

Identification of molecular mechanism controlling *P34* gene expression in soybean

Sung Cheol Koo · Jun Su Seo · Min Jung Park · Hyun Min Cho · Mi Suk Park ·
Cheol Woo Choi · Wook-Hun Jung · Kyoung Hee Lee · Byung Jun Jin · Seok Hyeon Kim ·
Sang In Shim · Jung-Sung Chung · Jong Il Chung · Min Chul Kim

Received: 6 September 2012 / Accepted: 17 October 2012 / Published online: 4 November 2012
© Korean Society for Plant Biotechnology and Springer Japan 2012

Abstract *P34* has long been known as one of major allergenic proteins in soybean. Recently, two low-*P34* soybean accessions, PI603570A and PI567476, were identified. In order to understand the molecular mechanism of a significant reduction of *P34* protein level in PI567476, we performed comparative analysis of the *P34* cDNAs and genomic sequences from low-*P34* and normal soybean accessions. While no sequence variation was detected in *P34* coding regions of the soybean accessions tested, the *P34* gene from PI567476 contains ‘ATGT’ 4-bp insertion in front of the start codon. We also found the length polymorphism of ‘TA’ repeats in the *P34* promoter regions among accessions. Northern and western blot analysis revealed that, even though mRNA levels of *P34* are similar in low-*P34* and normal soybean accessions, lower accumulation of *P34* protein is only detected in the low-*P34*

accession, suggesting that the genetic polymorphisms in *P34* promoters have a more significant effect on translation efficiency than transcription of the *P34* gene. Transient expression analysis showed that the *P34* promoter of PI567476 has significantly reduced activity compared to that of normal accession, cv. Clark. In addition, the chimeric promoter generated by introducing the 4-bp insertion in front of the start codon of the *P34* gene from normal Clark accession showed markedly reduced activity. The results suggest that lower accumulation of *P34* protein in low-*P34* accession is largely due to the reduced translation efficiency caused by the 4-bp insertion in the *P34* promoter, thus providing direct evidence for the molecular mechanism to explain the regulation of *P34* gene expression in various soybean accessions.

Keywords Allergen · Nucleotide insertion · *P34* · Polymorphism · TA repeats · Soybean

S. C. Koo and J. S. Seo contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s11816-012-0267-7) contains supplementary material, which is available to authorized users.

S. C. Koo
Department of Functional Crop, National Institute of Crop
Science, Rural Development Administration,
Miryang 627-803, Korea

J. S. Seo · H. M. Cho · W.-H. Jung · K. H. Lee ·
B. J. Jin · J. I. Chung · M. C. Kim (✉)
Division of Applied Life Science (BK21 Program), Research
Institute of Life Science, Gyeongsang National University,
Jinju 660-701, Korea
e-mail: mckim@gnu.ac.kr

M. J. Park · M. S. Park · C. W. Choi · S. H. Kim ·
S. I. Shim · J.-S. Chung · J. I. Chung · M. C. Kim
Department of Agronomy, Gyeongsang National University,
Jinju 660-701, Korea

Introduction

Soybean is a major source of edible oil and protein for human and animal nutrients. However, allergenic proteins that cause allergies to sensitive individuals limit its extensive use in the food industry. Recently, many soybean allergenic proteins have been identified and studied. In total, 38 soybean proteins have been reported as allergens in the Allergen Online database (<http://www.allergenonline.org>). These soybean allergens fall into five groups: the seed storage proteins Gly m 6 (glycinin G1, G2, G3, G4, and G5), Gly m 5 (β-conglycinin), Gly m TI (Kunitz trypsin inhibitor), Gly m Bd 30K (P34), and Gly m Bd 28K (Houston et al. 2011; L’Hocine and Boye 2007). These groups have conserved structural features related to biological activity.

Gly m Bd 30K is a relatively low-level seed protein, occupying <1 % of the total seed protein, but is consistently present in most soybean accessions evaluated (Joseph et al. 2006; Xu et al. 2007). Gly m Bd 30K is regarded as a major soybean allergen because more than 65 % of soy-sensitive patients react to this protein by IgE-binding studies (Helm et al. 2000; Ogawa et al. 1991). Gly m Bd 30K was characterized and identified as the soybean seed 34-kDa vacuolar protein, P34, which is initially synthesized as a larger precursor protein and is posttranslationally processed in seed (Kalinski et al. 1990, 1992). Sequence comparison and alignment indicate that P34 is a thiol proteinase of the papain family (Kalinski et al. 1990).

In view of food safety and the wide use of soybean in the food industry, identification of soybean accessions lacking soy allergens is very important. Recently, only two P34-low accessions, PI603570A and PI567476, had been identified through extensive screening by using about 16,000 soybean accessions from the USDA germplasm collections (Joseph et al. 2006). Previously, the comparative analysis of *P34* cDNAs from low-P34 accessions and wild-type accession (cv. Century) showed that six nucleotide changes resulting in four amino acid alterations were found in *P34* cDNAs from low-P34 accessions, suggesting that the significant reduction of P34 protein in low-P34 accessions might result from producing unstable P34 protein by four amino acid changes. In addition, Bilyeu et al. (2009) has reported that the *P34* gene from low-P34 accession contains a 'ATGT' 4-bp insertion in front of the start codon compared to that of cv. Williams 82, suggesting that the reduced level of the P34 protein in low-P34 accessions may be due to inefficient translation initiation. However, so far, there is no direct evidence to explain molecularly whether the significant reduction of the P34 protein level in P34-low soybean accessions is caused by four amino acid substitutions or the 4-bp insertion.

In the present study, we describe that *P34* promoters of low-P34 accessions contain two polymorphisms compared to normal soybean accessions, different numbers of TA repeats, and a 4-bp insertion in its promoter region, and furthermore, we provide molecular evidence that the 4-bp insertion is functionally distinguishable between low-P34 and normal accessions. Although mRNA expression level was similar to other soybean accessions tested, the significant reduction of the P34 protein was detected in P34-low accessions (PI567476). The reported gene analysis also showed the same result as the protein expression analysis. Especially, transient expression assay with chimeric P34 promoter shows that the 4-bp insertion is the major factor inhibiting P34 protein expression. Taken together, the results suggest that, although two polymorphisms have little effect on transcriptional regulation of the *P34* gene, regulation of the translation efficiency of the *P34* gene is

mainly mediated by the 4-bp insertion polymorphism, but also by the number of TA repeats in soybean accessions.

Materials and methods

Plant materials

The soybean accessions, PI567476, Clark, Jinpum2 (yellow soybean), and Gaechuk1 (black soybean) were used in this study. Soybean plants were grown under natural light condition in a greenhouse. Mature seeds were harvested, ground in liquid nitrogen, and stored at -80°C until used.

P34 cDNA cloning and northern blotting

Total RNA was isolated from mature soybean seeds by using Trizol reagent (Invitrogen). RNA samples were treated with DNase I at 37°C for 30 min. cDNA synthesis was performed using the SuperScript II Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions. *P34* cDNAs from various soybean accessions were amplified with gene-specific primers: forward primer: 5'-ATG GGT TTC CTT GTG TTG CTT C-3', and reverse primer: 5'-TCA AAG AGG AGA GTG ATC AAC TCT TCG-3'. PCR products were subcloned into pBluescriptII SK vector and sequenced.

For northern blot analysis, 15 μm of the total RNA extracted from seeds under various developmental stages was subjected to electrophoresis through a 1.5 % (w/v) formaldehyde agarose gel, transferred onto a nylon membrane (Hybond-N⁺; Amersham), and hybridized with a ³²P-labeled full-length *P34* cDNA probe. Hybridization was carried out at 65°C for 18 h in Church buffer (1 % BSA, 1 mM EDTA, 0.5 M NaHPO₄ pH 7.2, 7 % SDS). The membrane was finally washed in $0.1\times$ SSC and 0.1 % SDS at 50°C , and exposed to an X-ray film (Kodak).

Isolation of genomic DNA and PCR

Genomic DNA was extracted from soybean leaves as previously described (Murray and Thompson 1980). PCR reaction was carried out in 50 μl of reaction volume with 100 ng of genomic DNA as a template and specific primer sets, which were designed according to the P34 genomic sequence previously reported (GenBank accession number, AB013289): GFI forward: 5'-CTG GTC ATA AAG TGG GAG TAT GTG G-3', and GFI reverse: 5'-TAC ATG CCT CAT GAC ACC TAA GCA C-3'; GFII forward: 5'-TTC TTC ACT AAG TGT GCT TAG GTG T-3', and GFII reverse: 5'-CTC CTT TAA TGC AAT CTT CTC AAG G-3'; GFIII forward: 5'-TGA GAA GAT TGC ATT AAA GGA GGT C-3', and GFIII reverse: 5'-AAT GAT GAA

TCG TTG ATA AGC AGC G-3'; GFIV forward: 5'-TCT AGC AGT CTA GCT TAT GAG GCT C-3', and GFIV reverse: 5'-TGG AAC TAG AAG AGA GAC CTA AGA G-3'; GFV forward: 5'-ATG GGT TTC CTT GTG TTG CTT C-3', and GFV reverse: 5'-TGC GAC ACA TCC TTG GGA GCT TGC-3'; GFVI forward: 5'-AAG TTT GCT GAC ATC ACT CCT CAA G-3', and GFVI reverse: 5'-TGA CAT TAT TAG AGT TTC ATA TCC G-3'; GFVII forward: 5'-CAG AGC TAA AGA GGG TAG ATG CAA AGC-3', and GFVII reverse: 5'-TCA AAG AGG AGA GTG ATC AAC TCT TCG-3'. PCR products were subcloned into pBluescriptII SK vector and sequenced. The TA repeat region was amplified with gene-specific primers: forward primer: 5'-TAA AGG TGT AAC GTA TTT GTC AGG G-3', and reverse primer: 5'-TGA TGA ACT AAA TAA ACC AAG TTA GTT CTG-3'. PCR products were resolved by 2.5 % agarose gel electrophoresis.

Expression of recombinant P34 protein in *E. coli*, production of polyclonal antibody, and western blot analysis

The 743 bp of *P34* cDNA (nucleotide positions from 397 to 1,140 for Clark accession) corresponding to mature P34 protein (Joseph et al. 2006) was amplified by PCR with forward and reverse primers containing *EcoRI* and *SmaI* restriction sites, respectively: forward primer: 5'-AC GAA TTC GAC CAT CCA CCT GCA TCA TGG-3', and reverse primer: 5'-CT CCC GGG TCA AAG AGG AGA GTG ATC AAC TCT-3'. The PCR products were subcloned into pBluescriptII SK vector and confirmed by sequencing. The resulting clones were digested with *EcoRI/SmaI* or *EcoRI/NotI* and subcloned into same sites of pGEX-5X-1 or pET-28a vector, respectively. Expression of recombinant P34 protein in *E. coli*, production of polyclonal antibody, and western blot analysis were performed by described previously (Koo et al. 2011).

Transient expression assay using *Arabidopsis* protoplast

To analyze the promoter activity, we amplified 1,720 bp of promoter region of the *P34* gene containing TA repeat region and the start codon from Clark (nucleotide positions from -1,716 to 4 for Clark accession) and the corresponding regions from PI567476, respectively, by using specific primer sets: forward primer containing *KpnI* site for both Clark and PI567476 accessions: 5'-cgg tac cTA AAG GTG TAA CGT ATT TGT CAG GG-3', and reverse primer containing *BglIII* site for Clark accession: 5'-cag atc taC CAT AAC TTG GTG GAA GAA TTT TAT G-3', and reverse primer for PI567476 accession: 5'-cag atc taC CAT ACA TAA CTT GGT GGA AGA ATT TTA TG-3'. To

generate chimeric promoter containing the ATGT 4-bp insertion in front of the start codon of the *P34* promoter of Clark accession, we performed PCR using genomic DNA isolated from the Clark accession as a template with specific primers: forward primer: 5'-cgg tac cTA AAG GTG TAA CGT ATT TGT CAG GG-3', and reverse primer: 5'-cag atc taC CAT ACA TAA CTT GGT GGA AGA ATT TTA TG-3'. The PCR products were subcloned into the pBluescriptII SK vector and confirmed by sequencing. The resulting clones were digested with *KpnI* and *BglIII* and the fragments were subcloned into the same sites of the pCambia1301 vector, respectively.

Transient expression assay using *Arabidopsis* protoplast was performed as previously described (Choi et al. 2005). A *35S::LUC* (luciferase cDNA) construct was used as an internal control for normalizing the variations in transfection efficiency.

Results and discussion

Cloning of *P34* cDNAs from soybean accessions

Previously, two low-P34 soybean accessions (PI603570A and PI567476) have been identified from the screening of soybean germplasm. *P34* cDNAs of the two low-P34 accessions have six nucleotide substitutions, which are predicted to cause four amino acid changes, compared to the previously reported *P34* cDNA sequence from cv. Century (Joseph et al. 2006). Based on these sequence variations, it was suggested that alteration of amino acids may affect the stability of the P34 protein, resulting in less accumulation of this protein in seeds of those soybean accessions. In order to investigate the effect of amino acid alterations on P34 stability at the molecular level, we cloned *P34* cDNAs from both low-P34 (PI567476) and normal (cv. Clark) soybean accessions. Double-stranded cDNAs were synthesized using total RNAs extracted from maturing seeds where the *P34* gene is exclusively expressed (Kalinski et al. 1992). The 1,140 bp of *P34* cDNAs corresponding to the entire open reading frame were amplified from PI567476 and Clark soybean accessions, by RT-PCR using gene-specific primers, and the nucleotide sequences were determined (Fig. S1). As reported, *P34* cDNA from PI567476 (GenBank accession number, HQ153031) has six nucleotide substitutions compared to the previously reported *P34* cDNA sequence from cv. Century (GenBank accession number, J05560), and it is identical to that (GenBank accession number, DQ324851) of another low-P34 accession (PI603570A). But, interestingly, the *P34* cDNA sequence from the Clark accession (GenBank accession number, HQ153032) is identical not to that from Century but to those from low-P34 accessions (Fig. 1; Fig. S1). We performed two more rounds of independent

RT-PCR and sequencing with the same results (data not shown). To confirm these contradictory results, we determined *P34* cDNA sequences from another Korean soybean accession, Jinpum2 (yellow soybean; GenBank accession

number, HQ153033) and Gaechuk1 (black soybean; GenBank accession number, HQ153034) and obtained the same results as observed in Clark (Fig. 1; Fig. S1). The results reveal that there is no sequence variation in *P34* coding

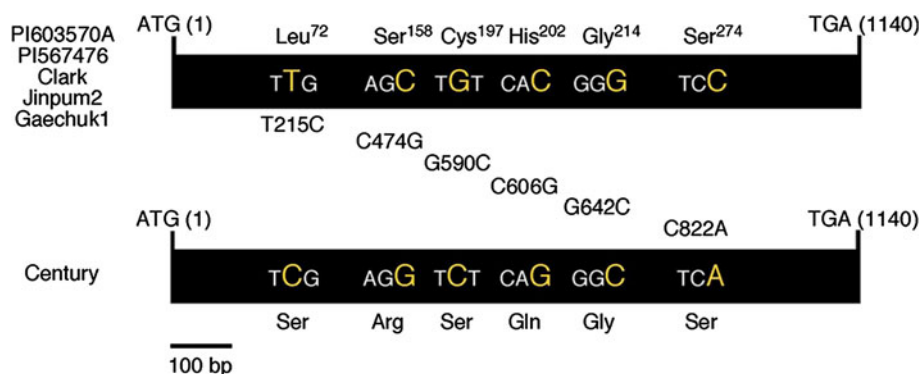


Fig. 1 Schematic representation of structural features of *P34* cDNAs identified from various soybean accessions. The only six variable nucleotides (highlighted in yellow) and accompanying codons are indicated. The Century soybean accession has six nucleotide substitutions compared to other accessions analyzed, including PI603570A, PI567476, Clark, Jinpum2, and Gaechuk1. The nucleotide sequences

and positions of the six variable nucleotides are indicated between two types of cDNAs. Four nucleotide changes (T215C, C474G, G590C, and C606G) resulted in amino acid alterations (L72S, S158R, C197S, and H202Q, respectively), but two substitutions (G642C and C822A) resulted in silent mutations at amino acid positions 214 and 274, respectively. Scale bar 100 bp (color figure online)

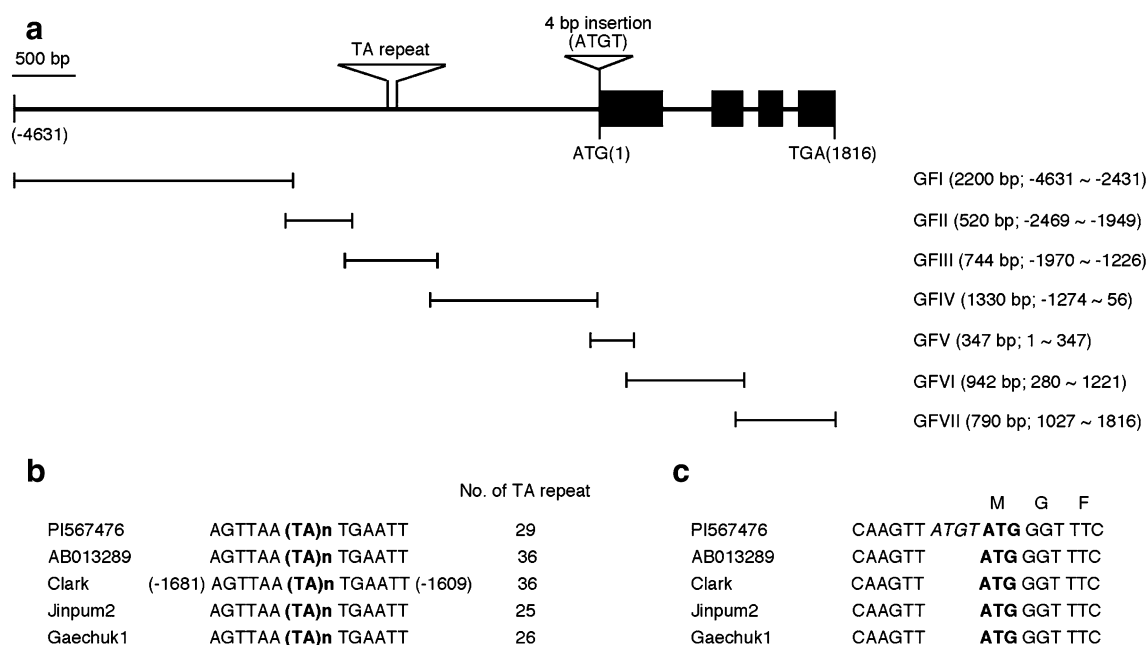


Fig. 2 Genomic organization of the *P34* gene for low-*P34* accession, PI567476, and normal soybean accessions, Clark, Jinpum2, and Gaechuk1. **a** Schematic representation of genomic structure of *P34* gene. Exons are represented as boxes and introns and promoter region are represented as lines. Two polymorphisms, TA repeat and 4-bp insertion, are indicated. To determine the nucleotide sequences of the *P34* gene, 6,447 bp of the *P34* genomic region containing both 4,631 bp of promoter region and 1,816 bp of the entire coding region were divided into seven parts (genomic fragment I–VII, designated GFI–GFVII, respectively), and each fragment was amplified by PCR with specific primer sets. The lengths and nucleotide positions of each genomic fragment, which correspond to the *P34* gene of the Clark

accession, are indicated in parentheses. Scale bar 500 bp. **b** Comparison of nucleotide sequences corresponding to TA repeat regions of the *P34* gene from PI567476, Clark, Jinpum2, Gaechuk1, and the previously reported AB013289. The nucleotide positions corresponding to the *P34* gene from the Clark accession are indicated in parentheses. The numbers of TA repeat, designated (TA)_n, are on the right of the sequences. **c** The polymorphism at the start codon of the *P34* gene. The ATGT 4-bp insertion in the *P34* gene from PI567476 is indicated by italics. The start codons are highlighted by bold and the amino acid sequences for first three codons are indicated above the corresponding nucleotide sequences

regions between the low-P34 accessions and the normal accessions analyzed, indicating that nucleotide polymorphisms in *P34* coding sequence is not the major cause determining the stability of the P34 protein, and the six nucleotide substitutions previously reported may be typical of Century accession.

Identification of *P34* genomic clones

To confirm the result from the cDNA sequence analysis, we determined the nucleotide sequences of *P34* genomic clones from soybean accessions, including PI567476, Clark, Jinpum2, and Gaechuk1. Based on the sequence of the previously reported *P34* genomic clone (GenBank accession number, AB013289), 6,447 bp of *P34* genomic region containing both 4,631 bp of promoter region and 1,816 bp of the entire coding region, which consisted of four exons separated by three introns, was divided into seven parts (genomic fragments I–VII, designated GFI–GFVII, respectively), and each fragment was amplified by PCR with specific primer sets (Fig. 2a) (data not shown for genomic PCR). Then, we determined the nucleotide sequences of *P34* genomic clones amplified from PI567476, Clark, Jinpum2, and Gaechuk1 accessions. Analysis of their genomic sequences in comparison with that of the previously reported *P34* gene (AB013289) indicated that there is no sequence variation in 1,816 bp of the entire coding regions among the soybean accessions analyzed, confirming the results from the *P34* cDNA sequence analysis (Fig. S2).

In contrast, two polymorphisms were identified within 4,631 bp of upstream regions from the start codon among the accessions. The first polymorphism was observed in the number of the dinucleotide TA repeat: *P34* genes both previously identified (AB013289) and identified from Clark have 36 TA repeats, and *P34* genes from PI567476, Jinpum2, and Gaechuk1 accessions contain 29, 25, and 26 TA repeats in their promoter regions, respectively (Fig. 2b; Fig. S2). It has been demonstrated that alternating purine–pyrimidine dinucleotide repeats such as TA repeats may affect the formation of alternative DNA including Z-DNA (Nordheim and Rich 1983; Wells 1988). and. moreover. TA repeats in the promoter region may be implicated in the regulation of gene expression (Liu et al. 2001; Tremblay et al. 2003). It has also been suggested that the length polymorphism in TA repeats, which is most likely caused by slipped-strand mispairing during DNA replication, is responsible for generating intra- and inter-species variability (Madsen et al. 1993; Pausova et al. 1995).

A recent report indicated that low-P34 alleles, PI567476 and PI603570A, contain the ATGT 4-bp insertion in front of the start codon of the *P34* gene compared to the normal P34 allele, cv. Williams 82, and that this insertion may cause less

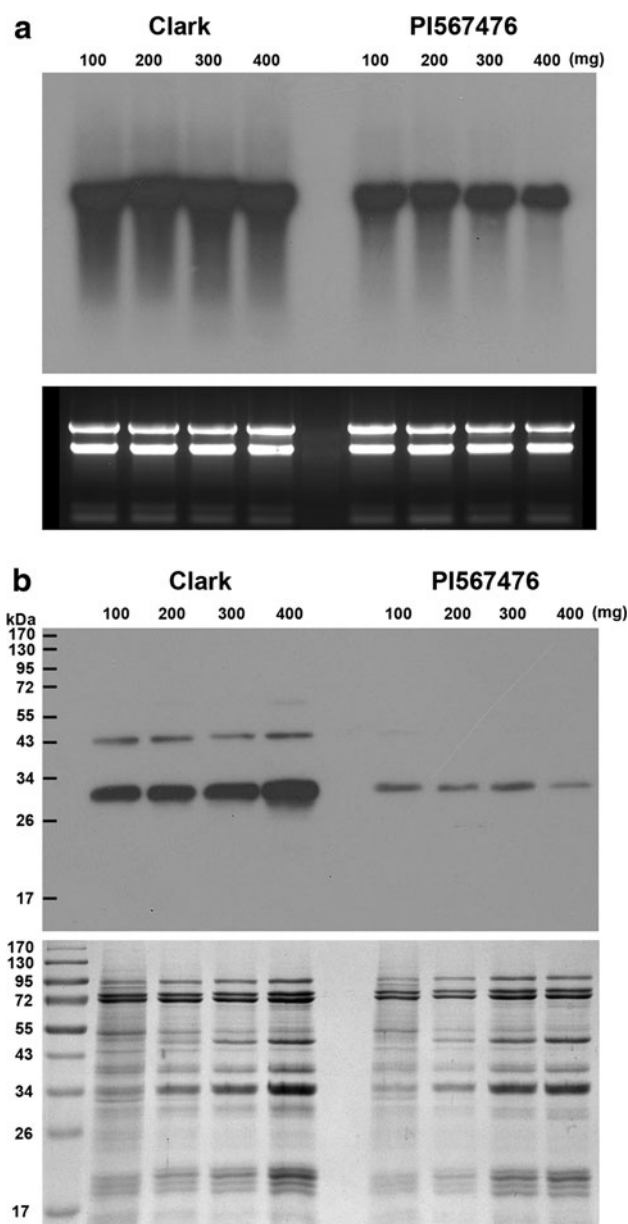


Fig. 3 Expression of *P34* gene in the seeds of various soybean accessions. **a** Northern blot analysis of the expression levels of *P34* gene in soybean accessions. Fifteen micrograms of total RNA extracted from dry seeds of PI567476 and Clark were used for northern blot analysis. **b** Western blot analysis of P34 protein level in dry seeds of PI567476 and Clark. Three micrograms of crude seed protein was used for western blot analysis

accumulation of the P34 protein in low-P34 alleles by inhibiting its efficient translation (Bilyeu et al. 2009). Consistently, we also observed this 4-bp insertion only in the low-P34 allele, PI567476, but not in other accessions we analyzed (Fig. 2c; Fig. S2). Taken together, the results indicate that the reduced level of the P34 protein in the low-P34 accession is mainly affecting the genetic polymorphisms found in the *P34* promoter region rather than the protein instability caused by amino acid substitution.

Comparison of *P34* expression between low-*P34* and normal soybean accessions

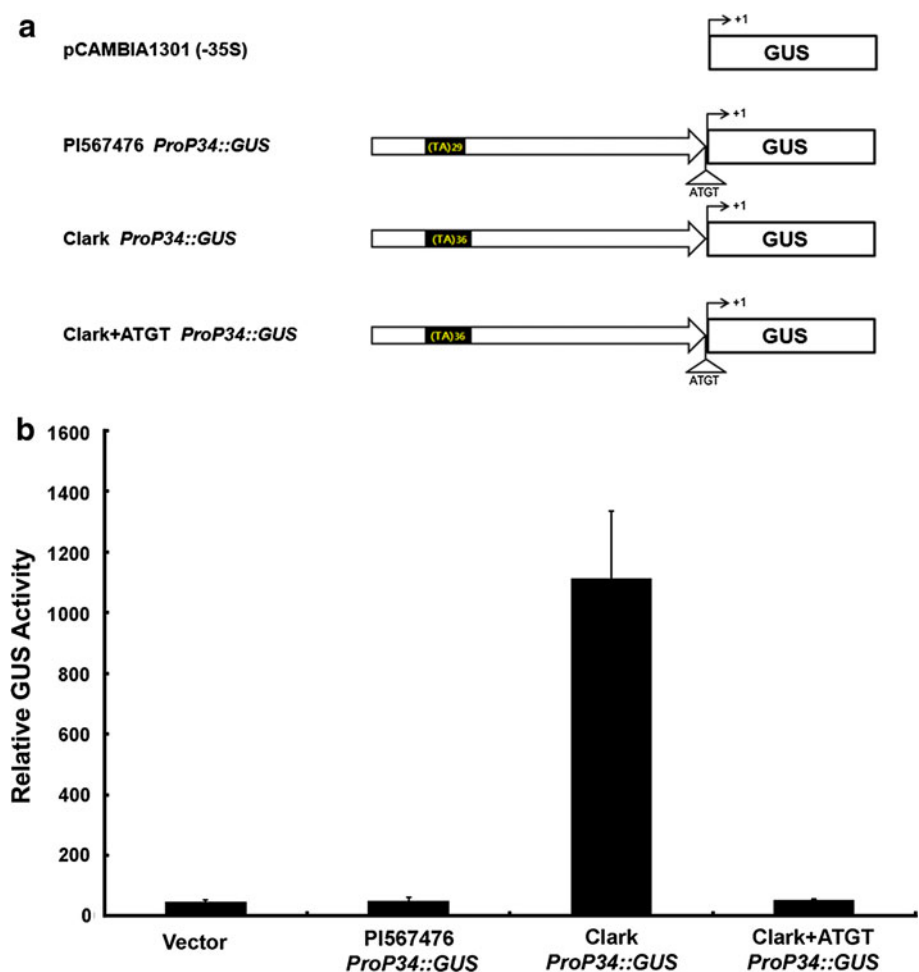
Since it has been known that the TA repeat sequence in the promoter region may have either a positive or a negative effect on the downstream gene expression (Liu et al. 2001; Nordheim and Rich 1983; Tremblay et al. 2003; Wells 1988) and it has also been proposed that the less accumulation of the *P34* protein in low-*P34* alleles may result from the 4-bp insertion at the start codon, we examined the effects of these two polymorphisms on *P34* gene expression by analyzing both mRNA and protein levels of *P34* in the seeds of low-*P34* (PI567476), and Clark as a representative of normal accessions. First, we analyzed the *P34* mRNA level in the seeds of both accessions by northern blotting. As shown in Fig. 3a, the *P34* mRNA level in the low-*P34* accession was slightly lower than that in the normal accession at the different seed developing stages. This indicates that two different polymorphisms, TA repeat and 4-bp insertion, have a weak effect on the transcriptional level of *P34*. Next, we also examined the *P34* protein level in both low-*P34* and normal accessions. In contrast to the transcript level, the *P34* protein level was greatly

reduced in the low-*P34* accession compared to the *P34* normal accession (Fig. 3b). These results suggest that the two polymorphisms showing in PI567476 exhibit a more significant effect on the translation than on the transcription of the *P34* gene.

Different *P34* promoter activity between low-*P34* and normal accessions

Polymorphisms within the promoter could determine gene expression levels (Bonner et al. 2012; Li et al. 2012; Ren et al. 2009). In order to further investigate the effects of the TA repeat region and the 4-bp insertion in the regulation of *P34* expression, we cloned about 1.7 kb promoter regions from PI567476 and Clark accessions and fused these fragments to the GUS reporter gene to generate *ProP34::GUS* constructs (Fig. 4a). Each construct was introduced into *Arabidopsis* protoplast for transient expression assay. As shown in Fig. 4b, the results showed that *P34* promoter activity of the low-*P34* allele, PI567476, was very low compared to that of the normal accession, Clark. These different *P34* promoter activities between low-*P34* and normal accession are consistent with the

Fig. 4 Comparison of relative *P34* promoter activity of low-*P34* and normal soybean accessions. **a** Schematic representation of Promoter::GUS fusion constructs used for the transient expression assays in *Arabidopsis* protoplasts. **b** Fluorometric GUS activity assays in *Arabidopsis* protoplasts transiently expressing the Promoter::GUS fusion constructs presented in (a), with the *CaMV35S* promoter-luciferase vector, pCambia1381. GUS activity was calculated relative to LUC activity. Values are mean $\pm$ SD from three independent biological repeats



different protein expression levels shown in Fig. 3. To dissect the effects of TA-repeats and 4-bp insertion on *P34* expression and also to determine whether 4-bp insertion can effect on translation efficiency of *P34* gene expression proposed previously (Bilyeu et al. 2009), we generated a chimeric *P34* promoter by adding the ATGT 4-bp into the 3'-end of the *P34* promoter Clark accession (Fig. 4a). The addition of ATGT 4-bp was sufficient to suppress the Clark *P34* promoter activity to a similar level as the PI567476 *P34* promoter activity. This result indicated that the 4-bp insertion plays a major role in the regulation of the *P34* gene expression, especially by affecting the translation efficiency of the *P34* gene.

In summary, we conducted comparative analyses of genetic variation in cDNA and genomic sequences, mRNA and protein levels, and promoter activity of *P34* genes identified from low- and normal-*P34* soybean accessions to identify the natural allelic variation that contributes to reduced *P34* expression in low-*P34* accessions compared to normal soybean accessions. From the comparative analyses, we found no sequence variation in the *P34* coding regions of the soybean accessions analyzed, but there were two polymorphisms in the *P34* promoter regions, such as length variation of 'TA' repeats among soybean accessions and ATGT 4-bp insertion in front of the start codon of the *P34* gene from the low-*P34* accession. The results indicate that the reduced level of the *P34* protein in the low-*P34* accession is not attributable to the protein instability caused by amino acid substitution, but to the genetic polymorphisms found in the *P34* promoter region. Of the two polymorphisms found in the promoter, the ATGT 4-bp insertion in front of the start codon was significantly associated with the reduced accumulation of *P34* protein in the seeds of the low-*P34* soybean accession, suggesting that the 4-bp insertion could be the major determinant controlling *P34* expression. This study cannot only help in understanding the regulatory mechanism of *P34* expression in various soybean accessions but can also provide valuable information for developing molecular markers for breeding low-*P34* elite soybean cultivars.

Acknowledgments This work was supported by a grant from the Next-Generation BioGreen 21 Program (SSAC, Grant#: PJ008025), Rural Development Administration, Republic of Korea. H.M.C., W.-H.J., K.H.L., and B.J.J. were supported by scholarships from the BK21 Program of the Ministry of Education, Science and Technology of Korea.

References

- Bilyeu K, Ren C, Nguyen HT, Herman E, Sleper DA (2009) Association of a four-basepair insertion in the *P34* gene with the low-allergen trait in soybean. *Plant Genome* 2:141–148
- Bonner SM, Pietropaolo SL, Fan Y, Chang Y, Sethupathy P, Morran MP, Beems M, Giannoukakis N, Trucco G, Palumbo MO, Solimena M, Pugliese A, Polychronakos C, Trucco M, Pietropaolo M (2012) Sequence variation in promoter of *Ica1* gene, which encodes protein implicated in type 1 diabetes, causes transcription factor autoimmune regulator (AIRE) to increase its binding and down-regulate expression. *J Biol Chem* 287:17882–17893
- Choi MS, Kim MC, Yoo JH, Moon BC, Koo SC, Park BO, Lee JH, Koo YD, Han HJ, Lee SY, Chung WS, Lim CO, Cho MJ (2005) Isolation of a calmodulin-binding transcription factor from rice (*Oryza sativa* L.). *J Biol Chem* 280:40820–40831
- Helm RM, Cockrell G, Connaughton C, West CM, Herman E, Sampson HA, Bannon GA, Burks AW (2000) Mutational analysis of the IgE-binding epitopes of P34/Gly m Bd 30K. *J Allergy Clin Immunol* 105:378–384
- Houston NL, Lee DG, Stevenson SE, Ladics GS, Bannon GA, McClain S, Privalle L, Stagg N, Herouet-Guicheney C, MacIntosh SC, Thelen JJ (2011) Quantitation of soybean allergens using tandem mass spectrometry. *J Proteome Res* 10:763–773
- Joseph LM, Hymowitz T, Schmidt MA, Herman EM (2006) Evaluation of glycine germplasm for nulls of the immunodominant allergen P34/Gly m Bd 30k. *Crop Sci* 46:1755–1763
- Kalinski A, Weisemann JM, Matthews BF, Herman EM (1990) Molecular cloning of a protein associated with soybean seed oil bodies that is similar to thiol proteases of the papain family. *J Biol Chem* 265:13843–13848
- Kalinski A, Melroy DL, Dwivedi RS, Herman EM (1992) A soybean vacuolar protein (P34) related to thiol proteases is synthesized as a glycoprotein precursor during seed maturation. *J Biol Chem* 267:12068–12076
- Koo SC, Bae DW, Seo JS, Park KM, Choi MS, Kim SH, Shim SI, Kim KM, Chung JI, Kim MC (2011) Proteomic analysis of seed storage proteins in low allergenic soybean accession. *J Korean Soc Appl Biol Chem* 54:332–339
- L'Hocine L, Boye JI (2007) Allergenicity of soybean: new developments in identification of allergenic proteins, cross-reactivities and hypoallergenization technologies. *Crit Rev Food Sci Nutr* 47:127–143
- Li Q, Yang X, Xu S, Cai Y, Zhang D, Han Y, Li L, Zhang Z, Gao S, Li J, Yan J (2012) Genome-wide association studies identified three independent polymorphisms associated with alpha-tocopherol content in maize kernels. *PLoS ONE* 7:e36807
- Liu R, Liu H, Chen X, Kirby M, Brown PO, Zhao K (2001) Regulation of CSF1 promoter by the SWI/SNF-like BAF complex. *Cell* 106:309–318
- Madsen CS, Ghivizzani SC, Hauswirth WW (1993) In vivo and in vitro evidence for slipped mispairing in mammalian mitochondria. *Proc Natl Acad Sci USA* 90:7671–7675
- Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* 8:4321–4325
- Nordheim A, Rich A (1983) Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences. *Nature* 303:674–679
- Ogawa T, Bando N, Tsuji H, Okajima H, Nishikawa K, Sasaoka K (1991) Investigation of the IgE-binding proteins in soybeans by immunoblotting with the sera of the soybean-sensitive patients with atopic dermatitis. *J Nutr Sci Vitaminol* 37:555–565
- Pausova Z, Morgan K, Fujiwara TM, Hendy GN (1995) Evolution of a repeat sequence in the parathyroid hormone-related peptide gene in primates. *Mamm Genome Off J Int Mamm Genome Soc* 6:408–414
- Ren Y, Yang S, Xu S, Gao M, Huang W, Gao T, Fang Q, Quan C, Zhang C, Sun L, Liang Y, Han J, Wang Z, Zhang F, Zhou Y, Liu J, Zhang X (2009) Genetic variation of promoter sequence

- modulates XBP1 expression and genetic risk for vitiligo. *PLoS Genet* 5:e1000523
- Tremblay J, Hum DH, Sanchez R, Dumas P, Pravenec M, Krenova D, Kren V, Kunes J, Pausova Z, Gossard F, Hamet P (2003) TA repeat variation, Npr1 expression, and blood pressure: impact of the Ace locus. *Hypertension* 41:16–24
- Wells RD (1988) Unusual DNA structures. *J Biol Chem* 263:1095–1098
- Xu C, Caperna TJ, Garrett WM, Cregan P, Bae H, Luthria DL, Natarajan S (2007) Proteomic analysis of the distribution of the major seed allergens in wild, landrace, ancestral, and modern soybean genotypes. *J Sci Food Agric* 87:2511–2518

Identification of Molecular Mechanism Controlling *P34* Gene Expression in Soybean

Century	ATGGGTTTCCTTGTTGCTTCTTTCTCCCTCTTAGGTCCTCTTCTAGTTCCAGCATA	60
PI603579	ATGGGTTTCCTTGTTGCTTCTTTCTCCCTCTTAGGTCCTCTTCTAGTTCCAGCATA	
PI567476	ATGGGTTTCCTTGTTGCTTCTTTCTCCCTCTTAGGTCCTCTTCTAGTTCCAGCATA	
Clark	ATGGGTTTCCTTGTTGCTTCTTTCTCCCTCTTAGGTCCTCTTCTAGTTCCAGCATA	
Jinpum2	ATGGGTTTCCTTGTTGCTTCTTTCTCCCTCTTAGGTCCTCTTCTAGTTCCAGCATA	
Gaechuk1	ATGGGTTTCCTTGTTGCTTCTTTCTCCCTCTTAGGTCCTCTTCTAGTTCCAGCATA	
	M G F L V L L L F S L L G L S S S S S I	(20)
Century	TCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTG	120
PI603579	TCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTG	
PI567476	TCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTG	
Clark	TCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTG	
Jinpum2	TCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTG	
Gaechuk1	TCAACTCATCGTTCCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTG	
	S T H R S I L D L D L T K F T T Q K Q V	(40)
Century	TCTTCACTGTTCCAACATATGGAAGAGTGAGCATGGACGTGTCTACCATAAACCACGAAGAA	180
PI603579	TCTTCACTGTTCCAACATATGGAAGAGTGAGCATGGACGTGTCTACCATAAACCACGAAGAA	
PI567476	TCTTCACTGTTCCAACATATGGAAGAGTGAGCATGGACGTGTCTACCATAAACCACGAAGAA	
Clark	TCTTCACTGTTCCAACATATGGAAGAGTGAGCATGGACGTGTCTACCATAAACCACGAAGAA	
Jinpum2	TCTTCACTGTTCCAACATATGGAAGAGTGAGCATGGACGTGTCTACCATAAACCACGAAGAA	
Gaechuk1	TCTTCACTGTTCCAACATATGGAAGAGTGAGCATGGACGTGTCTACCATAAACCACGAAGAA	
	S S L F Q L W K S E H G R V Y H N H E E	(60)
Century	GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTTGA	S
PI603579	GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTTGA	
PI567476	GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTTGA	
Clark	GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTTGA	
Jinpum2	GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTTGA	
Gaechuk1	GAGGCAAAGAGACTTGAGATTTTCAAGAATAACTTGA	
	E A K R L E I F K N N L N Y I R D M N A	(80)
Century	AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT	300
PI603579	AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT	
PI567476	AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT	
Clark	AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT	
Jinpum2	AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT	
Gaechuk1	AACAGAAAATCACCCCATTCTCATCGTTTAGGATTGAACAAGTTTGCTGACATCACTCCT	
	N R K S P H S H R L G L N K F A D I T P	(100)
Century	CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTGCGCAGCAAAATCAAAATG	360
PI603579	CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTGCGCAGCAAAATCAAAATG	
PI567476	CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTGCGCAGCAAAATCAAAATG	
Clark	CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTGCGCAGCAAAATCAAAATG	
Jinpum2	CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTGCGCAGCAAAATCAAAATG	
Gaechuk1	CAAGAGTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTGCGCAGCAAAATCAAAATG	
	Q E F S K K Y L Q A P K D V S Q Q I K M	(120)

Century GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT 420
 PI603579 GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT
 PI567476 GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT
 Clark GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT
 Jinpum2 GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT
 Gaechuk1 GCCAACAAGAAAATGAAGAAGGAACAATATTCTTGTGACCATCCACCTGCATCATGGGAT
 A N K K M K K E Q Y S C D H P P A S W D (140)

Century TGGAGGAAAAAAGGTGTGCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGTTGG 480
 PI603579 TGGAGGAAAAAAGGTGTGCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGTTGG
 PI567476 TGGAGGAAAAAAGGTGTGCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGTTGG
 Clark TGGAGGAAAAAAGGTGTGCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGTTGG
 Jinpum2 TGGAGGAAAAAAGGTGTGCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGTTGG
 Gaechuk1 TGGAGGAAAAAAGGTGTGCATCACCCAAGTAAAGTACCAAGGGGGCTGTGGAAGCGGTTGG
 W R K K G V I T Q V K Y Q G G C G S G W (160)

Century GCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTT 540
 PI603579 GCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTT
 PI567476 GCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTT
 Clark GCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTT
 Jinpum2 GCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTT
 Gaechuk1 GCGTTTTCTGCCACGGGAGCCATAGAAGCAGCACATGCAATAGCAACAGGAGACCTTGTT
 A F S A T G A I E A A H A I A T G D L V (180)

Century AGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGA 600
 PI603579 AGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGA
 PI567476 AGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGA
 Clark AGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGA
 Jinpum2 AGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGA
 Gaechuk1 AGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTCTTACAATGGA
 S L S E Q E L V D C V E E S E G C Y N G (200)

Century TGGCAGTATCAATCGTTTGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGAT 660
 PI603579 TGGCAGTATCAATCGTTTGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGAT
 PI567476 TGGCAGTATCAATCGTTTGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGAT
 Clark TGGCAGTATCAATCGTTTGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGAT
 Jinpum2 TGGCAGTATCAATCGTTTGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGAT
 Gaechuk1 TGGCAGTATCAATCGTTTGAATGGGTTTTAGAACATGGTGGCATTGCCACTGATGATGAT
 W H Y Q S F E W V L E H G G I A T D D D (220)

Century TATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACA 720
 PI603579 TATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACA
 PI567476 TATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACA
 Clark TATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACA
 Jinpum2 TATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACA
 Gaechuk1 TATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGATACAAGACAAGGTTACA
 Y P Y R A K E G R C K A N K I Q D K V T (240)

Century ATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAA 780
 PI603579 ATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAA
 PI567476 ATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAA
 Clark ATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAA
 Jinpum2 ATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAA
 Gaechuk1 ATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAA
 I D G Y E T L I M S D E S T E S E T E Q (260)

Century GCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCCATTGATGCAAAAAGATTTT 840
 PI603579 GCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCCATTGATGCAAAAAGATTTT
 PI567476 GCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCCATTGATGCAAAAAGATTTT
 Clark GCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCCATTGATGCAAAAAGATTTT
 Jinpum2 GCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCCATTGATGCAAAAAGATTTT
 Gaechuk1 GCGTTCTTAAGCGCCATCCTTGAGCAACCAATTAGTGTCTCCATTGATGCAAAAAGATTTT
 A F L S A I L E Q P I S V S I D A K D F (280)

Century CATTTATACACCGGGGGAATTTATGATGGAGAAAAGTGTACAAGTCCGTATGGGATTAAT 900
 PI603579 CATTTATACACCGGGGGAATTTATGATGGAGAAAAGTGTACAAGTCCGTATGGGATTAAT
 PI567476 CATTTATACACCGGGGGAATTTATGATGGAGAAAAGTGTACAAGTCCGTATGGGATTAAT
 Clark CATTTATACACCGGGGGAATTTATGATGGAGAAAAGTGTACAAGTCCGTATGGGATTAAT
 Jinpum2 CATTTATACACCGGGGGAATTTATGATGGAGAAAAGTGTACAAGTCCGTATGGGATTAAT
 Gaechuk1 CATTTATACACCGGGGGAATTTATGATGGAGAAAAGTGTACAAGTCCGTATGGGATTAAT
 H L Y T G G I Y D G E N C T S P Y G I N (300)

Century CACTTTGTTTTACTTGTGGGTTATGGTTACGCGGATGGTGTAGATTACTGGATAGCGAAA 960
 PI603579 CACTTTGTTTTACTTGTGGGTTATGGTTACGCGGATGGTGTAGATTACTGGATAGCGAAA

PI567476	CACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAA	
Clark	CACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAA	
Jinpum2	CACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAA	
Gaechuk1	CACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAA	
	H F V L L V G Y G S A D G V D Y W I A K	(320)
Century	AATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAACACGGGT	1020
PI603579	AATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAACACGGGT	
PI567476	AATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAACACGGGT	
Clark	AATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAACACGGGT	
Jinpum2	AATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAACACGGGT	
Gaechuk1	AATTCATGGGGAGAAGATTGGGGAGAAGATGGTTACATTTGGATCCAAAGAAACACGGGT	
	N S W G E D W G E D G Y I W I Q R N T G	(340)
Century	AATTTATTAGGAGTGTGTGGGATGAATTATTTTCGCTTCATACCCAACCAAGAGGAATCA	1080
PI603579	AATTTATTAGGAGTGTGTGGGATGAATTATTTTCGCTTCATACCCAACCAAGAGGAATCA	
PI567476	AATTTATTAGGAGTGTGTGGGATGAATTATTTTCGCTTCATACCCAACCAAGAGGAATCA	
Clark	AATTTATTAGGAGTGTGTGGGATGAATTATTTTCGCTTCATACCCAACCAAGAGGAATCA	
Jinpum2	AATTTATTAGGAGTGTGTGGGATGAATTATTTTCGCTTCATACCCAACCAAGAGGAATCA	
Gaechuk1	AATTTATTAGGAGTGTGTGGGATGAATTATTTTCGCTTCATACCCAACCAAGAGGAATCA	
	N L L G V C G M N Y F A S Y P T K E E S	(360)
Century	GAAACACTGGTGTCTGCTCGCGTTAAAGGTCATCGAAGAGTTGATCACTCTCCTCTTT TGA	1140
PI603579	GAAACACTGGTGTCTGCTCGCGTTAAAGGTCATCGAAGAGTTGATCACTCTCCTCTTT TGA	
PI567476	GAAACACTGGTGTCTGCTCGCGTTAAAGGTCATCGAAGAGTTGATCACTCTCCTCTTT TGA	
Clark	GAAACACTGGTGTCTGCTCGCGTTAAAGGTCATCGAAGAGTTGATCACTCTCCTCTTT TGA	
Jinpum2	GAAACACTGGTGTCTGCTCGCGTTAAAGGTCATCGAAGAGTTGATCACTCTCCTCTTT TGA	
Gaechuk1	GAAACACTGGTGTCTGCTCGCGTTAAAGGTCATCGAAGAGTTGATCACTCTCCTCTTT TGA	
	E T L V S A R V K G H R R V D H S P L *	(379)

Supplemental Fig. 1. Comparison of *P34* cDNA sequences identified from various soybean accessions. *P34* cDNA sequences determined from PI567476, Clark, Jinpum2 and Gaechuk1 were compared with those of previously reported *P34* cDNA from a low-*P34* accession (PI603570A, GenBank accession number; DQ324851) and a control accession (Century, GenBank accession number; J05560) by using Clustal W software. The variable nucleotide sequences are highlighted by shading and their accompanying codons are underlined. Four altered amino acids in Century accession by nucleotide substitutions are indicated above the corresponding codons. The numbers of nucleotides and deduced amino acids are on the right. Both start (ATG) and stop codons (TGA, also indicated by asterisk) are indicated by bold.

PI567476 CTGGTCATAAAGTGGGAGTATGTGGGATGAATTTTTTGGCTCTTGGCCAACCAAGAGAAATCAGCAACACTCGTGTCTGCTCGTGTAAAGCTGATCA -4522
AB013289 CTGGTCATAAAGTGGGAGTATGTGGGATGAATTTTTTGGCTCTTGGCCAACCAAGAGAAATCAGCAACACTCGTGTCTGCTCGTGTAAAGCTGATCA -4532
Clark CTGGTCATAAAGTGGGAGTATGTGGGATGAATTTTTTGGCTCTTGGCCAACCAAGAGAAATCAGCAACACTCGTGTCTGCTCGTGTAAAGCTGATCA -4532
Jinpum2 CTGGTCATAAAGTGGGAGTATGTGGGATGAATTTTTTGGCTCTTGGCCAACCAAGAGAAATCAGCAACACTCGTGTCTGCTCGTGTAAAGCTGATCA -4510
Gaechuck1 CTGGTCATAAAGTGGGAGTATGTGGGATGAATTTTTTGGCTCTTGGCCAACCAAGAGAAATCAGCAACACTCGTGTCTGCTCGTGTAAAGCTGATCA -4512

PI567476 AAGAGTGGATTACTCTCCTCTTTGAGCAGTGAAGGTTCAATTATAACGAGTGCTTGTTTTCTTAGAGACAAGCATCATATATTGTATATGATTTTGTGAC -4422
AB013289 AAGAGTGGATTACTCTCCTCTTTGAGCAGTGAAGGTTCAATTATAACGAGTGCTTGTTTTCTTAGAGACAAGCATCATATATTGTATATGATTTTGTGAC -4432
Clark AAGAGTGGATTACTCTCCTCTTTGAGCAGTGAAGGTTCAATTATAACGAGTGCTTGTTTTCTTAGAGACAAGCATCATATATTGTATATGATTTTGTGAC -4432
Jinpum2 AAGAGTGGATTACTCTCCTCTTTGAGCAGTGAAGGTTCAATTATAACGAGTGCTTGTTTTCTTAGAGACAAGCATCATATATTGTATATGATTTTGTGAC -4410
Gaechuck1 AAGAGTGGATTACTCTCCTCTTTGAGCAGTGAAGGTTCAATTATAACGAGTGCTTGTTTTCTTAGAGACAAGCATCATATATTGTATATGATTTTGTGAC -4412

PI567476 CATGCATGCCCCACGTAGTACTGACGTGAAAGCAAAAAATAAACACACAACATATATTTCTCAAAAAATAAAAAAGAACACTGCCACTGTTGCATGTG -4322
AB013289 CATGCATGCCCCACGTAGTACTGACGTGAAAGCAAAAAATAAACACACAACATATATTTCTCAAAAAATAAAAAAGAACACTGCCACTGTTGCATGTG -4332
Clark CATGCATGCCCCACGTAGTACTGACGTGAAAGCAAAAAATAAACACACAACATATATTTCTCAAAAAATAAAAAAGAACACTGCCACTGTTGCATGTG -4332
Jinpum2 CATGCATGCCCCACGTAGTACTGACGTGAAAGCAAAAAATAAACACACAACATATATTTCTCAAAAAATAAAAAAGAACACTGCCACTGTTGCATGTG -4310
Gaechuck1 CATGCATGCCCCACGTAGTACTGACGTGAAAGCAAAAAATAAACACACAACATATATTTCTCAAAAAATAAAAAAGAACACTGCCACTGTTGCATGTG -4312

PI567476 TTTTAATCAACATATATATTGCGAGATGGATTTTCCTTTGCTTCATGCCCTTACCTTGACTATGTGAAAGGAATAAATTAATGCCCTGTTTAATTAGTTTT -4222
AB013289 TTTTAATCAACATATATATTGCGAGATGGATTTTCCTTTGCTTCATGCCCTTACCTTGACTATGTGAAAGGAATAAATTAATGCCCTGTTTAATTAGTTTT -4232
Clark TTTTAATCAACATATATATTGCGAGATGGATTTTCCTTTGCTTCATGCCCTTACCTTGACTATGTGAAAGGAATAAATTAATGCCCTGTTTAATTAGTTTT -4232
Jinpum2 TTTTAATCAACATATATATTGCGAGATGGATTTTCCTTTGCTTCATGCCCTTACCTTGACTATGTGAAAGGAATAAATTAATGCCCTGTTTAATTAGTTTT -4210
Gaechuck1 TTTTAATCAACATATATATTGCGAGATGGATTTTCCTTTGCTTCATGCCCTTACCTTGACTATGTGAAAGGAATAAATTAATGCCCTGTTTAATTAGTTTT -4212

PI567476 TCCCATTAGCTAGCTAGCTAGCTACCGTCGAAAGGTTGATTTGTTTTAAAGACGGCACCCAATAATATGACAAGTACACATAAAAGCTTCAAGAGTCA -4122
AB013289 TCCCATTAGCTAGCTAGCTAGCTACCGTCGAAAGGTTGATTTGTTTTAAAGACGGCACCCAATAATATGACAAGTACACATAAAAGCTTCAAGAGTCA -4132
Clark TCCCATTAGCTAGCTAGCTAGCTACCGTCGAAAGGTTGATTTGTTTTAAAGACGGCACCCAATAATATGACAAGTACACATAAAAGCTTCAAGAGTCA -4132
Jinpum2 TCCCATTAGCTAGCTAGCTAGCTACCGTCGAAAGGTTGATTTGTTTTAAAGACGGCACCCAATAATATGACAAGTACACATAAAAGCTTCAAGAGTCA -4110
Gaechuck1 TCCCATTAGCTAGCTAGCTAGCTACCGTCGAAAGGTTGATTTGTTTTAAAGACGGCACCCAATAATATGACAAGTACACATAAAAGCTTCAAGAGTCA -4112

PI567476 AGCTTAGTCTGTAAGGGAGAGATCAAAATGACATGTAGCATTAACCTCTGATAGAGTAACAGCAAAATCCAAACTCTGAATACATCTTTATCATAATAAAAG -4022
AB013289 AGCTTAGTCTGTAAGGGAGAGATCAAAATGACATGTAGCATTAACCTCTGATAGAGTAACAGCAAAATCCAAACTCTGAATACATCTTTATCATAATAAAAG -4032
Clark AGCTTAGTCTGTAAGGGAGAGATCAAAATGACATGTAGCATTAACCTCTGATAGAGTAACAGCAAAATCCAAACTCTGAATACATCTTTATCATAATAAAAG -4032
Jinpum2 AGCTTAGTCTGTAAGGGAGAGATCAAAATGACATGTAGCATTAACCTCTGATAGAGTAACAGCAAAATCCAAACTCTGAATACATCTTTATCATAATAAAAG -4010
Gaechuck1 AGCTTAGTCTGTAAGGGAGAGATCAAAATGACATGTAGCATTAACCTCTGATAGAGTAACAGCAAAATCCAAACTCTGAATACATCTTTATCATAATAAAAG -4012

PI567476 TATGCAATTGAGCAGGAAAAATATTAGACATTTTTTTTCTATTTTTTTAATTCATTAAATTTATATAAATAATAAATTAAGATAATTAGAACCAATATATC -3922
AB013289 TATGCAATTGAGCAGGAAAAATATTAGACATTTTTTTTCTATTTTTTTAATTCATTAAATTTATATAAATAATAAATTAAGATAATTAGAACCAATATATC -3932
Clark TATGCAATTGAGCAGGAAAAATATTAGACATTTTTTTTCTATTTTTTTAATTCATTAAATTTATATAAATAATAAATTAAGATAATTAGAACCAATATATC -3932
Jinpum2 TATGCAATTGAGCAGGAAAAATATTAGACATTTTTTTTCTATTTTTTTAATTCATTAAATTTATATAAATAATAAATTAAGATAATTAGAACCAATATATC -3910
Gaechuck1 TATGCAATTGAGCAGGAAAAATATTAGACATTTTTTTTCTATTTTTTTAATTCATTAAATTTATATAAATAATAAATTAAGATAATTAGAACCAATATATC -3912

PI567476 TTCTAATTTGAAATAATTAATAATTATTATCAATATATATATGTTTTAATTAAGCTAATTACGATTATTTTATTACTTCAAATAGTCAATCTCTCA -3822
AB013289 TTCTAATTTGAAATAATTAATAATTATTATCAATATATATATGTTTTAATTAAGCTAATTACGATTATTTTATTACTTCAAATAGTCAATCTCTCA -3832
Clark TTCTAATTTGAAATAATTAATAATTATTATCAATATATATATGTTTTAATTAAGCTAATTACGATTATTTTATTACTTCAAATAGTCAATCTCTCA -3832
Jinpum2 TTCTAATTTGAAATAATTAATAATTATTATCAATATATATATGTTTTAATTAAGCTAATTACGATTATTTTATTACTTCAAATAGTCAATCTCTCA -3810
Gaechuck1 TTCTAATTTGAAATAATTAATAATTATTATCAATATATATATGTTTTAATTAAGCTAATTACGATTATTTTATTACTTCAAATAGTCAATCTCTCA -3812

PI567476 TTTCAAACCTTCATAACTAACCTAGCATCTCCCGATCGGAGAATGGCTAGTTGGTTAAATAATGTTAGTGATTGTTATAAACTCTGTTAATGTTTGAGTT -3722
AB013289 TTTCAAACCTTCATAACTAACCTAGCATCTCCCGATCGGAGAATGGCTAGTTGGTTAAATAATGTTAGTGATTGTTATAAACTCTGTTAATGTTTGAGTT -3732
Clark TTTCAAACCTTCATAACTAACCTAGCATCTCCCGATCGGAGAATGGCTAGTTGGTTAAATAATGTTAGTGATTGTTATAAACTCTGTTAATGTTTGAGTT -3732
Jinpum2 TTTCAAACCTTCATAACTAACCTAGCATCTCCCGATCGGAGAATGGCTAGTTGGTTAAATAATGTTAGTGATTGTTATAAACTCTGTTAATGTTTGAGTT -3710
Gaechuck1 TTTCAAACCTTCATAACTAACCTAGCATCTCCCGATCGGAGAATGGCTAGTTGGTTAAATAATGTTAGTGATTGTTATAAACTCTGTTAATGTTTGAGTT -3712

PI567476 TTATTACACTGGATAAAAAATCTAGCATTTCTCTCTTTCTCAGTATGATTAATAATACGATCCAGAGATATTATAAATAATCTTTTAAATGCTTTAATG -3622
AB013289 TTATTACACTGGATAAAAAATCTAGCATTTCTCTCTTTCTCAGTATGATTAATAATACGATCCAGAGATATTATAAATAATCTTTTAAATGCTTTAATG -3632
Clark TTATTACACTGGATAAAAAATCTAGCATTTCTCTCTTTCTCAGTATGATTAATAATACGATCCAGAGATATTATAAATAATCTTTTAAATGCTTTAATG -3632
Jinpum2 TTATTACACTGGATAAAAAATCTAGCATTTCTCTCTTTCTCAGTATGATTAATAATACGATCCAGAGATATTATAAATAATCTTTTAAATGCTTTAATG -3610
Gaechuck1 TTATTACACTGGATAAAAAATCTAGCATTTCTCTCTTTCTCAGTATGATTAATAATACGATCCAGAGATATTATAAATAATCTTTTAAATGCTTTAATG -3612

PI567476 ACTCAATACACGGCCAGTCAAGATCAGCCTTG6CCGTATAAACAACCCCTCATAACATGACATACAATTATAATGGAAAAATCTCATATAGCACAGTTATG -3522
AB013289 ACTCAATACACGGCCAGTCAAGATCAGCCTTG6CCGTATAAACAACCCCTCATAACATGACATACAATTATAATGGAAAAATCTCATATAGCACAGTTATG -3532
Clark ACTCAATACACGGCCAGTCAAGATCAGCCTTG6CCGTATAAACAACCCCTCATAACATGACATACAATTATAATGGAAAAATCTCATATAGCACAGTTATG -3532
Jinpum2 ACTCAATACACGGCCAGTCAAGATCAGCCTTG6CCGTATAAACAACCCCTCATAACATGACATACAATTATAATGGAAAAATCTCATATAGCACAGTTATG -3510
Gaechuck1 ACTCAATACACGGCCAGTCAAGATCAGCCTTG6CCGTATAAACAACCCCTCATAACATGACATACAATTATAATGGAAAAATCTCATATAGCACAGTTATG -3512

PI567476 AAGGGAAAAACATGGCACACAAAAGGTACTCTTTTTAACTATAAACGATTATGAAGTTTTTCATGGAGAATAGTATGTTGCTCTTTTAACTAACCTAGCAT -3422
AB013289 AAGGGAAAAACATGGCACACAAAAGGTACTCTTTTTAACTATAAACGATTATGAAGTTTTTCATGGAGAATAGTATGTTGCTCTTTTAACTAACCTAGCAT -3432
Clark AAGGGAAAAACATGGCACACAAAAGGTACTCTTTTTAACTATAAACGATTATGAAGTTTTTCATGGAGAATAGTATGTTGCTCTTTTAACTAACCTAGCAT -3432
Jinpum2 AAGGGAAAAACATGGCACACAAAAGGTACTCTTTTTAACTATAAACGATTATGAAGTTTTTCATGGAGAATAGTATGTTGCTCTTTTAACTAACCTAGCAT -3410
Gaechuck1 AAGGGAAAAACATGGCACACAAAAGGTACTCTTTTTAACTATAAACGATTATGAAGTTTTTCATGGAGAATAGTATGTTGCTCTTTTAACTAACCTAGCAT -3412

PI567476 TCTCCCAATAGGAGAATTAGTTGGTTAAAAAATGTTAATGATTGTTATAAACTGTTAATGTTTGAAGTTTTTATTACTATGGATAAAAAATCTAGCATTC -3322
AB013289 TCTCCCAATAGGAGAATTAGTTGGTTAAAAAATGTTAATGATTGTTATAAACTGTTAATGTTTGAAGTTTTTATTACTATGGATAAAAAATCTAGCATTC -3332
Clark TCTCCCAATAGGAGAATTAGTTGGTTAAAAAATGTTAATGATTGTTATAAACTGTTAATGTTTGAAGTTTTTATTACTATGGATAAAAAATCTAGCATTC -3332
Jinpum2 TCTCCCAATAGGAGAATTAGTTGGTTAAAAAATGTTAATGATTGTTATAAACTGTTAATGTTTGAAGTTTTTATTACTATGGATAAAAAATCTAGCATTC -3310
Gaechuck1 TCTCCCAATAGGAGAATTAGTTGGTTAAAAAATGTTAATGATTGTTATAAACTGTTAATGTTTGAAGTTTTTATTACTATGGATAAAAAATCTAGCATTC -3312

PI567476 TCTCTTTTCTCAGTATGATTAATAATACGATCCAGAAATTTTATAAAATTTTTTTTAAATGCTTAATGACTCAATACACGGCCAGTCAAGGTCACCT -3222
AB013289 TCTCTTTTCTCAGTATGATTAATAATACGATCCAGAAATTTTATAAAATTTTTTTTAAATGCTTAATGACTCAATACACGGCCAGTCAAGGTCACCT -3232
Clark TCTCTTTTCTCAGTATGATTAATAATACGATCCAGAAATTTTATAAAATTTTTTTTAAATGCTTAATGACTCAATACACGGCCAGTCAAGGTCACCT -3232
Jinpum2 TCTCTTTTCTCAGTATGATTAATAATACGATCCAGAAATTTTATAAAATTTTTTTTAAATGCTTAATGACTCAATACACGGCCAGTCAAGGTCACCT -3210
Gaechuck1 TCTCTTTTCTCAGTATGATTAATAATACGATCCAGAAATTTTATAAAATTTTTTTTAAATGCTTAATGACTCAATACACGGCCAGTCAAGGTCACCT -3212

PI567476 TGGTCGGATAAACCAACCTCATAACATGACATACAATTATAATGGAAAAATCTCATATAGCACAAATATGAAGGCAAAAACATGGCACACAAAAGGTACT -3122
AB013289 TGGTCGGATAAACCAACCTCATAACATGACATACAATTATAATGGAAAAATCTCATATAGCACAAATATGAAGGCAAAAACATGGCACACAAAAGGTACT -3132
Clark TGGTCGGATAAACCAACCTCATAACATGACATACAATTATAATGGAAAAATCTCATATAGCACAAATATGAAGGCAAAAACATGGCACACAAAAGGTACT -3132
Jinpum2 TGGTCGGATAAACCAACCTCATAACATGACATACAATTATAATGGAAAAATCTCATATAGCACAAATATGAAGGCAAAAACATGGCACACAAAAGGTACT -3110
Gaechuck1 TGGTCGGATAAACCAACCTCATAACATGACATACAATTATAATGGAAAAATCTCATATAGCACAAATATGAAGGCAAAAACATGGCACACAAAAGGTACT -3112

PI567476 TGTTTTAACTATAAACGATTATGAATTTTTCAAGGAAAAATAGCATGTTGCTCTTGATTCTCTTGAAGAACTTTTTAGGTACCTTTTACTAAGTTGGACGT -3022
AB013289 TGTTTTAACTATAAACGATTATGAATTTTTCAAGGAAAAATAGCATGTTGCTCTTGATTCTCTTGAAGAACTTTTTAGGTACCTTTTACTAAGTTGGACGT -3032
Clark TGTTTTAACTATAAACGATTATGAATTTTTCAAGGAAAAATAGCATGTTGCTCTTGATTCTCTTGAAGAACTTTTTAGGTACCTTTTACTAAGTTGGACGT -3032

[illegible]

PI567476 ATCAGACAAAAGTGCATGCAATCACTTGCAAGTAAATTAAGGATACTAATCTATTCCCTTCATTTATATGCTCCACTTTTATATAAAAAAATATACATT -936
AB013289 ATCAGACAAAAGTGCATGCAATCACTTGCAAGTAAATTAAGGATACTAATCTATTCCCTTCATTTATATGCTCCACTTTTATATAAAAAAATATACATT -932
Clark ATCAGACAAAAGTGCATGCAATCACTTGCAAGTAAATTAAGGATACTAATCTATTCCCTTCATTTATATGCTCCACTTTTATATAAAAAAATATACATT -932
Jinpum2 ATCAGACAAAAGTGCATGCAATCACTTGCAAGTAAATTAAGGATACTAATCTATTCCCTTCATTTATATGCTCCACTTTTATATAAAAAAATATACATT -932
Gaechuck1 ATCAGACAAAAGTGCATGCAATCACTTGCAAGTAAATTAAGGATACTAATCTATTCCCTTCATTTATATGCTCCACTTTTATATAAAAAAATATACATT -932

PI567476 ATTATATATGCAATTAATTAATTGTCAGTATTATGCTATTGGTTTTATGGCCCTGCTAAATAACCTAAATGAGTCTAACTATTGCATATGAATCAAAATGA -836
AB013289 ATTATATATGCAATTAATTAATTGTCAGTATTATGCTATTGGTTTTATGGCCCTGCTAAATAACCTAAATGAGTCTAACTATTGCATATGAATCAAAATGA -832
Clark ATTATATATGCAATTAATTAATTGTCAGTATTATGCTATTGGTTTTATGGCCCTGCTAAATAACCTAAATGAGTCTAACTATTGCATATGAATCAAAATGA -832
Jinpum2 ATTATATATGCAATTAATTAATTGTCAGTATTATGCTATTGGTTTTATGGCCCTGCTAAATAACCTAAATGAGTCTAACTATTGCATATGAATCAAAATGA -832
Gaechuck1 ATTATATATGCAATTAATTAATTGTCAGTATTATGCTATTGGTTTTATGGCCCTGCTAAATAACCTAAATGAGTCTAACTATTGCATATGAATCAAAATGA -832

PI567476 AGGAAGAATCATGATCTAAACCTGAGTACCCAATGCAATAAAATGCGTCCTATTACCTAAACCTCAAAACACACATTGCCATCGGACGTATAAATTAATGC -736
AB013289 AGGAAGAATCATGATCTAAACCTGAGTACCCAATGCAATAAAATGCGTCCTATTACCTAAACCTCAAAACACACATTGCCATCGGACGTATAAATTAATGC -732
Clark AGGAAGAATCATGATCTAAACCTGAGTACCCAATGCAATAAAATGCGTCCTATTACCTAAACCTCAAAACACACATTGCCATCGGACGTATAAATTAATGC -732
Jinpum2 AGGAAGAATCATGATCTAAACCTGAGTACCCAATGCAATAAAATGCGTCCTATTACCTAAACCTCAAAACACACATTGCCATCGGACGTATAAATTAATGC -732
Gaechuck1 AGGAAGAATCATGATCTAAACCTGAGTACCCAATGCAATAAAATGCGTCCTATTACCTAAACCTCAAAACACACATTGCCATCGGACGTATAAATTAATGC -732

PI567476 ATATAGATTATTTTGGAGAAAGAAAACATCAAAAGCTCTAAAACCTCTTTTAACTTTGAAATAAGCTGATAAAAAACGCTTTAAATCAACTGTGTGCTG -636
AB013289 ATATAGATTATTTTGGAGAAAGAAAACATCAAAAGCTCTAAAACCTCTTTTAACTTTGAAATAAGCTGATAAAAAACGCTTTAAATCAACTGTGTGCTG -632
Clark ATATAGATTATTTTGGAGAAAGAAAACATCAAAAGCTCTAAAACCTCTTTTAACTTTGAAATAAGCTGATAAAAAACGCTTTAAATCAACTGTGTGCTG -632
Jinpum2 ATATAGATTATTTTGGAGAAAGAAAACATCAAAAGCTCTAAAACCTCTTTTAACTTTGAAATAAGCTGATAAAAAACGCTTTAAATCAACTGTGTGCTG -632
Gaechuck1 ATATAGATTATTTTGGAGAAAGAAAACATCAAAAGCTCTAAAACCTCTTTTAACTTTGAAATAAGCTGATAAAAAACGCTTTAAATCAACTGTGTGCTG -632

PI567476 TATATAAGCTGCAATTTTTCACATTTTACCAAAACGAAACGAAGAATGGTAACAGTGAGGCAAAAATTTGAAAAATGTCCACTTTCACATTTCACATCAAAATTA -536
AB013289 TATATAAGCTGCAATTTTTCACATTTTACCAAAACGAAACGAAGAATGGTAACAGTGAGGCAAAAATTTGAAAAATGTCCACTTTCACATTTCACATCAAAATTA -532
Clark TATATAAGCTGCAATTTTTCACATTTTACCAAAACGAAACGAAGAATGGTAACAGTGAGGCAAAAATTTGAAAAATGTCCACTTTCACATTTCACATCAAAATTA -532
Jinpum2 TATATAAGCTGCAATTTTTCACATTTTACCAAAACGAAACGAAGAATGGTAACAGTGAGGCAAAAATTTGAAAAATGTCCACTTTCACATTTCACATCAAAATTA -532
Gaechuck1 TATATAAGCTGCAATTTTTCACATTTTACCAAAACGAAACGAAGAATGGTAACAGTGAGGCAAAAATTTGAAAAATGTCCACTTTCACATTTCACATCAAAATTA -532

PI567476 ATTACAACATAAATAAATCAATCGTGATTCAAGCAGTAATGAAAGTCGAAATCAGATAGAATATACACGTTTAAACATCAATTGAATTTTTTTTTTAAATG -436
AB013289 ATTACAACATAAATAAATCAATCGTGATTCAAGCAGTAATGAAAGTCGAAATCAGATAGAATATACACGTTTAAACATCAATTGAATTTTTTTTTTAAATG -432
Clark ATTACAACATAAATAAATCAATCGTGATTCAAGCAGTAATGAAAGTCGAAATCAGATAGAATATACACGTTTAAACATCAATTGAATTTTTTTTTTAAATG -432
Jinpum2 ATTACAACATAAATAAATCAATCGTGATTCAAGCAGTAATGAAAGTCGAAATCAGATAGAATATACACGTTTAAACATCAATTGAATTTTTTTTTTAAATG -432
Gaechuck1 ATTACAACATAAATAAATCAATCGTGATTCAAGCAGTAATGAAAGTCGAAATCAGATAGAATATACACGTTTAAACATCAATTGAATTTTTTTTTTAAATG -432

PI567476 GATATATACAAGTTTACTATTTTATATATAATGAAATTCATTTTGTGTTAGCACAAAACCTACAGAAAGAGATAAATTTTAAATAAAGAGAATATATC -336
AB013289 GATATATACAAGTTTACTATTTTATATATAATGAAATTCATTTTGTGTTAGCACAAAACCTACAGAAAGAGATAAATTTTAAATAAAGAGAATATATC -332
Clark GATATATACAAGTTTACTATTTTATATATAATGAAATTCATTTTGTGTTAGCACAAAACCTACAGAAAGAGATAAATTTTAAATAAAGAGAATATATC -332
Jinpum2 GATATATACAAGTTTACTATTTTATATATAATGAAATTCATTTTGTGTTAGCACAAAACCTACAGAAAGAGATAAATTTTAAATAAAGAGAATATATC -332
Gaechuck1 GATATATACAAGTTTACTATTTTATATATAATGAAATTCATTTTGTGTTAGCACAAAACCTACAGAAAGAGATAAATTTTAAATAAAGAGAATATATC -332

PI567476 CAATTTTATAATCCAAAATTAATCAAAATTAAGAAATATTGGCTAGATAGACCGGCTTTTTCACCTGCCCTGCTGGATAATGAAAATTCATATCAAAACAAAT -236
AB013289 CAATTTTATAATCCAAAATTAATCAAAATTAAGAAATATTGGCTAGATAGACCGGCTTTTTCACCTGCCCTGCTGGATAATGAAAATTCATATCAAAACAAAT -232
Clark CAATTTTATAATCCAAAATTAATCAAAATTAAGAAATATTGGCTAGATAGACCGGCTTTTTCACCTGCCCTGCTGGATAATGAAAATTCATATCAAAACAAAT -232
Jinpum2 CAATTTTATAATCCAAAATTAATCAAAATTAAGAAATATTGGCTAGATAGACCGGCTTTTTCACCTGCCCTGCTGGATAATGAAAATTCATATCAAAACAAAT -232
Gaechuck1 CAATTTTATAATCCAAAATTAATCAAAATTAAGAAATATTGGCTAGATAGACCGGCTTTTTCACCTGCCCTGCTGGATAATGAAAATTCATATCAAAACAAAT -232

PI567476 ACAGAAGTTCTAGTTTAAATAAAAAAGTTGGCAAACTGTCATCCCTGTTGGTTTTTAAAGCCAAATCACAATTCATACGTATCAGAAATTAATTTA -136
AB013289 ACAGAAGTTCTAGTTTAAATAAAAAAGTTGGCAAACTGTCATCCCTGTTGGTTTTTAAAGCCAAATCACAATTCATACGTATCAGAAATTAATTTA -132
Clark ACAGAAGTTCTAGTTTAAATAAAAAAGTTGGCAAACTGTCATCCCTGTTGGTTTTTAAAGCCAAATCACAATTCATACGTATCAGAAATTAATTTA -132
Jinpum2 ACAGAAGTTCTAGTTTAAATAAAAAAGTTGGCAAACTGTCATCCCTGTTGGTTTTTAAAGCCAAATCACAATTCATACGTATCAGAAATTAATTTA -132
Gaechuck1 ACAGAAGTTCTAGTTTAAATAAAAAAGTTGGCAAACTGTCATCCCTGTTGGTTTTTAAAGCCAAATCACAATTCATACGTATCAGAAATTAATTTA -132

PI567476 AACCAAATATATAGCTACGAGGGAACCTCTTCAGTCATTACTAGCTAGCTCAATCACTATATATACGACATGCTACAAGTGAAGTGACCATATCTTA -36
AB013289 AACCAAATATATAGCTACGAGGGAACCTCTTCAGTCATTACTAGCTAGCTCAATCACTATATATACGACATGCTACAAGTGAAGTGACCATATCTTA -32
Clark AACCAAATATATAGCTACGAGGGAACCTCTTCAGTCATTACTAGCTAGCTCAATCACTATATATACGACATGCTACAAGTGAAGTGACCATATCTTA -32
Jinpum2 AACCAAATATATAGCTACGAGGGAACCTCTTCAGTCATTACTAGCTAGCTCAATCACTATATATACGACATGCTACAAGTGAAGTGACCATATCTTA -32
Gaechuck1 AACCAAATATATAGCTACGAGGGAACCTCTTCAGTCATTACTAGCTAGCTCAATCACTATATATACGACATGCTACAAGTGAAGTGACCATATCTTA -32

PI567476 ATTTCAAATCATAAAATTTCTCCACCAAGTTATG67ATG66TTTTCCCTGTGTTGCTCTTTTCTCCCTCTTAGGTCCTCTCTTAGTTCACGACATATCAAC 65
AB013289 ATTTCAAATCATAAAATTTCTCCACCAAGTT - - - ATG66TTTTCCCTGTGTTGTTGCTCTTTTCTCCCTCTTAGGTCCTCTCTCTTAGTTCACGACATATCAAC 65
Clark ATTTCAAATCATAAAATTTCTCCACCAAGTT - - - ATG66TTTTCCCTGTGTTGTTGCTCTTTTCTCCCTCTTAGGTCCTCTCTCTTAGTTCACGACATATCAAC 65
Jinpum2 ATTTCAAATCATAAAATTTCTCCACCAAGTT - - - ATG66TTTTCCCTGTGTTGTTGCTCTTTTCTCCCTCTTAGGTCCTCTCTCTTAGTTCACGACATATCAAC 65
Gaechuck1 ATTTCAAATCATAAAATTTCTCCACCAAGTT - - - ATG66TTTTCCCTGTGTTGTTGCTCTTTTCTCCCTCTTAGGTCCTCTCTCTTAGTTCACGACATATCAAC 65

PI567476 TCATCGTTCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTGTCTTCACGTGTTCCAACATGGAAGAGTGAGCATGGACGTGTCTAC 165
AB013289 TCATCGTTCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTGTCTTCACGTGTTCCAACATGGAAGAGTGAGCATGGACGTGTCTAC 165
Clark TCATCGTTCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTGTCTTCACGTGTTCCAACATGGAAGAGTGAGCATGGACGTGTCTAC 165
Jinpum2 TCATCGTTCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTGTCTTCACGTGTTCCAACATGGAAGAGTGAGCATGGACGTGTCTAC 165
Gaechuck1 TCATCGTTCATATTGGACCTTGACCTAACCAAGTTTACCACACAGAAACAGGTGTCTTCACGTGTTCCAACATGGAAGAGTGAGCATGGACGTGTCTAC 165

PI567476 CATAACCACGAGAAGAGGCAAGAGACTTGAGATTTTCAAGAATAAATTTGAACATATATACAGGACATGAATGCAAAACAGAAAATCACCCCATTTCTCATC 265
AB013289 CATAACCACGAGAAGAGGCAAGAGACTTGAGATTTTCAAGAATAAATTTGAACATATATACAGGACATGAATGCAAAACAGAAAATCACCCCATTTCTCATC 265
Clark CATAACCACGAGAAGAGGCAAGAGACTTGAGATTTTCAAGAATAAATTTGAACATATATACAGGACATGAATGCAAAACAGAAAATCACCCCATTTCTCATC 265
Jinpum2 CATAACCACGAGAAGAGGCAAGAGACTTGAGATTTTCAAGAATAAATTTGAACATATATACAGGACATGAATGCAAAACAGAAAATCACCCCATTTCTCATC 265
Gaechuck1 CATAACCACGAGAAGAGGCAAGAGACTTGAGATTTTCAAGAATAAATTTGAACATATATACAGGACATGAATGCAAAACAGAAAATCACCCCATTTCTCATC 265

PI567476 GTTTAGGATTGAACAAGTTTGGTGACATCACTCCTCAAGAGTTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGACAGCAATCAAAATGGCCAA 365
AB013289 GTTTAGGATTGAACAAGTTTGGTGACATCACTCCTCAAGAGTTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGACAGCAATCAAAATGGCCAA 365
Clark GTTTAGGATTGAACAAGTTTGGTGACATCACTCCTCAAGAGTTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGACAGCAATCAAAATGGCCAA 365
Jinpum2 GTTTAGGATTGAACAAGTTTGGTGACATCACTCCTCAAGAGTTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGACAGCAATCAAAATGGCCAA 365
Gaechuck1 GTTTAGGATTGAACAAGTTTGGTGACATCACTCCTCAAGAGTTTCAGCAAAAAGTACTTGCAAGCTCCCAAGGATGTGTCGACAGCAATCAAAATGGCCAA 365

PI567476 CAAGAAAATGAAGAAGGAACAATATTCTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTTGCATCACCACCAAGTAAAGTACCAAGGGGGC 465
AB013289 CAAGAAAATGAAGAAGGAACAATATTCTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTTGCATCACCACCAAGTAAAGTACCAAGGGGGC 465
Clark CAAGAAAATGAAGAAGGAACAATATTCTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTTGCATCACCACCAAGTAAAGTACCAAGGGGGC 465
Jinpum2 CAAGAAAATGAAGAAGGAACAATATTCTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTTGCATCACCACCAAGTAAAGTACCAAGGGGGC 465
Gaechuck1 CAAGAAAATGAAGAAGGAACAATATTCTGTGACCATCCACCTGCATCATGGGATTGGAGGAAAAAAGGTTGCATCACCACCAAGTAAAGTACCAAGGGGGC 465

PI567476 TGTGgtatgtgaaaccatttaattgtttttactacgttaattaccactccttctatcttaaaataaacagttgtcttaattgttttatataaaatataaaat 565
AB013289 TGTGgtatgtgaaaccatttaattgtttttactacgttaattaccactccttctatcttaaaataaacagttgtcttaattgttttatataaaatataaaat 565
Clark TGTGgtatgtgaaaccatttaattgtttttactacgttaattaccactccttctatcttaaaataaacagttgtcttaattgttttatataaaatataaaat 565
Jinpum2 TGTGgtatgtgaaaccatttaattgtttttactacgttaattaccactccttctatcttaaaataaacagttgtcttaattgttttatataaaatataaaat 565
Gaechuck1 TGTGgtatgtgaaaccatttaattgtttttactacgttaattaccactccttctatcttaaaataaacagttgtcttaattgttttatataaaatataaaat 565

PI567476 aaaaataaaataaaataatattttttaccacactaaccttatttagtgaagagattaccatttaatactagattttaaanaactaaaaatatatttattagagtt 665
AB013289 aaaaataaaataaaataatattttttaccacactaaccttatttagtgaagagattaccatttaatactagattttaaanaactaaaaatatatttattagagtt 665
Clark aaaaataaaataaaataatattttttaccacactaaccttatttagtgaagagattaccatttaatactagattttaaanaactaaaaatatatttattagagtt 665
Jinpum2 aaaaataaaataaaataatattttttaccacactaaccttatttagtgaagagattaccatttaatactagattttaaanaactaaaaatatatttattagagtt 665
Gaechuck1 aaaaataaaataaaataatattttttaccacactaaccttatttagtgaagagattaccatttaatactagattttaaanaactaaaaatatatttattagagtt 665

PI567476 attggtgaaaaaaataatttaatttttagctgtaaaagttaaatgtgatattttattttagctacaatttttttttagtgtgacacttatttttggaaatgggaaaa 765
AB013289 attggtgaaaaaaataatttaatttttagctgtaaaagttaaatgtgatattttattttagctacaatttttttttagtgtgacacttatttttggaaatgggaaaa 765
Clark attggtgaaaaaaataatttaatttttagctgtaaaagttaaatgtgatattttattttagctacaatttttttttagtgtgacacttatttttggaaatgggaaaa 765
Jinpum2 attggtgaaaaaaataatttaatttttagctgtaaaagttaaatgtgatattttattttagctacaatttttttttagtgtgacacttatttttggaaatgggaaaa 765
Gaechuck1 attggtgaaaaaaataatttaatttttagctgtaaaagttaaatgtgatattttattttagctacaatttttttttagtgtgacacttatttttggaaatgggaaaa 765

PI567476 taattattaaacactaaccttttttttctgtcctgtggtatgtgaaatcattttgatatatagGAAGCGGTTGGGCGTTTTCTGCCACGGAGCCATAGAA 865
AB013289 taattattaaacactaaccttttttttctgtcctgtggtatgtgaaatcattttgatatatagGAAGCGGTTGGGCGTTTTCTGCCACGGAGCCATAGAA 865
Clark taattattaaacactaaccttttttttctgtcctgtggtatgtgaaatcattttgatatatagGAAGCGGTTGGGCGTTTTCTGCCACGGAGCCATAGAA 865
Jinpum2 taattattaaacactaaccttttttttctgtcctgtggtatgtgaaatcattttgatatatagGAAGCGGTTGGGCGTTTTCTGCCACGGAGCCATAGAA 865
Gaechuck1 taattattaaacactaaccttttttttctgtcctgtggtatgtgaaatcattttgatatatagGAAGCGGTTGGGCGTTTTCTGCCACGGAGCCATAGAA 865

PI567476 GCAGCAGATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTGTACAATGGATGGCACT 965
AB013289 GCAGCAGATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTGTACAATGGATGGCACT 965
Clark GCAGCAGATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTGTACAATGGATGGCACT 965
Jinpum2 GCAGCAGATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTGTACAATGGATGGCACT 965
Gaechuck1 GCAGCAGATGCAATAGCAACAGGAGACCTTGTTAGCCTTTCTGAACAAGAACTCGTAGACTGTGTGGAAGAAAGCGAAGGTTGTACAATGGATGGCACT 965

PI567476 ATCAATCGTTCGAATGGGTTTTAGAACATGTTGGGATTGCCACTGATGATGATTACCTTACAGAGCTAAAGAGGGTAGATGCAAGGCCAATAAGgtctg 1065
AB013289 ATCAATCGTTCGAATGGGTTTTAGAACATGTTGGGATTGCCACTGATGATGATTACCTTACAGAGCTAAAGAGGGTAGATGCAAGGCCAATAAGgtctg 1065

Clark	ATCAATCGTTCGAATGGGTTTTAGAACATGGTGGGATTGCCACTGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGgtctg	1065
Jinpum2	ATCAATCGTTCGAATGGGTTTTAGAACATGGTGGGATTGCCACTGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGgtctg	1065
Gaechuck1	ATCAATCGTTCGAATGGGTTTTAGAACATGGTGGGATTGCCACTGATGATTATCCTTACAGAGCTAAAGAGGGTAGATGCAAAGCCAATAAGgtctg	1065
PI567476	tgaatatataactgaaataatgcatgcatgcatgcatcccttttaataatttaacccttcacatcctttgttttaattttgtgcaacttttttttttttttct	1165
AB013289	tgaatatataactgaaataatgcatgcatgcatgcatcccttttaataatttaacccttcacatcctttgttttaattttgtgcaacttttttttttttttct	1165
Clark	tgaatatataactgaaataatgcatgcatgcatgcatcccttttaataatttaacccttcacatcctttgttttaattttgtgcaacttttttttttttttct	1165
Jinpum2	tgaatatataactgaaataatgcatgcatgcatgcatcccttttaataatttaacccttcacatcctttgttttaattttgtgcaacttttttttttttttct	1165
Gaechuck1	tgaatatataactgaaataatgcatgcatgcatgcatcccttttaataatttaacccttcacatcctttgttttaattttgtgcaacttttttttttttttct	1165
PI567476	gggagtagATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCCTTAAGCGC	1265
AB013289	gggagtagATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCCTTAAGCGC	1265
Clark	gggagtagATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCCTTAAGCGC	1265
Jinpum2	gggagtagATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCCTTAAGCGC	1265
Gaechuck1	gggagtagATACAAGACAAGGTTACAATTGACGGATATGAAACTCTAATAATGTCAGATGAGAGTACAGAATCAGAGACAGAGCAAGCGTTCCTTAAGCGC	1265
PI567476	CATCCTTGAGCAACCAATTAGTGTCTCCATTGATGCAAAAGATTTTCATTATACACCGGGtaagtaataatcttactatttgggaattattattatta	1365
AB013289	CATCCTTGAGCAACCAATTAGTGTCTCCATTGATGCAAAAGATTTTCATTATACACCGGGtaagtaataatcttactatttgggaattattattatta	1365
Clark	CATCCTTGAGCAACCAATTAGTGTCTCCATTGATGCAAAAGATTTTCATTATACACCGGGtaagtaataatcttactatttgggaattattattatta	1365
Jinpum2	CATCCTTGAGCAACCAATTAGTGTCTCCATTGATGCAAAAGATTTTCATTATACACCGGGtaagtaataatcttactatttgggaattattattatta	1365
Gaechuck1	CATCCTTGAGCAACCAATTAGTGTCTCCATTGATGCAAAAGATTTTCATTATACACCGGGtaagtaataatcttactatttgggaattattattatta	1365
PI567476	ttattatactatttttttatcggtcaattttgggtgggaaaagattcgaaatcatgatttcttcttttctcggtatcatcaaattaacctcatcaattcatt	1465
AB013289	ttattatactatttttttatcggtcaattttgggtgggaaaagattcgaaatcatgatttcttcttttctcggtatcatcaaattaacctcatcaattcatt	1465
Clark	ttattatactatttttttatcggtcaattttgggtgggaaaagattcgaaatcatgatttcttcttttctcggtatcatcaaattaacctcatcaattcatt	1465
Jinpum2	ttattatactatttttttatcggtcaattttgggtgggaaaagattcgaaatcatgatttcttcttttctcggtatcatcaaattaacctcatcaattcatt	1465
Gaechuck1	ttattatactatttttttatcggtcaattttgggtgggaaaagattcgaaatcatgatttcttcttttctcggtatcatcaaattaacctcatcaattcatt	1465
PI567476	attactgtacagtatactctatagattattgttattgtgtgtaattcattaattaatttgtgcagGGAATTTATGATGGAGAAAACGTACAAGTCCGT	1565
AB013289	attactgtacagtatactctatagattattgttattgtgtgtaattcattaattaatttgtgcagGGAATTTATGATGGAGAAAACGTACAAGTCCGT	1565
Clark	attactgtacagtatactctatagattattgttattgtgtgtaattcattaattaatttgtgcagGGAATTTATGATGGAGAAAACGTACAAGTCCGT	1565
Jinpum2	attactgtacagtatactctatagattattgttattgtgtgtaattcattaattaatttgtgcagGGAATTTATGATGGAGAAAACGTACAAGTCCGT	1565
Gaechuck1	attactgtacagtatactctatagattattgttattgtgtgtaattcattaattaatttgtgcagGGAATTTATGATGGAGAAAACGTACAAGTCCGT	1565
PI567476	ATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGA	1665
AB013289	ATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGA	1665
Clark	ATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGA	1665
Jinpum2	ATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGA	1665
Gaechuck1	ATGGGATTAATCACTTTGTTTTACTTGTGGGTTATGGTTCAGCGGATGGTGTAGATTACTGGATAGCGAAAAATTCATGGGGAGAAGATTGGGGAGAAGA	1665
PI567476	TGGTTACATTTTGGATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTTTCGCTTCATACCCAACCAAGAGGAATCAGAAACACTG	1765
AB013289	TGGTTACATTTTGGATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTTTCGCTTCATACCCAACCAAGAGGAATCAGAAACACTG	1765
Clark	TGGTTACATTTTGGATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTTTCGCTTCATACCCAACCAAGAGGAATCAGAAACACTG	1765
Jinpum2	TGGTTACATTTTGGATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTTTCGCTTCATACCCAACCAAGAGGAATCAGAAACACTG	1765
Gaechuck1	TGGTTACATTTTGGATCCAAAGAAACACGGGTAATTTATTAGGAGTGTGTGGGATGAATTTTCGCTTCATACCCAACCAAGAGGAATCAGAAACACTG	1765
PI567476	GTGTCGTGCTCGGTTAAAGGTCATCGAAGAGTTGATCACTCTCCTCTTTGA	1816
AB013289	GTGTCGTGCTCGGTTAAAGGTCATCGAAGAGTTGATCACTCTCCTCTTTGA	1816
Clark	GTGTCGTGCTCGGTTAAAGGTCATCGAAGAGTTGATCACTCTCCTCTTTGA	1816
Jinpum2	GTGTCGTGCTCGGTTAAAGGTCATCGAAGAGTTGATCACTCTCCTCTTTGA	1816
Gaechuck1	GTGTCGTGCTCGGTTAAAGGTCATCGAAGAGTTGATCACTCTCCTCTTTGA	1816

Supplemental Fig. 2. Comparison of *P34* genomic DNAs identified from various soybean accessions. Soybean genomic DNA was extracted from leaves of PI567476, Clark, Jinpum2 and Gaechuk1 accession. *P34* genomic DNAs containing promoter and entire coding sequence were compared that of previously reported *P34* genomic DNA (AB013289) by using Clustal W software. The variable nucleotide sequences are indicated by bold and underline.