Proteomic Analysis of Allergen and Antinutritional Proteins in Wild and Cultivated Soybean Seeds

Savithiry Natarajan[1]*, Chenping Xu[2], Hanhong Bae[3], Thomas J Caperna[4] and Wesley M Garrett[5]

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

[2]University of Maryland, Department of Natural Resource Sciences and Landscape Architecture, College Park, MD, 20742, USA

[3]USDA-ARS, Sustainable Perennial Crops Laboratory, Beltsville, MD 20705, USA

[4]USDA-ARS, Growth Biology Laboratory, Beltsville, MD 20705, USA

[5]USDA-ARS, Biotechnology and Germplasm Laboratory, Beltsville, MD 20705, USA

In this study, profiles of allergen and antinutritional proteins both in wild (*Glycine soja*) and cultivated (*Glycine max*) soybean seeds were compared. We used two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) for the separation of proteins at two different pH ranges and applied a combined matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS) and liquid chromatography mass spectrometry (LC-MS/MS) analysis for the identification of proteins. Although overall distribution patterns of the allergen (Gly m Bd 60K, Gly m Bd 30K, Gly m Bd 28K) and antinutritional proteins (trypsin inhibitors and lectin) appeared similar, there was remarkable variation in the number and intensity of the protein spots between wild and cultivated genotypes. The wild genotype showed fifteen polypeptides of Gly m Bd 60K and three polypeptides of trypsin inhibitors. The cultivated genotypes showed twelve polypeptides of Gly m Bd 60K and two polypeptides of trypsin inhibitors. In contrast, the cultivated genotype showed two polypeptides of Gly m Bd 30K and three polypeptides of lectin and the wild genotype showed two and one polypeptides of Gly m Bd 30K and lectin, respectively. Two polypeptides of Gly m Bd 28K were observed in both genotypes. This is the first study reporting the comparative analysis of allergen and antinutritional proteins in both wild and cultivated soybean genotypes using combined proteomic tools.

Key words: soybean, *Glycine soja*, *G. max*, 2D-PAGE, MALDI-TOF-MS, LC-MS/MS, conglycinin, glycinin, allergen proteins.

Soybean is a rich and inexpensive source of proteins for humans and animals. However, some people have severe allergic reactions to soy proteins (1). It has been estimated that nearly 2% of adults and 4-6% of children suffer from food allergies. Food allergies are defined as an adverse immunologically mediated reaction to antigenic molecules present in foods. The allergic response is mediated by reactions with an allergen (2). Soybean ranks among the eight most significant food allergens. Wide usage of soy protein in the food industry makes it difficult to eliminate from the diet. There are many reports of adverse reactions due to ingestion of soybean products (3). Recently, Ogawa *et al* (4) demonstrated the occurrence of fifteen protein

components that bind with IgE antibodies in the sera of people with soybean sensitive atopic dermatitis. Comprehensive analytical methods to separate and characterize allergen and antinutritional proteins are crucial for a better understanding of the proteins in soybean genotypes.

Application of combined proteomic technologies such as 2D-PAGE, MALDI-TOF-MS, and LC-MS/MS is a proven analytical solution for separating and accurately examining changes in protein composition (5). About 16 allergens have been identified in soybean; three allergen proteins, Gly m Bd 60K, Gly m Bd 30K (P34) and Gly m Bd 28K represent the main allergens in soybean seeds (6, 7). Recently, Moony and Thelen (8) studied the protein composition of a cultivated soybean genotype, *G. max cv* Jefferson using 2D-PAGE analysis with a pH range of 3.0-10.0 and reported only Gly m Bd 60K allergens. Herman *et al* (9) have applied 2D-PAGE methodology for protein separation and successfully analyzed allergen proteins in

---

cultivated soybean *cv* Jack, a soybean cultivar in which expression of P34 protein was suppressed. A comprehensive comparative analysis of specific allergen proteins in soybean seeds is lacking. Therefore, a comparative analysis of wild and cultivated seed allergen proteins is crucial for identifying and eliminating potential soybean allergens. Such an analysis of the soybean seeds could also prove useful for comparative analysis of the allergenicity of other legume species.

The objective of this investigation was to determine the variation of allergen proteins in wild and cultivated soybean genotypes, and to establish proteomics-based analytical methods for effectively analyzing the composition of complex allergen protein profiles in a wide range of soybean seeds. In this report, we have studied the profiles of major allergen and antinutritional proteins in two soybean genotypes, a wild *G. soja* and a cultivated *G. max*, using two different pH ranges for the first dimension protein separation in the 2D-PAGE technique. We characterized abundant and less abundant proteins by combining MALDI-TOF-MS and LC-MS/MS.

Materials and Methods

Plant materials — Soybean seeds of *G. soja* PI 393551 and *G. max* PI 423954 *cv* Shirome, were obtained from the USDA soybean germplasm collection, Urbana, Illinois.

Extraction, separation and identification of proteins — Protein was extracted by the modified TCA/acetone method (10). 2D-PAGE separation was performed according to Natarajan *et al* (10) using IPG strips with pH 3.0-10.0 and 4.0-7.0 ranges. Three separate samples were extracted individually and run on duplicate gels. The protein concentration was determined by the Bradford method (11). Protein spots were analyzed by MALDI-TOF-MS and LC-MS/MS using the protocols developed in our laboratory (10). Protein identification was performed by searching National Center for Biotechnology Information (NCBI) non-redundant database using the Mascot search engine (http://www.matrixscience.com). To qualify the MALDI-TOF-MS data as a positive identification, the molecular weight search (MOWSE) score was equal to or exceeded the minimum significant score of sixty-four. For LC-MS/MS, the positive identification required a minimum of two unique peptides with at least one peptide having a significant ion score.

Results and Discussion

In this study, we have compared the composition of allergen and antinutritional proteins in wild *G. soja* and cultivated *G. max* seeds using a 2D-PAGE separation technique with two pH ranges (3.0-10.0 and 4.0-7.0) and combined proteomic tools. The gel images presented in Figs. 1 and 2 are representative of the actual experimental data. Our results showed that the overall distribution pattern of allergen proteins is quite similar in both wild and cultivated genotypes (Fig. 1A, B). However, the number of spots and their intensities varied between these two genotypes. In this study, we focused on the characterization of three major allergen proteins, Gly m Bd 60K, Gly m Bd 30K, and Gly m Bd 28K, and two antinutritional factors, trypsin inhibitors and lectin (agglutinin).

Analysis of Gly m Bd 60K proteins — Soybean Gly m Bd 60K allergen proteins are grouped into two types, β-conglycinin and glycinin. β-conglycinin, a 7S globulin, consists of three subunits: α, α^2 and β subunits (12). Among these subunits, only the α subunit of β-conglycinin has allergenic reactions in about 25% of sera from soybean sensitive patients (4). Our study showed five polypeptides of the α subunit of β-conglycinin (spots # 1, 3, 4, 6 and 7) in the cultivated *G. max* genotype (Fig. 1B). In the wild genotype, there were seven α subunit polypeptides, which included the above five polypeptides and two additional polypeptides, spots #2 and 5 (Fig. 1A). The relative intensities of spots # 6 and 7 were higher in the wild genotype compared to the cultivated genotype. The sequence of tryptic peptide fragments from four spots (#1, 2, 3, and 4) matched with one gene (gil9967357) and two spots (#5 and 6) matched with another gene (gil15425633); spot 7 was from a different gene (gil51247837) (Table 1). Data listed in Table 1 consists of an assigned protein spot number, theoretical isoelectric point (pI) and molecular weight (Mr), protein identity, identification method, number of peptides matched, percent sequence coverage, MOWSE score, expected value, and the NCBI database accession number of the best match. In addition, it shows the proteins that are either present (+) or absent (-) in comparisons of the wild and cultivated soybean seeds. In 2D-PAGE images (Fig.1), some spots had different pIs and the same Mr (#1 and 2), while others had different pIs and Mr (#3, 4, 5, and 6). This could be explained by: 1) the existence of a multigene family that has not been identified yet, 2) alternate splicing from one gene, 3) co-existence of precursor and mature protein, 4) co-existence of degraded

Table 1. Proteins identified by MALDI-TOF-MS and LC-MS/MS in wild and cultivated soybean seeds

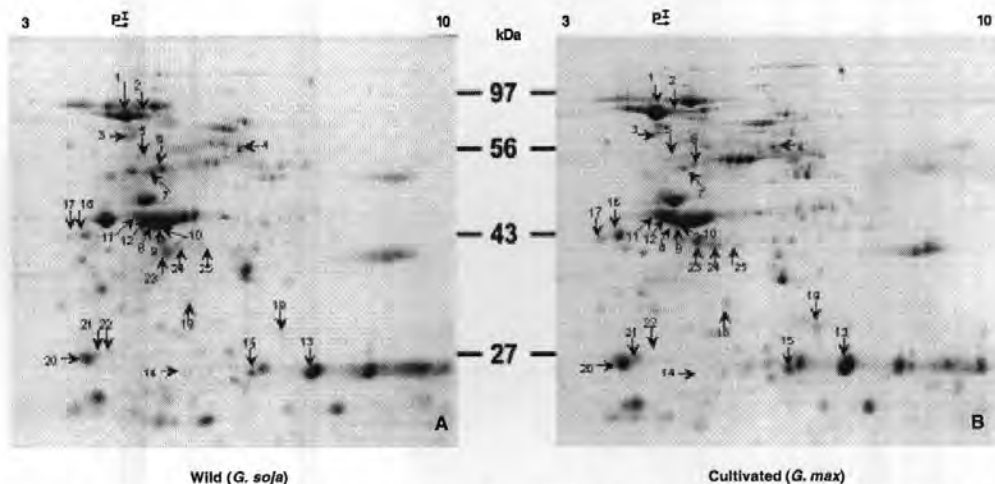| S. No.	Theoretical pI/Mr	Protein Identity	Identification method	Peptides matched	Sequence coverage	MOWSE score	Expect value	NCBI Accession #	Wild (G. soja)	Cultivated (G. max)
1	4.92/63127	α subunit of β conglycinin	MALDI-TOF-MS	25	39%	217	3.00E-17	gi|9967357	+	+
2	4.92/63127	α subunit of β conglycinin	MALDI-TOF-MS	25	41%	167	4.00E-12	gi|9967357	+	-
3	4.92/63184	α subunit of β conglycinin	MALDI-TOF-MS	27	43%	250	1.50E-20	gi|9967357	+	+
4	4.92/63184	α subunit of β conglycinin	MALDI-TOF-MS	20	34%	204	6.00E-16	gi|9967357	+	+
5	5.32/72717	α subunit of β conglycinin	MALDI-TOF-MS	9	18%	73	7.20E-03	gi|15425633	+	-
6	5.32/72717	α subunit of β conglycinin	MALDI-TOF-MS	14	24%	112	1.10E-03	gi|15425633	+	+
7	5.32/72717	α subunit of β conglycinin	LC-MS/MS	9	23%	238		gi|51247837	+	+
8	5.46/54927	Glycinin G2 precursor	MALDI-TOF-MS	9	19%	91	1.20E-04	gi|1212177	+	+
9	5.46/54927	Glycinin G2 precursor	MALDI-TOF-MS	9	19%	73	7.20E-03	gi|1212177	+	+
10	5.46/54927	Glycinin G2 precursor	MALDI-TOF-MS	8	15%	71	1.30E-02	gi|1212177	+	+
11	5.46/54927	Glycinin G2 precursor	MALDI-TOF-MS	9	14%	74	6.70E-03	gi|1212177	+	+
12	5.46/54927	Glycinin G2 precursor	MALDI-TOF-MS	12	21%	104	6.10E-06	gi|1212177	+	+
13	5.78/54047	Glycinin A2B1a precursor	MALDI-TOF-MS	6	37%	67	3.20E-02	gi|169967	+	+
14	5.56/54903	Glycinin A2B1a precursor	LC-MS/MS	3	9%	175		gi|72295	+	-
15	5.56/54903	Glycinin A2B1a precursor	LC-MS/MS	9	14%	313		gi|72295	+	+
16	5.74/43136	P34 probable thiol protease precursor	LC-MS/MS	4	9%	102		gi|129353	+	+
17	5.74/43136	P34 probable thiol protease precursor	LC-MS/MS	2	9%	60		gi|129353	-	+
18	5.73/52813	Allergen Gly m Bd 28K	LC-MS/MS	3	6%	187		gi|12697782	+	+
19	5.73/52780	Allergen Gly m Bd 28K	MALDI-TOF-MS	10	19%	97	2.90E-05	gi|12697782	+	+
20	4.61/20310	Soybean Kunitz trypsin inhibitor	MALDI-TOF-MS	15	53%	147	3.00E-10	gi|3318877	+	+
21	4.97/22817	Kunitz trypsin inhibitor	MALDI-TOF-MS	8	42%	72	1.10E-02	gi|125722	+	+
22	4.61/20310	Soybean Kunitz trypsin inhibitor	LC-MS/MS	9	22%	60	1.40E-01	gi|3318877	+	-
23	5.65/30909	Lectin precursor	LC-MS/MS	7	10%	244		gi|282898	+	+
24	5.15/27555	Soybean agglutinin	MALDI-TOF-MS	9	41%	90	1.50E-04	gi|3114258	+	+
25	5.15/27555	Soybean agglutinin	LC-MS/MS	7	30%	358		gi|3114258	-	+

Fig.1. A proteomic comparison of the allergen and antinutritional proteins of wild, *Glycine soja* (**A**) and cultivated, *G. max* (**B**). The first dimension was run using a pH gradient from 3.0-10.0 and the second dimension is a 12% SDS-PAGE. Numbered arrows indicate the polypeptides referred to in the text and table.

proteins, or 5) post-translational modifications. Co- and post-translational modifications of β-conglycinin have been reported in previous studies. For example, Davies (13) reported that β-conglycinin undergoes extensive co- and post-translational modification.

The second allergen protein, Gly m Bd 60 K (glycinin), is an 11S globulin, and is one of the most prevalent proteins in the soybean seed. It consists of five subunits, G1, G2, G3, G4 and G5 of acidic and basic polypeptides (14). Among these subunits, only acidic polypeptides of G1 and acidic and basic polypeptides of G2 are reported to be allergens (15). Therefore, we focused our studies only on acidic polypeptide (A1a) chain of G1 and acidic and basic polypeptides (A2/B1a) chain of G2 subunit. Initially, we used a pH range of 3.0-10.0 to analyze the glycinin subunits. But the individual composite chains could not be adequately separated using this pH range (Fig. 1). Therefore, we used a narrow pH range (pH 4.0-7.0) to separate these allergen proteins. The acidic polypeptide (A1a) of G1 (15 kD) was not observed in this investigation because of its low abundance and low molecular weight. We observed a total of eight spots (#8, 9, 10, 11, 12, 13, 14, and 15) of acidic and basic polypeptides of glycinin G2 in the wild genotype and seven spots in the cultivated genotype, which was missing spot #14 (Fig. 1A, B). Among these spots, spot #8-12 were identified as the A2 chain, and spots #13-15 were identified as B1a chain of glycinin subunit G2. We have also observed considerable variation in the intensity of glycinin polypeptides between the two

genotypes (Fig. 2A, B). Variation in the number of G2 polypeptides among cultivars of cultivated soybeans has been previously reported (8, 9). The variation in the distribution of protein spots with different pls and the same Mr is most likely due to post-translational modifications. Alternatively, the different isoforms could be products of different genes (8).

Analysis of Gly m Bd 30K/P34 — The Gly m Bd 30K protein was first identified by Kalinski *et al* (16) from the fractionated soybean oil body membrane and characterized as a 34 kD protein. It was later renamed Gly m 1 by Hessing *et al* (17). In our study, both wild and cultivated genotypes showed two spots, #16 and 17 of P34 (Fig. 1A, B). Similar numbers of P34 protein spots were also reported in *G.max cv* Jack using 2D-PAGE (9). These two spots showed very weak intensity in the wild genotype compared to the cultivated genotype (Fig. 1A, B). The similar Mr with the different pI observed in this investigation could be due to two possibilities. The protein could be modified by post-translational N-glycosylation, as has been reported by Bando *et al* (18). The other possibility is the presence of multiple genes. Kalinski *et al* (19) reported that P34 was encoded by more than one gene. The presence of two protein spots of P34 in this study was similar to the results of Gonzalez *et al* (20), who reported two isoforms using 2D-PAGE system. Yaklich *et al* (21) claimed there was no difference in P34 among wild and cultivated soybean genotypes based on one-dimensional SDS-PAGE analysis.

Analysis of Gly m Bd 28K proteins — Gly m Bd 28K is a less abundant protein of soybean but is designated as a major allergen and was originally identified as a 28 KD polypeptide in soybean seed flour (22). In our study, two spots (#18 and 19) of Gly m Bd 28K were identified by LC-MS/MS in both wild and cultivated genotypes. The Gly m Bd 28K allergen is processed into two smaller polypeptides in the soybean seeds, an N-terminal polypeptide and a C-terminal polypeptide (23). Xiang *et al* (22) reported that C-terminal fragments of Gly m Bd 28K showed slightly stronger binding to IgE as compared to the N-terminal fragments of the protein. Tsuji *et al* (23) reported that Gly m Bd 28K is about twice as large as anticipated from its initial discovery as a 28 kD polypeptide by one-dimensional SDS-PAGE. The different Mr and pI of the two spots (#18 and 19) from our study indicated that this could be due to alternate splicing, a degraded product or to multiple gene products.

Analysis of trypsin inhibitors — Kunitz trypsin inhibitors (KTIs) are the major trypsin inhibitors found in soybean. KTIs are also antinutritional factors and are involved in respiratory hypersensitivity reactions (6). In our study, three protein spots (#20, 21 and 22) for KTIs were detected in the wild genotypes (Fig. 2A), while only two spots (#20 and 21) were observed in the cultivated genotypes (Fig. 2B). The overall intensities of KTI spots were weak in the cultivated seeds when compared to wild genotypes. This finding raises the possibility that a greater abundance of KTIs in the wild soybean may enhance the plant's defenses by acting as a protective agent against soil pathogens and insects (24). Ritt *et al* (25) reported significant differences in the quantity of KTIs between different genotypes of soybeans. Three KTI genes have been reported (KTI1, KTI2 and KTI3). All three are expressed during embryogenesis and in the mature plant (26). Our 2D-PAGE results showed that spot #20 is the most abundant KTI, which was encoded by the KTI 3 gene (gil3318877). Jofuku and Goldberg (26) reported that the KTI3 transcript was detected only in the soybean seed, while KTI1 and KTI2 transcripts are expressed in leaf, root and stem. KTI1 and KTI2 genes have nearly identical nucleotide sequences and the KTI3 gene has diverged (20%) from the other 2 genes (26). Partial sequences of tryptic peptides from spot #21 (gil125722) matched to KTI1. On the other hand, spot #22 was only found in the wild genotype and matched to the KTI3 gene with low MOWSE score. Spot #22 has a different pI and Mr than spot #20. Therefore, there is a possibility that spot #22 could be a product of post-translational modification. Alternatively, the polypeptides could be encoded by different genes, which have not been reported to the database yet.

Analysis of lectin — Seed lectins are abundant allergen proteins and account for 10% of total protein in some legumes (27). Wild type soybeans showed two spots (#23 and 24, Fig. 2A) of lectin, whereas the cultivated genotype showed three spots (#23, 24 and 25, Fig 2B). Spots # 24 and 25 have similar Mr with different pI, which indicates that these two polypeptides may be the products of post-translational modification and matched with one gene (gil3114258). Spot # 23 has a higher Mr and a pI compared
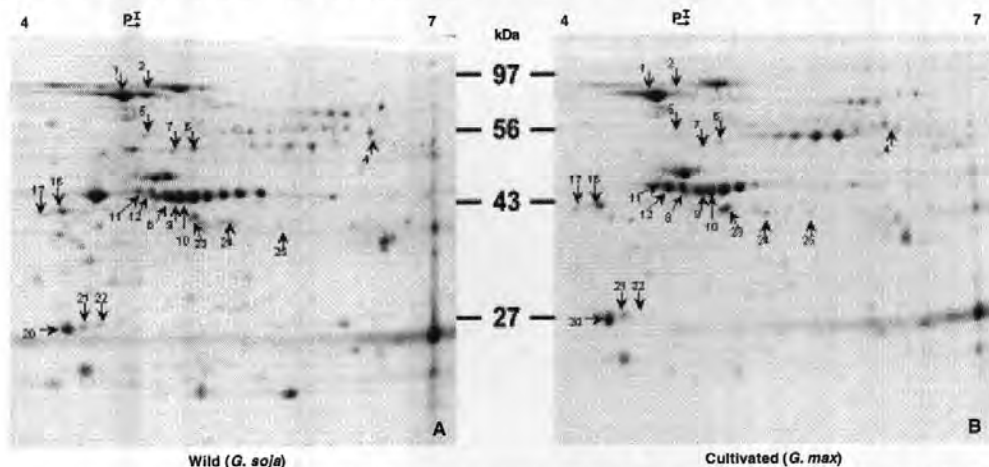

Fig.2. A proteomic comparison of the allergen and antinutritional proteins of wild, *Glycine soja* (**A**) and cultivated, *G. max* (**B**). The first dimension was run using a pH gradient from 4.0-7.0 and the second dimension is a 12% SDS-PAGE. Numbered arrows indicate the polypeptides referred to in the text and table.

to spot #24 and 25 and could be from a different gene (gil282898)

We have successfully separated and characterized most of the allergen and antinutritional proteins at the subunit level in both wild and cultivated soybean seeds using a proteomic approach. Our results show that analysis of proteins using two pH ranges in 2D-PAGE combined with a MALDI-TOF-MS/ LC-MS/MS is effective for the separation and identification of both abundant and non-abundant soybean allergen and antinutritional proteins. We observed that the wild genotype has more abundant allergenic and antinutritional proteins with a higher number of polypeptide subunits than the cultivated genotype. This may be because some cultivated genotypes have been bred to yield products with low allerginicity. These combined proteomic tools could be applied to accurately screen the allergenic proteins of a wide range of natural and genetically modified soybeans to identify those genotypes with the least potential allergenicity.

Acknowledgements

References

1   Bardare M, Magnolfi C & Zani G, *Allergy Immunol Paris*, **20** (2004) 66.

2   Metcalfe DD, Astwood JD, Townsend R, Sampson HA, Taylor SL & Fuchs RL, *Crit Rev Food Sci Nut Suppl*, **36** (1996) 165.

3   Senne GE, Crivellaro M, Bonadonna P, Dama A, Mezzelani P & Passalacqua G, *Allergy*, **53** (1998) 1106.

4   Ogawa T, Moriyama T, Samoto M & Takahashi K, *V11 World Soybean Research Proceedings*, (2004) 769.

5   Gorg AC, Obermaier G, Boguth A, Harder B, Scheibe R, Wildgruber & Weiss W, *Electrophoresis*, **21** (2000) 1037.

6   Besler M, Helm RM & Ogawa T, *Internet Sym On Food Allergens*, **2** (2000) p 35.

7   Ogawa T, Samoto M & Takahashi K, *J Nutr Sci Vitaminol*, **46** (2000) 271.

8   Mooney BP & Thelen JJ, *Phytochemistry*, **65** (2004) 1733.

9   Herman EM, Helm RM, Jung R & Kinney AJ, *Plant Physiol*, **132** (2003) 36.

10 Natarajan S, Xu C, Caperna TJ & Garrett WM, *Anal Biochem*, **342** (2005) 214.

11 Bradford MM, *Anal Biochem*, **72** (1976) 248.

12 Thanh VH & Shibasaki K, *J Agric Food Chem*, **24** (1976) 1117.

13 Davies H, *Isoelectric focusing*, (JB Arbuthnott, JA Beeley, Editors), Butterworths, London, (1975) p 97.

14 Helm RM, Cockrell G, West CM, Herman EM, Sampson HA, Bannon GA & Burks AW, *J Allergy Clin Immunol*, **105** (2000) 378.

15 Beardslee TA, Zeece MG, Sarath G & Markwell JP, *Int Arch Allergy Immunol*, **123** (2000) 299.

16 Kalinski A, Melroy DL, Dwivedi RS & Herman EM, *J Biol Chem*, **267** (1992) 12068.

17 Hessing M, Bleeker H, Tsuji H, Ogawa T & Vlooswijk RAA, *Agri Immunol*, **8** (1996) 51.

18 Bando N, Tsuji H, Yamanishi R, Nio N & Ogawa T, *Biosci Biotech Biochem*, **60** (1996) 347.

19 Kalinski A, Weisemann JM, Matthews BF & Herman EM, *J Biol Chem*, **265** (1990) 13843.

20 Gonzalez R, Polo F, Zapatero L, & Caravaca FC, *J Clin Exp Allergy*, **22** (1992) 748.

21 Yaklich R, Helm R& Herman E, *Crop Sci*, **39** (1999) 1444.

22 Xiang P, Haas EJ, Zeece MG, Markwell J & Sarath G, *Planta*, **220** (2004) 56.

23 Tsuji H, Hiemori M, Kimoto M, Yamashita H, Kobatake R, Adachi M, Fukuda T, Bando N, Okita M & Utsumi S, *Biochem Biophys Acta*, **1** (2001) 178.

24 Marchetti S, Delledonne M, Fogher C, Chiaba C, Chiesa F, Savazzina F & Giordano A, *Theor Appl Genet*, **101** (2000) 519.

25 Ritt AB, Mulinari F, Vasconcelos IM & Carlini CR, *J Sci Food & Agri*, **84** (2004) 263.

26 Jofuku KD & Goldberg RB, *The plant Cell*, **1** (1998) 1079.

27 Barnett D, Howden ME & Aller J, *Clin Immunol*, **80** (1987) 558.