FOOD STANDARDS
Australia New Zealand
Te Mana Kounga Kai – Ahitereiria me Aotearoa

9 November 2017
[32-17]

Supporting document 1

Safety assessment – Application A1147

Food derived from Herbicide-tolerant Cotton Line GHB811

Executive summary

Background

A genetically modified (GM) cotton line with OECD Unique Identifier BCS-GH811-4, hereafter referred to as GHB811, has been developed with tolerance to two herbicides: glyphosate and isoxaflutole. Tolerance to glyphosate is achieved through expression of the modified corn-derived gene *2mepsps*, which encodes the 2mEPSPS enzyme. Tolerance to isoxaflutole is achieved by the expression of a modified p-hydroxyphenyl pyruvate-dioxygenase (HPPD) enzyme, encoded by the *hppdPf W336* gene, derived from the soil bacterium *Pseudomonas fluorescens*.

In conducting a safety assessment of food derived from cotton line GHB811, a number of criteria have been addressed including: characterisation of the transferred genes including their origin, function and stability in the cotton genome; the nature of the introduced proteins and their potential to be either allergenic or toxic in humans; and compositional analyses and any resultant changes in the whole food. This approach evaluates intended and any unintended changes in the plant.

This safety assessment addresses only food safety and nutritional issues of the GM food *per se*. It therefore does not address:

- risks related to the environmental release of GM plants used in food production
- risks to animals that may consume feed derived from GM plants
- the safety of food derived from the non-GM (conventional) plant.

Food derived from the non-GM (conventional) plant with an accepted history of safe use is used as the benchmark for the comparative analysis.

History of use

The host organism is cultivated cotton (*Gossypium hirsutum* L.). Cotton is one of the oldest cultivated crops and is grown worldwide primarily as a fibre crop but also as a source of food products derived from the seed. Such products need to be highly processed because of the presence of natural toxicants (gossypol) and anti-nutrients (cyclopropenoid fatty acids) in unprocessed cottonseed. The main food products include cottonseed oil and linters.

Molecular characterisation

Cotton line GHB811 contains two genes introduced on a single expression cassette via *Agrobacterium*-mediated transformation. The *2mepsps* gene is a modified version of the *epsps* gene derived from *Zea mays* and encodes a modified version of 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), a critical enzyme in the synthesis of the aromatic amino acids. The modified protein 2mEPSPS maintains enzyme function in the presence of glyphosate. The *hppdPf W336* gene is a modified version of the *hppd* gene isolated from *Pseudomonas fluorescens* and encodes a modified p-hydroxyphenyl pyruvatedioxygenase (HPPD) enzyme, conferring tolerance to herbicides containing isoxaflutole. Detailed molecular analyses of cotton line GHB811 indicate that one complete copy of the two-gene expression cassette is present at a single insertion site. Plasmid backbone analysis shows no extraneous sequences, including antibiotic resistance genes derived from the plasmid, were transferred to the cotton line GHB811 genome.

The introduced genetic elements and the expression of new proteins in cotton line GHB811 were shown by phenotypic analysis and molecular techniques to be stably inherited from one generation to the next across multiple generations. The pattern of inheritance supports the conclusion that the herbicide-tolerance traits occur within a single locus in the cotton line GHB811 genome and are inherited in accordance with Mendelian principles.

Characterisation and safety assessment of new substances

Newly expressed proteins

A range of characterisation studies confirmed the identity of 2mEPSPS and HPPD*Pf* W336 expressed in cotton line GHB811 and their equivalence with the corresponding protein produced in a bacterial expression system. The plant 2mEPSPS and HPPD*Pf* W336 proteins had the expected molecular weights (47 kDa and 40 kDa respectively), immunoreactivity, lack of glycosylation, amino acid sequence and enzyme activity.

Both 2mEPSPS and HPPD*Pf* W336 were detected in all plant parts analysed, except for pollen. Expression of both proteins was highest in leaf tissue and lowest in the roots. The mean level of 2mEPSPS in seed from cotton line GHB811 was approximately 148 µg/g dw (0.0148%), corresponding to 0.0745 of total protein and that of HPPD W336 was approximately 28 µg/g dw (0.0028%), corresponding to 0.014% of total protein.

There are no concerns regarding the potential toxicity or allergenicity of the expressed proteins. Previous safety assessments of both 2mEPSPS and HPPD*Pf*W336 indicate that the proteins would be rapidly degraded in the stomach following ingestion and would be inactivated by heating. Additionally, updated bioinformatic studies considered in this assessment confirm the lack of any significant amino acid sequence similarity to known protein toxins or allergens.

Herbicide metabolites

The herbicide metabolites generated as a result of spraying cotton with glyphosate and isoxaflutole are expected to be the same as those found in other GM crops exposed to these herbicides. FSANZ has previously assessed the presence of herbicide metabolites for glyphosate sprayed onto cotton line GHB614 (A614) and isoxaflutole studies with soybean line FG72 (A1051).

Compositional analyses

Detailed compositional analyses were done on seed from cotton line GHB811, the control 'Coker 312' and seven commercial varieties grown under normal agricultural conditions at eight trial sites in cotton growing regions of the United States of America. The analyses included proximates (protein, fat, ash, moisture, carbohydrates by calculation), fibre components, fatty acids, amino acids, minerals, vitamins and anti-nutrients. The levels of 69 of these key analytes in cotton line GHB811 were compared to those in the control and to the ranges found in commercial non-GM cotton varieties grown concurrently in the same trial.

Statistically significant differences were found between seeds from cotton line GHB811 and the control for eleven of the analytes measured, however all differences were small in magnitude and were within the range established for existing commercial cotton varieties. Overall, the compositional data support the conclusion that there are no biologically significant differences in the levels of key constituents in seed from cotton line GHB811 compared to conventional cotton varieties available on the market.

Conclusion

No potential public health and safety concerns have been identified in the assessment of cotton line GHB811. On the basis of the data provided in the present Application, and other available information, food derived from GHB811 is considered to be as safe for human consumption as food derived from conventional cotton varieties.

Table of contents

Index of Figures

Index of Tables

List of abbreviations

a.i.	active ingredient
BC	backcrossed
BLOSUM	BLOcks SUbstitution Matrix
bp	base pairs
Da	dalton
DNA	deoxyribonucleic acid
DW	dry weight
ELISA	enzyme-linked immunosorbent assay
EMBOSS	European Molecular Biology open software suite
EPSP	5-enolpyruvylshikimate-3-phosphate
EPSPS	5-enolpyruvylshikimate-3-phosphate synthase
FARRP	Food Allergy Research and Resource Program
FASTA	Fast alignment search tool – all
FDA	U.S. Food and Drug Administration
FAO	Food and Agriculture Organization of the United Nations
FSANZ	Food Standards Australia New Zealand
FW	fresh weight
g	gram
GM	genetically modified
ha	hectare
HPPD	p-hydroxyphenyl pyruvatedioxygenase
kDa	kilodalton
LB	left border of T-DNA (*Agrobacterium tumefaciens*)
m	metres
mg	milligram
NCBI	National Centre for Biotechnology Information
OECD	Organisation for Economic Co-operation and Development
OGTR	Office of the Gene Technology Regulator
ORF	open reading frame
PCR	polymerase chain reaction
PEP	phosphoenolpyruvate
RB	right border of T-DNA (*Agrobacterium tumefaciens*)
RuBisCo	Ribulose-1,5-bisphosphate carboxylase/oxygenase
SAS	Statistical Analysis Software
T-DNA	transfer DNA
USA	United States of America
USDA	United States Department of Agriculture

1 Introduction

FSANZ has received an Application from Bayer CropScience Pty Ltd to vary Schedule 26 in the *Australia New Zealand Food Standards Code* (the Code). The variation is to add food derived from the genetically modified (GM) herbicide-tolerant cotton line GHB811 (hereafter referred to as GHB811), with the OECD Unique Identifier BCS-GH811-4. This line has dual-herbicide tolerance to glyphosate and isoxaflutole.

Tolerance to herbicides containing glyphosate is achieved with the introduction of a modified corn-derived gene *2mepsps*, which encodes the 2mEPSPS enzyme. FSANZ has previously approved food from GM lines containing the *2mepsps* gene in Applications: A362 – corn line GA21 (FSANZ 2000), A614 – cotton line GHB614 (FSANZ 2009), A1051 – soybean line FG72 (FSANZ 2011), A1073 – soybean line DAS-44406-6 (FSANZ, 2013) and A1112 – corn line MZHG0JG (FSANZ 2016). Tolerance to isoxaflutole is achieved by the expression of a modified p-hydroxyphenyl pyruvatedioxygenase (HPPD) enzyme, encoded by the *hppdPf W336* gene derived from the soil bacterium *Pseudomonas fluorescens*. FSANZ has previously approved food from one GM line containing the modified *hppdPf W336* gene in Application A1051 – soybean line FG72 (FSANZ 2011).

Cotton lines containing the GHB811 transformation event are intended initially for cultivation in the United States of America (USA) and Brazil but may later be grown in other cotton-producing countries. Food from GHB811 cotton will therefore enter the Australian and New Zealand food supply via imported products.

2 History of use

2.1 Host organism

The information provided here has been summarised from more detailed reports published by the Organisation for Economic Cooperation and Development (OECD, 2008), the Office of the Gene Technology Regulator (OGTR 2016), the Agricultural Biotechnology Council of Australia (ABCA, 2012), the United States Department of Agriculture Foreign Agricultural Service (USDA 2001, 2016, 2017), and the National Cottonseed Products Association (NCPA 2002, unknown date). Numerical and statistical data has been sourced from the FAOSTAT website from the Food and Agriculture Organization of the United Nations (FAO 2017).

The host organism *Gossypium hirsutum* L. is the most highly cultivated cotton species, and currently accounts for 90% of global cotton production. Cotton is predominantly grown for the textile industry, where evidence of human use has been found in archaeological sites in Pakistan, Egypt and Mexico dating back 5000 years. Beyond the textile use, cotton seed can be crushed to extract edible oil (cottonseed oil) and the remaining short fuzzy fibres that coat the seeds, also known as linters, can be processed into forms of cellulose that have use as food additives.

Cottonseed oil is commonly used for deep frying in restaurants and the fast food industry and can be used in margarines and salad oils. In Australia and New Zealand, cottonseed oil has been in the food supply for approximately 50-60 years and globally there is a long history of safe use, with commercialisation of the oil for human use beginning in the early 1900's. Cottonseed oil is imported into Australia and New Zealand to supplement domestic supply. In Australia, current domestic use of cottonseed oil is around 116 million tons per year.

Cotton linters are a source of pure cellulose that can be used in industry. In the food industry, cellulose from linters can be used to prepare the food additives microcrystalline cellulose

(INS# 460) and sodium carboxylmethyl cellulose (INS# 466). These additives are used in a variety of foods to perform different functions, for example as emulsifiers in low fat ice creams or as anti-caking agents in packaged shredded cheese. Cellulose from cotton linters is also used to bind solids in pharmaceutical preparations such as tablets, due to being white and flavourless.

The cotton plant is known to produce toxic compounds. Gossypol is a terpenoid aldehyde found in the pigment glands of root, leaves, flower buds and seeds. It is highly toxic to non-ruminant mammals, birds, insects and microorganisms as it can interfere in lysine metabolism and mitochondrial function. Plants that produce low levels of gossypol or have no gland tissue are highly susceptible to disease and pest infestations thus this toxin is considered to play an important part in the plants natural defence system. Another group of toxic compounds present in cotton seeds are the cyclopropenoid fatty acids, which are thought to deter insects. They are considered anti-nutrients because they can interfere in metabolism of saturated fats.

In cotton seed, the levels of gossypol range from 0.4-2.0% and the cyclopropenoid fatty acids range from 0.5-1%. During the processing of oil from the seeds, both compounds are inactivated resulting in minimal presence in human food products. Problems can arise when the partially processed cottonseed is used to make flour or meal for human foods, which is common in countries like Central America, India and Pakistan. Gland-less cotton that does not produce gossypol has been developed for these countries where cotton meal is used as a source of protein in the human diet.

Cotton is not known to be allergenic thus it is used widely in the pharmaceutical and medical industries, as well as the textile industry. Cotton dust generated during milling can cause an asthma-like condition in mill workers but generally consumers of cottonseed oil and linters are not exposed to cotton dust.

The cotton cultivar ‘Coker 312’ was used as the parental variety for the genetic modification described in this Application, and thus is regarded as the near-isogenic line for the purposes of the comparative assessment. It is no longer a commercially grown cultivar but is widely used to produce GM cotton lines because they can be readily cultured, transformed and regenerated in the laboratory. Traits introduced into ‘Coker’ cultivars are transferred to commercial cultivars by cross breeding.

2.2 Donor organisms

2.2.1 *Zea mays*

The DNA sequence encoding the EPSPS protein was originally isolated from the *Zea mays* variety Black Mexican Sweet. This variety of corn has been cultivated as a food crop in the USA since 1864 thus has a long history of safe use as food (reviewed in Herouet-Guicheney et al, 2009).

2.2.2 *Pseudomonas fluorescens*

The DNA sequence that encodes the HPPD enzyme was initially isolated from *Pseudomonas fluorescens* strain A32 (Sailland et al, 2001). This organism is considered a non-pathogenic, asporogenous saprophyte found in soil, water and on plant surface environments. This organism is generally considered non-pathogenic to humans because it has limited growth at body temperature but has been associated with opportunistic infections in immunocompromised patients or after infusion of contaminated whole blood or blood products (Scales et al, 2014).

2.2.3 *Arabidopsis thaliana*

A. thaliana is a member of the mustard family (*Brassicaceae*), which also includes cabbage, turnip and rapeseed, and is commonly known as mouse-ear or thale cress. Although this plant is not traditionally used as food, it is ubiquitous in the environment and is not known to be pathogenic, toxigenic or allergenic to humans.

2.2.4 *Helianthus annuus*

Sunflower has a long history of safe use, humans having consumed sunflower seeds as whole seeds, oil or protein-rich flour for centuries (OECD, 2004). The seeds do not contain significant amounts of anti-nutrients or toxins.

2.2.5 Other organisms

Genetic elements from several other organisms have been used in the genetic modification of GHB811 (refer to Table 1). These non-coding sequences are used to drive or enhance expression of the two new genes. None of the sources of these genetic elements are associated with toxic or allergenic responses in humans. The genetic elements derived from plant pathogens (Cassava vein mosaic virus) are not pathogenic in themselves and do not cause pathogenic symptoms in GHB811.

3 Molecular characterisation

Molecular characterisation is necessary to provide an understanding of the genetic material introduced into the host genome and helps to frame the subsequent parts of the safety assessment. The molecular characterisation addresses three main aspects:

- the transformation method together with a detailed description of the DNA sequences introduced to the host genome
- a characterisation of the inserted DNA, including any rearrangements that may have occurred as a consequence of the transformation
- the genetic stability of the inserted DNA and any accompanying expressed traits

The Applicant has submitted the following unpublished studies for the molecular characterisation of line GHB811.

Unpublished studies

1. Description of vector pTSIH09 (2015) Report M-388224-02, Bayer CropScience.
2. Description of the GHB811 cotton transformation methodology (2015) Report M-543362-01, Bayer CropScience.
3. Detailed insert characterization and confirmation of the absence of vector backbone sequence in cotton GHB811 (2016) Report M-572036-01, Bayer CropScience.
4. GHB811 cotton - inheritance of the insert over three generations (2016) Report M-547925-02, Bayer CropScience.
5. Structural stability analysis of cotton GHB811 (2016) Report M-548778-01, Bayer CropScience.
6. DNA sequence determination of the transgenic and insertion loci of cotton GHB811 (2015) Report M-533573-02, Bayer CropScience.
7. Bioinformatics analysis of the GHB811 cotton insertion locus (2017) Report M-581222-01, Bayer CropScience.
8. GHB811 cotton identification of open reading frames and homology search of sequences ≥30 amino acids to known allergens and toxins (2016) Report M-575144-01, Bayer CropScience.

3.1 Transformation Method

In order to create GHB811, plasmid pTSIH09 was transformed into the cotton variety Coker 312. Plasmid pTSIH09 contains two gene cassettes that encode the enzymes 2mEPSPS and HPPD*Pf* W336 (Figure 1).

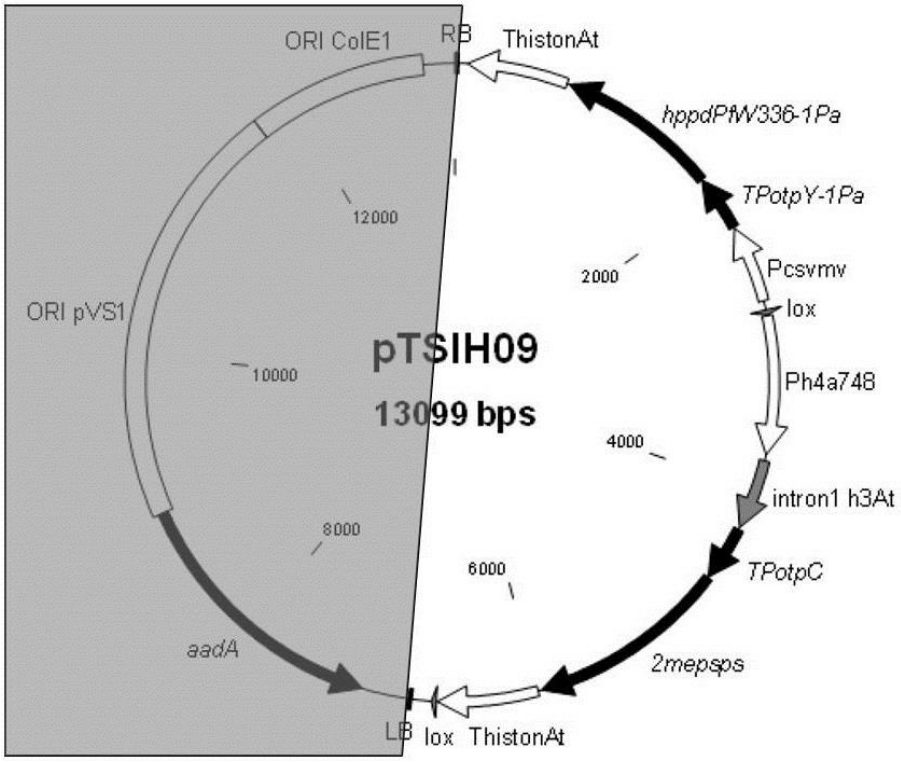

Figure 1: Plasmid map of TSIH09. The plasmid contains a T-DNA insert region that contains the two gene cassettes and is shown with a white background. The plasmid backbone region is shown with a grey background.

The transformation method involved infection of hypocotyl tissue, dissected from cotton seedlings, with a C58-derived strain of *Agrobacterium* C58C1 (van Larebeke et al., 1974) containing the pTSIH09 plasmid. Hypocotyl pieces were then grown on media containing the antibiotic ticarcillin, which suppresses the growth of the agrobacterium, and glyphosate, to select for positive (glyphosate-tolerant) transformants. After generation of T0 plantlets, further selection in the presence of HPPD-inhibitors was performed.

3.2 Detailed description of DNA to be introduced

The plasmid TSIH09 (Figure 1), used to generate GHB811, contains two gene cassettes in the T-DNA region. At the left border (LB) is the *2mepsps* gene cassette (Figure 2). Expression of *2mepsps* is under the control of the histone H4 promoter (P H4a748A) and terminator (3' h4AT) sequences from *Arabidopsis thaliana*. Histone genes are ubiquitously expressed in all tissues but the H4A748 promoter has been shown to be more highly expressed in meristematic tissues (Atanassova et al, 1992). Use of this promoter may allow higher expression in tissues undergoing growth and development, which in turn may ensure there is high expression of the herbicide genes during the growth stage when application of herbicides would more likely be used. The histone H4A748 promoter is followed by an intronic sequence from a histone H3 variant from *A. thaliana* (I h3At) (Chaubet et al, 1996). This intronic sequence has been included upstream of the transcriptional start site to improve the expression of the *2mepsps* gene (Rose 2004; Derose et al, 2011).

Figure 2: The 2mepsps gene cassette in pTSIH09.

The protein coding sequence in the *2mepsps* gene cassette begins with an optimised transit peptide (TP) based on the RuBisCo small subunit gene found in sunflower (*Helianthus annuus*) and corn (*Zea mays*) (Lebrun et al, 1996). This transit peptide ensures the enzyme is translocated to the plastids, such as chloroplasts, after protein translation. The transit peptide is followed by the mutated *epsps* gene from corn. The mutations to the *epsps* gene were introduced by site-directed mutagenesis and include a threonine to isoleucine change at position 102, a proline to serine change at position 106 and an additional methionine residue at the N-terminus of the protein coding region (Lebrun et al, 1996), to allow for efficient cleavage of the transit peptide. The amino acid changes within the *epsps* gene result in a protein with reduced affinity for glyphosate, enabling it to function in the presence of glyphosate.

At the right border (RB) is the *hppdPf W336* gene cassette (Figure 3). Expression of *hppdPf W336* is under the control of the Cassava vein mosaic virus promoter (P csvmv) and terminator sequences from the histone H4 promoter from *A. thaliana* (3' h4AT). The protein coding sequence begins with the optimised transit peptide (TP) used in the *2mepsps* gene cassette, followed by the mutated *hppd* gene from *Pseudomonas fluorescens*. This gene has been modified by site-directed mutagenesis to replace glycine at position 336 with tryptophan (Boudec et al., 2001). This amino acid change results in a protein with reduced affinity for isoxaflutole, enabling it to function in the presence of isoxaflutole-containing herbicides.

Figure 3: The hppdPf W336 gene cassette in pTSIH09.

There are intervening sequences present in the T-DNA region as outlined in Table 1. These sequences assist with cloning, mapping and sequence analysis. There are also two cre-lox sites which would allow future directed-insertion or deletion of DNA sequences at the site of integration.

Table 1: The genetic elements contained in the T-DNA region of pTSIH09, used to create GHB811.

Genetic element	Relative position	Orientation	Source	Description, Function & Reference
Right border (RB)	1 - 25		*Agrobacterium tumefaciens*	Right border repeat from the T-DNA (Zambryski, 1988)
Right border region	26 - 82		synthetic	Supports cleavage at RB and provides buffer for truncations
3′histonAt	83-749	counter clockwise	*Arabidopsis thaliana*	3´ untranslated region of the histone H4 gene containing the termination sequence (Chabouté et al, 1987)
Intervening sequence	750 - 765		synthetic	
hppdPfW336-1Pa	766 - 1842	counter clockwise	*Pseudomonas fluorescens*	coding sequence of the 4-hydroxyphenylpyruvate dioxygenase gene (Boudec et al., 2001)
TPotpY-1Pa	1843 - 2214	counter clockwise	*Helianthus annuus* *Zea mays*	coding sequence of an optimized transit peptide derivative from the RuBisCO small subunit gene (Lebrun et al., 1996)
Intervening sequence	2215 - 2222		Synthetic	
Pcsvmv	2223 - 2829	counter clockwise	*Cassava vein mosaic virus*	promoter sequence to drive expression of the *hppdPf W336* gene (Verdaguer et al., 1996)
Intervening sequence	2736 - 2795		synthetic	
lox	2796 - 3750	clockwise	*Bacteriophage P1*	Cre recombinase recognition site (Hoess and Abremski, 1985)
Intervening sequence	2830 - 2833		synthetic	
Ph4a748	2834 - 3750	clockwise	*Arabidopsis thaliana*	Promoter sequence of the histone H4 gene (Chabouté et al, 1987)
Intervening sequence	3751 - 3789		synthetic	
Intron1 h3At	3790 - 4255	clockwise	*Arabidopsis thaliana*	First intron of gene II of histone H3.III variant (Chaubet et al., 1992)
Intervening sequence	4256 - 4268		synthetic	
TPotpC	4269 - 4640	clockwise	*Helianthus annuus* *Zea mays*	coding sequence of an optimized transit peptide derivative from the RuBisCO small subunit gene (Lebrun et al., 1996)
2mepsps	4641 - 5978	clockwise	*Zea mays*	coding sequence of the double-mutant 5-enol-pyruvylshikimate-3-phosphate synthase gene (Lebrun et al., 1997)
Intervening sequence	5979 - 5998		synthetic	
3′histonAt	5999 - 6665	clockwise	*Arabidopsis thaliana*	3´ untranslated region of the histone H4 gene containing the termination sequence (Chabouté et al, 1987)
Intervening sequence	6666 - 6669		synthetic	
lox	6670 - 6703	clockwise	*Bacteriophage P1*	Cre recombinase recognition site (Hoess and Abremski, 1985)
Left border region	6704 - 6831		synthetic	Supports cleavage at LB and provides buffer for truncations
Left Border (LB)	6832 - 6856		*Agrobacterium tumefaciens*	Left border repeat from the T-DNA (Zambryski, 1988)

3.3 Development of the cotton line from original transformation

A breeding programme was undertaken for the purposes of:

- obtaining generations suitable for analysing the characteristics of GHB811
- ensuring that the GHB811 event is incorporated into an elite line for commercialisation of isoxaflutole- and glyphosate-tolerant cotton.

The breeding pedigree for the various generations is given in Figure 4.

Following development of the T0 plants (see Section 3.1) and initial selection for tolerance to HPPD-inhibitor, a series of self-pollination (selfing) and seed bulking proceeded up to generation T7. T0 plants were also crossed with a commercial Stoneville line to produce an F1 generation that was then selfed, or underwent backcrosses to Stoneville 457 (followed by selfing), to produce further F generations.

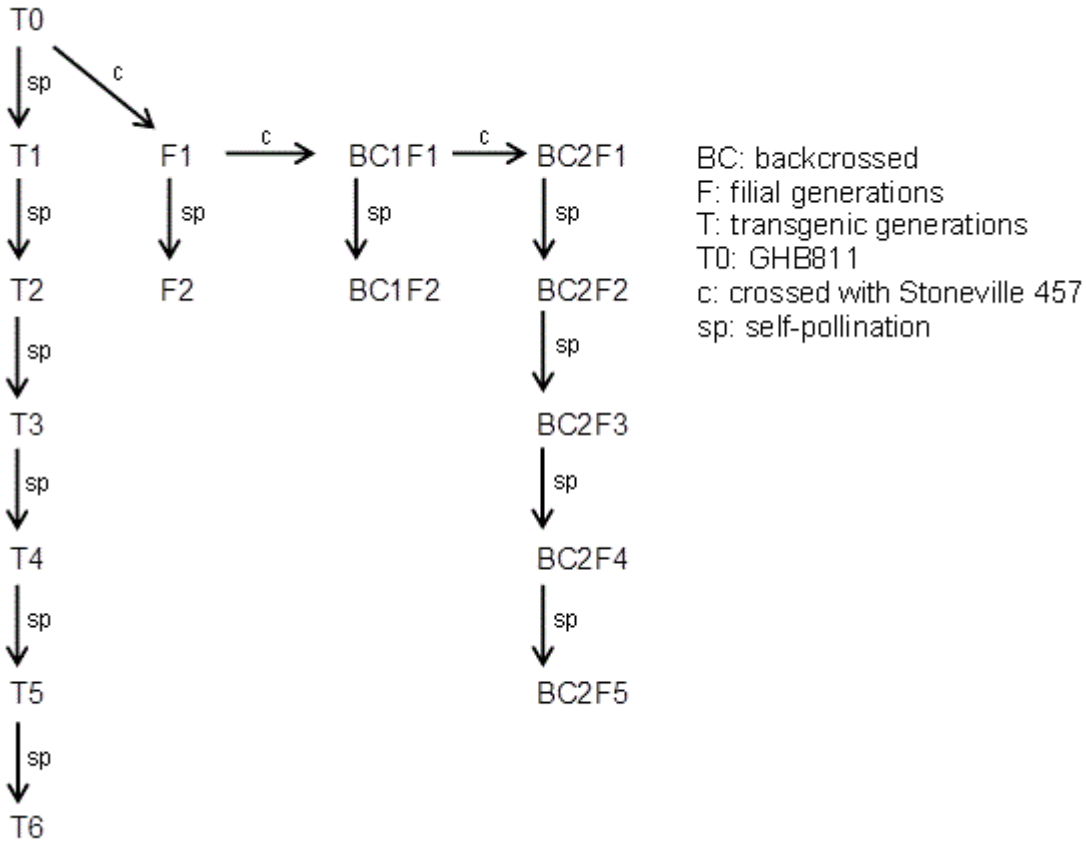

Figure 4: Breeding path used in the characterisation of the GHB811 line

3.4 Characterisation of the inserted DNA and site(s) of insertion

A range of analyses were undertaken to characterise the genetic modification in GHB811. These analyses focussed on the nature and stability of the insertion and whether any unintended re-arrangements or products may have occurred as a consequence of the transformation procedure. When characterising GHB811, different generations of plants were analysed and these are outlined in Table 2.

Table 2: Molecular characterisation studies performed in GHB811

Analysis	generation analysed	Control(s) used
Identifying the number of integration sites (Section 3.4.1)	T1	Coker312, Coker312 spiked with pTSIH09
Detection of backbone sequence (Section 3.4.2)	T1 and BC2F3	Coker312, Coker312 spiked with pTSIH09
Inheritance of the inserted DNA (Section 3.4.3)	F2, BC1F2, BC2F2	—
Genetic stability of the inserted DNA (Section 3.4.3)	T1, T3, T4, BC1F2, BC2F3	Coker312, Coker312 spiked with pTSIH09
Insert integrity and site of integration (Section 3.4.4)	T1	Coker312

3.4.1 Identifying the number of integration sites

Southern blot analysis was performed on leaf-derived genomic DNA from GHB811, the parental Coker 312 and Coker 312 spiked with pTSIH09. Genomic DNA was digested with various restriction enzymes and hybridised with a series of probes spanning the T-DNA region of TSIH09 (Figure 5). Comparison of hybridisation patterns between the parental control and GHB811 showed that a single integration event occurred.

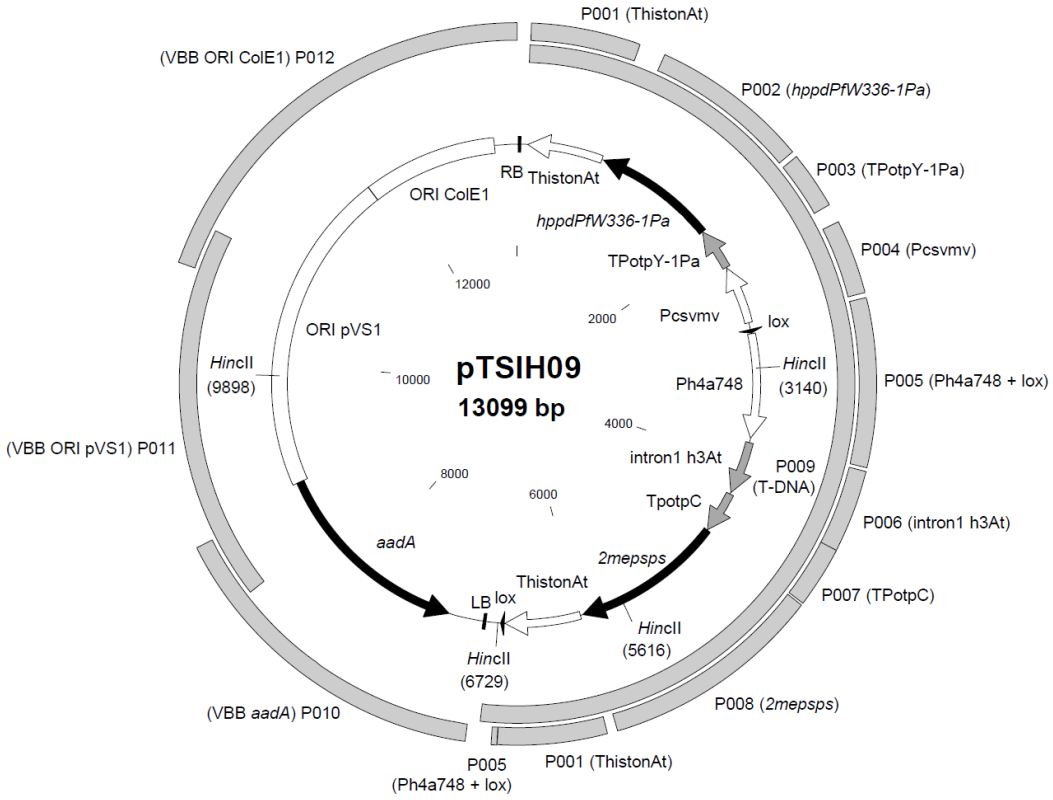

Figure 5: Map of the probes used for Southern blot analysis.

3.4.2 Detection of backbone sequence

Southern blot analysis was performed on leaf-derived genomic DNA from GHB811, the parental Coker 312 and Coker 312 spiked with pTSIH09. Genomic DNA was digested with numerous restriction enzymes and hybridised with a series of probes spanning the backbone region of TSIH09 (Figure 5). Hybridisation only occurred in the plasmid spiked Coker 312

control indicating that backbone sequences were absent in GHB811. Re-probing the Southern blot membranes with probes targeting the T-DNA region generated and the expected banding patterns, demonstrating that the experimental methodology was working and able to detect presence of targeted DNA.

3.4.3 Inheritance and genetic stability of the inserted DNA

Since it was demonstrated that the insert is present at a single locus in the GHB811 genome, there is the expectation that the genetic elements within this locus would be inherited according to Mendelian principles.

Chi-square (X^2) analysis was undertaken over several generations (as outlined in Figure 4 and Table 2) to confirm the segregation and stability of the T-DNA insert in GHB811. The inheritance pattern was assessed in the F2, BC1F2 and BC2F2 generations by quantitative real-time PCR analysis. Primers targeted the *2mepsps* and *hppdPfW336-1Pa* genes and an endogenous cotton gene.

The expected segregation ratio at the F2 generations was 1:2:1 and the critical value to reject the hypothesis of this ratio at the 5% confidence level was 5.99 (Strickberger, 1976). As the X^2 values calculated from these experiments were < 5.99, the results showed there were no significant differences between the observed and expected segregation ratios in any of the generations (Table 3). These data support the conclusion that the T-DNA is present at a single locus in GHB811 and was inherited predictably according to Mendelian principles in subsequent generations.

Table 3: Observed versus expected genotype for the novel genes in F2, BC1F2 and BC2F2 of GHB811

	F2		BC1F2		BC2F2	
	Observed	**Expected**	**Observed**	**Expected**	**Observed**	**Expected**
homozygous	19	21.25	24	23	45	57.25
hemizygous	49	42.50	50	46	116	114.5
null	17	21.25	18	23	68	57.25
X^2	2.08		1.48		4.66	

Southern blot analysis was also performed on leaf-derived genomic DNA from several generations of transgenic and back-crossed lines (Table 2). The DNA was digested to generate three distinguishable bands and hybridised using the P009 probe (Figure 5). The results showed that the inserted DNA was structurally stable in the GHB811 cotton line over several generations.

3.4.4 Insert integrity and site of integration

In order to identify rearrangements, deletions and insertions in the integrated DNA, leaf-derived genomic DNA samples from GHB811, Coker 312 and Coker 312 spiked with TSIH09 were mapped using probes spanning the T-DNA region by Southern blotting (Figure 5). High fidelity PCR was also performed to amplify the transgenic locus in GHB811 and the DNA sequence was compared to the T-DNA region of pTSIH09. The results confirmed that one complete copy of the T-DNA region was inserted in GHB811 and showed there have been no rearrangements, insertions or deletions.

DNA sequence analysis was also used to identify the site of integration. The location of integration for the pTSIH09 insert in GHB811 was chromosome A05. At the identified locus,

there are no known annotated genes thus the insertion should not have disrupted any genes. The insertion of DNA from pTSIH09 resulted in a 13 bp deletion of genomic DNA.

3.4.5 Open reading frame (ORF) analysis

The Applicant used the search program GetORF to identity all start-to-stop ORFs in both the inserted DNA and junctions between the insert and genomic DNA. All six reading frames were analysed. ORFs of 3 or more amino acids were initially captured identifying a total of 549 ORFs. From this pool, 126 potential proteins were identified with a minimum length of 30 amino acids. Proteins of 30 or more amino acids meet the minimum requirements of a 35% match over an 80 amino acid sequence (Codex 2009). These 126 potential proteins were then used as query sequences in homology searches for known allergens and toxins.

3.4.5.1 Bioinformatic analysis for potential allergenicity

The Applicant has provided the results of *in silico* analyses comparing the 126 amino acid sequences identified as ORFs to known allergenic proteins in the Food Allergy Research and Resource Program (FARRP) dataset, which is available through AllergenOnline[1] (University of Nebraska). At the date of the search, there were 1956 sequences in the allergen database. Three types of analyses were performed for this comparison:

(a) Full length sequence search – a FASTA alignment was performed comparing the whole sequence to the database entries. Significant homology was determined when there was more than 50% similarity between the query protein and database entry (BLOSUM50), with the E-value threshold very conservatively set at 1.

(b) 80-mer sliding window search – a FASTA alignment was performed comparing all contiguous 80 amino acids within the ORF to the database entries. Matches were identified if there was greater than 35% homology.

(c) 8-mer exact match search – A FASTA alignment was performed comparing contiguous 8 amino acids within the ORF to the database entries. Matches were identified if there was 100% homology. The search was performed using SeqMatchAll from the European Molecular Biology Open Software Suite.

Of the 126 potential ORFs used to query the FARRP database, no similarities were found to any of the known allergenic proteins.

3.4.5.2 Bioinformatic analysis for potential toxicity

The Applicant provided results from *in silico* analyses comparing the 126 potential proteins identified as ORFs to known protein toxins identified in the NCBI non-redundant protein database. A FASTA algorithm was used with a BLOSUM50 scoring matrix and the E-value threshold set to 0.1.

Of the 126 potential ORFs, there was no homology to any biologically relevant toxins. There was homology to plant proteins associated with resistance to toxigenic plant pathogens and to the EPSPS and HPPD enzymes that are targets for the herbicides glyphosate and isoxaflutole respectively. These herbicides are considered toxigenic to plants. Neither the plant immune proteins nor the EPSPS and HPPD enzymes are considered toxins themselves.

[1] www.allergenonline.org

3.4.6 Conclusion

The data provided by the Applicant showed that a single integration event has occurred at a specific locus with no disruption to endogenous genes. The complete T-DNA region from pTSIH09, containing the *2mepsps* and *hppdPf W336* gene cassettes, has been inserted, without rearrangement, deletions or insertions. Furthermore, no backbone sequences from the transforming plasmid were incorporated, including any antibiotic resistance genes. The introduced DNA was shown to be stably inherited from one generation to the next. Several ORFs encoding potential proteins were identified but were shown to have no homology to known allergens or toxins and would be unlikely to be biologically active due to the absence of required elements that would enable expression of a protein.

4 Characterisation and safety assessment of novel substances

In considering the safety of novel proteins it is important to consider that a large and diverse range of proteins are ingested as part of the normal human diet without any adverse effects. Only a small number of dietary proteins have the potential to impair health, because they have anti-nutrient properties or they can cause allergies in some consumers (Delaney et al. 2008). As proteins perform a wide variety of functions, different possible effects have to be considered during the safety assessment including potential toxic, anti-nutrient or allergenic effects.

To effectively identify any potential hazards, knowledge of the characteristics, concentration and localisation of all newly expressed proteins in the organism as well as a detailed understanding of their biochemical function and phenotypic effects is required. It is also important to determine if the newly expressed protein is expressed in the plant as expected, including whether any post-translational modifications have occurred.

The Applicant has submitted the following unpublished studies regarding the molecular characterisation of the novel proteins expressed in the GHB811 cotton line.

Unpublished studies

1. GHB811 Cotton - protein expression analyses of field samples grown in the USA during 2015 (2017) Report M-574232-01, Bayer CropScience.
2. Characterization of the recombinant 2mEPSPS protein batch 1417_2mEPSPS (2014) Report M-497839-01, Bayer CropScience.
3. Characterization of 2mEPSPS protein purified from GHB811 cotton and comparability with the recombinant 2mEPSPS protein batch 1417_2mEPSPS (2016) Report M-568145-01, Bayer CropScience.
4. 2mEPSPS protein: amino acid sequence homology search with known allergens and known toxins (2016) Report M-445681-04, Bayer CropScience.
5. Safety assessment of the double mutant 5-enol-pyruvylshikimate-3-phosphate synthase (2mEPSPS) protein (2006) Report M-278169-01, Bayer CropScience.
6. Characterization of the recombinant HPPD W336 protein batch 1411_HPPD W336 (2014) Report M-497842-01, Bayer CropScience.
7. Characterization of HPPD W336 protein purified from GHB811 cotton and comparability with the recombinant HPPD W336 protein batch 1411_HPPD W336 (2016) Report M-576569-02, Bayer CropScience.
8. HPPD W336 protein: amino acid sequence homology search with known allergens and known toxins (2016) Report M-445678-04, Bayer CropScience.

4.1 Description of the 2mEPSPS protein

5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) is a critical enzyme in the synthesis of the aromatic amino acids (phenylalanine, tryptophan and tyrosine) and other aromatic compounds. This synthesis pathway, known as the shikimate pathway, occurs exclusively in plants and microorganisms, including fungi (Herrmann, 1995; Tzin and Galili, 2010). Inhibition of EPSPS by glyphosate blocks synthesis of the aromatic amino acids, leading to plant death. Humans and other mammals do not have this pathway so are dependent on dietary intake of these amino acids from plant sources. As EPSPS is present in all plants commonly used as food, this protein has a long history of safe human consumption.

By introducing two amino acid substitutions into the EPSPS protein found in corn, threonine to isoleucine at position 102 and proline to serine at position 106 (Lebrun et al., 1997), and expressing this in plants, the resulting mutant enzyme is still able to function but cannot bind glyphosate, thus there is no inhibition of the shikimate pathway and the plant survives (Herouet-Guicheney et al, 2009). This occurs even when there is co-expression of the wild type endogenous EPSPS protein.

The modified 2mEPSPS protein shares 99.6% sequence homology to wild type corn EPSPS. This sequence difference poses no safety concern considering the sequence variation that exists across the common plant foods that form part of the human diet. For example, rice EPSPS shares 86% homology and tomato shares 75% homology to corn EPSPS (Herouet-Guicheney et al, 2009). FSANZ has also approved several plant lines engineered to express the 2mEPSPS protein (see Introduction).

The *2mepsps* gene prepared by the Applicant encodes a protein of 569 amino acids, consisting of a 124 amino acid transit peptide and a mature protein of 445 amino acids. After cleavage of the plastid transit peptide, the mature protein will have an expected molecular mass of 47.5 kDa.

4.1.1 Expression of the 2mEPSPS protein in cotton tissues

Protein expression in plant tissues was determined by an enzyme-linked immunosorbent assay (ELISA). The ELISA system used is a commercially available product (Envirologix) and is specific to the 2mEPSPS protein being expressed in cotton tissues. Equivalent plant tissue extracts from the parental Coker 312 variety was used as a background control in the ELISA. The standard curve was generated using microbially expressed 2mEPSPS, of which the characterisation is described in section 4.1.3.

In order to determine the sites of accumulation of the 2mEPSPS protein, samples were collected from GHB811 (generation T6) grown in the presence or absence of glyphosate and isoxaflutole, from three field-trial sites in the USA. For each tissue sample analysed, four samples were processed from each field-trial site. Specific tissues were collected at different growth stages. The growth stage has been indicated by way of the extended BBCH scale for cotton (Munger et al, 1998).

The results from the protein analysis (Table 4) showed the highest expression of 2mEPSPS was in the foliage. This was demonstrated in leaf material collected at the leaf development (14-16), inflorescence emergence (51-55) and post-flowering (60-69) stages. This was expected considering the presence of the transit peptide in the *2mepsps* cassette facilitates protein accumulation in tissues that contain plastids, and leaf tissue generally contains cells with the highest number of plastids, such as chloroplasts. The lowest expression was seen in pollen, fuzzy seed and root tissue. 2mEPSPS was also found in the developing flower and fertilised seed, from the flower bud or square phase, to the boll and final fuzzy seed stage.

Expression levels decreased from flower bud to boll to seed stage, which is likely associated with the loss of plastid-containing tissue. Treatment of the plants with both glyphosate and isoxaflutole-containing herbicides did not significantly change the expression of 2mEPSPS.

Table 4: Expression of 2mEPSPS in plant tissue samples

Tissue	Growth Stage	Treatment[1]	2mEPSPS (µg/g DW)[2]		2mEPSPS (µg/g FW)[3]	
			Mean	Range	Mean	Range
Leaf	14-16	-	968.03	211-1424	161.89	37-236
		+	874.63	347-1418	144.25	64-193
Leaf	51-55	-	1344.37	840-1651	255.91	174-380
		+	1269.39	1030-1686	252.37	177-343
Leaf	60-69	-	1422.12	1117-1763	307.05	239-384
		+	1267.95	756-1600	273.05	156-352
Root	14-16	-	169.25	97-250	24.56	16-30
		+	163.76	118-219	23.08	15-29
Pollen[4]	60-69	-	n/a	n/a	24.69	13-34
		+	n/a	n/a	27.68	21-33
Square	60-67	-	591.00	486-689	126.23	106-151
		+	506.64	382-659	106.20	78-125
Bolls	60-67	-	474.77	360-576	80.29	58-123
		+	437.00	319-522	71.26	56-86
Fuzzy Seed	83-97	-	145.11	76-221	129.79	65-206
		+	150.88	87-199	132.94	81-163

1. Treatment refers to plants grown in the absence (-) or presence (+) of herbicides. 2. DW - dry weight. 3. FW - fresh weight. 4. Protein can only be measured in fresh not dried pollen.

The presence of 2mEPSPS was also examined in edible fractions of the plant using ELISA (Table 5). The highest levels of 2mEPSPS were in the fuzzy and delinted seeds. The linters contained about 10-15% the level of 2mEPSPS as the fuzzy seeds. The linters used in this analysis were first-cut linters as they contain approximately 3% total protein (see Section 5.3.1). Because linters are actually made of pure cellulose, the protein is likely to be a contaminant carried through from seed or other vegetative material.

Cottonseed hull and meal contained ~30% and ~20% respectively the level of 2mEPSPS found in delinted seeds. Toasted meal had no detectable 2mEPSPS. The subsequent decrease in 2mEPSPS is expected considering that during the preparation of the meal fractions, the samples are exposed to temperatures greater than 150°C for no less than 1 hour, and toasted meal fractions are heated to above 220°C for no less than 45 min. At these temperatures, most large polypeptides would be denatured and be partially degraded.

2mEPSPS was not detected in the crude or refined oil fractions. This is not unexpected considering the mean level of 2mEPSPS in cottonseed is approximately 148 µg/g dw (0.0148%). As the mean percent dry weight of total protein in GHB811 cottonseed is approximately 20% dw (see Table 8), the amount of 2mEPSPS protein was calculated to be 0.074% of total protein. In view of the low levels of 2mEPSPS in cottonseed, their presence in the crude and refined oil will be difficult to detect.

Table 5: Expression of 2mEPSPS in edible plant fractions

Tissue	Treatment[1]	2mEPSPS	
		(µg/g DW)[2]	(µg/g FW)[3]
Fuzzy Seed	-	150.66	138.12
	+	123.48	112.99
Linters	-	15.69	14.73
	+	15.53	14.64
Delinted Seed	-	209.98	193.22
	+	209.87	193.60
Hull	-	67.01	60.65
	+	62.66	56.67
Untoasted Meal	-	28.49	25.56
	+	58.88	52.58
Toasted Meal	-	BD[4]	BD
	+	BD	BD
Crude Oil	-	n/a	BD
	+	n/a	BD
Refined Oil	-	n/a	BD
	+	n/a	BD

1. Treatment refers to plants grown in the absence (-) or presence (+) of herbicides. 2. DW - dry weight. 3. FW - fresh weight. 4. BD – below detection.

4.1.2 Characterisation of 2mEPSPS protein expressed in cotton and equivalence of bacterially-produced forms

2mEPSPS was affinity purified from leaf tissue collected from greenhouse grown plants. The protein was shown by electrophoresis to have an apparent molecular mass just below 50 kDa. Mass spectrometry (MS) identified two polypeptides with masses of 47284.9 Da (minor peak) and 47551.5 Da (major peak). N-terminal sequencing confirmed the presence of two polypeptides and showed that they differed in the junction between the transit peptide and mature protein. The major peak protein had the N-terminal sequence XMAGAEEIVL whereas the minor peak protein started at AGAEEIVL, without the expected methionine at position 1. Potential identification of the X amino acid in the major peak protein was determined by analysis of the molecular mass difference between the two polypeptides and knowledge that the final amino acid in the transit peptide was a cysteine residue. The analysis suggested that X was an oxidised version of cysteine, known as cysteine sulfinic acid.

The difference in N-terminal sequence between polypeptides was also confirmed by peptide mapping. Peptide mapping also showed that the protein being expressed in cotton was 2mEPSPS, with 88% sequence coverage achieved (Figure 6). Western blotting and ELISA demonstrated that the cotton-produced protein was immunoreactive to 2mEPSPS antibodies. The protein was also able to convert phosphoenolpyruvate (PEP) and shikimate-3-phosphate into 5-enolpyruvylshikimate-3-phosphate (EPSP), thus confirming the enzyme had specificity and was functional. Finally, the protein was shown by a modified Periodic Acid-Schiff staining procedure to be non-glycosylated, even though two putative glycosylation sites have been identified in corn EPSPS (Herouet-Guicheney et al, 2009). This indicates that no unintentional glycosylation has occurred in the modified plant expressed protein.

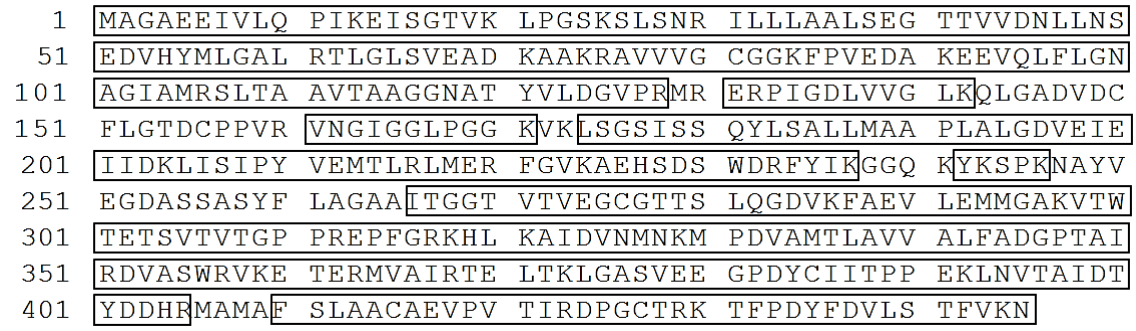

Figure 6: Tryptic Peptide Map of the GHB811-produced 2mEPSPS protein. The deduced amino acid sequence is 445 amino acids as shown. Boxed regions correspond to peptide sequence coverage achieved using MALDI-TOF MS, from the plant-produced protein sample.

The mature 2mEPSPS protein was also expressed in *Escherichia coli*. The bacterially-derived protein was shown by electrophoresis to have the expected mass of 47.4 kDa. Mass spectrometry further revealed the mass was 47288 Da. Peptide mapping and N-terminal sequencing confirmed the sequence matched 2mEPSPS (Figure 6) with no N-terminal methionine. The bacterially-expressed protein was shown by western blotting to be immunoreactive to 2mEPSPS antibodies and was functional in the PEP to EPSP enzymatic assay. These data demonstrated that the bacterially-derived protein was appropriate for use as a standard in the 2mEPSPS ELISA described in Section 4.1.1.

4.1.3 Safety of the introduced 2mEPSPS

FSANZ has assessed the safety of 2mEPSPS in five previous applications: A362 – corn line GA21 (FSANZ 2000), A614 – cotton line GHB614 (FSANZ 2009), A1051 – soybean line FG72 (FSANZ 2011), A1073 – soybean line DAS-44406-6 and A1112 – corn line MZHG0JG (FSANZ 2016). Results in the literature also support the safety of 2mEPSPS (Herouet-Guicheney et al. 2009; Fard et al. 2013; Hammond et al. 2013; Shafer et al. 2016)).

A summary of these previous characterisations is provided in Table 6. For information, a reference is provided to the application in which the most recent detailed study or information was considered by FSANZ and is available on the FSANZ website. For the bioinformatic studies, which analyse sequence similarity to known protein toxins and allergens, where the Applicant provided searches (see provided studies listed below) using an updated (and hence larger) database, the results did not alter conclusions reached previously.

Table 6: Summary of consideration of 2mEPSPS in previous FSANZ safety assessments

Consideration	Sub-section	2mEPSPS
Potential toxicity	Amino acid sequence similarity to protein toxins	This application – using search updated in March 2016
	In vitro digestibility	A1112 (FSANZ 2016)
	Stability to heat	A1112 (FSANZ 2016)
	Acute oral toxicity	A1051 (FSANZ 2011)
Potential allergenicity	Source of the protein	A362 (FSANZ 2000)
	Amino acid sequence similarity to allergens	This application – using search updated in March 2016

Taken together, the evidence indicates that 2mEPSPS is not likely to be toxic or allergenic to humans.

4.1.4 Conclusion

A range of characterisation studies were performed on plant-derived 2mEPSPS confirming the identity and equivalence to the corresponding protein produced in a bacterial expression system. It was noted that two polypeptides were being produced in GHB811, a predominant 444 amino acid protein and a minor 446 amino acid protein with an extra N-terminal cysteine and methionine residue. Except for the minor differences in molecular mass and N-terminal sequence, these polypeptides showed the expected immunoreactivity, lack of glycosylation, and enzyme activity.

It was found that the expression of 2mEPSPS in GHB811 was highest in plastid-containing tissues such as the leaves. 2mEPSPS was not detected in the crude and refined cottonseed oil fractions and was expressed at low levels in the hull and untoasted meal, which are mainly used for animal feed. Toasted meal, which can be used in human food, also did not contain detectable 2mEPSPS.

There are no concerns regarding the potential toxicity or allergenicity of the expressed proteins. Previous safety assessments of 2mEPSPS have indicated this protein would be rapidly degraded in the digestive system following ingestion and would be inactivated by heating. Additionally, updated bioinformatic studies considered in this assessment confirm the lack of any significant amino acid sequence similarity to known protein toxins or allergens.

4.2 Description of the HPPD*Pf* W336 protein

4-hydroxyphenylpyruvate dioxygenase (HPPD) is a ubiquitously expressed enzyme involved in the catabolism of tyrosine to homogentisate. The fate of homogentisate differs between photosynthetic and non-photosynthetic organisms (Shaner, 2003; Moran, 2005). In photosynthetic organisms, homogentisate is converted to plastoquinones and alpha-tocopherol (vitamin E). Plastoquinones have a few roles in plants, from acting as an electron carrier between photosystem II and the cytochrome b6f complex in photosynthesis, to being a co-factor for the enzyme phytoene desaturase, involved in carotenoid formation. In animals, homogentisate is converted to fumarate and acetoacetate, which can be directed into the citric acid cycle for cellular energy generation.

Isoxaflutole is an inhibitor of HPPD resulting in a decrease in plastoquinones (Pallett et al, 1998; Little and Pallett, 2003). By reducing plastiquinone levels, there is a direct impact on photosynthesis and the activity of the enzyme phytoene desaturase. Blocking the function of phytoene desaturase limits the formation of carotenoids, reducing their photo-protective effect on chlorophylls (Anderson and Robertson, 1966; Mayfield et al, 1986). This in turn leads to photo-bleaching of the foliage and a reduction in the plants ability to undergo further photosynthesis, resulting in plant death.

The HPPD protein sourced from *Pseudomonas fluorescens* was the first HPPD protein characterised. Studies comparing mutant forms of the protein identified a glycine to tryptophan change at position 336 in the mature protein produced an enzyme with a slightly lower Km compared to the unmodified protein yet was insensitive to the inhibitory effect of isoxaflutole (Boudec et al, 2001; Sailland et al, 2001; Matringe et al, 2005).

Analysis performed by FSANZ revealed that the HPPD protein from *P. fluorescens* has low sequence homology to plant-derived forms however the similarity in sequence between common food crops is also highly variable. Related plants share high homology, such as a comparison of the predicted potato and tomato HPPD proteins that show > 95% homology. Comparison of the HPPD sequence from potato with cotton, soybean and corn showed sequence variation from 55-75%. Humans have therefore been exposed to many different forms of the HPPD protein via the food supply. FSANZ has also previously assessed the HPPD*Pf* W336 protein in soybean line FG72 (A1051; FSANZ 2011).

The *hppd W336* gene prepared by the Applicant encodes a protein of 482 amino acids, consisting of a 124 amino acid transit peptide and a mature protein of 358 amino acids. After cleavage of the plastid transit peptide, the final mature protein will have an expected molecular mass of 40.4 kDa.

4.2.1 Expression of the HPPD*Pf* W336 protein in cotton tissues

Protein expression in plant tissues was determined by ELISA. The ELISA system used is a commercially available product (Envirologix) and is specific to the HPPD*Pf* W336 protein. Equivalent plant tissue extracts from the parental Coker 312 variety were used as a background control in the ELISA. The standard curve was generated using microbially expressed HPPD*Pf* W336, of which the characterisation is described in section 4.2.3.

In order to determine the sites of accumulation of the HPPD*Pf* W336 protein, samples were collected from GHB811 (generation T6) grown in the presence or absence of glyphosate and isoxaflutole, from three field-trial sites in the USA. For each tissue sample analysed, four samples were processed from each field-trial site. Specific tissues were collected at different growth stages.

The results from the protein analysis (Table 7) showed the highest expression of HPPD*Pf* W336 was in the foliage, reaching a peak in expression around the inflorescence emergence (51-55) stage. The lowest expression was seen in fuzzy seed and root tissue and was not detected in pollen. Expression was also detected in the developing flower tissue and the eventual seed and the levels decreased over this growth period, likely associated with the loss of plastid-containing tissue. Treatment of the plants with both glyphosate and isoxaflutole-containing herbicides did not significantly change the expression of HPPD*Pf* W336.

The presence of HPPD*Pf* W336 was also examined in the seed and processed fractions of the seed using ELISA (Table 8). The highest levels of HPPD*Pf* W336 were in the fuzzy and delinted seeds. The linters contained about 20-31% the level found in the fuzzy seeds. Cotton seed meal and hull contained about 12-37% the amount of the protein found in the delinted seeds and toasted meal had no detectable HPPD*Pf* W336. The sequential loss of detectable HPPD*Pf* W336 is expected considering that during the preparation of the meal fractions, the samples are exposed to very high temperatures, which are likely to denature and partially degrade proteins.

Table 7: Expression of HPPD*Pf* W336 in plant tissue samples

Tissue	Growth Stage	Treatment[1]	HPPD*Pf* W336 (µg/g DW)[2]		HPPD*Pf* W336 (µg/g FW)[3]	
			Mean	Range	Mean	Range
Leaf	14-16	-	668.06	136-1338	116.29	23-244
		+	808.10	377-1403	142.10	59-232
Leaf	51-55	-	1043.64	717-1674	198.51	125-328
		+	956.75	723-1232	188.78	125-245
Leaf	60-69	-	862.75	516-1226	184.73	120-251
		+	781.28	563-1013	166.51	116-221
Root	14-16	-	22.12	12-43	3.40	2-7
		+	25.42	11-46	3.59	2-6
Pollen[4]	60-69	-	n/a	n/a	BD[5]	-
		+	n/a	n/a	BD	-
Square	60-67	-	304.97	269-338	65.30	52-83
		+	284.52	236-365	60.88	45-96
Bolls	60-67	-	181.03	117-242	30.59	21-46
		+	125.62	70-194	20.29	12-27
Fuzzy Seed	83-97	-	29.61	11-62	26.45	9-56
		+	27.01	11-44	23.82	10-39

1. Treatment refers to plants grown in the absence (-) or presence (+) of herbicides. 2. DW -dry weight. 3. FW - fresh weight. 4. Protein can only be measured in fresh not dried pollen. 5. BD – below detection.

Table 8: Expression of HPPD*Pf* W336 in edible plant fractions

Tissue	Treatment[1]	HPPD*Pf* W336	
		(µg/g DW)[2]	(µg/g FW)[3]
Fuzzy Seed	-	42.50	38.97
	+	28.58	26.15
Linters	-	8.38	7.87
	+	8.82	8.32
Delinted Seed	-	37.67	34.66
	+	31.10	28.69
Untoasted Meal	-	4.54	4.07
	+	11.62	10.38
Toasted Meal	-	BD[4]	BD
	+	BD	BD
Hull	-	16.62	15.04
	+	14.46	13.07
Crude Oil	-	n/a	BD
	+	n/a	BD
Refined Oil	-	n/a	BD
	+	n/a	BD

1. Treatment refers to plants grown in the absence (-) or presence (+) of herbicides. 2. DW - dry weight. 3. FW - fresh weight. 4. BD – below detection.

No HPPD*Pf* W336 was detected in the purified oil fractions. The mean level of HPPD*Pf* W336 in the major food component from GHB811, cottonseed, is approximately 28 µg/g dw (0.0028%). As the mean percent dry weight of total protein in GHB811 cottonseed is approximately 20% dw (see Table 8), the amount of HPPD*Pf* W336 protein was calculated to be 0.014% of total protein. Given this low level of HPPD*Pf* W336 in cottonseed, the protein is highly unlikely to be detectable in refined oil.

4.2.2 Characterisation of the expressed HPPD*Pf* W336 protein in cotton and equivalence of bacterially-produced forms

HPPD*Pf* W336 was affinity purified from leaf tissue collected from greenhouse grown plants. The apparent molecular weight of the protein was shown by electrophoresis to be just higher than 37 kDa. Similar to that observed with the 2mEPSPS protein, mass spectrometry identified two polypeptides with masses of 40179.9 Da (minor peak) and 40446.8 Da (major peak). N-terminal sequencing confirmed the presence of two polypeptides and showed that they also differed in the region around the cleavage site between the transit peptide and mature protein. The major peak protein had the N-terminal sequence XMADLYENPM whereas the minor peak protein started at ADLYENPM, without the methionine at position 1. This difference in the N-terminal sequence between the polypeptides was also confirmed by peptide mapping. Peptide mapping also showed that the protein being expressed in cotton was HPPD*Pf* W336, with 98.6% sequence coverage achieved (Figure 7).

```
  1  MADLYENPMG LMGFEFIEFA SPTPGTLEPI FEIMGFTKVA THRSKNVHLY
 51  RQGEINLILN NEPNSIASYF AAEHGPSVCG MAFRVKDSQK AYNRALELGA
101  QPIHIDTGPM ELNLPAIKGI GGAPLYLIDR FGEGSSIYDI DFVYLEGVER
151  NPVGAGLKVI DHLTHNVYRG RMVYWANFYE KLFNFREARY FDIKGEYTGL
201  TSKAMSAPDG MIRIPLNEES SKGAGQIEEF LMQFNGEGIQ HVAFLTDDLV
251  KTWDALKKIG MRFMTAPPDT YYEMLEGRLP DHGEPVDQLQ ARGILLDGSS
301  VEGDKRLLLQ IFSETLMGPV FFEFIQRKGD DGFGEWNFKA LFESIERDQV
351  RRGVLTAD
```

Figure 7: Tryptic Peptide Map of the GHB811-produced HPPDPf W336 protein. The deduced amino acid sequence is 358 amino acids as shown. Boxed regions correspond to peptide sequence coverage achieved using MALDI-TOF MS, from the plant-produced protein sample.

Western blotting and ELISA demonstrated that the cotton-produced protein was immunoreactive to HPPD*Pf* W336 antibodies. The protein was shown to be functional due to the demonstrated resistance to isoxaflutole during selection of transformants (section 3.1). Finally, the protein was shown by a modified Periodic Acid-Schiff staining procedure to be non-glycosylated. This data indicates that no unintentional glycosylation has occurred in the modified plant expressed protein.

The mature HPPD*Pf* W336 protein was also expressed in *E. coli*. The bacterially-derived protein was shown by electrophoresis to have a similar mass as the protein extracted from GHB811. Mass spectrometry further revealed the mass was 40181 Da. Peptide mapping and N-terminal sequencing confirmed the sequence matched HPPD*Pf* W336 with no N-terminal methionine. Western blotting showed the protein was immunoreactive to HPPD*Pf* W336 antibodies and a colorimetric assay measuring the conversion of 4-hydroxyphenylpyruvate to homogentisate confirmed the enzyme was functional. The expressed protein was also shown to be non-glycosylated. These data demonstrated that the bacterially-derived protein was appropriate for use as a standard in the HPPD*Pf* W336 ELISA described in Section 4.2.1.

4.2.3 Safety of the introduced HPPD*Pf* W336

FSANZ has previously assessed the safety of HPPD*Pf* W336 in a previous application: A1051 – soybean line FG72 (FSANZ 2011) and the conclusion was that HPPD*Pf* W336 was not likely to be toxic or allergenic to humans. The safety characterisation of HPPD*Pf* W336 presented in A1051 included *in vitro* digestibility, stability to heat, acute toxicity and bioinformatic studies. For this Application, the Applicant provided bioinformatic search results for homology to known toxins and allergens using an updated (and hence larger) database. The results did not alter conclusions reached previously.

4.2.4 Conclusion

A range of characterisation studies were performed on plant-derived HPPD*Pf* W336 confirming the identity and equivalency to the corresponding protein produced in a bacterial expression system. Similarly to 2mEPSPS, two polypeptides of HPPD*Pf* W336 are produced in GHB811, a predominant 357 amino acid protein and a minor 359 amino acid protein with an extra N-terminal cysteine and methionine residue. Except for the minor differences in molecular mass and N-terminal sequence, these polypeptides showed the expected immunoreactivity, lack of glycosylation, and enzyme activity.

Expression of HPPDPf W336 in GHB811 was highest in plastid-containing tissues such as the leaves. HPPD*Pf* W336 was not detected in the crude and refined cottonseed oil fractions and was expressed at low levels in the hull and untoasted meal, which are mainly used for animal feed. Toasted meal also did not contain HPPD*Pf* W336.

No safety concerns were identified regarding the potential toxicity or allergenicity of HPPD*Pf* W336. Previous safety assessments showed this protein would be rapidly degraded in the digestive system following ingestion and would be inactivated by heating. Additionally, updated bioinformatic studies considered in this assessment confirm the lack of any significant amino acid sequence similarity to known protein toxins or allergens.

4.3 Further characterisation of the N-terminal sequence discrepancies in the novel proteins

In order to better understand why two versions of the 2mEPSPS and HPPD*Pf* W336 proteins are being produced in GHB811, FSANZ investigated the predicted transit peptide cleavage sites using the ChloroP server[2] at the Center for Biological Sequence Analysis. These analyses identified two likely cleavage sites around the junction between the transit peptide and mature protein. The most prominent site was directly N-terminal of cysteine (C) 124 in the transit peptide sequence (Figure 8). A secondary cleavage site was also identified directly N-terminal to alanine (A) 126, which is also present in both proteins. This information matches what is observed for the two novel proteins and the two polypeptides being produced.

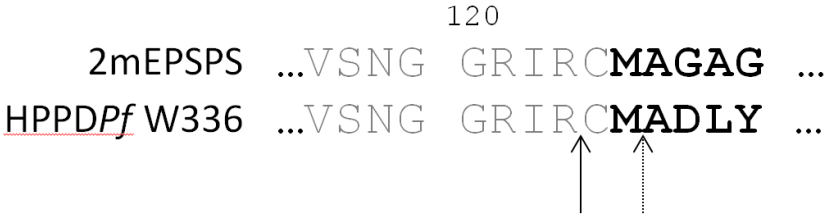

Figure 8: N-terminal processing of the novel proteins expressed in cotton. Both novel proteins contain a transit peptide (shown in grey) to facilitate translocation into plastid organelles. After translocation, the signal peptide is cleaved away, leaving the mature protein (shown in bold). Two cleavage sites around the junction between the transit peptide and mature protein were identified by ChloroP and are shown by the arrows.

Evidence from studies trying to identify sequence patterns in transit peptides have reported two scenarios to explain the potential of multiple cleavage sites. Firstly, it has been suggested that after removal of the transit peptide, some trimming of a few N-terminal amino acids can occur (Emanuelsson et al, 1999). Secondly, the cleavage site for the plastid processing enzyme (stromal processing peptidase; SPP) may not be optimal in the system

[2] http://www.cbs.dtu.dk/services/ChloroP/

used in GHB811 and two sites may be targeted (Rowland et al, 2015). The more favoured site could be the most likely site (full arrow) identified using ChloroP and the second site may be the secondary site (dotted arrow) shown in Figure 8. Although two versions of each protein are produced *in planta*, the differences are located in the first few amino acids and both proteins have been shown to be functionally equivalent and expected to be safe.

4.4 Novel herbicide metabolites in GM herbicide-tolerant plants

FSANZ has previously assessed the novel herbicide metabolites for glyphosate and isoxaflutole. In the glyphosate-tolerant cotton event GHB614, which was genetically modified to express the 2mEPSPS protein (A614; FSANZ 2009), both glyphosate and the metabolite aminomethylphosphonic acid were found in delinted seed and were present at levels that were below the established residue tolerances for glyphosate. It is expected that no new glyphosate metabolites would be generated in cotton event GHB811.

In the isoxaflutole-tolerant soybean event FG72 expressing the HPPD*Pf* W336 protein, (A1051; FSANZ 2011), the parental isoxaflutole and three major metabolites were identified. The same metabolites were also identified in non-GM corn varieties exposed to isoxaflutole (Pallett et al, 1998; 2001), indicating that the metabolites are consistent across plant species. It is expected that no new isoxaflutole metabolites would be generated in cotton event GHB811.

5 Compositional analysis

The main purpose of compositional analyses is to determine if, as a result of the genetic modification, an unexpected change has occurred to the food. These changes could take the form of alterations in the composition of the plant and its tissues and thus its nutritional adequacy. Compositional analyses can also be important for evaluating the intended effect where there has been a deliberate change to the composition of the food.

The classic approach to the compositional analyses of GM food is a targeted one. Rather than analysing every possible constituent, which would be impractical, the aim is to analyse only those constituents most relevant to the safety of the food or that may have an impact on the whole diet. Important analytes therefore include the key nutrients, toxicants and anti-nutrients for the food in question. The key nutrients and anti-nutrients are those components in a particular food that may have a substantial impact in the overall diet. They may be major constituents (fats, proteins, carbohydrates or enzyme inhibitors such as anti-nutrients) or minor constituents (minerals, vitamins). Key toxicants are those toxicologically significant compounds known to be inherently present in an organism, such as compounds whose toxic potency and level may be significant to health.

5.1 Key Components

Cottonseed oil is the primary cotton product used for human consumption. The key components to be analysed for the comparison of transgenic and conventional cotton are outlined in the OECD Consensus Document on Compositional Considerations for New Varieties of Cotton, and include proximates and fibre (cottonseed only), fatty acids, tocopherol (vitamin E), gossypol and the cyclopropenoid fatty acids - malvalic, sterculic and dihydrosterculic acids (OECD, 2009).

5.2 Study design

Fifteen successful field trials were conducted for GHB811 in the USA during the 2014 and 2015 growing seasons. Of the 15 sites, eight sites[3] were selected for compositional analysis, to match the typical geographical and field management styles of the commercial cotton growing regions. The agronomic practices and pest control measures used were location-specific and were typical for all aspects of cotton cultivation including soil preparation, fertiliser application, irrigation and pesticide-based control methods. The materials tested in the field trials included GHB811, the parental control (Coker 312) and a range of reference varieties. Three reference varieties were grown at each site and were selected from Acala Maxxa, DP399, FM989, FM958, FM966, ST457 and ST468.

The field trials were established in a randomised complete block design, with four replicates of each plot. The minimum plot size was 27.9 m^2. Plots were separated by a combination of fallow alleyways and planted buffers of conventional non-GM cotton. Replicate plots of GHB811 were cultivated with and without dual herbicide treatment (one application of each herbicide). Isoxaflutole was applied just before or shortly after the emergence stage at 100.3-115.2 g active ingredient (ai) per hectare (ha) and glyphosate was applied at the 6-9 leaf-stage at 1067-1222 g ai/ha. The spray volumes were 95.6 to 166 L/ha.

Cottonseed samples from all plots were harvested at maturity and delinted before despatch to an analytical laboratory under full identity labelling. The analyses were performed at EPL Bio Analytical Services. The compositional analyses were based on internationally recognised procedures including official methods specified by the Association of Official Analytical Chemists (AOAC), the USDA and published articles or technical notes from industrial-based sources.

A total of 69 different analytes were measured. Statistical analyses were performed using SAS 9.3 (SAS Institute, Cary, NC). For each analyte, 'descriptive statistics' (mean and standard error) were generated. A linear mixed model analysis of variance was then applied for combined data, which included multiple test years and locations, covering the eight replicated field trial sites. The mixed model analysis was also applied to the data from each site separately. In assessing the significance of any difference between the mean analyte value for GHB811 and the parental control, a P-value of 0.05 was used. Where statistically significant differences were observed in the combined data from all sites (presented in Tables 9-13), analysis of the data from each site was used to determine if the differences were common to the majority of sites.

In order to complete the statistical analysis for any component in this study, it was deemed that more than 2/3rd of the values must be greater than the assay LOQ. If analytes had more than 2/3rd of observations below the LOQ for that assay, they were excluded from the overall summary analysis. Values for all components were expressed on a dry weight basis with the exception of moisture, expressed as percent fresh weight, and fatty acids, expressed as percent of total fatty acids.

Compositional data from the non-GM reference varieties grown concurrently in the same trial as GHB811 and the parental control, were combined across all sites and used to calculate a 99% tolerance interval for each component to define the natural variability in cottonseed varieties that have a history of safe consumption. Any statistically significant differences between GHB811 and the control Coker 130 were also compared to this tolerance interval to assess whether the differences were likely to be biologically meaningful.

[3] The location of the eight field trial sites: Fresno, CA; Tift, GA; Rapides, LA; Washington, MS; Barnwell, SC; Tom Green, TX; Perquimans, NC; Hale, TX.

Unpublished studies

1. GHB811 Cotton – Field Production in the USA during 2014 and 2015 (2016) Report M-558563-01, Bayer CropScience.
2. GHB811 Cotton - Composition Assessment of GHB811 Cotton Grown in the USA during 2014 and 2015 (2016) Report M-566678-01, Bayer CropScience.
3. GHB811 Cotton – Production, processing and analysis of resultant fractions, 2014/2015 (2017) Report M-574125-01, Bayer CropScience.

5.3 Analyses of key components in fuzzy seed

5.3.1 Proximates and fibre

Analysis of the proximate and fibre levels in both untreated and treated GHB811 compared to the parental control (Table 9), showed statistically significant differences only in crude fat. There was also a statistically significant difference in the neutral detergent fibre content between the treated GHB811 and parental control. While there was a statistically significant change, the magnitude of these differences was actually small (~3-4%) and the mean levels fall within the tolerance interval, indicating that the differences are not biologically significant.

Table 9: Comparison of Proximates and Fibre (% DW)

Parameter	Non-GM counterpart	GHB811 untreated	GHB811 treated	Non-GM reference varieties	Tolerance Interval
	Mean ± SD	Mean ± SD	Mean ± SD	Min-max	lower-upper
Moisture[1]	10.8 ± 1.9	11.1 ± 2.8	11.5 ± 2.7	8.26 - 23.7	2.51 – 20.2
Ash	3.80 ± 0.3	3.80 ± 0.3	3.79 ± 0.4	3.07 – 5.16	2.66 – 4.99
Carbohydrate	52.3 ± 2.8	53.3 ± 2.6	53.4 ± 2.6	46.3 – 61.0	46.5 – 63.0
Crude protein	20.5 ± 2.8	20.4 ± 2.5	20.5 ± 2.2	13.7 – 24.4	13.6 – 25.6
Crude fat	23.3 ± 2.6	22.5 ± 2.4[2]	22.3 ± 2.6	15.8 – 28.7	14.9 – 28.8
Acid Detergent Fibre	43.0 ± 5.6	42.9 ± 4.2	43.3 ± 3.7	36.6 – 54.1	33.1 – 54.2
Neutral Detergent Fibre	45.6 ± 3.2	46.7 ± 3.1	46.9 ± 3.0	41.2 – 58.6	39.2 – 57.7
Total Dietary Fibre	43.7 ± 8.8	47.4 ± 5.3	47.6 ± 6.6	25.2 – 77.0	30.3 – 64.9

1. Moisture content is expressed as % fresh weight (% FW) whereas the other values are expressed as % dry weight (% DW). 2. Cells highlighted in blue show statistically significant data.

5.3.2 Amino acids

Statistically significant decreases were observed for cysteine and methionine, in the herbicide-treated GHB811 (Table 10). The magnitude of the change ranged from 6-7% and the mean levels fall within the tolerance interval, indicating that the differences are not biologically significant. Furthermore, amino acid content is not a key component for cottonseed oil and the data would only be applicable for cotton seed meal, used in animal feed.

5.3.3 Minerals and alpha tocopherol (vitamin E)

A statistically significant decrease was observed for vitamin E (alpha-tocopherol) in both untreated and treated GHB811 (Table 11). The magnitude of the change ranged from 9-10% however the mean levels fall within the tolerance interval, indicating that this difference was not biologically significant. This decrease was not unexpected considering the production of alpha-tocopherol is dependent on HPPD activity. The modified HPPD W336 enzyme that has been used to gain isoxaflutole-tolerance, has previously been shown to have lower enzyme

activity compared to the unmodified enzyme (Boudec et al, 2001; Sailland et al, 2001; Matringe et al, 2005).

Table 10: Comparison of Amino Acids (% DW)

Parameter	Non-GM counterpart	GHB811 untreated	GHB811 treated	Non-GM reference varieties	Tolerance Interval
	Mean ± SD	Mean ± SD	Mean ± SD	Min-max	lower-upper
Alanine	0.91 ± 0.09	0.90 ± 0.09	0.93 ± 0.09	0.63 - 1.07	0.61 - 1.14
Arginine	2.43 ± 0.39	2.40 ± 0.35	2.44 ± 0.35	1.45 - 3.18	1.33 - 3.29
Aspartic Acid	2.10 ± 0.30	2.12 ± 0.29	2.12 ± 0.31	1.32 - 2.58	1.22 - 2.78
Cysteine	0.44 ± 0.08	0.42 ± 0.08	0.41 ± 0.07[1]	0.27 - 0.57	0.22 - 0.61
Glutamic Acid	4.53 ± 0.63	4.50 ± 0.58	4.55 ± 0.64	2.97 - 5.70	2.73 - 6.06
Glycine	1.01 ± 0.10	0.99 ± 0.10	1.01 ± 0.11	0.67 - 1.26	0.65 - 1.27
Histidine	0.68 ± 0.08	0.67 ± 0.07	0.68 ± 0.07	0.46 - 0.97	0.42 - 0.89
Isoleucine	0.74 ± 0.08	0.74 ± 0.08	0.74 ± 0.08	0.52 - 0.87	0.49 - 0.94
Leucine	1.38 ± 0.15	1.37 ± 0.14	1.39 ± 0.13	0.94 - 1.64	0.91 - 1.74
Lysine	1.05 ± 0.10	1.05 ± 0.12	1.07 ± 0.12	0.75 - 1.30	0.69 - 1.35
Methionine	0.32 ± 0.05	0.31 ± 0.06	0.30 ± 0.05	0.19 - 0.38	0.19 - 0.41
Phenylalanine	1.29 ± 0.17	1.26 ± 0.16	1.29 ± 0.15	0.80 - 1.57	0.75 - 1.68
Proline	0.88 ± 0.10	0.87 ± 0.09	0.88 ± 0.09	0.60 - 1.08	0.57 - 1.12
Serine	1.05 ± 0.11	1.03 ± 0.11	1.04 ± 0.14	0.71 - 1.66	0.63 - 1.39
Threonine	0.78 ± 0.07	0.77 ± 0.07	0.78 ± 0.07	0.56 - 0.89	0.55 - 0.95
Tryptophan	0.24 ± 0.03	0.25 ± 0.04	0.25 ± 0.03	0.18 - 0.29	0.16 - 0.31
Tyrosine	0.50 ± 0.07	0.47 ± 0.08	0.49 ± 0.07	0.18 – 0.61	0.25 – 0.69
Valine	1.04 ± 0.12	1.03 ± 0.12	1.04 ± 0.11	0.72 – 1.24	0.67 – 1.32

1. Cells highlighted in blue show statistically significant data.

Table 11: Comparison of Minerals and Vitamin E (mg/kg DW)

Parameter	Non-GM counterpart	GHB811 untreated	GHB811 treated	Non-GM reference varieties	Tolerance Interval
	Mean ± SD	Mean ± SD	Mean ± SD	Min-max	lower-upper
Calcium	983 ± 288	1022 ± 296	1039 ± 270	702 - 1960	456 - 1930
Copper	6.19 ± 2.35	6.21 ± 2.55	6.15 ± 2.27	1.99 - 11.5	0 - 13.3
Iron	37.2 ± 6.7	36.1 ± 6.1	37.1 ± 6.8	22.1 - 60.1	13.2 - 60.0
Magnesium	4013 ± 489	4002 ± 486	3917 ± 355	2848 - 5328	2380 - 5438
Manganese	12.9 ± 2.5	12.5 ± 2.0	12.4 ± 2.2	9.94 - 25.2	5.35 - 23.4
Phosphorus	6613 ± 1298	6749 ± 960	6449 ± 1277	3888 - 9333	3137 - 10251
Potassium	12287 ± 1158	12508 ± 1385	12193 ± 943	10263 - 15727	9269 - 15027
Sodium	30.3 ± 15.5	37.3 ± 29.8	32.5 ± 24.9	10.0 - 148	0 - 74.9
Zinc	33.7 ± 5.4	33.7 ± 5.8	33.2 ± 5.8	21.5 - 49.6	17.2 - 50.3
Vitamin E	121.6 ± 19.1	109.8 ± 18.1[1]	109.1 ± 15.0	49.8 - 141	37.0 - 157

1. Cells highlighted in blue show statistically significant data.

5.3.4 Fatty Acids

The following fatty acids were excluded from the statistical summary as they had more than 2/3rd of observations below the LOQ: 8:0 caprylic acid, 10:0 capric acid, 12:0 lauric acid, 14:1 myristoleic acid, 15:0 pentadecanoic acid, 15:1 pentadecenoic acid, 18:4 stearidonic acid, 20:2 eicosadienoic acid, 20:3 eicosatrienoic acid, 20:4 arachidonic acid, 20:5 eicosapentaenoic acid, C22:1 erucic acid, C22:5 N3 docosapentaenoic acid, C22:5 N6 docosapentaenoic acid and C22:6 docosahexaenoic acid.

Statistically significant differences were observed for palmitoleic (↑ 6-8% change) and stearic acid (↓ 4-5% change) in both the untreated and treated GHB811 lines (Table 12) and a statistically significant decrease was observed in arachidic acid (↓ 4% change) in treated GHB811. For palmitoleic acid, the significant increase in GHB811 was noted in seven out of the eight sites. The magnitude of these changes for the three fatty acids are small and the mean values fall within the tolerance interval, indicating that this differences are not biologically significant.

Table 12: Comparison of Fatty Acids (% Total Fatty Acids)

Parameter	Non-GM counterpart	GHB811 untreated	GHB811 treated	Non-GM reference varieties	Tolerance Interval
	Mean ± SD	Mean ± SD	Mean ± SD	Min-max	lower-upper
C14:0 Myristic Acid	0.65 ± 0.15	0.67 ± 0.15	0.67 ± 0.14	0.47 - 1.17	0.22 - 1.11
C16:0 Palmitic Acid	23.2 ± 1.8	23.2 ± 1.8	23.4 ± 1.6	19.7 - 27.6	17.7 - 27.7
C16:1 Palmitoleic Acid	0.49 ± 0.05	0.52 ± 0.05[1]	0.53 ± 0.04	0.39 - 0.64	0.33 - 0.62
C17:0 Heptadecanoic Acid	0.08 ± 0.01	0.08 ± 0.01	0.08 ± 0.01	0.07 - 0.12	0.05 - 0.12
C17:1 Heptadecenoic Acid	0.04 ± 0.01	0.05 ± 0.01	0.05 ± 0.01	0.02 - 0.08	0.01 - 0.09
C18:0 Stearic Acid	2.37 ± 0.16	2.28 ± 0.17	2.26 ± 0.15	2.00 - 2.97	1.75 - 3.07
C18:1 Oleic Acid	14.2 ± 1.0	14.2 ± 1.0	14.2 ± 1.0	13.4 - 20.8	9.2 - 21.3
C18:2 Linoleic Acid	57.3 ± 2.9	57.5 ± 3.0	57.3 ± 2.8	46.3 - 60.7	45.2 - 67.8
C18:3 Linolenic Acid	0.23 ± 0.12	0.23 ± 0.11	0.21 ± 0.11	0.03 - 0.55	0 - 0.56
C20:0 Arachidic Acid	0.25 ± 0.03	0.25 ± 0.04	0.24 ± 0.03	0.20 - 0.38	0.12 - 0.39
C20:1 Eicosenoic Acid	0.07 ± 0.02	0.07 ± 0.02	0.07 ± 0.01	0.04 - 0.10	0.03 - 0.11
C22:0 Behenic Acid	0.15 ± 0.02	0.15 ± 0.03	0.15 ± 0.02	0.11 - 0.21	0.07 - 0.22
C24:0 Lignoceric Acid	0.12 ± 0.05	0.10 ± 0.06	0.11 ± 0.07	0.03 - 0.29	0 - 0.29

1. Cells highlighted in blue show statistically significant data.

5.3.5 Anti-nutrients

As shown in Table 13, there was a statistically significant decrease in gossypol found in the GHB811 line compared to the parental Coker 312 line however the change was minor and not biologically significant. Although there was a statistically significant decrease in dihydrosterculic acid in the herbicide-treated GHB811 line compared to the parental line, the change was also minor and considered to be not biologically significant.

Table 13: Comparison of Anti-nutrients

Parameter	Non-GM counterpart	GHB811 untreated	GHB811 treated	Non-GM reference varieties	Tolerance Interval
	Mean ± SD	Mean ± SD	Mean ± SD	Min-max	lower-upper
Gossypol (% DW)					
Free gossypol	0.59 ± 0.15	0.53 ± 0.12[1]	0.53 ± 0.12	0.27 - 0.94	0.11 – 1.15
Total gossypol	0.80 ± 0.17	0.70 ± 0.16	0.74 ± 0.15	0.35 – 1.34	0.15 – 1.53
Cyclopropenoic fatty acids (% total fatty acids)					
Malvalic acid	0.38 ± 0.18	0.38 ± 0.15	0.36 ± 0.17	0.01 – 0.98	0.00 – 1.06
Sterculic acid	0.15 ± 0.07	0.14 ± 0.06	0.13 ± 0.06	0.01 – 0.37	0.00 – 0.42
Dihydrosterculic acid	0.20 ± 0.05	0.19 ± 0.06	0.17 ± 0.05	0.13 – 0.41	0.10 – 0.44

1. Cells highlighted in blue show statistically significant data.

5.4 Analyses of key components in linters and oil

A field trial was performed in Argentina over the 2014-2015 growing season using a randomised complete block design, with four replicates of each plot. The material planted included GHB811 and the parental control Coker 312. Replicate plots of GHB811 were cultivated with and without dual herbicide treatment (one application of each herbicide). Isoxaflutole was applied at the pre-emergence stage at 98.4 g ai/ha and glyphosate was applied at the 7 leaf-stage at 1109.9 g ai/ha.

Cottonseed samples from all plots were harvested at maturity and shipped to GLP Technologies in the USA for processing. Fuzzy seeds were mechanically delinted to prepare delinted seed and linter fractions. Delinted seed was then further processed to prepare a crude oil fraction. Crude oil was further refined, bleached and deoderised (RBD) to prepare a commercial-grade cottonseed oil product (RBD oil).

The compositional analyses was then performed on the linter and cottonseed oil fractions, based on internationally recognised procedures including official methods specified by the Association of Official Analytical Chemists (AOAC), the USDA and published articles or technical notes from industrial-based sources.

Hull, untoasted meal and toasted meal fractions were also prepared and compositional analyses data was provided. As these products are predominantly used in animal feed in Australia and New Zealand, no data for these products is provided in this assessment.

5.4.1 Proximates and fibre in linters

Linters comprise approximately 98% cellulose and would therefore be expected to be rich in carbohydrate. The analysis of GHB811 presented in Table 14 confirmed this and indicated the presence of only residual protein (~3%) and fat (<1%). Subsequent purification steps during normal processing would remove the residual protein and fat resulting in a pure fraction of cellulose.

Table 14: Comparison of Proximates and Fibre in linters (% DW)

Parameter	Non-GM counterpart	GHB811 untreated	GHB811 treated
	Mean ± SD	Mean ± SD	Mean ± SD
Moisture[1]	6.43 ± 0.04	6.08 ± 0.21	5.73 ± 0.16
Ash	1.69 ± 0.09	1.85 ± 0.18	2.23 ± 0.05
Carbohydrate	94.5 ± 0.2	94.3 ± 0.3	94.2 ± 0.1
Crude protein	2.95 ± 0.23	3.16 ± 0.14	3.00 ± 0.09
Crude fat	0.82 ± 0.09	0.83 ± 0.05	0.73 ± 0.05

1. Moisture content is expressed as % fresh weight (% FW) whereas the other values are expressed as % dry weight (% DW).

5.4.2 Fatty acid levels in processed fractions

The fatty acid levels were compared for fuzzy seed, delinted seed, crude oil and RBD oil fractions for the untreated and treated GHB811 and parental Coker 312 lines. Twelve fatty acids were consistently measured with linoleic (55-60%), palmitic (20-25%) and oleic (14-15%) being the prominent fatty acids (Table 15). The data showed there was no loss of fatty acids during processing and the fatty acid levels in the fuzzy seed was representative of the fatty acid levels in the final refined oil fraction.

Table 15: Comparison of Fatty Acids in processed fractions (% Total Fatty Acids)

Parameter	Non-GM counterpart	GHB811 untreated	GHB811 treated
	Mean ± SD	Mean ± SD	Mean ± SD
Fuzzy Seed			
C18:2 Linoleic Acid	56.4 ± 0.1	60.5 ± 0.2	61.0 ± 0.1
C16:0 Palmitic Acid	24.5 ± 0.0	21.1 ± 0.1	20.4 ± 0.1
C18:1 Oleic Acid	14.2 ± 0.1	13.9 ± 0.1	14.3 ± 0.1
Delinted Seed			
C18:2 Linoleic Acid	56.6 ± 0.1	59.9 ± 0.1	60.4 ± 0.2
C16:0 Palmitic Acid	24.4 ± 0.1	21.5 ± 0.1	20.7 ± 0.1
C18:1 Oleic Acid	14.0 ± 0.0	13.9 ± 0.1	14.3 ± 0.1
Crude Oil			
C18:2 Linoleic Acid	55.8 ± 0.1	59.7 ± 0.0	59.5 ± 0.2
C16:0 Palmitic Acid	24.4 ± 0.0	21.1 ± 0.1	20.7 ± 0.1
C18:1 Oleic Acid	14.9 ± 0.1	14.6 ± 0.1	15.1 ± 0.0
RBD Oil			
C18:2 Linoleic Acid	57.5 ± 0.3	60.0 ± 0.1	60.1 ± 0.1
C16:0 Palmitic Acid	20.3 ± 0.1	21.2 ± 0.0	20.7 ± 0.1
C18:1 Oleic Acid	13.8 ± 0.1	14.3 ± 0.1	14.9 ± 0.1

5.4.3 Alpha-tocopherol levels in refined oil

The presence of vitamin E in the commercial oil product increases the shelf life of the oil due to its antioxidant properties (OECD, 2009). The vitamin E levels were measured in the final refined oil fraction (RBD oil) and found to be 505 mg/kg in Coker 312, 365 mg/kg in untreated GHB811 and 431 mg/kg in treated GHB811. These levels would provide a degree of antioxidant protection to the oil.

5.4.4 Anti- nutrient levels in refined oil

Gossypol was not detected in the crude or refined oil fractions prepared from Coker 312 or GHB811. The levels of cycloprenoid fatty acids (Table 16) were in the range typical for refined cottonseed oil fractions (OECD, 2009). This data indicate that the level of anti-nutrients found in the refined cottonseed oil fraction of the modified cotton line GHB811 is within the range found in standard non-GM cotton varieties.

Table 16: Comparison of Anti-nutrients in refined oil

Parameter	Non-GM counterpart	GHB811 untreated	GHB811 treated	Typical Levels
	Mean ± SD	Mean ± SD	Mean ± SD	lower-upper
Cyclopropenoic fatty acids (% total fatty acids)				
Malvalic acid	0.263 ± 0.001	0.340 ± 0.001	0.351 ± 0.003	0.22 – 1.44
Sterculic acid	0.120 ± 0.002	0.146 ± 0.002	0.151 ± 0.001	0.08 – 0.58
Dihydrosterculic acid	0.153 ± 0.002	0.193 ± 0.001	0.194 ± 0.003	0.00 – 0.22

5.5 Conclusion

Of the 69 analytes measured in cotton fuzzy seed, mean values were provided for 54 analytes. A summary of the eleven analytes that showed a significant difference between cotton line GHB811 and the parental Coker 312 is provided in Table 17.

For each of the analytes presented in Table 17, the differences were small in magnitude, and were within the tolerance interval. Data were also within the ranges reported in the published literature (Berberich et al, 1996; Nida et al, 1996; Hamilton et al, 2004; OECD, 2009; Codex, 2017; ILSI, 2017). As the composition of cotton can vary significantly due to the cultivation site and agricultural practices, the differences reported here most likely reflect the normal biological variability that exists in cotton.

Overall, the compositional data are consistent with the conclusion that there are no biologically significant differences in the levels of key constituents in cotton line GHB811 when compared with conventional cotton cultivars already available in agricultural markets.

Table 17: Summary of statistically significant analyte means from GHB811 and parental control Coker 312.

Parameter	Non-GM counterpart	GHB811 untreated	GHB811 treated	Are values within the reference ranges?	Are values within the literature range?
	Mean ± SD	Mean ± SD	Mean ± SD	Yes / No	Yes / No
Crude fat	23.3 ± 2.6	22.5 ± 2.4[1]	22.3 ± 2.6	yes	yes
Neutral Detergent Fibre	45.6 ± 3.2	46.7 ± 3.1	46.9 ± 3.0[2]	yes	yes
Cystine	0.44 ± 0.08	0.42 ± 0.08	0.41 ± 0.07	yes	yes
Methionine	0.32 ± 0.05	0.31 ± 0.06	0.30 ± 0.05	yes	yes
Vitamin E	121.6 ± 19.1	109.8 ± 18.1	109.1 ± 15.0	yes	yes
Palmitoleic Acid	0.49 ± 0.05	0.52 ± 0.05	0.53 ± 0.04	yes	yes
Stearic Acid	2.37 ± 0.16	2.28 ± 0.17	2.26 ± 0.15	yes	yes
Arachidic Acid	0.25 ± 0.03	0.25 ± 0.04	0.24 ± 0.03	yes	yes
Free gossypol	0.59 ± 0.15	0.53 ± 0.12	0.53 ± 0.12	yes	yes
Total gossypol	0.80 ± 0.17	0.70 ± 0.16	0.74 ± 0.15	yes	yes
Dihydrosterculic acid	0.20 ± 0.05	0.19 ± 0.06	0.17 ± 0.05	yes	yes

1. Cells highlighted in red show data where GHB811 is significantly lower than the parental Coker 312 line.
2. Cells highlighted in green show data where GHB811 is significantly higher than the parental Coker 312 line.

6 Nutritional impact

In assessing the safety of a GM food, a key factor is the need to establish that the food is nutritionally adequate and will support typical growth and wellbeing. In most cases, this can be achieved through a detailed understanding of the genetic modification and its consequences, together with an extensive compositional analysis of the food, such as that presented in Section 5 of this report.

Where a GM food has been shown to be compositionally equivalent to conventional varieties, the evidence to date indicates that feeding studies using target livestock or other animal species will add little to the safety assessment (Bartholomaeus et al., 2013; OECD, 2003). If the compositional analysis indicates biologically significant changes, either intended or unintended, to the levels of certain nutrients in the GM food, additional nutritional studies should be undertaken to assess the potential impact of the changes on the whole diet.

Cotton line GHB811 is the result of a simple genetic modification to confer dual herbicide tolerance, with no intention to significantly alter nutritional parameters in the food. The compositional analyses have demonstrated that the genetic modification has not altered the nutritional adequacy of cotton line GHB811 as a source of food when compared with that of conventional cotton varieties. The introduction of foods derived from cotton line GHB811 into the food supply is therefore expected to have negligible nutritional impact.

7 References

ABCA (2012) GM cotton in Australia: a resource guide. Resource guide 3, Agricultural Biotechnology Council of Australia. http://www.abca.com.au/wp-content/uploads/2012/ 09/ABCA_Resource_Guide _3_v2.pdf (accessed August 2017)

Anderson IC, Robertson DS (1960) Role of carotenoids in protecting chlorophyll from photodestruction. Plant Physiol 35(4):531-534

Atanassova R, Chaubet N, Gigot C (1992) A 126 bp fragment of a plant histone gene promoter confers preferential expression in meristems of transgenic Arabidopsis. Plant J 2(3):291-300

Bartholomaeus A, Parrott W, Bondy G, Walker K (2013) The use of whole food animal studies in the safety assessment of genetically modified crops: Limitations and recommendations. Crit Rev Toxicol 43(S2):1-24

Berberich SA, Ream JE, Jackson TL, Wood R, Harvey P, Patzer S, Stipanovic R, Fuchs RL (1996) The composition of insect-protected cottonseed is equivalent to that of conventional cottonseed. J Agric Food Chem 44:365-371

Boudec P, Rodgers M, Dumas F, Sailland A, Bourdon H (2001) Mutated hydroxyphenylpyruvate dioxygenase, DNA sequence and isolation of plants which contain such a gene and which are tolerant to herbicides (patent US6245968)

Chaboute M, Chaubet N, Philipps G, Ehling M, Gigot C (1987) Genomic organization and nucleotide sequences of two histone H3 and two histone H4 genes of Arabidopsis thaliana. Plant Mol.Biol. 8:179-191

Chaubet N, Clement B, Gigot C (1992) Genes encoding a histone H3.3-like variant in Arabidopsis contain intervening sequences. J Mol Biol 225(2):569-574

Chaubet N, Gigot C, Freyssinet G, Leroux B (1996) Chimeric gene comprising the arabidopsis histone H4 promoter for the transformation of plants (patent US5491288)

Codex (2009) Foods derived from modern biotechnology, 2nd edition. Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme, Food and Agriculture Organization of the United Nations, Rome.

Codex (2017) Standard for named vegetable oils. Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme, Food and Agriculture Organization of the United Nations, Rome www.fao.org/input/download/standards/336/CXS_210e_2015.pdf

Delaney B, Astwood JD, Cunny H, Eichen Conn R, Herouet-Guicheney C, MacIntosh S, Meyer LS, Privalle LS, Gao Y, Mattsson J, Levine M, ILSI (2008) Evaluation of protein safety in the context of agricultural biotechnology. Food Chem Toxicol 46:S71-S97

Derose R, Chaubet N, Gigot C (2011) Chimeric gene comprising intron from histone H3.3 gene (patent US7982092)

Emanuelsson O, Nielsen H, Heijne G von (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Science 8:978-984

Fard NA, Minuchehr Z, Mousavi A (2013) Allergenicity study of genetically modified herbicide resistant crops (bioinformatics assessment). Bulletin of Environment, Pharmacology and Life Sciences 2(3):24–32

FAO (2017) FAOSTAT. Production: Crops. Food and Agriculture Organization of the United Nations, Rome. www.fao.org/faostat/en/#data/QC/visualize (accessed May 2017)

FSANZ (2000) Application A362 - Food derived from herbicide-tolerant corn line GA21. Report prepared by Food Standards Australia New Zealand. http://www.foodstandards.gov.au/code/applications/pages/applicationa362foodd952.aspx

FSANZ (2009) Application A614 - Food derived from herbicide-tolerant cotton line GHB614. Report prepared by Food Standards Australia New Zealand. http://www.foodstandards.gov.au/code/applications/pages/applicationa614foodd3782.aspx

FSANZ (2011) Application A1051 - Food derived from herbicide-tolerant soybean line FG72. Report

prepared by Food Standards Australia New Zealand.
http://www.foodstandards.gov.au/code/applications/pages/applicationa1051food4902.aspx

FSANZ (2013) Application A1073 - Food derived from herbicide-tolerant soybean line DAS-44406-6. Report prepared by Food Standards Australia New Zealand.
http://www.foodstandards.gov.au/code/applications/Pages/a1073.aspx

FSANZ (2016) Application A1112 - Food derived from herbicide-tolerant corn line MZHG0JG. Report prepared by Food Standards Australia New Zealand.
http://www.foodstandards.gov.au/code/applications/Pages/A1112-GM-CornLineMZHG0JG.aspx

Hamilton KA, Pyla PD, Breeze M, Olson T, Li M, Robinson E, Gallagher SP, Sorbet R, Chen Y (2004) Bollgard II cotton: compositional analysis and feeding studies of cottonseed from insect-protected cotton (*Gossypium hirsutum* L.) producing the Cry1Ac and Cry2Ab2 proteins. J Agric Food Chem 52:6969-6976

Hammond B, Kough J, Herouet-Guicheney C, Jez JM (2013) Toxicological evaluation of proteins introduced into food crops. Critical Reviews in Toxicology 43(S2):25–42

Herouet-Guicheney C, Rouquié D, Freyssinet M, Currier T, Martone A, Zhou J, Bates EEM, Ferullo J, Hendrickx K, Rouan D (2009) Safety evaluation of the double mutant 5-enol pyruvylshikimate-3-phosphate synthase (2mEPSPS) from maize that confers tolerance to glyphosate herbicide in transgenic plants. Regul Toxicol Pharmacol 54(2):143-153

Herrmann KM (1995) The shikimate pathway: early steps in the biosynthesis of aromatic compounds. Plant Cell 7:907-919

Hoess RH, Abremski K (1985) Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system. J Mol Biol 181(3):351-362

ILSI (2017) International Life Sciences Institute Crop Composition Database Version 6
https://www.cropcomposition.org/query/index.html (accessed August 2017)

Lebrun M, Leroux B, Sailland A (1996) Chimeric gene for the transformation of plants (patent US5510471)

Lebrun M, Sailland A, Freyssinet G, DeGryse E (1997) Mutated 5-enolpyruvylshikimate-3-phosphate synthase, gene coding for said protein and transformed plants containing said gene (patent WO 97/04103)

Little JP. Pallett KE (2003) Herbicides, Inhibiting 4-Hydroxyphenylpyruvate Dioxygenase (HPPD). Encyclopedia of Agrochemicals

Matringe M, Sailland A, Pelisser B, Rolland A, Zink O (2005) p-Hydroxyphenylpyruvate dioxygenase inhibitor-resistant plants. Pest Manag Sci 61:269-276

Mayfield SP, Nelson T, Taylor WC (1986) The fate of chloroplast proteins during photooxidation in carotenoid-deficient maize leaves. Plant Physiol 82(3):760-764

Munger P, Bleiholder H, Hack H, Hess M, Stauss R, van den Boom T, Weber E (1998) Phenological Growth Stages of the Cotton plant (*Gossypium hirsutum* L.) Codification and Description according to the BBCH Scale – with figures. J Agro Crop Sci 180, 143 - 149.

Moran GR (2005) 4-hydroxyphenylpyruvate dioxygenase. Arch Biochem Biophys 433:117-128

NCPA (2002) Cottonseed and its products. National Cottonseed Products Association.
http://www.cottonseed.org/publications/cottonseedanditsproducts.asp (accessed May 2017)

NCPA (unknown) Functional properties of cottonseed oil. National Cottonseed Products Association
http://www.cottonseedoiltour.com/pdf/PropertiesofCSO_whitepaper.pdf (accessed May 2017)

Nida DL, Patzer S, Harvey P, Stipanovic R, Wood R, Fuchs RL (1996) Glyphosate-tolerant cotton: the composition of the cottonseed is equivalent to that of conventional cottonseed. J Agric Food Chem 44:1967-1974

OECD (2003) Considerations for the Safety Assessment of Animal Feedstuffs Derived from Genetically Modified Plants. Series on the Safety of Novel Foods and Feeds, No. 9. Organisation for Economic Cooperation and Development, Paris.

OECD (2004) Consensus document on the biology of *Helianthus annuus* L. (sunflower). Series on

Harmonisation of Regulatory Oversight in Biotechnology, no. 31. Organisation for Economic Co-operation and Development, Paris.

OECD (2008) Consensus document on the biology of cotton (Gossypium spp.). Series on Harmonisation of Regulatory Oversight in Biotechnology, no. 45. Organisation for Economic Co-operation and Development, Paris.

OECD (2009) Consensus document on compositional considerations for new varieties of cotton (*Gossypium hirsutum and Gossypium barbadense*): key food and feed nutrients and anti-nutrients. Organisation for Economic Co-operation and Development, Paris.

OGTR (2008) The Biology of *Gossypium hirsutum* L. and *Gossypium barbadense* L. (cotton). Office of the Gene Technology Regulator, Canberra.

Pallett KE, Little JP, Sheekey M, Veerasekaran P (1998) The mode of action of isoxaflutole. I. Physiological effects, metabolism and selectivity. Pesticide Biochemistry and Physiology 62(2):113-124

Pallett KE, Cramp SM, Little JP, Veerasekaran P, Crudace AJ, Slater AE (2001) Isoxaflutole: the background to its discovery and the basis of its herbicidal properties. Pest Management Science 57:133-142

Rose AB (2004) The effect of intron location on intron-mediated enhancement of gene expression in Arabidopsis. Plant J 40(5):744-751

Rowland E, Kim J, Bhuiyan NH, van Wijk KJ (2015) The Arabidopsis chloroplast stromal N-terminome: complexities of amino-terminal protein maturation and stability. Plant Physiol 169(3):1881-1896

Sailland A, Rolland A, Matringe M, Pallett K (2001) DNA sequence of a gene of hydroxy-phenyl pyruvate dioxygenase and production of plants containing a gene of hydroxy-phenyl pyruvate dioxygenase and which are tolerant to certain herbicides (patent US6268549)

Scales BS, Dickson RP, LiPuma JJ, Huffnagle GB (2014) Microbiology, genomics, and clinical significance of the *Pseudomonas fluorescens* species complex, an unappreciated colonizer of humans. Clin Microbiol Rev 27(4):927–48

Schafer BW, Embrey SK, Herman RA (2016) Rapid simulated gastric fluid digestion of in-seed/grain proteins expressed in genetically engineered crops. Regul Toxicol Pharmacol 81:106-112

Shaner DL (2004) Herbicide safety relative to common targets in plants and mammals. Pest Manag Sci 60(1):17-24

Strickberger MW. 1976. Probability and statistical testing in Genetics, 2nd ed. New York: Macmillan Publishing Company. pp. 140-163.

Tzin V, Galili G (2010) New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. Mol Plant 3(6):956-972

USDA (2001) Oilseeds: World markets and trade. United States Department of Agriculture: Foreign Agricultural Service. Cotton, Oilseeds, Tobacco and Seeds Division, Washington DC

USDA (2016) Australia: Oilseeds and Products Annual. United States Department of Agriculture: Foreign Agricultural Service. Global Agricultural Information Network, Washington DC

USDA (2017) Oilseeds: World markets and trade. United States Department of Agriculture: Foreign Agricultural Service. Office of Global Analysis, Washington DC

van Larebeke N, Engler G, Holsters S, van den Elsacker S, Zaenen I, Schilperoort RA, Schell J (1974) Large plasmid in Agrobacterium tumefaciens essential for crown gall-inducing ability. Nature 252:169-170

Verdaguer B, Kochko A de, Beachy RN, Fauquet C (1996) Isolation and expression in transgenic tobacco and rice plants, of the cassava vein mosiac virus (CVMV) promoter. Plant Mol Bio 31:1129-1139

Zambryski P (1988) Basic processes underlying Agrobacterium-mediated DNA transfer to plant cells. Annu Rev Genet 22:1-30