

Study Title

EXPURGATED

Bioinformatics Analysis of the Potential Fusion Proteins at DNA Junctions in Canola (*Brassica napus*) for Omega 3 Fatty Acids: Identity Comparison to Allergens and Toxins

Author

[REDACTED]

Study Completed On

2 January 2017

Performing Laboratory

[REDACTED]

Laboratory Project ID

Study Number: REG NuSeed 2
Nuseed Report N° 2016-004

EXECUTIVE SUMMARY

The transgenic canola event, elite event B0050-027, called DHA canola (OECD ID NS-B50027-4), which could produce docosahexaenoic acid (DHA) [REDACTED] in *Brassica napus* L. seeds, includes [REDACTED] DNA junctional sites, one at each end of the two DNA inserts [REDACTED]

[REDACTED] Potential open reading frames (stop-to-stop codons) at each site were predicted using Expasy Tools and the coding regions covering the junctions were evaluated for matches to sequences of known allergens and toxins. The sequences of the entire inserts and short-host-plant flanking DNA surrounding the inserts were provided by the sponsor. The sponsor identified plant DNA and the sequence of the insert construct. [REDACTED]

The individual junction DNA sequences were evaluated using ExPASy Tools Translation module to identify potential Open Reading Frames (ORFs) in all six potential reading frames. The peptide sequences that overlapped the insert-junctions [REDACTED] [REDACTED] were compared to the AllergenOnline.org database version 16, using full-length FASTA3, sliding window of 80 amino acids and exact word match for eight amino acid (AA) identity matches to known and putative allergens that are considered here for potential identity matches to allergens and toxins. This report summarizes a study of potential allergenicity and toxicity of the potential (putative) peptides constructed by the insert.

None of the results from the bioinformatics searches of the potential fusion protein amino acid sequences, compared to known and putative allergens or toxins, identified any significant sequence identity match to a protein likely to cause an adverse effect in consumers. Thus, there is no need for toxicity tests or further evaluation of potential allergenicity or toxicity.

STATEMENT OF CONFIDENTIALITY

Claim of confidentiality under FIFRA section 10(d)(1)(A), (B), or (C).

Information claimed confidential on the basis of its falling within the scope of FIFRA section 10(d)(1)(A), (B), or (C)

.

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

3 January 2017

Date

These Data May Be Considered Confidential In Countries Outside The United States.

GOOD LABORATORY PRACTICE COMPLIANCE STATEMENT

This study was not conducted and reported in compliance with the requirements of the Good Laboratory Practice Standards (40 CFR Part 160) of the Code of Federal Regulations of the United States of America. This is a characterization assessment of the similarity of the introduced proteins to known and putative allergens based on source of the genes and the sequences of the proteins. There is no test system. However, raw data including PubMed searches and bioinformatics comparisons were archived in PDF format in the Authors laboratory with a copy given to the sponsor.

[REDACTED]

Date: 3 January 2017

STUDY NUMBER: REG – NuSeed 2

Title: Bioinformatics Analysis of the Potential Fusion Proteins at DNA Junctions in Canola (*Brassica napus*) for Omega 3 Fatty Acids: Identity Comparison to Allergens and Toxins

Facility:



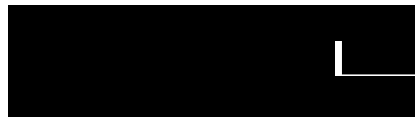
Principle Investigator:



Study Start Date: 2 January 2016

Study Completion Date: 2 January 2017

Records Retention: All study specific raw data and a copy of the final report will be retained at the Food Allergy Research and Resource Program, University of Nebraska.



Date



TABLE OF CONTENTS

SECTION	PAGE
TITLE PAGE	1
SUMMARY	2
STATEMENT OF NO CONFIDENTIALITY	3
GOOD LABORATORY PRACTICE COMPLIANCE STATEMENT.....	4
SIGNATURE OF FINAL REPORT APPROVAL.....	5
Table of Contents.....	6
Abbreviations.....	9
1. Introduction	10
2. Purpose.....	10
3. Methods.....	11
3.1 Amino acid sequences of potential proteins at [REDACTED] DNA junction sites.....	11
3.2 Sequence database search strategies	12
3.2.1 FASTA3 overall search of AllergenOnline.....	23
3.2.2 FASTA3 of AllergenOnline by 80 aa segments	23
3.2.3 Exact word match of AllergenOnline by 8-contiguous aa	23
3.2.4 BLASTP of NCBI Entrez without keyword limit	23
3.2.5 BLASTP of NCBI Entrez with “allergen” as keyword limit	23
3.2.6 BLASTP of NCBI Entrez with “toxin” and “toxic” as keyword limit.	24
4. Results and Discussion.....	24
4.1 Sequence comparison of the putative peptides at DNA junctions in DHA canola to allergens and toxins	24
4.2.1 Full length FASTA3 to AllergenOnline with putative peptides.....	24
4.2.2 Sliding 80-amino acid window FASTA3 vs. AllergenOnline.org version 16 with putative peptides	50





4.2.3 Eight-contiguous amino acid identity matches with AllergenOnline.org version 16	50
4.2.4 BLASTP of NCBI without keyword limit with each of the eight proteins introduced in canola	51
4.2.5 BLASTP of NCBI Entrez using keywords “allergen”	51
4.2.6 BLASTP of NCBI Entrez with “toxin” and “toxic”	51
4.3 Bioinformatics summary for the eight proteins produced by the genes inserted in canola	51
5. Conclusions	52
6. References	52
7. Unpublished References	53

TABLES

Table 1. Junction DNA sequences in DHA canola.	11
Table 2. Hypothetical Open Reading Frames (ORFs) at insert junctions	13
Table 3. Chromosome [REDACTED] Junction #1, Reading frame 1: 5’-3’	26
Table 4. Chromosome [REDACTED] Junction #1, Reading frame 2: 5’-3’	27
Table 5. Chromosome [REDACTED] Junction #1, Reading frame 3: 5’-3’	28
Table 6. Chromosome [REDACTED] Junction #1, Reading frame 1: 3’-5’	29
Table 7. Chromosome [REDACTED] Junction #1, Reading frame 2: 3’-5’	30
Table 8. Chromosome [REDACTED] Junction #2, Reading frame 2: 5’-3’	31
Table 9. Chromosome [REDACTED] Junction #2, Reading frame 3: 5’-3’	32
Table 10. Chromosome [REDACTED] Junction #2, Reading frame 2: 3’-5’	33
Table 11. Chromosome [REDACTED] Junction #2, Reading frame 3: 3’-5’	34
Table 12. Chromosome [REDACTED] Junction #3, Reading frame 1: 5’-3’	35
Table 13. Chromosome [REDACTED] Junction #3, Reading frame 3: 5’-3’	36
Table 14. Chromosome [REDACTED] Junction #3, Reading frame 1: 3’-5’	37
Table 15. Chromosome [REDACTED] Junction #3, Reading frame 2: 3’-5’	38
Table 16. Chromosome [REDACTED] Junction #3, Reading frame 3: 3’-5’	39
Table 17. Chromosome [REDACTED] Junction #4, Reading frame 2: 5’-3’	40
Table 18. Chromosome [REDACTED] Junction #4, Reading frame 3: 5’-3’	41

Table 19. Chromosome [REDACTED] Junction #4, Reading frame 1: 3'-5'	42
Table 20. Chromosome [REDACTED] Junction #4, Reading frame 2: 3'-5'	43
Table 21. Chromosome [REDACTED] Junction #4, Reading frame 3: 3'-5'	44
Table 22. Chromosome [REDACTED], Reading frame 1: 5'-3'	45
Table 23. Chromosome [REDACTED], Reading frame 2: 5'-3'	46
Table 24. Chromosome [REDACTED], Reading frame 3: 5'-3'	47
Table 25. Chromosome [REDACTED], Reading frame 1: 3'-5'	48
Table 26. Chromosome [REDACTED], Reading frame 2: 3'-5'	49
Table 27. Chromosome [REDACTED], Reading frame 3: 3'-5'	50

ABBREVIATIONS

aa	Amino acid
	
	
AOL v16	http://www.AllergenOnline.com/ database version 16
BLASTP	Algorithm used to find local high scoring alignments between a pair of protein sequences (using databases on Entrez)
CODEX	CODEX Alimentarius Commission Guideline for the safety assessment of food derived from biotechnology
DHA	Docosahexaenoic acid
Entrez NCBI	A public genetic database maintained by the National Center for Biotechnology Information (NCBI) at the National Institutes of Health, Bethesda, MD. Protein entries in the Entrez search and retrieval system are maintained by the NCBI of the National Institutes of Health (U.S.A.)
FAO	Food and Agricultural Organization
FARRP	Food Allergy Research and Resource Program, University of Nebraska
FASTA3	Algorithm used to find local high scoring alignments between a pair or protein sequences (using the AllergenOnline database)
GE	Genetic engineering
GI	A unique identification number assigned by NCBI to each sequence in the database
ORF	Open reading frame
PubMed	A public information database of scientific journal articles and abstracts maintained by the National Library of Medicine, National Institutes of Health (U.S.A.)
RB	Right Border
UTR	Untranslated Region
8mer	Exact word search for segments of eight amino acid matches between the query protein and proteins in AllergenOnlin.org

1.0 INTRODUCTION

Nuseed Pty, Ltd has developed a new canola variety that has been genetically engineered (GE) to produce high levels of docosahexaenoic acid. The transgenic canola event, elite event B0050-027, called DHA canola (OECD ID NS-B50027-4, produced docosahexaenoic acid (DHA) [REDACTED] in *Brassica napus* L. seeds. The canola was obtained through gene transformation of seven microalgae and yeast genes as well as the Phosphinothricin N-Acetyltransferase (PAT) gene as a selectable marker. Through construct-captured, whole-genome and PCR-amplicon sequencing, DHA canola was characterized to have one insert on chromosome [REDACTED] and another insert on chromosome [REDACTED]

[REDACTED] This report summarizes a study of potential allergenicity and toxicity of potential expressed proteins identified as Open Reading Frames (ORFs) at the individual DNA junctions. The methods used for the safety assessment are consistent with the process outlined by the CODEX Alimentarius Commission (2009) for evaluation of the potential safety of crops developed through genetic engineering. The results are summarized in this report along with my opinion related to food safety for these proteins.

2.0 Purpose

The purpose of this study is to perform an evaluation of the potential allergenicity and toxicity of potential proteins that might be expressed from unexpected transcription and translation of junctional regions of DNA junctions in canola line called NS-B50027-4. The sequences of potential proteins are predicted by computer predicted translation of DNA overlapping the junctions at the insertion sites between plant and DNA inserts of the two insertions, [REDACTED]

[REDACTED] Potential proteins are identified as stop-to-stop open reading frames overlapping the DNA junctions. Those potential protein sequences were evaluated for sequence identity matches to known allergens and toxins to identify any need for additional safety tests.

3.0 Methods

3.1 Amino acid sequences of the potential proteins at DNA junctions. junctions were identified and evaluated for potential ORFs. The DNA sequences are shown in Table 1. The 5' and 3' ends of the insert in chromosome and the 5' and 3' ends of the insert in chromosome were evaluated. Potential ORFs were defined by stop-to-stop codons using the ExPASy Translation tool program (web.expasy.org/translate/). The potential six reading frames of each of the DNA sequences were examined to identify the junction DNA fusion sites). The stop-to-stop predicted peptides that include the junction are the only ones evaluated for potential allergenicity and potential identity matches with toxins.


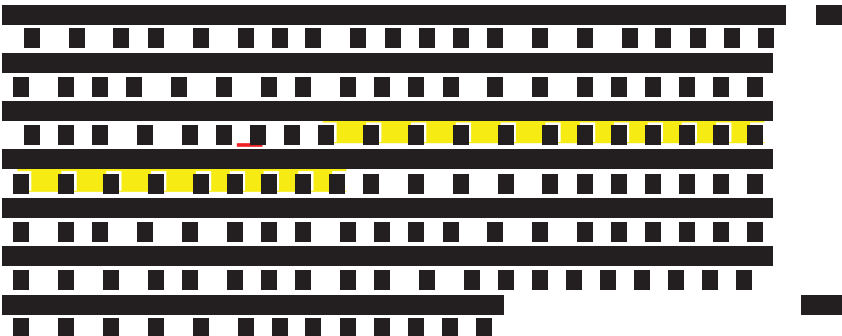

Table 1. Junction DNA sequences in DHA canola. Approximately 200 base pairs of DNA (nucleotide sequences) at each junction of this transformed event in canola are shown. The sequences highlighted in green are segments of plant DNA. Sequences highlighted in red are from the insert. Potential open reading frames (ORFs) were predicted by ExPASy Translation Tool. The proteins encoded by the eight genes introduced into canola to allow efficient production of high concentrations of DHA in oil. The full genomic DNA sequences of the inserts of these events were verified and are reported in a Molecular Report (see Report N° 2016-002). The full sequences are made for the enzymes and the PAT marker protein.

Insertion Site	DNA sequence of at the insert, plant DNA highlighted in green, insert DNA highlighted in red.
Enzyme name	Sequences are shown from 5' to 3'.
Sequence ID 1, plant to insert	
Sequence ID 2, Insert to plant plant to insert	

Sequence ID 3, Plant to insert plant to insert	
Sequence ID 4, Insert to plant	

3.2 Sequence database search strategies. The AllergenOnline version 16, December 2016 (<http://www.allergenonline.org/>) and the NCBI Entrez Protein (<http://www.ncbi.nlm.nih.gov/BLAST/>) databases were used as the protein amino acid data sources for the sequence comparisons for allergens and toxins in December 2016. The AllergenOnline database was updated in 27 January 2016 and is maintained by the Food Allergy Research and Resource Program of the University of Nebraska. The sequences of putative (potential) peptides (Table 2) predicted for junctions were tested using all six methods described here. The peptides were again searched for identity matches against AllergenOnline version 16 on 22 December 2016 and in NCBI Protein database. Protein entries in the Entrez search and retrieval system is compiled and maintained by the NCBI of the National Institutes of Health (U.S.A.). The database is potentially updated or modified daily, and therefore the date of sequence searches by BLASTP is relevant to the dataset used in the BLASTP searches. BLASTP and FASTA3 are unique computer algorithms that provide similar local alignments and results if the appropriate scoring matrices and criteria are used.



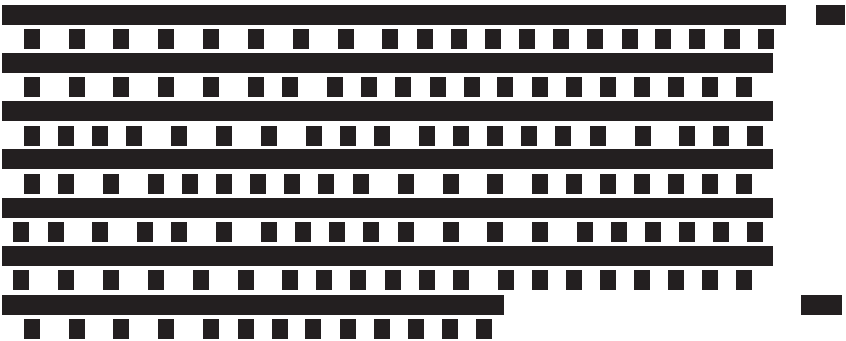
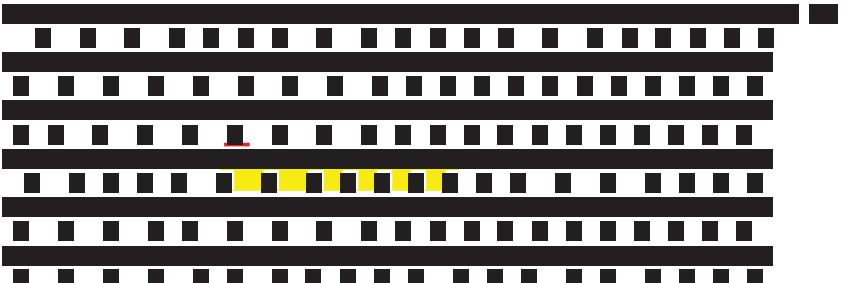
Table 2. Hypothetical Open Reading Frames (ORFs) at insert junctions. Potential open reading frames for the four total insertion junctions from two insertion sites using Expassy Translation tool to identify stop-to-stop coding sequences. Red highlight indicates the junction of plant and insert DNA. Yellow highlight indicates the peptide coding sequence of the hypothetical ORF that overlaps the junction. Junctions that match a stop codon at intersection of sequences do not result in a junctional hypothetical ORF and are therefore not assessed (e.g. Insert 2, ████ frame 3, insert #1 reading frame 3 3'-5')


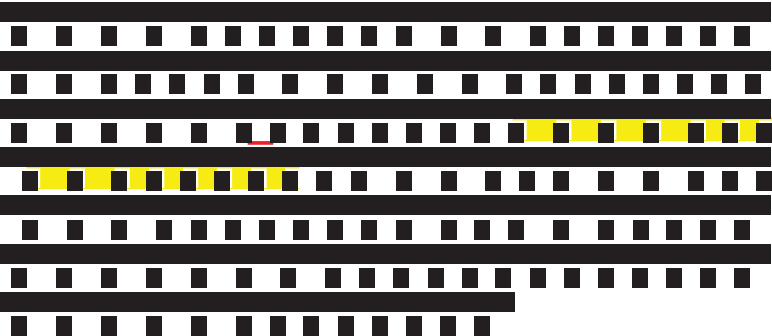
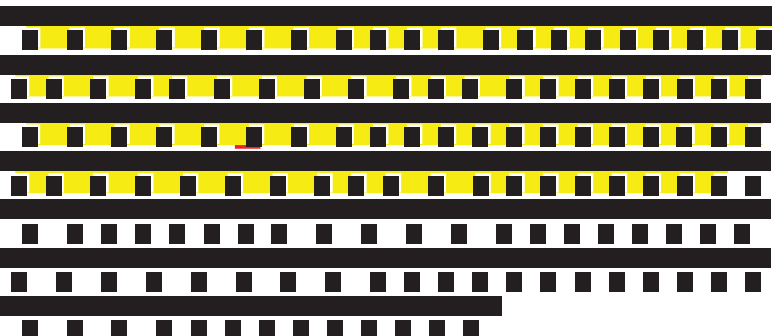

Junction Insertion site	Reading Frame & DNA direction	Sequences Nucleotides of DNA, then amino acids or '-' stop codons
<div data-bbox="344 643 415 667">#1</div> <div data-bbox="323 695 436 737">Canola to tDNA</div>	<div data-bbox="527 643 625 662">1 5'-3'</div>	
<div data-bbox="344 1029 415 1053">#1</div>	<div data-bbox="527 1029 625 1049">2 5'-3'</div>	
		

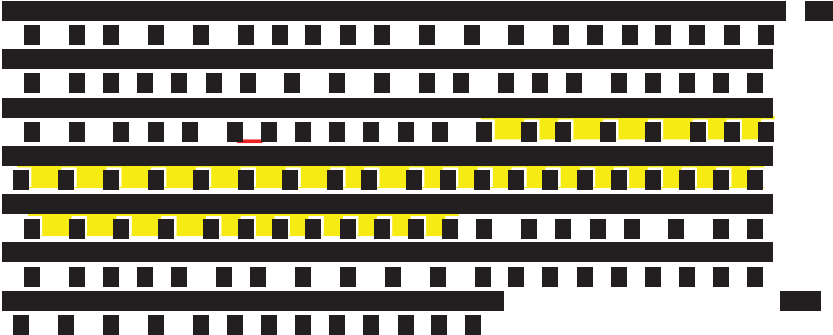
#1	3 5'-3'	
#1	1 3'-5'	
#1	2 3'-5'	
#1	3 3'-5'	



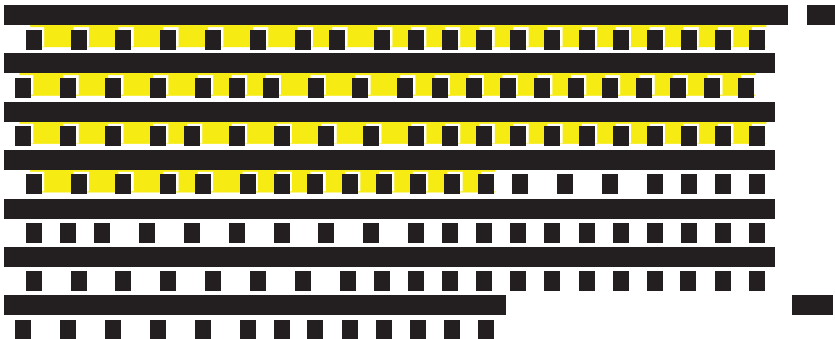

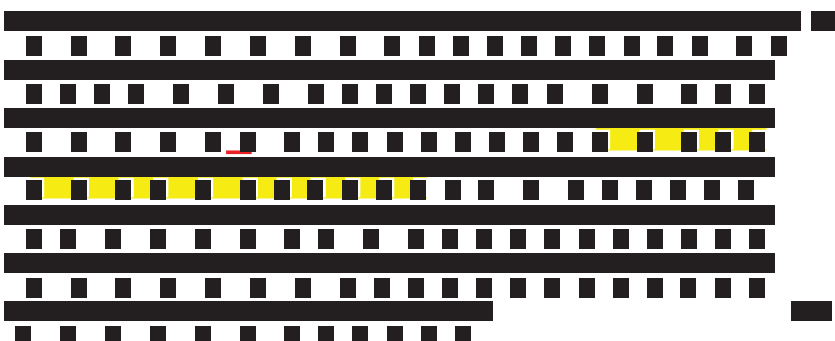

#2	1 5'-3'	
#2	2 5'-3'	
#2	3 5'-3'	

#2	1 3'-5'	
#2	2 3'-5'	
#2	3 3'-5'	

		
#3	1 5'-3'	
#3	2 5'-3'	
#3	3 5'-3'	

		
#3	1 3'-5'	
#3	2 3'-5'	
#3	3 3'-5'	

 #4	1 5'-3'	
 #4	2 5'-3'	
 #4	3 5'-3'	

 #4	1 3'-5'	
 #4	2 3'-5'	
 #4	3 3'-5'	
		

	<p>1 5'-3'</p>	<p>- L E S R N Y T F N T R - K T K Y S A Q</p> 
	<p>2 5'-3'</p>	
	<p>3 5'-3'</p>	
	<p>1 3'-5'</p>	

	2 3'-5'	
	3 3'-5'	

3.2.1 FASTA3 overall search of AllergenOnline. The potential sequential and inferred structural similarities of the putative peptides were evaluated using version 16 of AllergenOnline.org with *E* scores of 10 using FASTA, version 35.04 (15 January 2009).

3.2.2 FASTA3 of AllergenOnline by 80 aa segments. This short (80-amino acid) segment search is based on the recommendation of Codex (2003). The rationale is that this might help in identifying structural motifs, much shorter than the intact protein, which might contain a conformational IgE binding epitope. It should also help to identify potentially cross-reactive proteins that are not true homologues of an allergen that have significant local identities that might provide an immunological target for IgE antibodies in those with allergies to the matched allergen. A match of >35% with a known allergen will suggest further testing for possible cross-reactivity although matches using the sliding 80 amino acid window search that are not also identified by overall FASTA3 search may represent an irrelevant alignment. All putative peptide sequences were searched using this search criterion.

3.2.3 Exact word match of AllergenOnline by 8-contiguous aa.

A word/string search routine on AllergenOnline.org was used to identify any eight contiguous amino acid sequences of the putative peptide sequences that exactly matched any 8 amino acid segment of any of the allergen sequences in database. The rationale for identifying identity matches in very short sequences is an assumption that individual epitopes may be represented by peptide segments as short as eight amino acids (Metcalf et al., 1996).

3.2.4 BLASTP of NCBI Entrez without keyword limit. The BLASTP is available on the NCBI Entrez website (<http://www.ncbi.nlm.nih.gov/BLAST/>). The current version is BLASTP 2.6.0 (13 December 2016). The purpose of this BLASTP search is to compare the putative peptide sequences were all evaluated against all protein sequences to determine the prevalence of common homologues.

3.2.5 BLASTP of NCBI Entrez with “allergen” as keyword limit. BLASTP search was used comparing the putative peptide sequences against the entire Entrez Protein database, with a limit option selected to query entries for “allergen”, to align only with proteins identified as allergens or allergy. The purpose of this BLASTP search is to ensure that a significant match with a newly discovered allergenic sequence that has not yet been entered into

AllergenOnline is not overlooked. Evaluation of the *E* value, the length of the alignment and the percent identity of any identified match is necessary to judge the significance of any alignment using BLASTP.

3.2.6 BLASTP of NCBI Entrez with “toxin” and “toxic” as keywords limit. The purpose of this BLASTP search is to identify matches to known toxic proteins (toxins) and if alignments share significant identities, to determine potential risks that would require further testing for all putative peptides.

4.0 Results and Discussion.

The summary results for the proteins compared to AllergenOnline and the NCBI Protein database search using search terms, and the amino acid sequences of twenty five putative peptides that are coded by junctional regions at the inserts are presented.

4.1 Sequence comparison of the putative peptides at the DNA junctions in DHA canola to allergens and toxins. The DNA sequences of the junctions of plant-insert and insert-insert are shown in Table 1. Predicted putative peptides expressed in the canola are shown highlighted in yellow in Table 2. All putative peptide sequences were compared to known allergens using both a full-length FASTA alignment search and a sliding window of 80 comparisons and exact 8 amino acid match against AllergenOnline.org, version 16. Additionally, a BLASTP search was performed against the NCBI database using keyword search limits of “allergen”, “toxin” and “toxic”. The complete results for all comparisons for each putative peptide are shown in a single table, with separate tables for each peptide.

4.2.1 Full length FASTA3 vs. AllergenOnline with putative peptides. Results of the full length FASTA3 searches of putative peptides against AllergenOnline.org, version 16 is the most important step for uncovering potential risks of allergy. None of the full-length FASTA alignments were significant in terms of uncovering any risk of potential allergenicity or cross-reactivity (Table 3-Table 28). The length of putative peptides ranged from 12 to 52 aa, making it unlikely that they could elicit a reaction even if bound by IgE. Scoring results for the putative peptides showing alignments with *E* scores less than 10 are shown and demonstrate no significant matches with any allergen. Their identities (%) are markedly below the level that is likely to indicate cross-reactivity (< 50% identity, Aalberse, 2000) and it is also below the 35% identity level suggested by Codex (2003) as a match that may possibly be cross-reactive. Thus, there is only a small likelihood that any

of the eight proteins are sufficiently similar to an allergen to suspect they might trigger allergic responses in allergic subjects due to cross-reactivity. There is no reason to suggest serum IgE tests would be useful to evaluation safety further.

Table 3. Chromosome [REDACTED] Junction #1, Reading frame 1: 5'-3'. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
Frame 1: 5'-3'	[REDACTED]
AllergenOnline Full-FASTA3 alignment	No matches
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	[REDACTED]
BLASTP "allergen" (best matches)	[REDACTED]
BLASTP "toxin" (best match)	[REDACTED]
BLASTP "toxic" (best match)	[REDACTED]

Table 4. Chromosome [REDACTED] Junction #1, Reading frame 2: 5'-3'. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
Frame 2: 5'-3'	[REDACTED]
AllergenOnline Full-FASTA3 alignment	No matches
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	[REDACTED]
BLASTP "allergen" (best matches)	[REDACTED]
BLASTP "toxin" (best match)	[REDACTED]
BLASTP "toxic" (best match)	[REDACTED]

Table 5. Chromosome [REDACTED] Junction #3, Reading frame 3: 5'-3' Alignment of Peptides. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
Frame 3: 5'-3'	[REDACTED]
AllergenOnline Full-FASTA3 alignment	No matches
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	[REDACTED]
BLASTP "allergen" (best matches)	[REDACTED]
BLASTP "toxin" (best match)	[REDACTED]
BLASTP "toxic" (best match)	[REDACTED]

Table 6. Chromosome [REDACTED] Junction #1, Reading frame 1: 3'-5' Alignment of Peptides. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
Frame 3: 3'-5'	[REDACTED]
AllergenOnline Full-FASTA3 alignment	No matches
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	[REDACTED]
BLASTP "allergen" (best matches)	No match
BLASTP "toxin" (best match)	No match
BLASTP "toxic" (best match)	No match

Table 7. Chromosome [REDACTED] Junction #1, Reading frame 2: 3'-5' Alignment of Peptides. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
Frame 3: 3'-5'	[REDACTED]
AllergenOnline Full-FASTA3 alignment out of 3 alignments	[REDACTED]
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	No match
BLASTP "allergen" (best matches)	No match
BLASTP "toxin" (best match)	[REDACTED]
BLASTP "toxic" (best match)	No match

Table 8. Chromosome [REDACTED] Junction #2, Reading frame 2: 5'-3'. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
Frame 2: 5'-3'	[REDACTED]
AllergenOnline Full-FASTA3 alignment	No matches
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	[REDACTED]
BLASTP "allergen" (best matches)	[REDACTED]
BLASTP "toxin" (best match)	[REDACTED]
BLASTP "toxic" (best match)	[REDACTED]

Table 9. Chromosome [REDACTED] Junction #2, Reading frame 3: 5'-3'. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
Frame 3: 5'-3'	[REDACTED]
AllergenOnline Full-FASTA3 alignment	No matches
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	[REDACTED]
BLASTP “allergen” (best matches)	[REDACTED]
BLASTP “toxin” (best match)	[REDACTED]
BLASTP “toxic” (best match)	[REDACTED]

Table 10. Chromosome [REDACTED] Junction #2, Reading frame 2: 3'-5'. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
Frame 2: 3'-5'	[REDACTED]
AllergenOnline Full-FASTA3 alignment	No matches
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]
BLASTP "allergen" (best matches)	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]
BLASTP "toxin" (best match)	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]
BLASTP "toxic" (best match)	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]

Table 11. Chromosome [REDACTED] Junction #3, Reading frame 3: 3'-5'. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
Frame 3: 3'-5'	[REDACTED]
AllergenOnline Full-FASTA3 alignment	No matches
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	[REDACTED]
BLASTP "allergen" (best matches)	[REDACTED]
BLASTP "toxin" (best match)	[REDACTED]
BLASTP "toxic" (best match)	[REDACTED]

Table 12. Chromosome [REDACTED] Junction #3, Reading frame 1: 5'-3'. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
Frame 1: 5'-3'	[REDACTED]
AllergenOnline Full-FASTA3 alignment	[REDACTED]
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	[REDACTED]
BLASTP "allergen" (best matches)	[REDACTED]
BLASTP "toxin" (best match)	[REDACTED]
BLASTP "toxic" (best match)	[REDACTED]

Table 13. Chromosome [REDACTED] Junction #3, Reading frame 3: 5'-3'. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
Frame 3: 5'-3'	[REDACTED] [REDACTED] [REDACTED]
AllergenOnline Full-FASTA3 alignment	No matches
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]
BLASTP "allergen" (best matches)	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]
BLASTP "toxin" (best match)	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]
BLASTP "toxic" (best match)	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]

Table 14. Chromosome [REDACTED] Junction #3, Reading frame 1: 3'-5'. No significant matches to allergens or toxins.

Reading frame: nucleotide range	Amino acid sequence (number of amino acids)
Frame 1: 3'-5'	[REDACTED]
AllergenOnline Full-FASTA3 alignment	No match
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	[REDACTED]
BLASTP "allergen" (best match)	[REDACTED]
BLASTP "toxin" (best match)	[REDACTED]
BLASTP "toxic" (best match)	[REDACTED]

Table 15. Chromosome [REDACTED] Junction #3, Reading frame 2: 3'-5'. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
Frame 2: 3'-5'	[REDACTED]
AllergenOnline Full-FASTA3 alignment of 3 matches	[REDACTED]
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	No matches
BLASTP "allergen" (best match)	No matches
BLASTP "toxin" (best match)	[REDACTED]
BLASTP "toxic" (best match)	No matches

Table 16. Chromosome [REDACTED] Junction #3, Reading frame 3: 3'-5' Alignment of Peptides. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
Frame 3: 3'-5'	[REDACTED] [REDACTED] [REDACTED]
AllergenOnline Full-FASTA3 alignment	No matches
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]
BLASTP "allergen" (best match)	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]
BLASTP "toxin" (best match)	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]
BLASTP "toxic" (best match)	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]

Table 17. Chromosome [REDACTED] Junction #4, Reading frame 2: 5'-3'. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
Frame 2: 5'-3'	[REDACTED]
AllergenOnline Full-FASTA3 alignments	No matches
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best matches)	No matches
BLASTP "allergen" (best matches)	<div>[REDACTED]</div> <div>[REDACTED]</div> <div>[REDACTED]</div> <div>[REDACTED]</div> <div>[REDACTED]</div> <div>[REDACTED]</div> <div>[REDACTED]</div> <div>[REDACTED]</div> <div>[REDACTED]</div> <div>[REDACTED]</div>
BLASTP "toxin" (best match)	No match
BLASTP "toxic" (best match)	No match

Table 18. Chromosome [REDACTED] Junction #4, Reading frame 3: 5'-3'. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
Frame 3: 5'-3' 8	[REDACTED]
AllergenOnline Full-FASTA3 alignment (of 3)	[REDACTED]
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best matches)	No matches
BLASTP "allergen" (best match)	No matches
BLASTP "toxin" (best match)	[REDACTED]
BLASTP "toxic" (best match)	No match

Table 20. Chromosome [REDACTED] Junction #4, Reading frame 2: 3'-5'. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
Frame 2 : 3'-5'	[REDACTED]
AllergenOnline Full-FASTA3 alignments	[REDACTED]
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	No matches
BLASTP "allergen" (best match)	No matches
BLASTP "toxin" (best match)	No matches
BLASTP "toxic" (best match)	No matches

Table 21. Chromosome 1 Junction #4, Reading frame 3: 3'-5'. No significant matches to allergens or toxins.


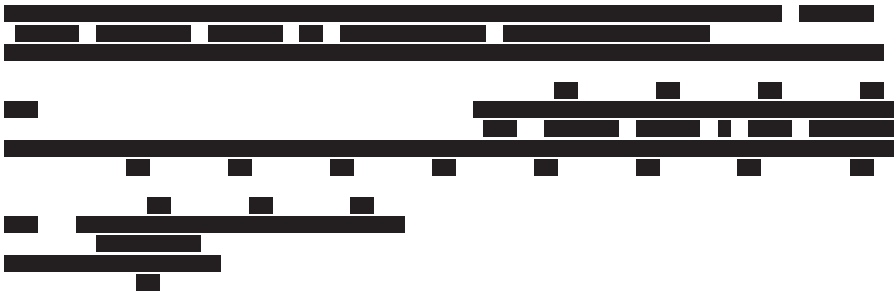




Reading frame:	Amino acid sequence (number of amino acids)
Frame 3: 3'-5'	
AllergenOnline Full-FASTA3 alignments	
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best matches)	
BLASTP "allergen" (best matches)	
BLASTP "toxin" (best match)	
BLASTP "toxic" (best match)	

Table 22. Chromosome [REDACTED] Reading frame 1: 5'-3'. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
[REDACTED] Frame 1: 5'-3'	[REDACTED]
AllergenOnline Full-FASTA3 alignment	[REDACTED]
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	[REDACTED]
BLASTP "allergen" (best matches)	[REDACTED]
BLASTP "toxin" (best match)	No significant matches to a sequence with keyword toxin
BLASTP "toxic" (best match)	No significant matches to a sequence with keyword toxic

Table 23. Chromosome [REDACTED] Reading frame 2: 5'-3'. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
[REDACTED] Frame 2: 5'-3'	[REDACTED]
AllergenOnline Full-FASTA3 alignment	[REDACTED]
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	[REDACTED]
BLASTP "allergen" (best matches)	[REDACTED]
BLASTP "toxin" (best match)	[REDACTED]
BLASTP "toxic" (best match)	[REDACTED]

Table 24. Chromosome [REDACTED] Reading frame 3: 5'-3'. No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
[REDACTED] Frame 3: 5'-3'	[REDACTED]
AllergenOnline Full-FASTA3 alignment	[REDACTED]
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	No significant match found
BLASTP "allergen" (best matches)	No significant match found
BLASTP "toxin" (best match)	No significant match found
BLASTP "toxic" (best match)	No significant match found

Table 25. Chromosome [REDACTED] Reading frame 1: 3'-5'. Alignment of Peptides [REDACTED]
[REDACTED] No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
[REDACTED] Frame 1: 3'-5'	[REDACTED]
AllergenOnline Full-FASTA3 alignment	[REDACTED]
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	[REDACTED]
BLASTP "allergen" (best matches)	[REDACTED]
BLASTP "toxin" (best match)	No significant matches to a sequence with keyword toxin
BLASTP "toxic" (best match)	No significant matches to a sequence with keyword toxic

Table 26. Chromosome [REDACTED] Reading frame 2: 3'-5'. Alignment of Peptides [REDACTED]
[REDACTED] No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
[REDACTED] Frame 2: 3'-5'	[REDACTED] [REDACTED] [REDACTED]
AllergenOnline Full-FASTA3 alignment	[REDACTED] [REDACTED]
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]
BLASTP "allergen" (best matches)	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]
BLASTP "toxin" (best match)	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]
BLASTP "toxic" (best match)	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]

Table 27. Chromosome [REDACTED] Reading frame 3: 3'-5'. Alignment of Peptides [REDACTED]

No significant matches to allergens or toxins.

Reading frame:	Amino acid sequence (number of amino acids)
[REDACTED] Frame 3: 3'-5'	[REDACTED]
AllergenOnline Full-FASTA3 alignment	[REDACTED]
Sliding 80mer	No matches
Exact 8 aa match	No matches
BLASTP No keyword (best match)	No significant match found
BLASTP "allergen" (best matches)	No significant match found
BLASTP "toxin" (best match)	No significant match found
BLASTP "toxic" (best match)	No significant match found

4.2.2 Sliding 80-amino acid window FASTA3 vs. AllergenOnline.org version 16 with eight new proteins. Results of the comparisons of the amino acid sequences of the putative peptides against all of the sequences in AllergenOnline.org version 16 database were negative. This is a very stringent bioinformatics evaluation for potential risks of allergy and cross-reactivity based on the CODEX Alimentarius guidelines (2009). The lack of any match for each protein indicates low risk for allergy from these proteins (Tables 3-28).

4.2.3 Eight contiguous amino acid identity matches with AllergenOnline.org version 16. Each protein aa sequences of the putative peptides was searched for 8 contiguous amino acid identity matches as a conservative indicator of possible cross reactivity, based on the CODEX Alimentarius guidelines (2009). The lack of any match for any of the proteins indicates a very low probability of any shared IgE binding or potential cross-reactivity (Tables 3-28).

4.2.4 BLASTP of NCBI without keyword limit with each of the putative peptides in canola. The full-length sequences of the putative peptides were compared to all sequences in NCBI-Entrez database to find the most evolutionarily conserved proteins with results presented (Tables 3-27). The scoring alignments with *E* scores of the top three protein alignments identified by BLASTP were considered in some detail to determine if there is significant homology to proteins of sources with likely safe human exposure or unsafe exposure, and when compared to results from searches with keywords (allergen, allergy, toxin, toxic), provides a relative evaluation of potential risks. The results from BLASTP comparison to all proteins were neutral, but the ubiquitous nature of the proteins without obvious indications of harm suggesting they are generally safe, abundant enzymes.

4.2.5 BLASTP of NCBI Entrez using keywords “allergen”. The full-length amino acid sequences of the putative peptides were compared to sequences in NCBI Entrez, which were designated as “allergen” in the NCBI database on 19 December 2016. The alignment results with keyword “allergen” returned a few results with an *E* score of less than 10 e.g. Table 22). However, the *E* scores and identity matches did not show results that suggest high sequence identities and the matched proteins have tenuous connections to “allergen”, none are in the WHO/IUIS www.allergen.org database or in the www.AllergenOnline.org database. The aligned matches with keyword “allergy” (Tables 3-27) were not significant. Thus the probability of allergy to the eight proteins or of cross-reactivity is extremely small based on observations of Aalberse (2000) and Goodman et al. (2008).

4.2.6 BLASTP of NCBI Entrez with “toxin” and “toxic”. The putative peptide sequences from the junctions of DNA in canola were compared to sequences in NCBI-Entrez, which were designated as “toxin” or “toxic” in the NCBI database on 22 December 2016. The alignment results with keyword “toxic” did not return any significant alignments. There does not appear to be a basis to suspect that the eight proteins introduced into canola are likely toxins.

4.3 Bioinformatics summary for the putative peptides encoded at the [REDACTED] DNA junctions due to insertion of the genes in canola. None of the results from the bioinformatics searches of the amino acid sequences from the putative peptides at the junctions of inserted DNA in canola carry significant risks of allergy or toxicity compared to commonly consumed proteins from a diverse variety of food sources.

5.0 Conclusions

Bioinformatics analyses were performed on putative peptides encoded at the [REDACTED] DNA junctions present due to insertion of the DNA in two chromosomes in GE canola line to produce DHA. The function of this bioinformatics evaluation was to determine whether there might be some safety concerns for foods produced with this protein included as ingredients. All new potential ORFs were analyzed and no significant homologies were identified at the junctions of the introduced DNA and the endogenous canola DNA. Based on the evidence and my knowledge of cross-reactive IgE binding and food toxicology, there is not a scientifically justifiable reason to perform serum IgE binding studies with the peptides or to perform any toxicity tests. There is no known risk and no “at-risk” population (Goodman, 2008).

Sequence comparisons of the seven proteins to known toxins identified some statistically significant, but short and modest identity matches to diverse proteins that are from toxic sources or are associated with “toxins” or “toxicity” based on unclear associations (Report N° 2016-017). There were no clear links with pathogenic or toxic effects of the proteins that were matched. The homologous proteins have not been identified as toxins. Thus, *in vitro* and *in vivo* toxicity testing should not be required to evaluate food safety in our opinion.

Taken together, this bioinformatics analyses demonstrate that the development of the genetically modified DHA canola (elite event B0050-027) has not produced any new open reading frames, or new proteins, which would raise any safety concerns.

6.0 References

Aalberse RC. (2000). Structural biology of allergens. *J Allergy Clin Immunol* 106:228-238.

Codex Alimentarius Commission. (2003). Alinorm 03/34: Joint FAO/WHO Food Standard Programme, Codex Alimentarius Commission, Twenty-Fifth Session, Rome, Italy 30 June-5 July, 2003. 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, pp. 47-60.

Goodman RE. (2008). Performing IgE serum testing due to bioinformatics matches in the allergenicity assessment of GM crops. *Food Chem Toxicol* 46:S24-S34

Goodman RE, Vieths S, Sampson HA, Hill D, Ebisawa M, Taylor SL, van Ree R. (2008). Allergenicity assessment of genetically modified crops—what makes sense? *Nat Biotechnol* 26(1):73-81.

Metcalf, D.D., Astwood, J.D., Townsend, R., Sampson, H.A., Taylor, S.L., Fuchs, R.L. (1996). Assessment of the allergenic potential of foods derived from genetically modified crop plants. *Crit. Rev. Food Sci. Nutr.* 36, 165-186.

7.0 Unpublished References

Report N° 2016-002.

Nuseed Pty Ltd.

Report N° 2016-017.

Nuseed Pty Ltd.