorable conditions. They are also the main source of nutrients for humans and animals. Important aims of plant studies are therefore to increase seed size and number and to decrease seed shattering and the length of the dormancy period, as such modifications can facilitate increases in crop yield, improvements in plant systems for the molecular farming of pharmaceutical proteins, and modifications in the composition and content of carbohydrates or lipids in seeds. To this end, the development of promoter systems which are expressed specifically in seeds or in particular constituents or tissues/compartments of seeds is indispensable.

Angiosperm (flowering plant) seeds consist of three genetically distinct components, i.e., the endosperm, embryo, and seed coat (testa). Three types of endosperm (a nutrient storage tissue) can be distinguished – the micropylar, chalazal and peripheral endosperms – especially in *Arabidopsis*. During seed maturation and subsequent germination, the testa ruptures at the micropylar endosperm. The chalazal endosperm, found at the non-micropylar end of the seed, is located at the base of an ovule bearing an embryo sac surrounded by integuments. Both the micropylar and chalazal endosperms are involved in the nutrient flow to an embryo. Over-proliferation of those tissues/compartments has been shown to result in increased nutrient flow to the embryo (Scott et al., 1998), suggesting that efficient seed-engineering could be attained by developing the appropriate micropylar endosperm- or chalazal endosperm-specific promoter. Endosperm-specific promoters have been studied and their respective value in crop production or industry reported. Hu et al. (2011) found that the *PzsS3a* promoter drives the preferentially stronger expression of a luciferase reporter gene in the endosperm than the embryo, but not in the leaf or root. The endosperm-specific *LPAAT* gene promoter isolated from coconut has been reported to induce high expression of the *GUS* reporter gene in the endosperm of transgenic rice seeds (Xu et al, 2010). A *cis*-acting element, ESP, contributes to high-level endosperm-specific expression in a globulin promoter (Vickers et al., 2006). Barley endosperm-specific

Abbreviations: Arabidopsis, *Arabidopsis thaliana*; dNTP, deoxyribonucleoside triphosphate; RT-PCR, reverse transcription-polymerase chain reaction; GUS, β-glucuronidase; NCBI, National Center for Biotechnology Information; GEO, Gene Expression Omnibus; MEV, MultiExperiment Viewer; UTR, untranslated region.

* Corresponding author.
E-mail address: jsshin@korea.ac.kr (J.S. Shin).

http://dx.doi.org/10.1016/j.gene.2014.09.051
0378-1119/

hordein promoters have been reported to drive the long-term stability of transgene expression in transgenic barley (Choi et al., 2003).

A seed plant embryo consists of precursor tissues for the shoot (plumule) and the root (radicle) and one or more cotyledons, and will eventually grow into a new plant. The seed coat plays pivotal roles in seed dormancy, germination, and protection against pathogens and chemical/mechanical damage. Various constituent-specific promoters have been reported. The *Arabidopsis At4g12960* promoter (*AtGILTpro*) functions as a seed coat outer integument-specific promoter in *Brassica napus* (Wu et al, 2011). *Arabidopsis* promoter of the *DIRIGENT PROTEIN1* (*DIP*) gene has been identified as a seed coat-specific promoter (Esfandiari et al., 2013). An embryo-specific promoter has also been applied in suppressing the soybean *MRP* gene to produce low levels of phytic acid seeds (Shi et al., 2007). All of these promoters can be used to finely regulate the expression of transgenes in genetically modified crops. Although many promoters which are specifically expressed in the three seed components have been identified, many more seed tissue/compartment-specific promoters are needed.

We report here seven novel *Arabidopsis* seed tissue/compartment-specific promoters. These were identified following an initial search of *Arabidopsis* Affymetrix 25K (ATH1) array data from the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) database of the National Center for Biotechnology Information (NCBI) (Barrett et al., 2011), which revealed 32 potential candidate genes, and subsequent examination of the specific expression of these 32 candidate genes in seeds by reverse transcription-polymerase chain reaction (RT-PCR). Based on the RT-PCR results, we selected seven genes, cloned their promoters, and analyzed β-glucuronidase (GUS) expression driven by these promoters in the different tissues/compartments of seeds.

2. Materials and methods

2.1. Meta-expression analysis

Anatomical expression data were obtained using five Affymetrix *Arabidopsis* ATH1 microarray datasets (GSE5630, GSE5631, GSE5632, GSE5633, and GSE5634) deposited in the GEO database at the NCBI (Barrett et al., 2011). The Affymetrix microarray data were normalized by the MAS 5.0 method implemented in the R package *affy* and the normalized values converted to $\log_2$ scale (Bolstad et al., 2003). A heat map was generated with the log-transformed data by using MultiExperiment Viewer (MEV; http://www.tm4.org/mev/).

2.2. Plant material and growth conditions

Arabidopsis plants used in this study were of the ecotype Columbia (Col-0). All *Arabidopsis* plants were grown in soil (Sunshine Mix #5; Sun Gro Horticulture, Vancouver, BC, Canada) at 22 °C and 50% relative humidity in a growth chamber under the long-day condition (16/8 h, light/dark) after 4 days at 4 °C in the dark.

2.3. RNA extraction and RT-PCR

Total RNA was isolated using a NucleoSpin® RNA Plant kit (Macherey-Nage, Düren, Germany). Total RNA (5 μg) was reverse transcribed with a Transcriptor First Strand cDNA Synthesis kit (Roche Applied Science, Indianapolis, IN) using anchored oligo$(dT)_{18}$ primers. The total PCR reaction volume (20 μl) contained 1 μl of the cDNA and 0.5 μl of each oligonucleotide primer (10 pmol). The PCR amplification regimen consisted of an initial denaturation for 5 min at 95 °C, followed by 30 cycles of 20 s at 94 °C, 20 s at 55 °C, and 30 s at 72 °C, with a final extension for 5 min at 72 °C. The RT-PCR primers are listed in Supplementary Table S1. *Arabidopsis EF1a* DNA was amplified for 22 cycles with primers (5′-GCACTGTCATTGATGCTCC-3′ and 5′-GTCAAGAGCCTC AAGGAGAG-3′) as an internal equal loading control.

2.4. Construction of the promoter:GUS vector and transformation into Arabidopsis

To generate the promoter:GUS constructs, we first amplified the 645-, 728-, 493-, 1434-, 882-, 552-, and 716-bp promoter regions of *Pro-at3g03230*, *Pro-at4g27530*, *Pro-at4g31830*, *Pro-at5g10120*, *Pro-at5g16460*, *Pro-at5g53100*, and *Pro-at5g54000*, respectively, from *Arabidopsis* genomic DNA by PCR using the primers listed in Supplementary Table S1. These fragments were then digested with *Hin*dIII, *Xba*I, *Bam*HI, or *Sal*I (Supplementary Fig. S2) and then cloned into the corresponding site of the pBI101 binary vector (Clontech, Mountain View, CA). The constructs were introduced into *Arabidopsis* plants using *Agrobacterium tumefaciens* GV3101 (Koncz and Schell, 1986) by the floral dip method (Clough and Bent, 1998). Transgenic lines were selected on a solid MS medium (Duchefa Biochemie, Haarlem, The Netherlands) containing 100 mg L^{-1} carbenicillin and either 50 mg L^{-1} kanamycin or 50 mg L^{-1} hygromycin and then screened for the presence of the transgene by PCR. The PCR amplification regimen consisted of an initial denaturation at 95 °C for 5 min, followed by 35 cycles of 95 °C for 20 s, 55 °C for 20 s, and 72 °C for 1 min, with a final 7-min extension at 72 °C. Anatomical patterns of promoter activity were analyzed in T_2 plants.

2.5. Histochemical GUS assay

For histochemical GUS staining, fresh samples from various tissues were incubated overnight in X-Gluc solution [1 mM 5-bromo-4-chloro-3-indoly-β-D-glucuronide, 100 mM sodium phosphate (pH 7.0), 10 mM Na2EDTA, 0.5 mM potassium ferricyanide, 0.5 mM potassium ferrocyanide, 0.1% (v/v) Triton X-100] at 37 °C. The staining buffer was then carefully removed, and the samples were washed three times (2-h washes) with 70% ethanol and stored for 16 h in 100% ethanol. The assay was slightly modified for transformed *Arabidopsis* seeds which were stained with X-Gluc solution for 24 h at 37 °C and fixed in a solution of ethanol:acetic acid (9:1, v/v) for 24 h at 25 °C after the X-Gluc solution had been removed. The samples were then washed twice with 90% ethanol (30-min washes) at 25 °C and once with 1 ml 30% glycerol containing 2.5 g chloral hydrate for 24 h at 25 °C. Tissues were observed and photographed by microscopy (model Olympus BX-51; Olympus Corp., Tokyo, Japan).

3. Results

3.1. Identification of genes specifically expressed in the seed using the Arabidopsis microarray database

To identify genes specifically expressed in the developing or mature seeds, we used Affymetrix 25 k (ATH1) array data containing approximately 22,000 *Arabidopsis* genes downloaded from the GEO database of the NCBI. The five Affymetrix array datasets GSE5630, GSE5631, GSE5632, GSE5633, and GSE5634, which consisted of full gene arrays expressed in five different organs (leaf, stem, root, flower, silique, and seed), were used. The initial search was for candidate genes exhibiting seed-preferential expression patterns, which led to the selection of 32 candidate genes and the subsequent analysis of their expression patterns in six different tissues (flower, leaf, stem, root, developing seed and silique, and mature seed). Based on the ATH1 microarray data, we categorized the 32 genes into two groups (I and II in Fig. 1) depending on whether the pattern was expressed generally in the whole plant tissues or expressed specifically in the silique and seed. Group I was then divided into two subgroups based on relative expression in the whole plant tissues: subgroup 1 (*At1g68220*, *At1g61140*, and *At1g17510*) and subgroup 2 (*At2g44070* and *At1g62520*), with subgroup 1 having higher expression than subgroup 2. Based on the ATH1 microarray dataset GSE5634, which contains gene expression information on silique and seed tissues, we subdivided group II into four subgroups, based on