SUMMARY

(In accordance with 40 CFR part 152, this summary is available for public release after registration)

STUDY TITLE

Bioinformatics Evaluation of the Putative Reading Frames across the Whole T-DNA Insert and Junctions in DAS-44406-6 Soybean for Potential Protein Allergenicity and Toxicity

DATA REQUIREMENTS

Not Applicable

AUTHOR(S)

S. Guttikonda

STUDY COMPLETED ON

29-Mar-2012

PERFORMING LABORATORY

Regulatory Sciences and Government Affairs—Indianapolis Lab
Dow AgroSciences LLC
9330 Zionsville Road
Indianapolis, Indiana 46268-1054

LABORATORY STUDY ID

120481

Bioinformatics Evaluation of the Putative Reading Frames across the Whole T-DNA Insert and Junctions in DAS-44406-6 Soybean for Potential Protein Allergenicity and Toxicity

SUMMARY

Soybean (*Glycine max* L.) event DAS-44406-6 was generated by integration of the T-DNA from plasmid pDAB8264 through *Agrobacterium*-mediated transformation of a variety "Maverick". The T-DNA of plasmid pDAB8264 contains three plant transcription units (PTUs): 1) Histone H4A748 promoter, *2mepsps* gene, and Histone H4A748 3' UTR; 2) AtUbi10 promoter, *aad*-12 gene, and AtuORF23 3' UTR; 3) CsVMV promoter, *pat* gene, and AtuORF1 3' UTR. In addition, a RB7 matrix attachment region (RB7 MAR) is located at the 5' end of the T-DNA insert. Molecular characterization indicated that DAS-44406-6 soybean contained a single insert including three intact expression cassettes for 2mEPSPS, AAD-12, and PAT respectively. DNA sequences flanking the insert in event DAS-44406-6 soybean were also cloned and characterized. The DNA sequence of the insert is identical to the corresponding portion in the T-DNA insert of plasmid pDAB8264 except for an extra 3 bp insertion at the 5' junction. The whole T-DNA insert and its flanking borders were screened for putative reading frames within the T-DNA insert and spanning the junctions between the insert and its borders. A total of 12 across-junction putative reading frames and total of 651 within T-DNA putative reading frames were identified. All of the putative reading frames were evaluated for potential allergenicity and toxicity using bioinformatics tools. Searches of those putative reading frames against a peer reviewed allergen database (FARRP Allergen Database Version 12, Released in February, 2012) did not generate any significant amino acid sequence similarities with known allergens. Similarly, the search against the GenBank non-redundant (nr) protein dataset did not detect any significant protein sequence similarity with toxic proteins harmful to humans or animals.

STUDY TITLE

Bioinformatics Evaluation of the Putative Reading Frames across the Whole T-DNA Insert and Junctions in DAS-44406-6 Soybean for Potential Protein Allergenicity and Toxicity

DATA REQUIREMENTS

Not Applicable

AUTHOR(S)

S. Guttikonda 317-337-4493
[SKGuttikonda@dow.com]

STUDY COMPLETED ON

29-Mar-2012

PERFORMING LABORATORY

Regulatory Sciences and Government Affairs—Indianapolis Lab
Dow AgroSciences LLC
9330 Zionsville Road
Indianapolis, Indiana 46268-1054

LABORATORY STUDY ID

120481

STATEMENT OF NO DATA CONFIDENTIALITY CLAIMS

Compound: DAS-44406-6 Soybean

Title: Bioinformatics Evaluation of the Putative Reading Frames across the Whole T-DNA Insert and Junctions in DAS-44406-6 Soybean for Potential Protein Allergenicity and Toxicity

- STATEMENT OF NO DATA CONFIDENTIALITY CLAIMS:

No claim of confidentiality, on any basis whatsoever, is made for any information contained in this document. I acknowledge that information not designated as within the scope of FIFRA sec. 10(d)(1)(A), (B), or (C) and which pertains to a registered or previously registered pesticide is not entitled to confidential treatment and may be released to the public, subject to the provisions regarding disclosure to multinational entities under FIFRA sec. 10(g).

Company: Dow AgroSciences LLC

Company Agent: L. Taglaini

Title: Regulatory Manager

Signature: 

Date: 23-Mar-2012

THIS DATA MAY BE CONSIDERED CONFIDENTIAL IN COUNTRIES OUTSIDE THE UNITED STATES.

STATEMENT OF COMPLIANCE WITH GOOD LABORATORY PRACTICE STANDARDS

Title: Bioinformatics Evaluation of the Putative Reading Frames across the Whole T-DNA Insert and Junctions in DAS-44406-6 Soybean for Potential Protein Allergenicity and Toxicity

Study Initiation Date: 27/Feb/2012

This report represents data generated after the effective date of the EPA FIFRA Good Laboratory Practice Standards.

United States Environmental Protection Agency
Title 40 Code of Federal Regulations Part 160
FEDERAL REGISTER, August 17, 1989

Organisation for Economic Co-Operation and Development
ENV/MC/CHEM(98)17, Paris January 26, 1998

At the time this study was conducted, it was not subject to the Good Laboratory Practice Standards and was, therefore, not monitored by the quality assurance unit.

L. Tagliani
Sponsor
Dow AgroSciences LLC

23-Mar-2012
Date

L. Tagliani
Submitter
Dow AgroSciences LLC

23-Mar-2012
Date

S. Guttikonda
Study Director/Author
Dow AgroSciences LLC

29-Mar-2012
Study Completion Date

QUALITY ASSURANCE STATEMENT

Compound: DAS-44406-6 Soybean

Title: Bioinformatics Evaluation of the Putative Reading Frames across the Whole T-DNA Insert and Junctions in DAS-44406-6 Soybean for Potential Protein Allergenicity and Toxicity

Study Initiation Date: 27/Feb/2012 Study Completion Date: 29/Mar/2012

NON-GLP STUDY

SIGNATURE PAGE

22-Mar-2012
Date

S. Guttikonda
Author
Dow AgroSciences LLC

29-Mar-2012
Date

P. Song
Peer Reviewer
Dow AgroSciences LLC

22 Mar 2012
Date

M. Zhuang
Peer Reviewer
Dow AgroSciences LLC

22-Mar-2012
Date

R. A. Herman
Peer Reviewer
Dow AgroSciences LLC

22 Mar 2012
Date

G. Shan
Science Leader
Dow AgroSciences LLC

22-Mar-2012

S. P. Kumpatla
Global Leader, Biotechnology Regulatory Sciences
Dow AgroSciences LLC

STUDY PERSONNEL

Title: Bioinformatics Evaluation of the Putative Reading Frames across the Whole T-DNA Insert and Junctions in DAS-44406-6 Soybean for Potential Protein Allergenicity and Toxicity

Principal Analyst: S. Guttikonda
(Principle Investigator)

Analysts:

TABLE OF CONTENTS

Bioinformatics Evaluation of the Putative Reading Frames across the Whole T-DNA Insert and Junctions in DAS-44406-6 Soybean for Potential Protein Allergenicity and Toxicity

ABSTRACT

Soybean (*Glycine max* L.) event DAS-44406-6 was generated by integration of the T-DNA from plasmid pDAB8264 through *Agrobacterium*-mediated transformation of a variety "Maverick". The T-DNA of plasmid pDAB8264 contains three plant transcription units (PTUs): 1) Histone H4A748 promoter, *2mepsps* gene, and Histone H4A748 3' UTR; 2) AtUbi10 promoter, *aad*-12 gene, and AtuORF23 3' UTR; 3) CsVMV promoter, *pat* gene, and AtuORF1 3' UTR. In addition, a RB7 matrix attachment region (RB7 MAR) is located at the 5' end of the T-DNA insert. Molecular characterization indicated that DAS-44406-6 soybean contained a single insert including three intact expression cassettes for2mEPSPS, AAD-12, and PAT respectively. DNA sequences flanking the insert in event DAS-44406-6 soybean were also cloned and characterized. The DNA sequence of the insert is identical to the corresponding portion in the T-DNA insert of plasmid pDAB8264 except for an extra 3 bp insertion at the 5' junction. The whole T-DNA insert and its flanking borders were screened for putative reading frames within the T-DNA insert and spanning the junctions between the insert and its borders. A total of 12 across-junction putative reading frames and total of 651 within T-DNA putative reading frames were identified. All of the putative reading frames were evaluated for potential allergenicity and toxicity using bioinformatics tools. Searches of those putative reading frames against a peer reviewed allergen database (FARRP Allergen Database Version 12, Released in February, 2012) did not generate any significant amino acid sequence similarities with known allergens. Similarly, the search against the GenBank non-redundant (nr) protein dataset did not detect any significant protein sequence similarity with toxic proteins harmful to humans or animals.

INTRODUCTION

Soybean (*Glycine max* L.) event DAS-44406-6 was generated by integration of the T-DNA from plasmid pDAB8264 through *Agrobacterium*-mediated transformation of a variety "Maverick". The T-DNA of plasmid pDAB8264 contains three plant transcription units (PTUs): 1) Histone H4A748 promoter, *2mepsps* gene, and Histone H4A748 3' UTR; 2) AtUbi10 promoter, *aad*-12 gene, and AtuORF23 3' UTR; 3) CsVMV promoter, *pat* gene, and AtuORF1 3' UTR. In addition, a RB7 matrix attachment region (RB7 MAR) is located at the 5' end of the T-DNA insert. Molecular characterization indicated that the DAS-44406-6 soybean contained a single insert including three intact expression cassettes for 2mEPSPS, AAD-12 and PAT, respectively (Poorbaugh, 2011). DNA sequences flanking the insert in event DAS-44406-6 soybean were also cloned and characterized (Guttikonda, 2011). The DNA sequence of the insert is identical to the corresponding portion in the T-DNA insert of plasmid pDAB8264 except for an extra 3 bp insertion at the 5' junction.

In the safety assessment of transgenic crops, one of the concerns is that should "novel" proteins be expressed they might have a potential to elicit allergic or toxic reactions in humans. Therefore, putative reading frames spanning the junctions between the transgene insert and its borders, and within the transgene insert itself, can be analyzed for sequence similarity to known allergens or toxins as an indication of a safety concern should those reading frames actually be expressed. For this study, putative reading frames are defined very conservatively as any reading frame spanning the junctions and inside the whole T-DNA insert regardless of the presence of a start codon and the number of amino acid residues.

To assess potential allergenicity using bioinformatics tools, two criteria for evaluating structural similarities between query proteins and known allergens are currently used based on amino acid sequence alignments (Codex Alimentarius Commission, 2003; Ladics, 2008). The first criterion is a search over 80-amino-acid stretches (sliding window search) to detect >35% identity between a query protein and known allergens. The window size of 80 amino acids was selected to correspond with a typical domain size in a protein, and recognizes that single protein domains

may contain epitopes that mediate antibody binding. The second criterion involves evaluating short amino-acid stretches for identity between the query protein and known allergens. As stated in the report of Codex Ad Hoc Working Group on Allergenicity (Codex Alimentarius Commission, 2003) "the size of the contiguous amino acid search should be based on a scientifically justified rationale in order to minimize the potential for false negative or false positive results". Window sizes of 6 to 8 amino acids have been suggested based on hypothetical epitope sizes, however, use of window sizes of less than 8 amino acids have been largely abandoned based on the high probability of random alignments that are of no predictive value (Stadler and Stadler, 2003; Silvanovich *et al.*, 2006). The use of any short-alignment criteria for predicting the allergenic potential of proteins has also been recently criticized (Goodman *et al.*, 2008; Thomas *et al.*, 2008; Cressman and Ladics, 2009; Herman *et al.*, 2009; EFSA, 2010).

For evaluation of potential protein toxicity, structural similarity between a query protein and known protein toxins are identified using local sequence alignment search tools such as BLAST and FASTA algorithms against a database of all available protein sequences.

The purpose of this study is to identify putative reading frames spanning the border junctions and within the whole T-DNA insert in DAS-44406-6 soybean and evaluate them for potential allergenicity and toxicity using bioinformatics tools along with updated allergen and non-redundant protein databases.

METHODS

Search and Extraction of Putative Reading Frames

DNA sequences including the whole insert and its border regions of DAS-44406-6 soybean were analyzed with an in-house Perl script to search six-frame translations from stop codon to stop codon across all the identified junctions (Figure 1). For each reading frame (RF), the exact locations of 5' and 3' stop codons were identified. Those RFs spanning the junctions were evaluated for any potential protein allergenicity or toxicity using bioinformatics tools.

DNA sequences of the whole T-DNA insert in DAS-44406-6 soybean were analyzed with an in-house Perl script to search six-frame translations from stop codon to stop codon across whole T-DNA insert without flanking border regions (Figure 1). For each reading frame (RF), the exact locations of 5' and 3' stop codons were identified. Those RFs within the T-DNA insert including the RFs containing gene of interest were evaluated for any potential protein allergenicity or toxicity using bioinformatics tools.

Query Sequence Preparation

Each putative reading frame sequence was prepared in FASTA format for the use with BLASTp search programs.

Allergenicity Assessment

For the allergenicity assessment, the amino acid sequence of each RF was compared with a peer-reviewed database containing 1603 known and putative allergens as well as celiac-induction protein sequences residing in the FARRP dataset (Version 12, Released in February 2012, University of Nebraska, http://www.allergenonline.org/). Potential identities between the RF peptide sequences and proteins in the allergen database were evaluated with the FASTA program (v34) using the default algorithm parameters (Matrix = BLOSUM50; Expect = 10; Gap Penalties = -12/-2; *ktup*=2). The FASTA search was run by an in-house Perl script on a UNIX computer with a Linux operation system. If a query sequence is longer than 80 amino acids, the script parses the query sequence into a complete (overlapping) set of 80 amino acid long fragments and each fragment is subjected to a FASTA search. A greater than 35% identity threshold over any 80 or more amino acid sequences between a query sequence and an allergen was used to indicate the potential for cross-reactivity. To ensure that high identity over a short stretch (for example, 80% over 60 amino acids) will not be overlooked, a calculation, (Identity% × number of overlapped amino acids)/80, was implemented as a conversion to check the criteria of >35% over 80 amino acids when the FASTA alignment (overlapped amino acids) is less than 80 amino acids. Reading frames shorter than 29 amino acids were not evaluated using FASTA search since >35% identity requires at least a match of 29 amino acids over 80 amino acids. RF peptide

sequences were also screened for any matches of 8 contiguous amino acids to the allergens contained in the database noted above as long as an RF is equal to or longer than 8 amino acids. This was done using an in-house Perl script that generates all sequentially possible (overlapping) 8-residue peptides from a query protein, followed by Fuzzpro program (Emboss Package v2.10.0) search that compares each query "word" with all allergen sequences in the database for perfect matches.

Toxicity Assessment

To assess potential toxicity of the *in silico* translated peptides of the RFs, a similarity search was conducted using the BLASTp algorithm. Reading frames were queried using the BLASTp 2.2.21 algorithm against GenBank protein sequences (nr) database (update to March 05, 2012), which incorporates non-redundant entries from all GenBank and RefSeq nucleotide translations, including non-redundant GenBank CDS translation along with protein sequences from SWISS-PROT (http://www.expasy.org/sprot/), PIR (http://pir.georgetown.edu/), PRF (http://www.prf.or.jp/aboutdb-e.html), and PDB (http://www.wwpdb.org/). BLASTp searches were done in the NCBI (National Center of Biotechnology Information) BLAST website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). BLASTp searches were performed in an internal UNIX computer using the default setting of algorithm parameters (Matrix = BLOSUM 62, Gap Costs: Existence: 11, Extension: 1, Word Size=3) except that the low complexity filter was set to "off" and Expectation was set to 1. Although a statistically significant sequence similarity generally requires an alignment with an expectation value less than 0.01, a threshold of E-value < 1.0 ensures that proteins with even limited similarity will not be overlooked in the search (Pearson, 2000).

RESULTS AND CONCLUSIONS

Allergenicity and Toxicity Assessment of Reading Frames across Border Junctions

A total of 12 reading frames spanning the junctions across the insert and its border regions in DAS-44406-6 soybean were identified (Table 1). When the amino acid sequences of the 12

reading frames were compared with the FARRP allergen dataset (Version 12, February 2012), no matches of eight or greater contiguous amino acids were observed in any of the translated sequences. Of those 12 reading frames, six RFs (1_+2, 2_+2, 2_+3, 2_-1, 2_-2 and 2_-3) were subject to searches against the allergen database using the FASTA program. The remaining six reading frames were less than 29 amino acids and thereby incapable of meeting the ">35%-identity over 80 amino-acid" threshold. No over threshold identities (greater than 35% identity over greater than or equal to 80 amino acid residues) were detected in the FASTA search outputs when the peptide sequences deduced from the 6 RFs were used as query sequences (Table 2, Appendix 1).

When the 12 reading frames were subjected to BLASTp search against the GenBank non-redundant protein dataset, no alignments with E-values less than 1 were returned (Table 2, Appendix 2).

Allergenicity and Toxicity Assessment of Reading Frames across Whole T-DNA Insert

A total of 651 reading frames within the whole T-DNA insert in DAS-44406-6 soybean were identified (Appendix 3). When the amino acid sequences of the 651 reading frames including the reading frames containing genes of interest within T-DNA insert were compared with the FARRP allergen dataset (Version 12, February 2012), no matches of eight or greater contiguous amino acids were observed in any of the translated sequences. Of those 651 reading frames, 210 RFs were subject to searches against the allergen database using the FASTA program. The remaining 441 reading frames were less than 29 amino acids. No over threshold identities (greater than 35% identity over greater than or equal to 80 amino acid residues) were detected in the FASTA search outputs when the peptide sequences deduced from the 210 RFs were used as query sequences (Appendix 4).

When all the reading frames greater than or equal to eight amino acids were subjected to BLASTp search against the GenBank non-redundant protein dataset, 10 reading frames that

showed alignments with E() < 1.0 were returned. Number of BLASTp hits and description of alignments for 10 reading frames were shown in Table 3. Out of the 10 reading frames, RF_+2-69, RF_+2-83 and RF_2-69 returned alignments with AAD-12, PAT and 2mEPSPS proteins respectively. RF_+1-111 returned alignment with *Agrobacterium* related proteins. RF_+2-29 returned only one weak alignment with a flagellar hook-length control protein from *Desulfomicrobium baculatum* DSM 4028. RF_-2-12 returned alignment with Photosystem II protein D1related proteins. RF_-3-25 returned alignment with Mas2' proteins from *Agrobacterium*. RF_-3-34, RF_-3-35 and RF_-1-176 returned alignments with unnamed proteins from *Ostreococcus tauri* and unknown protein from *Zea mays* respectively (Appendix 5). None of the significant sequence alignments are related to any known protein toxins.

In conclusion, bioinformatics evaluation of the putative reading frames from DAS-44406-6 soybean did not generate any significant amino acid sequence similarities with known allergens or toxic proteins that are harmful to humans or animals.

REFERENCES

Codex Alimentarius Commission, 2003. Guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants. Annex 1: Assessment of possible allergenicity. CAC/GL 45-2003. Annex 1: Assessment of Possible Allergenicity, pp. 9-11.

Cressman, R.F., Ladics, G., 2009. Further evaluation of the utility of "sliding window" fasta in predicting cross-reactivity with allergenic proteins. Regulatory Toxicology and Pharmacology 54, S20-S25.

EFSA, 2010. Scientific opinion on the assessment of allergenicity of gm plants and microorganisms and derived food and feed. EFSA Journal 8, 1700.

Goodman, R.E., Vieths, S., Sampson, H.A., Hill, D., Ebisawa, M., Taylor, S.L., van Ree, R., 2008. Allergenicity assessment of genetically modified crops - what makes sense? Nature Biotechnology 26, 73 - 81.

Guttikonda, S.K., 2011. Cloning and characterization of the DNA sequence for the insert and its flanking border regions of DAS-44406-6 soybean. Study ID 102117. Dow AgroSciences LLC.

Herman, R., Song, P., ThirumalaiswamySekhar, A., 2009. Value of eight-amino-acid matches in predicting the allergenicity status of proteins: An empirical bioinformatic investigation. Clinical and Molecular Allergy 7, 9.

Ladics, G.S., 2008. Current codex guidelines for assessment of potential protein allergenicity. Food and Chemical Toxicology 46, S22 - S23.

Pearson, W.R., 2000. Flexible sequence similarity searching with the fasta3 program package. Methods Mol Biol 132, 185-219.

Poorbaugh, J., 2011. Molecular characterization of DAS-44406-6 soybean. Study ID 101947. Dow AgroScience LLC.

Silvanovich, A., Nemeth, M.A., Song, P., Herman, R., Tagliani, L., Bannon, G.A., 2006. The value of short amino acid sequence matches for prediction of protein allergenicity. Toxicological Sciences 90, 252 - 258.

Stadler, M.B., Stadler, B.M., 2003. Allergenicity prediction by protein sequence. The FASEB Journal 17, 1141 - 1143.

Thomas, K., Herouet-Guicheney, C., Ladics, G., McClain, S., MacIntosh, S., Privalle, L., Woolhiser, M., 2008. Current and future methods for evaluating the allergenic potential of proteins: International workshop report 23–25 october 2007. Food and Chemical Toxicology 46, 3219-3225.

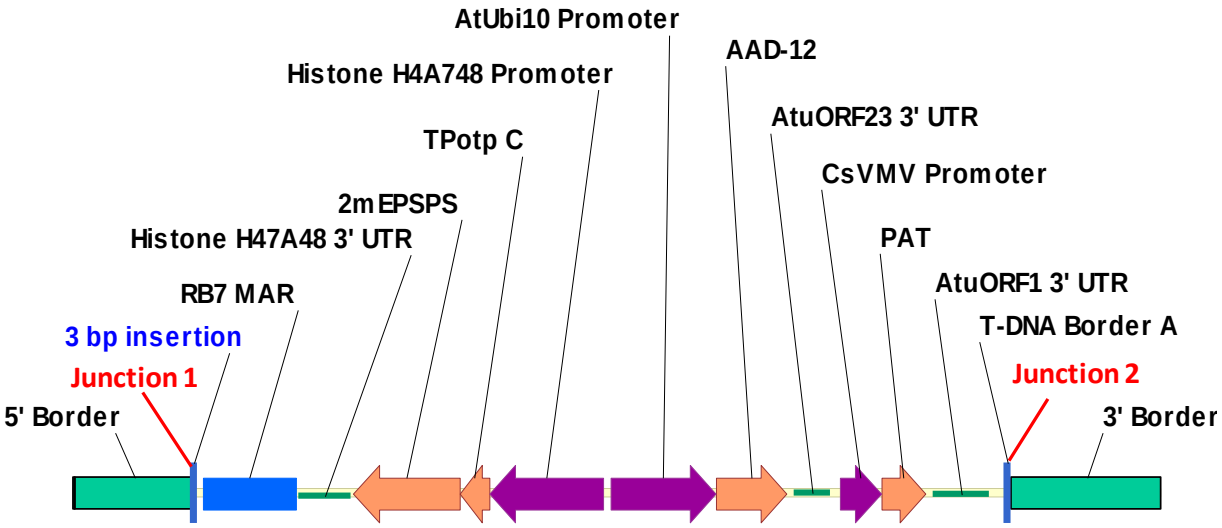

Figure 1. Diagram of the Insert, its Flanking Borders, and Junction Sites in DAS-44406-6 Soybean

Table 1. Deduced Amino Acid Sequences of Reading Frames across the border Junctions in DAS- 44406-6 Soybean

Reading Frame Name (Junction_Frame)	Nucleotide location	Number of amino acids	Deduced amino acid sequence
1_+1	1483−1554	24	GHHGGPNSLKLESSQLRSTGQIRS_
1_+2	1307−1507	67	RVGPDIVACYWGFLSVACVLHYCMGLAHPTIQCIF MCDNVMGFYCSCCFLFRNLHVNGKVIMEVRIV_
1_+3	1476−1502	9	TVRSSWRSE_
1_-1	11943−11773	25	FQTIRTSMMTLPFTCKFLKRKQQEQ_
1_-2	1553−491	21	ERIWPVDLNCELSNFKLFGPP_
1_-3	1534−1469	22	TSIASFLISNYSDLHDDLTVYM_
2_+1	11737−11790	18	IYPAPASQQLDLQRTNVV_
2_+2	11726−12019	98	FTIEYILPQPANSSIYRERMSCDMWNKATTTTYMNL TIESGSPSCDVIHGMDMVADRKRKKKCMYMCENE SFFYPNNKKKLIIYPKNYLHDRYVHFFP_
2_+3	11682−11825	48	KRPQCVIKLSKRQFDLQLNISCPSQPTARFTENECRV ICGTRQRQQHT_
2_-1	11943−11773	57	KKLSFSHIYMHFFFLFRSATMSMPWITSQLGDPDSIV RFMYVVVVALFHISHDIRSL_
2_-2	11831−11730	34	DSCMLLSLPCSTYHTTFVLCKSSCWLAGAGYIQL_
2_-3	11965−11720	82	FFFIIWIKKTLIFTHIHAFLFSFSIGHHVHAMDYITTR RPRLYCEIHVCCCRCLVPHITRHSFSVNRAVGWLGQ DIFNCKSN_

Table 2. Summary of Results from BLASTp Search for Sequence Similarities of Putative Reading Frames across Border Junctions in DAS-44406-6 Soybean

Reading Frame (Junction_Frame)	Length (aa)	Match of 8 or more residues with known allergens	FASTA Search (>35% identity over ≥ 80 residues)	Number of BLASTp hits (E()<1)
1_+1	24	No	N/A	0
1_+2	67	No	No	0
1_+3	9	No	N/A	0
1_-1	25	No	N/A	0
1_-2	21	No	N/A	0
1_-3	22	No	N/A	0
2_+1	18	No	N/A	0
2_+2	98	No	No	0
2_+3	48	No	No	0
2_-1	57	No	No	0
2_-2	34	No	No	0
2_-3	82	No	No	0

N/A = Not applicable;

Table 3. Summary of Results from BLASTp Search for Sequence Similarities of Putative Reading Frames across Whole T-DNA Insert in DAS-44406-6

Reading Frame (Junction_ Frame)	Length (aa)	Match of 8 or more residues with known allergens	FASTA Search (>35% identity over ≥ 80 residues)	Number of BLASTp hits (E()<1)	Description
RF_+1-111	56	No	No	8	*Agrobacterium* related proteins; E()=7×10^{-13}~0.087
RF_+2-29	149	No	No	1	*Desulfomicrobium baculatum* flagellar hook-length control protein DSM 4028 (Accession: YP_003158929); E()=0.28
RF_+2-67	298	No	No	2683	S-2,4-dichlorophenoxypropionate/alpha-ketoglutara (AAD-12); E()=1×10^{-168}~0.99
RF_+2-83	185	No	No	2139	Phosphinothricin-N-acetyltransferase (PAT); E()=1×10^{-103}~0.98
RF_+3-34	65	No	No	1	Unnamed protein product from *Ostreococcus tauri* (Accession : XP_003083033.1; E()=3×10^{-04})
RF_+3-35	116	No	No	2	Unnamed protein product from *Ostreococcus tauri* (Accession : XP_003083033; E()=0.002 ; Accession : XP_003521865 ; E() =0.003)
RF_-1-76	127	No	No	1	*Zea mays* unknown protein (Accession: ACR35606.1; E()=0.007)
RF_-2-12	41	No	No	18	*Palmaria decipiens*Photosystem II protein D1; (Accession: ACR35606.1; E()=0.030)
RF_-2-69	580	No	No	4777	5-enolpyruvylshikimate-3-phosphate synthase (EPSPS); E()=1×10^{-99}~0
RF_-3-25	50	No	No	5	Mas2' *Agrobacterium*; E()=5×10^{-15}~2×10^{-6}
RF_+1-111	56	No	No	8	*Agrobacterium* related proteins; E()=7×10^{-13}~0.087
RF_+2-29	149	No	No	1	*Desulfomicrobium baculatum* flagellar hook-length control protein DSM 4028 (Accession: YP_003158929); E()=0.28

APPENDIX

APPENDIX 1 FASTA Search Outputs of the Putative Reading Frames (<29 aa) across Junctions in DAS-44406-6 Soybean against the Allergen Database V12

APPENDIX 2 Results from BLASTp Search for Sequence Similarities of all Putative Reading Frames with Amino Acid Sequences across Junctions in DAS-44406-6 Soybean

APPENDIX 3 Deduced Amino Acid Sequences of Reading Frames across the border Junctions in DAS- 44406-6 Soybean

APPENDIX 4 FASTA Search Outputs of the Putative Reading Frames (<29 aa) across Whole T-DNA Insert in DAS-44406-6 Soybean against the Allergen Database V12 FASTA Output File was more than 3000 pages and subgroup into 3 files. Appendix 4a, Appendix 4b, Appendix 4c

APPENDIX 5 Results from BLASTp Search for Sequence Similarities of all the Putative Reading Frames with Amino acid sequences across Whole T-DNA Insert in DAS-44406-6

Output files of all the appendices with a combined size of more than thousands of pages were generated. These files are electronically stored in a secured computer in Dow AgroSciences and are available for viewing in PDF format.