



Natural variability in abundance of prevalent soybean proteins

Savithiry S. Natarajan*

USDA-ARS, Soybean Genomics and Improvement Laboratory, PSI, Beltsville, MD 20705, USA

ARTICLE INFO

Article history:
Available online 13 August 2010

Keywords:

Soybean
G. soja
G. max
2D-PAGE
MALDI-TOF-MS
LC-MS
Conglycinin
Glycinin

ABSTRACT

Soybean is an inexpensive source of protein for humans and animals. Genetic modifications (GMO) to soybean have become inevitable on two fronts, both quality and yield will need to improve to meet increasing global demand. To ensure the safety of the crop for consumers it is important to determine the natural variation in seed protein constituents as well as any unintended changes that may occur in the GMO as a result of genetic modification. Understanding the natural variation of seed proteins in wild and cultivated soybeans that have been used in conventional soybean breeding programs is critical for determining unintended protein expression in GMO soybeans. In recent years, proteomic technologies have been used as an effective analytical tool for examining modifications of protein profiles. We have standardized and applied these technologies to determine and quantify the spectrum of proteins present in soybean seed. We used two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS), and liquid chromatography mass spectrometry (LC-MS) for the separation, quantification, and identification of different classes of soybean seed proteins. We have observed significant variations in different classes of proteins, including storage, allergen and anti-nutritional protein profiles, between non-GMO cultivated and wild soybean varieties. This information is useful for scientists and regulatory agencies to determine whether the unintended expression of proteins found in transgenic soybean is within the range of natural variation.

© 2010 Published by Elsevier Inc.

1. Introduction

Soybean [*Glycine max* L. (Merr.)], is an important commercial crop grown as a high-protein feed supplement for its edible oil and for its other phytonutrients (Boulter et al., 1980). This has generated significant interest by major agricultural research institutes and biotech industries in enhancing the quality of nutrients in soybeans as well as in increasing resistance against various pests. This has resulted in the introduction of several new GMO cultivars in the US. As more transgenic cultivars become available, it has become important to identify and characterize the natural variation of proteins in non-genetically modified soybean genotypes as a baseline against which to compare the new cultivars.

At present, proteomics tools such as two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry, and liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) provide alternative methods to effectively examine protein profiles. These tools can be used to evaluate protein expression and potential protein modifications, as well as the

relative abundance of proteins in non-GMO and GMO crops (Corpillo et al. 2004; Lehesranta et al. 2005). However, protein extraction remains a challenge in the analysis of proteins because of contaminants that could affect the performance of the 2D-PAGE, MALDI-TOF and LC-MS/MS analyses. In our laboratory, we carried out a comparison of four commonly used protein extraction methods described in the literature (Hurkman and Tanaka, 1986; Berkelman et al., 1998; Herman et al. 2003) for 2D-PAGE analysis, and subsequently selected an efficient extraction method for our analysis (Natarajan et al. 2005). This report summarizes some of the results obtained in our laboratory over the past few years. We have reported the variation of different classes of soybean seed protein profiles in cultivated and wild soybean genotypes. The wild genotypes are completely cross compatible with cultivated soybean by conventional plant breeding. Results from the wild genotypes can be compared with the amounts of storage, allergen and anti-nutritional proteins in GMO soybeans to determine if the levels of these proteins in the GMO soybeans are within the limits of natural variation.

2. Proteomic analysis of soybean storage proteins

The soybean storage protein families, β -conglycinin and glycinin, are classified based on their sedimentation coefficients.

* Address: Soybean Genomics and Improvement Laboratory, PSI-ARS-USDA, 10300, Baltimore Avenue, Beltsville, MD 20705, USA. Fax: 301-504-5728.

E-mail address: savi.natarajan@ars.usda.gov

β -conglycinins are encoded by two mRNA groups, the first mRNA group encodes α and α' β -conglycinin subunits and the second mRNA group encodes the β -subunit of β -conglycinin (Schuler et al. 1982; Koshiyama, 1983). Initially, we used pH 3.0–10.0 in the first dimension to separate the subunits of storage proteins, however, this did not adequately separate glycinin storage protein subunits. Therefore, we currently used two narrow pH strips, 4.0–7.0 for acidic proteins (Fig. 2A and B) and pH 6.0–11.0 range for basic proteins (Fig. 3A and B). Triplicate samples were used for soybean seed protein extraction and 2D-PAGE analysis. Densitometry images were analyzed with Image Master 2D-Elite software (version 4.01) (GE Healthcare, Piscataway, NJ). Image analysis included the following parameters such as spot detection, background subtraction, spot measurement, and spot matching. A total of forty-seven different protein spots could be excised from 2D-PAGE gels using all three pH ranges and were characterized using mass spectrometry. Protein spots that could not be identified by MALDI-TOF-MS were further analyzed by LC-MS. All the proteins identified by MALDI-TOF-MS and LC-MS in the NCBI non-redundant database are listed in Table 1, which gives: an assigned protein spot number and its theoretical isoelectric point (pI) and molecular weight (Mr), protein identity, number of peptides matched, percent sequence coverage, MOWSE score, expect value, the NCBI database accession number of the best match, and a database that yielded concurrent identification.

Comparison of proteins from the wild and the cultivated soybean genotypes by 2D-PAGE using pH 3.0–10.0 IPG strips (Fig. 1A and B) highlight three subunits of β -conglycinin, the α , α' and β -subunits (Natarajan et al. 2006, 2007a). The α subunit showed five protein spots (#1, 3, 4, 6 and 7) in the cultivated genotype (Fig. 1B) and seven spots (#1–7, Fig. 1A) were resolved in the wild genotype. The intensity of #6 and #7 spots were higher in the wild genotype compared to the cultivated genotype. Some of the spots demonstrated heterogeneity in molecular weight and isoelectric point (pI), and these differences could be the result of post-translational modifications, proteolysis, and/or alternate splicing. The α' subunit of β -conglycinin showed one spot (#8) in both wild and cultivated genotypes (Table 1), but the intensity of this spot was lower in the wild (Fig. 1A) compared with the cultivated genotype (Fig. 1B). The β -subunit of β -conglycinin resolved into four protein spots, (#9, 11–13) in the wild genotype and five protein spots (#9–14) in the cultivated genotype. Schuler et al. (1982) reported that β -conglycinin subunits are products of a multigene family, and the

variation in the distribution of protein spots in our study could also be due to post-translational modifications (Koshiyama, 1983; Davies et al. 1985).

Glycinin is composed of five subunits, G1, G2, G3, G4 and G5, the precursors of which are encoded by five non-allelic genes, Gy1, Gy2, Gy3, Gy4 and Gy5, respectively (Nielsen et al., 1989). On pH 3.0–10.0 gels, it could be separated into acidic and basic polypeptides. However, many did not separate well. We now use narrow pH range IEF strips (4.0–7.0 and 6.0–11.0) to separate and identify the acidic and basic chains of glycinin subunits. The G1 subunit showed three basic polypeptides (spot #15, 16 and 17) in the pH 6.0–11.0 range that were similar in both wild and cultivated genotypes (Fig. 3A and B). The glycinin G2 showed eight spots (#18–25) of acidic and basic polypeptides in the wild genotype and seven spots in cultivated genotype. G3 glycinin acidic and basic polypeptides resolved into seven spots (#26–32) in the wild genotype, and the cultivated genotype showed an absence of two spots (#29 and 30), weak intensity of spot #28, and the presence of an additional spot #33 (Figs. 2 and 3 and Table 1). Two acidic polypeptides, which were identified as glycinin G4 (#34, 35), were present in the wild genotype and absent in the cultivated genotype (Fig. 1A and B). Glycinin G4 basic polypeptide separated into five spots (#36–40), was only present in the wild type seeds, and showed strong intensity (Fig. 1). The absence of G4 subunits in these particular cultivated genotypes could be due to the absence of the gene(s) encoding G4 subunits. Nielsen et al. (1989) reported the absence of G4 in soybean cv. Raiden, and suggested that heterogeneity of the G4 subunit may not have functional relevance. The glycinin G5 polypeptide (spot #41) that was clustered in the pH 3.0–10.0 gels (Fig. 1A and B) was clearly separated into three polypeptides (spot #42–44) in the narrow pH 4.0–7.0, in both wild and cultivated genotypes (Fig. 2). Three basic G5 polypeptides (spot #45–47) showed strong intensity in the cultivated genotype and weak intensity in the wild genotype. Fukazawa et al. (1985) reported the diversity of the glycinin G5 mRNA subunit between *G. max* and *G. soja*. Our study showed considerable variations of both acidic and basic glycinin polypeptides between wild and cultivated genotypes. The large amount of variation in seed protein composition between wild and cultivated genotypes may be due to a different complement of genes in the wild genotype that control expression of β -conglycinin and glycinin protein composition compared with the cultivated genotypes (Kwanyuen et al. 1997).

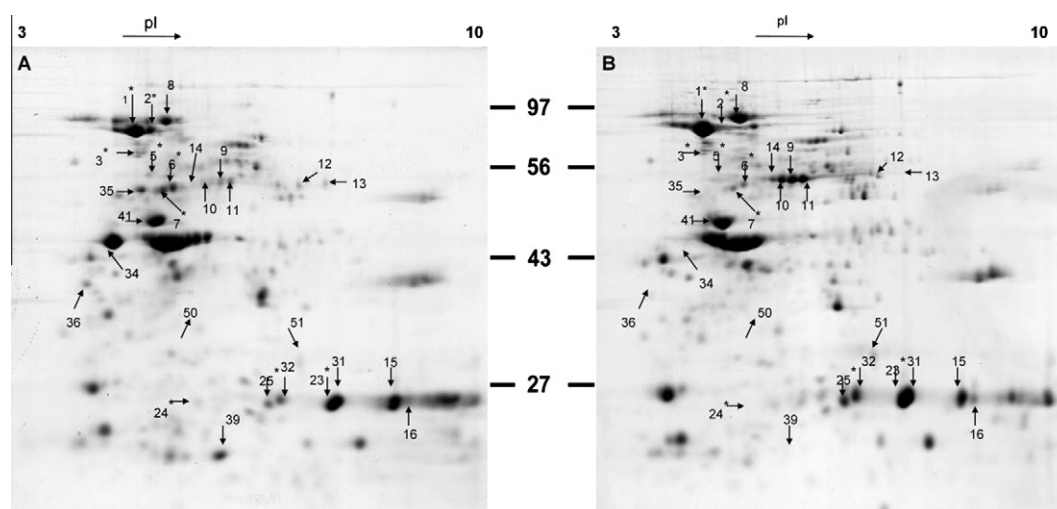


Fig. 1. Two-dimensional electrophoresis profile of soybean seed proteins of wild, *G. soja* (1A) and cultivated, *G. max* (1B). The first dimension IEF was performed using pH 3.0–10.0 linear IPG strips. Numbered arrows indicate the polypeptides referred to in the text and table.

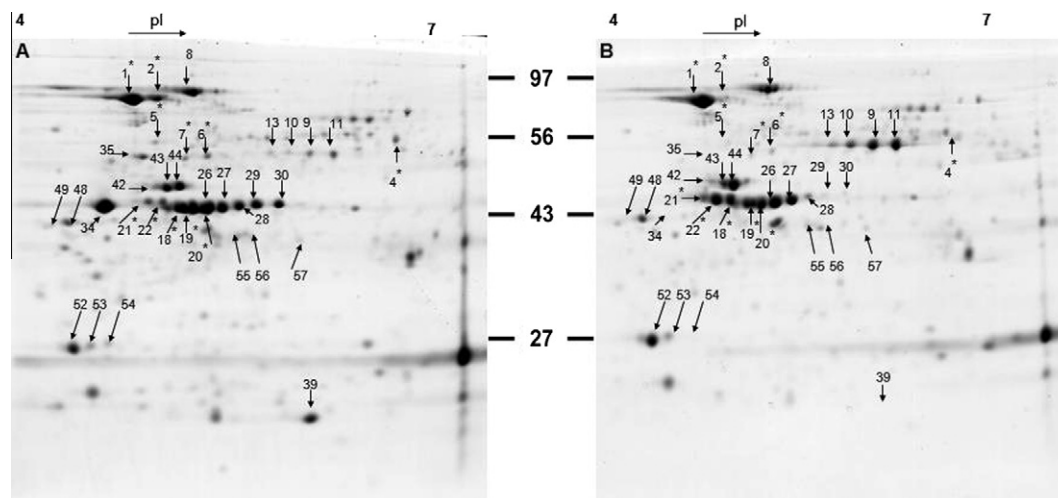


Fig. 2. Two-dimensional electrophoresis profile of soybean seed proteins of wild, *G. soja* (2A) and cultivated, *G. max* (2B). The first dimension IEF was performed using pH 4.0–7.0 linear IPG strips. Numbered arrows indicate the polypeptides referred to in the text and table.

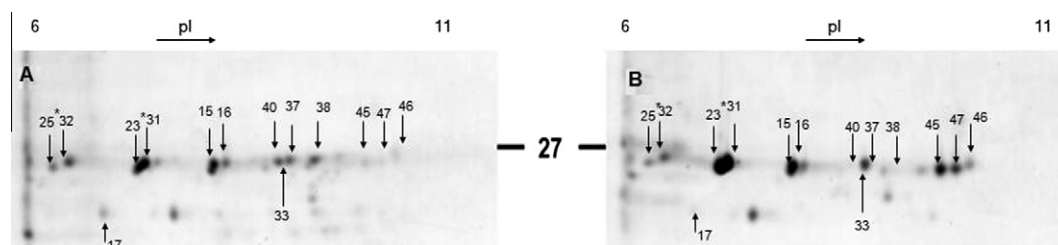


Fig. 3. Two-dimensional electrophoresis profile of soybean seed proteins of wild, *G. soja* (3A) and cultivated, *G. max* (3B). The first dimension IEF was performed using pH 6.0–11.0 linear IPG strips. Numbered arrows indicate the polypeptides referred to in the text and table.

3. Proteomic analysis of soybean allergen proteins

Soybean Gly m Bd 60 K has been described as storage seed proteins and is one of the major allergens in soybean. Gly m Bd 60 K includes β -conglycinin and glycinin. Only the α subunit of β -conglycinin is reported to be allergenic for some consumers (Wilson et al. 2005). The acidic polypeptides of G1 and all G2 subunits of glycinin are also reported to be allergenic (Beardslee et al. 2000). The cultivated and wild genotypes showed seven spots (# 1–7) of α subunits (Xu et al. 2007). The glycinin G2 allergens showed eight spots (#18–25) and the distribution was very similar between wild and cultivated genotypes, although there was variation in protein intensity. Gly m Bd 30 K is another protein of soybean designated as a major allergen. This protein was previously described as the 34-kD vacuolar protein, P34 (Ogawa et al. 1993). Using a monoclonal antibody against P34, Yaklich et al. (1999) found that P34 showed genetic diversity in soybean with respect to protein quantity. In our analyses, P34 (#48, 49) is present in both wild and cultivated soybean genotypes, and the spot intensity varied. Joseph et al. (2006) screened 16,266 accessions of USDA soybean germplasm, and identified 12 genotypes as P34 null lines. Gly m Bd 28 K is a less abundant allergen protein of soybean that was originally isolated from soybean meal as a 28 kDa glycosylated protein (Tsuji et al. 1997). In our study, two spots (#50 and 51) of Gly m Bd 28 K were identified. The Gly m Bd 28 K allergen is probably processed into two smaller polypeptides of 240 and 212 amino acids in the soybean seed, and both portions are expected to be present in soybean-derived foods (Tsuji et al. 2001; Xiang et al. 2004). These two spots of Gly m Bd 28 K probably come from the post-translational processing of the same gene.

4. Proteomic analysis of anti-nutritional proteins

Kunitz trypsin inhibitor (KTI) is one of the abundant anti-nutritional proteins in soybean seed, that inhibits trypsin, an important animal digestive enzyme (Baur et al. 1996). In addition, KTIs have been characterized as food allergens in humans (Burks et al. 1994; Quirce et al. 2002) and have 32% sequence homology with a rye grass pollen allergen (Vanvan Ree et al. 1995). In soybean, three KTI genes (KTI1–KTI3) have been reported, and both transcriptional and post-transcriptional processes regulate KTI gene expression (Jofuku and Goldberg, 1989). The KTI3 transcript was detected only in the soybean seed, while KTI1 and KTI2 transcripts are expressed in soybean leaf, root and stem (Jofuku and Goldberg, 1989). The genes encoding KTI1 and KTI2 have nearly identical nucleotide sequences, while the KTI3 gene has diverged (20%) from the other two genes (Jofuku and Goldberg, 1989). Our study showed that although the overall distribution patterns of KTI protein spots are quite similar in wild and cultivated, the number of protein spots and their intensities varied between these genotypes (Natarajan et al. 2007b). Three protein spots (spots 52–54) of KTI were identified by MALDI-TOF-MS and database searches in wild soybean varieties, and cultivated genotypes showed only two protein spots (#52 and 53). Our results were similar to the previous reports, in which polymorphism had been observed among the KTIs of wild soybeans by Western blotting and PAGE analysis (Stahlhut and Hymowitz, 1983; Becker-Ritt et al. 2004; Wang et al. 2005). Soybean agglutinins are another class of anti-nutritional proteins present abundantly in soybeans. The presence of this protein accounts for about 10% of total protein in some legumes. The wild and cultivated soybeans showed three spots

Table 1

Soybean seed proteins identified by MALDI-TOF-MS and LC-MS analysis.

Spot ID	Theoretical pI/Mr	Protein identity	Peptides matched	Sequence coverage (%)	MOWSE score	Expect value	NCBI accession No.	Method/MS
<i>Soybean Storage</i>								
1–4*	4.92/63127	α subunit of β -conglycinin	25	39	217	3.00E-17	gi 19967357	MALDI-TOF
5–6*	5.32/72717	α subunit of β -conglycinin	9	18	73	7.20E-03	gi 15425633	MALDI-TOF
7*	5.32/72717	α subunit of β -conglycinin	9	23	238		gi 51247837	LC-MS
8	5.23/65160	α' subunit of β -conglycinin	20	41	194	5.90E-15	gi 19967361	MALDI-TOF
9–14	5.67/48358	β -conglycinin β -subunit	26	47	216	5.10E-17	gi 63852207	MALDI-TOF
15–16	5.89/56299	Glycinin G1 subunit	7	15	145		gi 18635	LC-MS
17	6.15/56134	Glycinin G1 subunit	4	9	254		gi 72296	LC-MS
18–22*	5.46/54927	Glycinin G2 precursor	9	19	91	1.20E-04	gi 1212177	MALDI-TOF
23*	5.78/54047	Glycinin G2 precursor	6	37	67	3.20E-02	gi 169967	MALDI-TOF
24–25*	5.56/54903	Glycinin G2 precursor	3	9	175		gi 72295	LC-MS
26–32	5.78/54047	Glycinin G3 subunit	8	19	72	1.00E-02	gi 15988117	MALDI-TOF
33	5.73/54835	Glycinin G3 subunit	2	4	148		gi 18639	LC-MS
34–35	5.38/64097	Glycinin G4/precursor	13	16	76	4.20E-03	gi 99910	MALDI-TOF
36	4.46/24349	Glycinin G4 precursor	10	38	110	1.50E-06	gi 6015515	MALDI-TOF
37–38	6.47/31065	Glycinin G4 precursor	10	45	142	9.50E-10	gi 81785	MALDI-TOF
39–40	5.38/64136	Glycinin G4 precursor	6	8	160		gi 99910	LC-MS
41–44	5.46/55850	Glycinin G5 subunit	13	26	96	3.50E-05	gi 33357661	MALDI-TOF
45–46	5.69/26938	Glycinin G5 subunit	7	30	65	5.00E-02	gi 541941	MALDI-TOF
47	9.64/21482	Glycinin G5 subunit	3	12	109		gi 625538	LC-MS
<i>Soybean allergen</i>								
48	5.74/43136	P34	4	9	102		gi 129353	LC-MS
49	5.74/43136	P34	2	9	60		gi 129353	LC-MS
50	5.73/52813	Allergen Gly m Bd 28 K	3	6	187		gi 12697782	LC-MS
51	5.73/52780	Allergen Gly m Bd 28 K	10	19	97	2.90E-05	gi 12697782	MALDI-TOF
52	4.61/20310	Kunitz trypsin inhibitor	15	53	147	3.00E-10	gi 3318877	MALDI-TOF
53	4.97/22817	Kunitz trypsin inhibitor	8	42	72	1.10E-02	gi 125722	MALDI-TOF
54	4.61/20310	Kunitz trypsin inhibitor	9	44	60	1.40E-01	gi 3318877	MALDI-TOF
<i>Soybean anti-nutritional</i>								
55	5.65/30909	Soybean agglutinin	7	10	244	1.50E-04	gi 282898	LC-MS
56	5.15/27555	Soybean agglutinin	9	41	90		gi 3114258	MALDI-TOF
57	5.15/27555	Soybean agglutinin	7	30	358		gi 3114258	LC-MS

* Gly m Bd 60 K allergens.

(#55–57) of soybean agglutinins, and we observed minor variation in the distribution.

We identified major and minor variations in amount and number of storage, allergen and anti-nutritional proteins among various wild and cultivated soybean genotypes. This variation may be due to selection of high oil and protein genotypes by breeders. We demonstrated that changes in protein composition and its level produced during generation and selection of new soybean varieties. In addition, this research provides useful data for scientists developing soybean with improved storage protein quality and for the education of consumer about soybean proteins available on the market.

Acknowledgments

We thank Dr. Janet Slovin for critical review of this manuscript. Funding for this research was provided by ARS project 1275-21000-223-00D.

References

- Baur, X., Pau, M., Czuppon, A., Fruhmman, G., 1996. Allergy 51, 326–330.
- Beardslee, T.A., Zeece, M.G., Sarath, G., Markwell, J.P., 2000. Int. Arch. Allergy Immunol. 123, 299–307.
- Becker-Ritt, A.B., Mulinari, F., Vasconcelos, I.M., Carlini, C.R., 2004. J. Sci. Food Agric. 84, 263–270.
- Berkelman, T., Stenstedt, T., Bjellqvist, B., Laird, N., McDowell, M., Olsson, I., Westermeier, R., 1998. 2D Electrophoresis Using Immobilized PH Gradients: Principles and Methods. Amersham Biosciences, Piscataway, NJ.
- Advanced legume science, 1980. In: Boulter, D., Summerfield, R.J., Bunting, A.H. (Eds.). Proc. International Conference. Royal Botany Garden Publishing Kew, UK, pp. 127–134.
- Burks, A.W., Cockrell, G., Connaughton, C., Guin, J., Allen, W., Helm, R.M., 1994. Int. Arch. Allergy Immunol. 105, 143–149.
- Corpillo, D., Gardini, G., Vaira, A.N., Basso, M., Aime, S., Accotto, G.P., Fasano, M., 2004. Proteomics 4, 193–200.
- Davies, C.S., Coates, J.B., Nielson, N.C., 1985. Theor. Appl. Genet. 71, 351–358.
- Fukazawa, C., Momma, H., Hirano, H., Harada, K., Uda, K., 1985. J. Biol. Chem. 260, 6234–6239.
- Herman, E.M., Helm, R.M., Jung, R., Kinney, A.J., 2003. Plant Physiol. 132, 36–43.
- Hurkman, W.J., Tanaka, C.K., 1986. Plant Physiol. 81, 802–806.
- Jofuku, K.D., Goldberg, R.B., 1989. Plant Cell 1, 1079–1093.
- Joseph, L.M., Hymowitz, T., Schmidt, M.A., Herman, E.M., 2006. Crop Sci. 46, 1755–1763.
- Koshiyama, I., 1983. In: seed proteins Biochemistry, Genetics, Nutritive Value. 427–450.
- Kwanyuen, P., Pantalone, V.R., Burton, J.W., Wilson, R.F., 1997. J. Am. Oil Chem. Soc. 74, 983–987.
- Lehesranta, S.J., Davies, H.V., Shepherd, L.V.T., Nunan, N., McNicol, J.W., Auriola, S., Koistinen, K.M., Suomalainen, S., Kokko, H.I., Karenlampi, S.O., 2005. Plant Physiol. 138, 1690–1699.
- Natarajan, S.S., Xu, C., Caperna, T.J., Garrett, W.M., 2005. Anal. Biochem. 342, 214–220.
- Natarajan, S.S., Xu, C., Bae, H., Caperna, T.J., Garrett, W.M., 2006. J. Agric. Food Chem. 54, 3114–3120.
- Natarajan, S.S., Xu, C., Bae, H., Cregan, P., Caperna, T.J., Bailey, B.A., Garrett, W.M., Luthria, D., 2007a. Plant Physiol. Biochem. 45, 436–444.
- Natarajan, S.S., Xu, C., Bae, H., Bailey, B.A., 2007b. J. Plant Physiol. 164, 756–763.
- Nielsen, N.C., Dickinson, C.D., Cho, T.J., Thanh, V.H., Scallion, B.J., Fischer, R.L., Sims, T.L., Goldberg, G.N., 1989. Plant Cell 1, 313–328.
- Ogawa, T., Tsuji, H., Bando, N., Kitamura, Zhu, Y.L., Hirano, H., Nishikawa, K., 1993. Biosci. Biotechnol. Biochem. 57, 1030–1033.
- Quirce, S., Fernandez-Nieto, M., Polo, F., Sastre, J., 2002. J. Allergy Clin. Immunol. 109, 178.
- Schuler, M.A., Schmitt, E.S., Beachy, R.N., 1982. Nucleic Acids Res. 10, 8225–8243.
- Stahlhut, R.W., Hymowitz, T., 1983. Crop Sci. 23, 766–769.
- Tsuji, H., Bando, N., Hiemori, M., Yamanishi, R., Kimoto, M., Nishikawa, K., Ogawa, T., 1997. Biosci. Biotechnol. Biochem. 61, 942–947.
- Tsuji, H., Hiemori, M., Kimoto, M., Yamashita, H., Kobatake, R., Adachi, M., Fukuda, T., Bando, N., Okita, M., Utsumi, S., 2001. Biochem. Biophys. Acta 1518, 178–182.
- Vanvan Ree, R., Hoffman, D.R., van Dijk, W., Brodard, V., Mahieu, K., Koeleman, C.A., Grande, M., van Leeuwen, W.A., Aalberse, R.C., 1995. J. Allergy Clin. Immunol. 95, 970–978.
- Wang, K.J., Takahata, Y., Kono, Y., Kaizuma, N., 2005. Theor. Appl. Genet. 112, 66–71.
- Wilson, S., Blaschek, K., de Mejia, E., 2005. Nutr. Rev. 63, 47–58.
- Xiang, P., Haas, E.J., Zeece, M.G., Markwell, J., Sarath, G., 2004. Planta 220, 56–63.
- Xu, C., Caperna, T.J., Garrett, W.M., Cregan, P., Bae, H., Luthria, D., Natarajan, S.S., 2007. J. Sci. Food Agric. 87, 2511–2518.
- Yaklich, R.W., Helm, R.M., Cockrell, G., Herman, E.M., 1999. Crop Sci. 39, 1444–1447.