



In silico assessment of the potential allergenicity of transgenes used for the development of GM food crops

Ankita Mishra^a, S.N. Gaur^b, B.P. Singh^a, Naveen Arora^{a,*}

^a Institute of Genomics and Integrative Biology (CSIR), North Campus, Delhi 110007, India

^b Department of Respiratory Medicine, Vallabhbhai Patel Chest Institute, University of Delhi, Delhi 110007, India

ARTICLE INFO

Article history:

Received 30 March 2011

Accepted 4 February 2012

Available online 14 February 2012

Keywords:

Food allergy

Genetically modified foods

Transgenes

Sequence homology

Safety assessments

ABSTRACT

Genetically modified (GM) crops require allergenicity and toxicity assessment of the novel protein(s) to ensure complete safety to the consumers. These assessments are performed in accordance with the guidelines proposed by [Codex \(2003\)](#) and [ICMR \(2008\)](#). The guidelines recommend sequence homology analysis as a preliminary step towards allergenicity prediction, later *in vitro* experiments may be performed to confirm allergenicity. In the present study, an *in silico* approach is employed to evaluate the allergenic potential of six transgenes routinely used for the development of GM food crops. Among the genes studied, manganese superoxide dismutase (MnSOD) and osmotin shares greater than 90% identity with Hev b 10 and Cap a 1w, respectively. Chitinase shares greater than 70% identity with allergens namely Pers a 1 and Hev b 11, and fungal chitinase showed significant IgE binding with 7 of 75 patients' sera positive to different food extracts. Glucanases (alfalfa, wheat) and glycine betaine aldehyde dehydrogenase gene share 50% homology with allergens like – Ole e 9, Cla h 10 and Alt a 10. The results demonstrate the allergenic potential of six genes and can serve as a guide for selection of transgenes to develop GM crops.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Food allergies are estimated to affect 1–2% of the adult population and 6–8% of children ([Sampson, 2004](#)). It is an adverse immune response to a food component (protein) mediated by IgE antibody and is characteristic of type 1-hypersensitivity reaction. The allergenic protein binds to IgE antibodies on the surface of mast cells or basophils, thereby stimulating histamine release and other mediators. Recent studies show that allergenic proteins from different food sources possess similarity in sequence and structure, which accounts for cross reactivity ([Schein et al., 2007](#)). Studies suggest that milk, egg, peanut, tree nuts, fish, shellfish, soy and wheat account for 90% of all food-allergic reactions worldwide. They can be different depending upon the dietary habit, like celery in some European countries, buckwheat in Southeast Asia and

Abbreviations: 3-D, three dimensional; BADH, betaine aldehyde dehydrogenase gene; FAO/WHO, Food and Agriculture Organization of the United Nations/World Health Organisation; FARRP, Food Allergy Research and Resource Program; gbs A, glycine betaine aldehyde dehydrogenase gene; IGMORIS, Indian GMO Research Information System; MnSOD, manganese superoxide dismutase gene; OECD, Organization for Economic Cooperation and Development; PR, pathogenesis-related; SDAP, Structural Database of Allergenic Proteins; SPT, skin prick test.

* Corresponding author. Address: Room 509, Allergy and Immunology Section, Institute of Genomics and Integrative Biology, Delhi University Campus, Mall Road, Delhi 110007, India. Tel.: +91 (0) 11 27666157; fax: +91 (0) 11 27667471.

E-mail address: naveen@igib.res.in (N. Arora).

sesame in Middle East ([Wuthrich et al., 1990](#)). GM foods are developed for human or animal consumption by introducing a foreign gene from diverse sources using recombinant DNA technology. They are required to undergo safety evaluation that includes the assessment of the potential allergenicity of the novel proteins expressed in the GM products. However, it is advised not to use any transgene from an allergenic source. The organizations like OECD, 1993 ([Organization for Economic Cooperation and Development, 1993](#)) FAO/WHO, 2001 ([Food and Agriculture Organization/World Health Organization, 2001](#)); Codex, 2003 ([Codex Alimentarius Commission, 2003](#)) and ICMR, 2008 ([Indian Council of Medical Research, guidelines on assessing GM crops, 2008](#)) have defined guidelines and protocols to be followed for the safety evaluation. In previous studies, bioinformatics tools were employed for predicting the potential allergenicity of novel proteins ([Goodman and Wise, 2006](#); [Singh et al., 2006](#)). Amino acid sequence homology is the preliminary step towards allergenicity assessment that provides information regarding the cross reactivity between the query sequence and the known allergen(s) ([Singh et al., 2008](#)). The present study was aimed to assess the allergenic potential of transgenes using *in silico* approach. These transgenes were randomly selected from IGMORIS (Indian GMO research information system, <http://igmoris.nic.in/>); a database which keeps a record of all the GMO(s) being developed in India. Among the transgenes assessed, osmotin, glucanases and chitinase belong to the family of pathogenesis-related (PR) proteins associated with defense of the plant against

fungal infection. While superoxide dismutase and glycine betaine aldehyde dehydrogenase genes provide resistance against oxidative and osmotic stress, respectively.

2. Materials and methods

Sequence homology approach involves comparison of the query sequence to each of the member sequence(s) in the allergen databases. Here three databases namely Allermatch (Fiers et al., 2004), Allergenonline (FARRP) and Structural Database of Allergenic Proteins (SDAP) (Ivanciuc et al., 2003) were used for allergenicity prediction of transgenes.

2.1. Selection of the query protein sequences

Six (transgene) protein sequences *Bacillus subtilis* glycine betaine aldehyde dehydrogenase (gbs A) (Accession No. ABS84664); *Triticum aestivum* beta-1, 3-glucanase (ACF33176); *Medicago sativa* beta-1, 3-glucanase (ABD91577); *Oryza sativa* chitinase (AAC37516); *Nicotiana plumbaginifolia* mitochondrial manganese superoxide dismutase (MnSOD) (P11796) and *Nicotiana tabacum* osmotin (P14170) served as query sequences. The amino acid sequences for the genes were retrieved from NCBI (<http://www.ncbi.nlm.nih.gov/>).

2.2. Search mode used for allergenicity prediction studies

FAO/WHO (2001) and Codex (2003) recommended the use of 80 amino acid sliding window search for the allergenicity prediction. The search modes used in the present study are described below:

2.2.1. Sequence database search for full length FASTA alignment

Full length FASTA alignment is not recommended by FAO/WHO and Codex guidelines but it is performed along with 80 amino acid sliding window as a supplement to the analysis. In full length FASTA alignment, we evaluated the extent of similarity on the basis of percent identity and *E*-value (expectation value). The *E*-value reflects the degree of similarity between pairs of sequences based on matches of identical or functionally similar amino acids. Very small *E*-values indicate probable evolutionary homology, and structural similarity. Proteins sharing less than 50% identity across the full length of the amino acid sequences are less likely to be cross reactive, while proteins sharing greater than 70% identity cause cross reactions (Aalberse, 2000).

2.2.2. Sequence database search for 80 amino acid sliding window

The query protein sequences were evaluated for the potential cross reactivity with known allergens in databases. FASTA alignment was performed to compare all possible contiguous amino acid segments of each of the 6 proteins against the sequences listed in the databases. Each contiguous 80 amino acid sequence of individual protein was searched starting with 1–80 aa, then 2–81 aa, and so on until the last 80 amino acid segment of each protein is compared with database. In this search, % (percent) identities were calculated to evaluate possible cross allergenicity. An alignment of >35% identity over segments as short as 80 amino acids with known allergen(s) depicts that the query sequence may be a potential allergen and should be subjected to further testing.

2.3. The allergen databases visited

2.3.1. Allermatch

(<http://allermatch.org/>) is build around the FASTA package version 3.4t21; <ftp://ftp.virginia.edu/pub/fasta/>, running with default parameters (ktup = 2, matrix = Blossum50 (Henikoff and Henikoff, 1992), Gap open = –10, Gap extend = –2). This web tool and the web interface are written in Python and run on a Suse Linux Enterprise server with an Apache web server (version 1.3.26). This database contains sequences from three databases i.e. SwissProt/TrEMBL, WHO-IUIS and a combination of these two databases. SwissProt database was last updated on December 20, 2005 and contains 368 sequences derived from SwissProt 48.7. While the WHO-IUIS database was last updated on September 12, 2005 and contains 681 sequences including 430 on the allergen list and 251 on the isoallergen list. The Allermatch database was last updated on January 10, 2006 and contains 1049 sequences.

2.3.2. Food Allergy Research and Resource Program (FARRP) AllergenOnline database version 11.0 (updated on 8 February 2011)

Is a peer-reviewed database, updated every year (Goodman and Hefle, 2005) and is available at <http://www.Allergenonline.org/>. This curated database is maintained by the Food Allergy Research and Resource Program of the University of Nebraska. FASTA 35 algorithm is used as a sequence alignment tool, with the default scoring matrix BLOSUM 50. The penalty for each gap inserted is calculated as $-q + -r * k$,

where q (10) is an initial penalty for each independent gap, r (2) is a penalty for each amino acid position within the gap and k is the number of amino acid positions within the gap (Reese and Pearson, 2002). The default word size (ktup) is two.

2.3.3. Structural database of allergenic proteins

(SDAP) is a web server controlled by a set of CGI scripts. The database component of SDAP <http://fermi.utmb.edu/SDAP/> is implemented with MySQL under Linux. The list of allergens in the database are assembled from all allergens in IUIS (International Union of Immunological Societies) website, literature, and from major sequence databases – Swiss-Prot/TrEMBL, PIR-ALN, NCBI taxonomy browser and Protein Data Bank. In addition to FAO/WHO allergenicity testing for new proteins, SDAP database evaluates the potential cross reactivity between known allergens and also predicts the IgE-binding potential of genetically modified food proteins. This allergen database is updated every three months.

2.4. Allergen extracts and sera collection

Food extracts for allergy diagnosis were procured from All Cure Pharma Private Limited, New Delhi, approved by Drug Controller. Patients with allergic rhinitis and asthma aged 15–50 years were skin prick tested with various pollens, fungi, insects, and food allergen extracts to detect sensitization. The study protocol was approved by human ethics committee of the institute. The diagnosis of asthma was made following guidelines of the American Thoracic Society. The patients having any two of the symptoms, viz. sneezing, rhinorrhea, nasal blockage, postnasal drip, etc., for most of the time in the last 2 years were diagnosed with rhinitis. Blood samples (sera) were collected from patients showing marked positive skin reactions to different food extracts at the outpatient department, V.P. Chest Institute, Delhi, a referral chest hospital.

2.5. Chitinase protein

Trichoderma viride chitinase used for IgE serum screening studies was obtained from SIGMA-Aldrich (St. Louis, MO, USA) having a percent purity of 84%.

2.6. Specific IgE estimation

Specific IgE against chitinase protein was estimated by ELISA (enzyme-linked immunosorbent assay). Briefly, microtiter plates (Nunc, Roskilde, Denmark) were coated with 5 µg of chitinase protein in 100 µl per well in carbonate buffer (pH 9.6) overnight at 4 °C. After they were washed with PBST (phosphate-buffered saline-Tween 20) (0.05%), the nonspecific sites were blocked with 3% defatted milk for 3 h at 37 °C. The plate was washed again and incubated with diluted (1:10 v/v) patients' sera showing positive SPT to food extracts overnight at 4 °C. The serum from non allergic individuals served as negative control. The plate was washed and incubated with antihuman IgE – horse radish peroxidase (1:1000 v/v, SIGMA-Aldrich (St. Louis, MO, USA) for 3 h at 37 °C. The plate was washed with PBS-Tween 20 and followed by PBS, and color was developed with orthophenylene diamine. The reaction was stopped by adding 5 N H₂SO₄, and the absorbance/O.D. (optical density) was read at 492 nm.

2.7. Statistical analysis

Statistical analysis for the specific IgE values of the patients showing positive SPT to food extracts and normal human sera (taken as control) against chitinase was performed using Epiinfo. A *P* value ≤0.05 was considered as statistically significant.

3. Results

Bioinformatic tools were used to analyze query sequences (proteins) in three databases. Sequence identity of >35% between query sequence and the known allergens was considered significant as per FAO/WHO, 2001 and Codex, 2003 guidelines.

3.1. Manganese superoxide dismutase gene

The length of the protein sequence is 228 amino acids. Full length FASTA alignment in three databases showed sequence identity of 85.09%, 86.207% and 86.20% with *Hevea brasiliensis* (Table 1). Thus the query protein may be a potent allergen since it showed very high identity with Hev b 10.0101, Hev b 10.0102, Hev b 10.0103 and *H. brasiliensis* MnSOD. Different isoforms of *H. brasiliensis* – Hev b 10 (Hev b 10.0102 & Hev b 10.0103) and *Aspergillus*

Table 1

Nicotiana plumbaginifolia Manganese (Mn) Superoxide dismutase gene shares sequence homology with the reported allergens.

Allergens	Full length FASTA alignment		80 aa sliding window – sequence identity %
	% Identity	E-value	
SDAP			
<i>Hevea brasiliensis</i> *	85.09 (194/228)	3.2e–90	91.25
Hev b 10.0101			
<i>Hevea brasiliensis</i>	76.75 (175/228)	2.3e–84	88.75
Hev b 10.0102			
<i>Hevea brasiliensis</i>	76.75 (175/228)	2.7e–84	–
Hev b 10.0103			
<i>Aspergillus fumigatus</i> #	43.42 (99/228)	8.6e–42	–
Asp f 6			
<i>Aspergillus fumigatus</i> Asp f 6	42.98 (98/228)	8.2e–42	–
<i>Pistacia vera</i>	81.14 (185/228)	5.2e–89	88.75
Pis v 4.0101			
<i>Malassezia sympodialis</i>	42.98 (98/228)	2.5e–41	–
Mala s 11			
* <i>Hevea brasiliensis</i> (Three isoforms)			
# <i>Aspergillus fumigatus</i> (Two isoform)			
Allermatch			
<i>Hevea brasiliensis</i> *	82.906	1.2e–91	91.25
Hev b 10.0101			
<i>Hevea brasiliensis</i>	86.207	1.1e–85	90.00
Hev b 10.0102			
<i>Hevea brasiliensis</i>	86.207	1.4e–85	90.00
Hev b 10.0103			
<i>Aspergillus fumigatus</i>	49.246	1.1e–42	61.25
Asp f 6			
<i>Malassezia sympodialis</i>	43.946	4.7e–42	55.00
Mala s 11			
* <i>Hevea brasiliensis</i> (Two isoforms)			
FARRP			
<i>Hevea brasiliensis</i>	86.20	1.5e–086	90.00
MnSOD			
<i>Hevea brasiliensis</i> superoxide dismutase (manganese)	82.9	1.4e–092	91.20
<i>Hevea brasiliensis</i> IgE-binding protein	86.2	1.8e–086	90.00
MnSOD			
<i>Aspergillus fumigatus</i> manganese superoxide dismutase	49.20	7.5e–043	61.30
<i>Aspergillus fumigatus</i> Superoxide dismutase	49.2	5.1e–043	61.30
Asp f 6			
<i>Malassezia sympodialis</i> MnSOD	43.90	2.2e–042	55.00
MnSOD			
* <i>Hevea brasiliensis</i> (Three isoforms)			
# <i>Aspergillus fumigatus</i> (Two isoform)			

fumigatus – Asp f 6 allergens are known and included in FARRP and SDAP databases. E-values for different alignments between the query sequence and known allergens are 1.2e–91 (Hev b 10), 5.2e–89 (Pis v 4), 1.1e–42 (Asp f 6) and 4.7e–42 (Mala s 11). These values are indicative of high homology in the primary sequence of MnSOD query sequence and reported allergens like Hev b 10.0101, Pis v 4.0101, Asp f 6 and Mala s 11. The 80 amino acid sliding window search showed that query sequence shares greater than 35% identity with *H. brasiliensis* (90% and above), *Pistacia vera* (88.75%), *A. fumigatus* (61.25% and above) and *Malassezia sympodialis* (55%) (Table 1). All the search parameters used for sequence homology analysis in three databases suggest mitochondrial manganese superoxide dismutase to be an allergen having clinically

important cross reactivity with reported allergens and may be considered as a candidate for pan allergen group MnSOD.

3.2. Rice chitinase gene

The length of the protein sequence is 175 amino acids. FASTA alignment (full length) showed sequence identity of 68.57% (Hev b 11.0102), 63.102% (Pers a 1) and 63.40% (chitinase Ib) with known allergens (Table 2). Thus analysis showed high degree of identity between rice chitinase and allergens from rubber and chestnut, class I chitinases and avocado endochitinase. E-values of 2.8e–47, 1e–47 and 1.7e–046 for different alignments between query sequence and reported allergens indicate probable evolutionary homology between the sequences compared. High sequence homology is predictive of homologous 3D structures and thereby potential cross reactivity. The 80 amino acid sliding window search showed that query sequence shares high identity (>35%) with chitinase from *H. brasiliensis*, *Persea americana* (avocado) and *Castanea sativa* (chest nut) allergens (Table 2).

3.3. Glycine betaine aldehyde dehydrogenase gene

The length of the protein sequence is 490 amino acids. FASTA alignment (full length) showed sequence identity values of 40.00%, 41.02% and 43.70% with known allergens (Table 3). The allergens namely Cla h 10 and Alt a 10 encode for aldehyde dehydrogenase enzyme from different fungal sources – *Cladosporium herbarum* and *Alternaria alternata*, respectively. E values for the sequence alignment between the query sequence and the fungal

Table 2

Rice chitinase gene share sequence identity with the allergens listed in three databases namely – SDAP, Allermatch and FARRP.

Allergens	Full length FASTA alignment		80 aa sliding window – sequence identity %
	% Identity	E-value	
SDAP			
<i>Persea americana</i>	67.43 (118/175)	2.8e–47	76.25
Pers a 1			
<i>Musa acuminata</i>	65.14 (114/175)	1.7e–45	61.25
Mus a 2.0101			
<i>Hevea brasiliensis</i> *	68.57 (120/175)	2.2e–44	72.50
Hev b 11.0102			
<i>Hevea brasiliensis</i>	68.00 (119/175)	1.0e–43	–
Hev b 11.0101			
* <i>Hevea brasiliensis</i> (Two isoforms)			
Allermatch			
<i>Persea americana</i>	63.102	1.6e–49	76.25
Pers a 1			
<i>Castanea sativa</i>	63.441	1e–47	71.25
Cas s 5			
<i>Hevea brasiliensis</i> *	63.83	1.7e–46	70.37
Hev b 11.0102			
<i>Hevea brasiliensis</i>	63.298	8.6e–46	69.14
Hev b 11.0101			
* <i>Hevea brasiliensis</i> (Two isoforms)			
FARRP			
<i>Persea americana</i>	63.10	4.2e–045	76.25
Endochitinase			
<i>Castanea sativa</i> chitinase Ib	63.40	1.8e–043	71.30
<i>Hevea brasiliensis</i> *	63.80	2.5e–042	70.40
class I chitinase			
<i>Hevea brasiliensis</i> putative class I chitinase	63.3	1.1e–041	69.10
<i>Cryptomeria japonica</i> class IV chitinase	34.9	4.5e–014	41.24
* <i>Hevea brasiliensis</i> (Two isoforms)			

Table 3

Bioinformatics analysis of gbsA gene using full length FASTA alignment and 80 amino acid sliding window search.

Allergens	Full length FASTA alignment		80 aa sliding window – sequence identity %
	% Identity	E-value	
SDAP			
<i>Alternaria alternata</i>	40.00 (196/490)	1.8e–77	52.50
Alt a 10			
<i>Cladosporium herbarum</i>	39.80 (195/490)	8.1e–79	50.00
Cla h 10			
<i>Harmonia axyridis</i>	20.61 (101/490)	7.6e–27	–
Har a 2.0101			
Allermatch			
<i>Cladosporium herbarum</i> *	41.02	2.6e–81	51.25
Cla h 10			
<i>Alternaria alternata</i>	41.35	6.3e–80	51.85
Alt a 10			
<i>Cladosporium herbarum</i>	41.02	2.6e–81	51.25
Cla h 3			
*Cla h 3 is an isoform of Cla h 10 and Allermatch shows the same accession No. for both.			
FARRP			
<i>Alternaria alternata</i> (NAD+)	43.70	2.1e–080	54.99
<i>Davidiella tassiana</i>	42.9	5e–080	51.29
Cla h 3			
<i>Plantago lanceolata</i>	27.7	0.99	–
Pla l 1.0101			

Table 4

Analysis of the allergenic potential of Alfalfa glucanase using two search modes – full length FASTA alignment and 80 amino acid sliding window.

Allergens	Full length FASTA alignment		80 aa sliding window – sequence identity %
	% Identity	E-value	
SDAP			
<i>Olea europea</i>	34.32 (174/507)	6.0e–74	51.25
beta-1,3-glucanase			
<i>Hevea brasiliensis</i>	26.04 (132/507)	3.3e–51	37.50
beta-1,3-glucanase			
<i>Olea europea</i>	7.89 (40/507)	8.4e–16	45.00
glycosyl hydrolase			
Allermatch			
<i>Olea europea</i>	38.242	1e–68	51.25
beta-1,3-glucanase			
<i>Hevea brasiliensis</i>	38.957	4e–46	49.38
beta-1,3-glucanase			
<i>Olea europea</i>	40.86	1.4e–14	45.00
Ole e 10			
FARRP			
<i>Olea europea</i>	38.20	5.9e–074	51.25
beta-1,3-glucanase			
<i>Olea europea</i>	40.90	1.2e–015	45.00
Ole e 10			
<i>Olea europea</i>	34.60	2.8e–013	40.00
Chain A, C-Terminal			
Domain Ole e 9			
<i>Hevea brasiliensis</i> *	39.10	4.6e–052	49.40
beta-1,3-glucanase			

Note: Seven isoforms are reported for *Hevea brasiliensis* (FARRP matches) and all have the same value for 80 amino acid sliding window search, therefore the best match is reported here for convenience.

allergens – Cla h 10 (2.6e–81), Cla h 3 (5e–080) and Alt a 10 (6.3e–80) indicate the significance of the match and structural similarity between the sequences. The 80 amino acid sliding window search of gbs A protein revealed that query sequence shares $\geq 50\%$ identity with aldehyde dehydrogenase allergens (Table 3).

Table 5

Wheat glucanase gene share sequence homology with reported allergens. The parameters assessed are % identity and E – score using full length FASTA alignment and 80 amino acid sliding window.

Allergens	Full length FASTA alignment		80 aa sliding window – sequence identity %
	% Identity	E-value	
SDAP			
<i>Olea europea</i>	33.83 (135/399)	8.8e–51	50.00
beta-1,3-glucanase			
<i>Hevea brasiliensis</i>	29.82 (119/399)	1.5e–40	38.75
beta-1,3-glucanase			
Allermatch			
<i>Olea europea</i>	41.195	8.4e–47	50.00
beta-1,3-glucanase			
<i>Hevea brasiliensis</i>	36.335	8.8e–37	42.50
beta-1,3-glucanase			
FARRP			
<i>Olea europea</i>	41.2	1.2e–045	50.02
beta-1,3-glucanase			
<i>Hevea brasiliensis</i> *	37.4	4.3e–038	43.80
beta-1,3-glucanase			

Note: Seven isoforms are reported for *Hevea brasiliensis* (FARRP matches) therefore the best match is reported here, for convenience.

3.4. Alfalfa beta-1, 3-glucanase gene

The length of the protein sequence is 507 amino acids. Full length FASTA alignment in three databases revealed identity values of 34.32%, 38.96%, 40.86% and 40.90% with known allergens (Table 4). Ole e 9 (*Olea europea*) and Hev b 2 (*H. brasiliensis*) are major allergens encoding for beta-1, 3-glucanase enzyme, while Ole e 10 is a homolog of glycosyl hydrolase. E-values for different alignments between alfalfa glucanase and the reported allergens in all three databases are 6.0e–74 (Ole e 9), 4e–46 (Hev b 2), 5.9e–074 (Ole e 9) and 1.2e–015 (Ole e 10) indicating similarity between the aligned sequences. The 80 amino acid sliding window search showed that query sequence shares identity greater than 35% with beta-1, 3-glucanase enzyme allergens (Table 4). These findings indicate that query sequence (protein) may be a cross reactive allergen, but its putative allergenicity should be validated further.

3.5. Wheat beta-1, 3-glucanase gene

The length of the protein sequence is 399 amino acids. Full length FASTA alignment showed identity values of 33.83%, 36.33% and 41.20% with known allergens (Table 5). Ole e 9 and Hev b 2, are major allergens encoding for beta-1, 3-glucanase enzyme from *O. europea* (Olive pollen) and *H. brasiliensis* (rubber latex), respectively. The E-values for the wheat glucanase and olive glucanase sequence alignment are 8.8e–51, 8.4e–47 and 1.2e–045 in all three databases illustrating significance of the match and possible cross reactivity of wheat glucanase with olive allergens. Similarly E values for the alignment between query protein and the *H. brasiliensis* glucanase gene showed significant match and thereby cross reactivity. The 80 amino acid sliding window search of protein revealed that query sequence shares greater than 35% identity with beta-1, 3-glucanase enzyme allergens (Table 5).

3.6. Osmotin gene

The length of the protein sequence is 246 amino acids. Full length FASTA alignment showed identity values of 91.87% (Lyc e NP24), 88.21% (Cap a 1w) and 66.50% (Act d 2) with known allergens (Table 6). E-values for different alignments i.e. 1.2e–93, 3.4e–91 and 9.4e–061 indicate that the query sequence may share homology with the reported allergens. Further, 80 amino acid

Table 6

Nicotiana tabacum Osmotin gene shares sequence homology with the reported allergens.

Allergens	Full length FASTA alignment		80 aa sliding window – sequence identity %
	% Identity	E-value	
SDAP			
<i>Lycopersicon esculentum</i> Lyc e NP24	91.87 (226/246)	1.2e−93	97.50%
<i>Capsicum annuum</i> Cap a 1w	88.21 (217/246)	3.4e−91	–
<i>Actinidia deliciosa</i> Act d 2	60.57 (149/246)	7.9e−60	–
Allermatch			
<i>Capsicum annuum</i> Cap a 1	88.496	1.6e−90	93.75
<i>Juniperus ashei</i> Jun a 3	54.146	3e−45	63.75
<i>Actinidia chinensis</i> Act c 2	70.000	9.1e−06	–
FARRP			
<i>Actinidia deliciosa</i> * thaumatin-like protein	66.50	9.4e−061	75.00
<i>Juniperus ashei</i>	50.90	4.5e−043	63.79
<i>Cryptomeria japonica</i>	52.30	1.5e−043	62.51
<i>Cupressus sempervirens</i> Cup s 3.2	52.60	2.4e−042	61.30
<i>Malus x domestica</i> # thaumatin-like protein	42.30	1.1e−019	49.97

* *Actinidia deliciosa* (Two isoforms)

Malus x domestica (Five isoform)

Table 7

Specific IgE binding of Chitinase against Patients' sera.

Subject No.	Allergen (History, SPT positive)	Specific IgE (OD 492 nm)	
		Chitinase	Corresponding Allergen
1	Apple	0.482	0.452
2	Cowpea	0.409	0.392
3	Cowpea	0.437	0.421
4	Mustard	0.363	0.354
5	Orange	0.379	0.369
6	Tomato	0.376	0.421
7	Tomato	0.382	0.368
8	Control	0.121	0.119
9	Control	0.096	0.102
10	Control	0.114	0.113

Subject No. 1–7: Patients showing positive ELISA to Chitinase protein.

Subject No. 8–10: Non allergic healthy individuals.

sliding window search also showed that osmotin, which belongs to the PR5 family of pathogenesis related proteins share significant homology with food allergens like Cap a 1 (bell pepper), Lyc e NP24 (tomato) and Act d 2 (kiwi) (Table 6).

3.7. In vitro IgE reactivity

Specific IgE binding of chitinase (*T. viride*) protein was assessed with food allergic patients' sera by ELISA. A total of 75 patients' sera were screened and 7 samples showed specific IgE levels, i.e. ≥ 3 times of negative control (Table 7). Statistically significant difference ($P \leq 0.05$) was observed between IgE binding of the patients and the control sera samples.

4. Discussion

Food allergens have gained importance because of the severe systemic reactions upon inhalation or ingestion in many cases

(Helm and Burks, 2000). Food allergy, in particular with peanut has shown an increasing trend in North America and Europe (Grundy et al., 2002; Sicherer et al., 2003). Food allergy appears to be on the rise in Asia (Chandra, 2003), but there has been lack of data from Asia on the local prevalence and clinical presentation of food associated allergic diseases (Shek and Lee, 2006). The factors accounting for the allergenicity of proteins are B-cell and T-cell epitopes, resistance to proteolysis, degree of glycosylation and the inherent enzymatic activity. Genetic engineering aims at developing crops having an edge over the conventionally breed crops. The selection of transgenes is an important parameter and is determined by the function of the gene. However, the allergenic potential of novel proteins should be explored to ensure the safety of individuals from GM crop products. In the present study, the allergenic potential of six transgenes used for GM crops was assessed by *in silico* methods.

Sequence homology analysis was first proposed by ILSL/IFBC in 1996, to identify allergenic structures in novel proteins (Metcalfe et al., 1996). Allergen databases play a key role in homology studies and serve as reference for comparison with the query sequence. However, the databases have a limited number of reported allergens. SDAP has 1396 allergens and isoallergens whereas FARRP, Allergen Online Database, version 10.0 contains a list of 1471 known and putative allergenic proteins (food, airway, venom/salivary and contact). Similarly Allermatch has 368 sequences listed in the Swiss Prot database and 681 protein sequences listed in the WHO–IUIS database. In order to lessen the probability of missing out the known allergens, for the present study, three allergen databases were used to analyze the query sequence(s). The analysis of a protein by FASTA algorithm predicts clinical cross reactivity with reported allergens (Hileman et al., 2002). In the present study, data obtained by full length FASTA alignment and 80 amino acid sliding window search correspond to potential IgE cross reactivity with known allergens. Both the approaches illustrate that superoxide dismutase, osmotin and rice chitinase gene share high identity with known allergens in the databases. Superoxide dismutase gene shares high identity with known allergens like Hev b 10, Pis v 4, Asp f 6 and Mala s 11. Previous studies have experimentally verified cross reactivity of Asp f 6, the MnSOD from *A. fumigatus*, human MnSOD and rHev b 10 (Wagner et al., 2001). Andersson and co-workers (Andersson et al., 2004) showed Mala s 11 shares high sequence homology with MnSOD from *Homo sapiens* (50%) and *A. fumigatus* (56%). Similarly, experimental studies have been performed to evaluate the allergenic potential of osmotin protein. Osmotin shares identity with common food allergens like Lyc e NP24, Cap a 1w and Act d 2 and *in vitro* IgE binding assay showed that it shares cross reactive epitopes with apple and tomato allergen (Sharma et al., 2011). Rice chitinase shares homology with rubber latex, avocado and chestnut allergens. Class I chitinases are major pan allergens associated with latex-fruit syndrome, while Pers a 1 possess IgE binding capacity and recognizes 15/20 of avocado- and/or latex-allergic patients (Andersson et al., 2004). Further, studies revealed chitinase-related proteins (the basic 25–35-kDa variants, PR-3 and PR-4) to be major source of cross-reactivity between rubber latex and fruits such as banana, avocado (Pers a 1), and chestnut (Alenius et al., 1995; Breiteneder and Ebner, 2000; Sowka et al., 1998). Around 50% of latex-allergic patients show hypersensitivity to fruits, while 86% of fruit allergic patients show sensitization to latex (Blanco et al., 1999; Garcia Ortiz et al., 1998). Alfalfa and wheat glucanase genes share homology with major Ole e 9 and Hev b 2. Yagami and co-workers reported that Ole e 9 and Hev b 2 share cross reactive epitopes and exhibit IgE binding with sera of patients allergic to tomato, potato and banana (Yagami et al., 1998).

The potential allergenicity of novel gene products can be determined to a reasonable degree by IgE serum screening studies. Here,

chitinase from *T. viride* demonstrated significant IgE binding with patients' sera positive to different foods extracts. These findings suggest that chitinase from fungi may share cross reactive epitopes with food allergens. However, the available information on the epitopic regions of chitinases is limited in accurately predicting its cross reactivity with novel proteins. Therefore, further experimental studies are required to confirm clinically important cross reactivity.

Studies suggest that most of the proteins showing cross reactivity with allergens have demonstrable homology. However, Beyer et al., reported lack of clinical cross reactivity in 11S globulins sequenced from hazelnut, peanut, brazil nut, cashew; having 45–55% identity (Beyer et al., 2002). Thus the bioinformatics data need to be validated with relevant experiments. However, the research groups involved in GM crop development should perform an initial homology search for allergenicity evaluation of the transgene. Other parameters of allergenicity assessment can be worked out later to confirm the findings of *in silico* analysis. In conclusion, manganese superoxide dismutase, osmotin, chitinase, glucanase and glycine betaine aldehyde dehydrogenase genes showed potential allergenic cross reactivity by *in silico* analysis. Further studies are required to validate these findings experimentally.

Conflict of Interest

The authors have no financial conflict of interest.

References

- Aalberse, R.C., 2000. Structural biology of allergens. *J. Allergy Clin. Immunol.* 106, 228–238.
- Alenius, H., Kalkkinen, N., Lukka, M., Reunala, T., Turjanmaa, K., Mäkinen-Kiljunen, S., Yip, E., Palosuo, T., 1995. Prohevein from the rubber tree (*Hevea brasiliensis*) is a major latex allergen. *Clin. Exp. Allergy* 25, 659–665.
- Andersson, A., Rasool, O., Schmidt, M., Kodzius, R., Flückiger, S., Zargari, A., Cramer, R., Scheynius, A., 2004. Cloning, expression and characterization of two new IgE-binding proteins from the yeast *Malassezia sympodialis* with sequence similarities to heat shock proteins and manganese superoxide dismutase. *Eur. J. Biochem.* 271, 1885–1894.
- Beyer, K., Grishina, G., Bardina, L., Grishin, A., Sampson, H.A., 2002. Identification of an 11S globulin as a major hazelnut food allergen in hazelnut-induced systemic reactions. *J. Allergy Clin. Immunol.* 110, 517–523.
- Blanco, C., Diaz-Perales, A., Collada, C., Sánchez-Monge, R., Aragoncillo, C., Castillo, R., Ortega, N., Alvarez, M., Carrillo, T., Salcedo, G., 1999. Class I chitinases as potential panallergens involved in the latex-fruit syndrome. *J. Allergy Clin. Immunol.* 103, 507–513.
- Breiteneder, H., Ebner, C., 2000. Molecular and biochemical classification of plant derived food allergens. *J. Allergy Clin. Immunol.* 106, 27–36.
- Chandra, R.K., 2003. Food hypersensitivity and allergic disease: a new threat in India. *Indian Pediatr.* 40, 99–101.
- Codex Alimentarius Commission. 2003. Alinorm 03/34: Joint FAO/WHO Food Standard Programme, Codex Alimentarius Commission, Appendix III, Guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants and Appendix IV, Annex on the assessment of possible allergenicity, 25th Session, Rome, Italy 30 June–5 July, 2003. pp. 47–60. http://ftp.fao.org/esn/food/guide_plants_en.p.
- Fiers, M. W., Kleter, G. A., Nijland, H., Peijnenburg, A. A., Nap, J. P., & van 279 Ham, R. C., 2004. Allermatch, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. *BMC Bioinformatics* 5, 133–138.
- Food and Agriculture Organization/World Health Organization. 2001. Evaluation of the Allergenicity of Genetically Modified foods: Report of a Joint FAO/WHO Expert Consultation, Rome, Italy, 22–25 January. <http://www.fao.org/docrep/007/y0820e/y0820e00.HTM>.
- Garcia Ortiz, J.C., Moyano, J.C., Alvarez, M., Bellido, J., 1998. Latex allergy in fruit – allergic patients. *Allergy* 53, 532–536.
- Goodman, R.E., Hefle, S.L., 2005. Gaining perspective on the allergenicity assessment of genetically modified food crops. *Expert Rev. Clin. Immunol.* 1, 561–578.
- Goodman, R.E., & Wise, J., 2006. Bioinformatics analysis of proteins in Golden rice2 to assess potential allergenic cross-reactivity. In: Food Allergy Research & Resource Program, Univ. of Nebraska (Study no. BIO-02-2006), 1–24.
- Grundy, J., Matthews, S., Bateman, B., Dean, T., Arshad, S.H., 2002. Rising prevalence of allergy to peanut in children: data from 2 sequential cohorts. *J. Allergy Clin. Immunol.* 110, 784–789.
- Helm, R.M., Burks, A.W., 2000. Mechanisms of food allergy. *Curr. Opin. Immunol.* 12, 647–653.
- Henikoff, S., Henikoff, J.G., 1992. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* 89, 10915–10919.
- Hileman, R.E., Silvanovich, A., Goodman, R.E., Rice, E.A., Hollerschak, G., Astwood, J.D., Hefle, S.L., 2002. Bioinformatic methods for allergenicity assessment using a comprehensive allergen database. *Int. Arch. Allergy Immunol.* 128, 280–291.
- Indian guidelines on assessing GM crops, published in 2008 by the Indian Council of Medical Research. <http://igmoris.nic.in/files%5CCoverpage.pdf>.
- Ivanciuc, O., Schein, C.H., Braun, W., 2003. SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res.* 31, 359–362.
- Metcalfe, D. D., Astwood, J. D., Townsend, R., Sampson, H. A., Taylor, 306 S. L., & Fuchs, R. L., 1996. Assessment of the allergenic potential of foods from genetically engineered crop plants. *Crit. Rev. Food Sci. Nutr.* 36, 165–186.
- OECD. (1993). Safety evaluation of foods derived by modern biotechnology concepts and principles. Organisation for Economic Cooperation and Development, Paris.
- Reese, J.T., Pearson, W.R., 2002. Empirical determination of effective gap penalties for sequence comparison. *Bioinformatics* 18, 1500–1507.
- Sampson, H.A., 2004. Update on food allergy. *J. Allergy Clin. Immunol.* 113, 805–819.
- Schein, C.H., Ivanciuc, O., Braun, W., 2007. Bioinformatics approaches to classifying allergens and predicting cross-reactivity. *Immunol. Allergy Clin. North Am.* 27, 1–27.
- Sharma, P., Singh, A.K., Singh, B.P., Gaur, S.N., Arora, N., 2011. Allergenicity assessment of osmotin, a pathogenesis-related protein, used for transgenic crops. *J. Agric. Food Chem.* 59, 9990–9995.
- Shek, L.P., Lee, B.W., 2006. Food allergy in Asia. *Curr. Opin. Allergy Clin. Immunol.* 6, 197–201.
- Sicherer, S.H., Munoz-Furlong, A., Sampson, H.A., 2003. Prevalence of peanut and tree nut allergy in the United States determined by means of a random digit dial telephone survey: a 5-year follow up study. *J. Allergy Clin. Immunol.* 112, 1203–1207.
- Singh, A.K., Mehta, A.K., Sridhara, S., Gaur, S.N., Singh, B.P., Sarma, P.U., Arora, N., 2006. Allergenicity assessment of transgenic mustard (*Brassica juncea*) expressing bacterial codA gene. *Allergy* 61, 491–497.
- Singh, A.K., Singh, B.P., Prasad, G.B., Gaur, S.N., Arora, N., 2008. Safety assessment of bacterial choline oxidase protein introduced in transgenic crops for tolerance against abiotic stress. *J. Agric. Food Chem.* 56, 12099–12104.
- Sowka, S., Hsieh, L.S., Krebitz, M., Akasawa, A., Martin, B.M., Starrett, D., Peterbauer, C.K., Scheiner, O., Breiteneder, H., 1998. Identification and cloning of Prs a 1, a 32-kDa endochitinase and major allergen of avocado, and its expression in the yeast *Pichia pastoris*. *J. Biol. Chem.* 273, 28091–28097.
- Wagner, S., Sowka, S., Mayer, C., Cramer, R., Focke, M., Kurup, V.P., Scheiner, O., Breiteneder, H., 2001. Identification of a *Hevea brasiliensis* latex manganese superoxide dismutase (Hev b 10) as a cross reactive allergen. *Int. Arch. Allergy Immunol.* 125, 120–127.
- Wuthrich, B., Stager, J., Johansson, S.G., 1990. Celery allergy associated with birch and mugwort pollinosis. *Allergy* 45, 566–571.
- Yagami, T., Sato, M., Nakamura, A., Komiyama, T., Kitagawa, K., Akasawa, A., Ikezawa, Z., 1998. Plant defense-related enzymes as latex antigens. *J. Allergy Clin. Immunol.* 101, 379–385.