Application to Food Standards Australia New Zealand for the Inclusion of Cotton MON 88702 in Standard 1.5.2 - Food Derived from Gene Technology

Submitted by:

Monsanto Australia Limited
Level 12 / 600 St Kilda Road,
Melbourne Victoria 3004

16th August 2017

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

UNPUBLISHED REPORTS BEING SUBMITTED

[redacted]2017. Amended Report for MSL0028045: The Effect of Heat Treatment on the Functional Activity of Cry51Aa2.834_16 Protein. **MSL0029144**. Monsanto Company.

[redacted] 2016. Assessment of Cry51Aa2.834_16 Protein Levels in Cotton Tissues Collected from MON 88702 Produced in United States Field Trials During 2015. **MSL0027766**. Monsanto Company.

[redacted]2017. Amended Report for MSL0027259: Molecular Characterization of Insect Protected Cotton MON 88702. **MSL0028391**. Monsanto Company.

[redacted] 2016. Characterization of the Cry51Aa2.834_16 Protein Purified from the Cotton Seed of MON 88702 and Comparison of the Physicochemical and Functional Properties of the Plant Produced and Bt Produced Cry51Aa2.834_16 Proteins. **MSL0027791**. Monsanto Company.

[redacted] 2017a. Amended Report for MSL0028693: Bioinformatics Evaluation of the DNA Sequences Flanking the 5′ and 3′ Junctions of the MON 88702 Insert: Assessment of Putative Polypeptides Utilizing the AD_2017, TOX_2017, and PRT_2017 Databases. **MSL0028798**. Monsanto Company.

[redacted] 2017b. Bioinformatics Evaluation of the Transfer DNA Insert in MON 88702 Utilizing the AD_2017, TOX_2017, and PRT_2017 Databases. **MSL0028694**. Monsanto Company.

[redacted]2017. Amended Report for MSL0029033: Compositional Analyses of Cottonseed from MON 88702 Grown in the United States During the 2015 Season. **MSL0029119**. Monsanto Company.

[redacted] 2017. Amended from MSL0027647: An Acute Oral Gavage Toxicity Study of Cry51Aa2.834_16 Protein in CD-1 Mice. **MSL0028578**. Monsanto Company.

[redacted]2016. Amended Report for MSL0026822: Segregation Analysis of the T-DNA Insert in Insect-protected Cotton MON 88702 Across Three Generations. **MSL0027485**. Monsanto Company.

[redacted] 2017. Bioinformatics Evaluation of the mCry51Aa2 Protein in MON 88702 Utilizing the AD_2017, TOX_2017, and PRT_2017 Databases. **MSL0028423**. Monsanto Company.

[redacted] 2016. Assessment of the in vitro Digestibility of Cry51Aa2_834.16 Protein by Pepsin and Pancreatin. **MSL0027977**. Monsanto Company.

[redacted] 2016. Demonstration of the Presence of Cry51Aa2.834_16 Protein in Lygus Cotton Leaf Samples across Multiple Generations of MON 88702. **MSL0027352**. Monsanto Company.

CHECKLIST

General Requirements (3.1)	Reference
3.1.1 Form of application	
☑ Executive Summary	*Executive Summary*
☑ Relevant sections of Part 3 identified	
☑ Pages sequentially numbered	
☑ Electronic + 2 hard copies	
☑ Electronic and hard copies identical	
☑ Hard copies capable of being laid flat	
☑ All references provided	
3.1.2 Applicant details	*Page 1*
3.1.3 Purpose of the application	*Page 1*
3.1.4 Justification for the application	*Page 2*
3.1.5 Information to support the application	*12 Monsanto unpublished reports as provided*
3.1.6 Assessment procedure	*Page 3*
☑ General	
☐ Major	
☐ Minor	
3.1.7 Confidential Commercial Information	
☑ Confidential material separated in both electronic and hard copy	
☑ Justification provided	
3.1.8 Exclusive Capturable Commercial Benefit	*Page 4*
3.1.9 International and Other National Standards	*Page 4*
3.1.10 Statutory Declaration	*Page 108*
3.1.11 Checklist/s provided with Application	
☑ Checklist	
☑ Any other relevant checklists for Sections 3.2 – 3.7	*Checklist 3.5.1*

Foods Produced using Gene Technology (3.5.1)	
☑ A.1 Nature and identity of the GM food	*Page 5*
☑ A.2 History of use of the host and donor organisms	*Page 5*
☑ A.3 The nature of the genetic modification	*Page 9*
☑ B.1 Characterisation and safety assessment of new substances	*Page 47*
☑ B.2 New proteins	*Page 62*
☑ B.3 Other (non-protein) new substances	*Page 62*
☑ B.4 Novel herbicide metabolites in GM herbicide-tolerant plants	*Page 66*
☑ B.5 Compositional analyses of the food produced using gene technology	*Page 85*
☑ C information related to the nutritional impact of the food produced using gene technology	*Page 107*
☑ D Other information	*Page 107*

ABBREVIATIONS AND DEFINITIONS[1]

ADF	Acid Detergent Fiber
BSA	Bovine Serum Albumin
Bt	*Bacillus thuringiensis*
BW	Body Weight
CaMV	Cauliflower Mosaic Virus
Cry	Crystal
DNA	Deoxyribonucleic Acid
Dw	Dry Weight
ELISA	Enzyme-linked Immunosorbent Assay
EPA	Environmental Protection Agency
ETS	Excellence Through Stewardship
FA	Fatty Acid
FDA	Food and Drug Administration (U.S.)
FMV	Figwort Mosaic Virus
FSANZ	Food Standards Australia New Zealand
HOSU	History of Safe Use
LOD	Limit of Detection
LOQ	Limit of Quantitation
MALDI-TOF	Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry
NDF	Neutral Detergent Fiber
NGS	Next Generation Sequencing
NOAEL	No Observable Adverse Effect Level
OECD	Organization for Economic Co-operation and Development
ORF	Open Reading Frame
OSL	Over Season Leaf
OSR	Over Season Root
PCR	Polymerase Chain Reaction
RBD	Refined, Bleached and Deodorized
RP	Recurrent Parent
SD	Standard Deviation
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
T-DNA	Transfer Deoxyribonucleic Acid
TDF	Total Dietary Fiber
USDA	United States Department of Agriculture
WTP	Western Tarnished Plant bug

[1] Alred, G.J., C.T. Brusaw, and W.E. Oliu. 2003. Handbook of Technical Writing, 7th edn., pp. 2-7. Bedford/St. Martin's, Boston, MA.

PART 1 GENERAL INFORMATION

1.1 Applicant Details

(a) Applicant's name/s	[redacted].
(b) Company/organisation name	Monsanto Australia Limited
(c) Address (street and postal)	Level 12 / 600 St Kilda Road, Melbourne, Victoria, 3004 PO Box 6051, St Kilda Road Central, Victoria, 8008
(d) Telephone number	[redacted]
(e) Email address	[redacted]
(f) Nature of applicant's business	Technology Provider to the Agricultural and Food Industries
(g) Details of other individuals, companies or organisations associated with the application	

1.2 Purpose of the Application

This application is submitted to Food Standards Australia New Zealand by Monsanto Australia Limited on behalf of Monsanto Company.

The purpose of this submission is to make an application to vary **Standard 1.5.2** – *Food Produced Using Gene Technology* of the *Australia New Zealand Food Standards Code* to seek the addition of cotton line MON 88702 and products containing cotton line MON 88702 (hereafter referred to as MON 88702) to the Table to Clause 2 (see below).

Food derived from gene technology	**Special requirements**
Food derived from cotton line MON 88702	None

1.3 Justification for the Application

1.3(a) The need for the proposed change

Monsanto Company has developed insect-protected cotton MON 88702, which produces a modified Cry51Aa2 insecticidal crystal (Cry) protein derived from *Bacillus thuringiensis* (*Bt*) that protects against feeding damage caused by targeted hemipteran and thysanopteran insect pests. The modified Cry51Aa2 protein has been assigned the unique name Cry51Aa2.834_16 (herein referred to as mCry51Aa2).

1.3(b) The advantages of the proposed change over the status quo, taking into account any disadvantages

In recent years, hemipteran and thysanopteran insects have become some of the most economically detrimental pests in cotton production.

Beyond the use of chemical insecticides (including both seed treatments and foliar sprays) to control these hemipteran and thysanopteran insects, additional approaches include crop rotation, variety selection, and judicious application of fertilizer. As an additional tool for minimizing crop damage from these insect pests, Monsanto Company has developed insect-protected cotton MON 88702. MON 88702 produces the mCry51Aa2 insecticidal crystal (Cry) protein derived from *Bacillus thuringiensis* (*Bt*), which protects against feeding damage caused by targeted hemipteran and thysanopteran insect pests, including two species of tarnished plant bugs (*Lygus hesperus*, and *Lygus lineolaris*), cotton fleahopper (*Pseudatomoscelis seriatus*), and thrips (*Frankiniella* spp.).

MON 88702 offers cotton growers an additional choice for insect pest management, and may be combined through traditional breeding methods with other insect-protected and herbicide-tolerant biotechnology traits. These next-generation cotton products will provide greater crop management choices for growers, to help meet the needs of Australian and global food, feed and fiber markets.

1.4 Regulatory Impact Information

1.4(a) Costs and benefits

If the draft variation to permit the sale and use of food derived from MON 88702 is approved, possible affected parties may include consumers, industry sectors and government. The consumers who may be affected are those particularly concerned about the use of biotechnology. Industry sectors affected may be food importers and exporters, distributors, processors and manufacturers. Lastly, government enforcement agencies may be affected.

A cost/benefit analysis quantified in monetary terms is difficult to determine. In fact, most of the impacts that need to be considered cannot be assigned a dollar value. Criteria would need to be deliberately limited to those involving broad areas such as trade, consumer information and compliance. If the draft variation is approved:

Consumers:

- There would be benefits in the broader availability of cotton products.
- There is unlikely to be any significant increase in the prices of foods if manufacturers are able to use comingled cotton products.
- Consumers wishing to do so will be able to avoid GM cotton products as a result of labeling requirements and marketing activities.

Government:

- Benefit that if cotton MON 88702 was detected in food products, approval would ensure compliance of those products with the Code. This would ensure no potential for trade disruption on regulatory grounds.
- Approval of cotton MON 88702 would ensure no potential conflict with WTO responsibilities.
- In the case of approved GM foods, monitoring is required to ensure compliance with the labeling requirements, and in the case of GM foods that have not been approved, monitoring is required to ensure they are not illegally entering the food supply. The costs of monitoring are thus expected to be comparable, whether a GM food is approved or not.

Industry:

- Sellers of processed foods containing cotton derivatives would benefit as foods derived from cotton MON 88702 would be compliant with the Code, allowing broader market access and increased choice in raw materials. Retailers may be able to offer a broader range of cotton products or imported foods manufactured using cotton derivatives.
- Possible cost to food industry as some food ingredients derived from cotton MON 88702 would be required to be labelled.

1.4(b) Impact on international trade

If the draft variation to permit the sale and use of food derived from MON 88702 was rejected it would result in the requirement for segregation of any cotton derived products containing MON 88702 from those containing approved cotton, which would be likely to increase the costs of imported cotton derived foods.

It is important to note that if the draft variation is approved, cotton MON 88702 will not have a mandatory introduction. The consumer will always have the right to choose not to use/consume this product.

1.5 Assessment Procedure

Monsanto Australia is submitting this application in anticipation that it will fall within the General Procedure category.

1.6 Exclusive Capturable Commercial Benefit

This application is likely to result in an amendment to the Code that provides exclusive benefits and therefore Monsanto intends to pay the full cost of processing the application.

1.7 International and Other National Standards

1.7(a) International standards

Monsanto makes all efforts to ensure that safety assessments are aligned, as closely as possible, with relevant international standards such as the Codex Alimentarius Commission's *Principles for the Risk Analysis of Foods Derived from Modern Biotechnology* and supporting *Guideline for the Conduct of Food Safety Assessment of Foods Derived from Recombinant-DNA Plants* (Codex Alimentarius, 2009).

In addition, the composition analysis is conducted in accordance with OECD guidelines and includes the measurement of OECD-defined cotton nutrients and anti-nutrients based on conventional commercial cotton varieties (OECD, 2009).

1.7(b) Other national standards or regulations

Monsanto has submitted an application for an experimental use permit and a petition for a termporary tolerance exemption to the United States Environmental Protection Agency (US EPA), for which approval was obtained on January 10, 2017. Furthermore, an application for a Section 3 Seed Increase registration for MON 88702 and a petition for a permanent tolerance exemption for the mCry51Aa2 protein were submitted to the US EPA. Monsanto has submitted a food and feed safety and nutritional assessment summary for MON 88702 to the United States Food and Drug Administration (FDA) and has also requested a Determination of Nonregulated Status for MON 88702, including all progenies derived from crosses between MON 88702 and other cotton, from the Animal and Plant Health Inspection Service (APHIS) of the U.S. Department of Agriculture (USDA). Further, Monsanto has also submitted a dossier for food and feed safety of MON 88702 to the Canadian Food Inspection Agency and Health Canada.

Consistent with our commitments to the Excellence Through Stewardship® (ETS) Program[2], regulatory submissions have been or will be made to countries that import significant cotton or food and feed products derived from U.S. cotton and have functional regulatory review processes in place.

[2] Excellence Through Stewardship is a registered trademark of Excellence Through Stewardship, Washington, DC. (http://www.excellencethroughstewardship.org)

PART 2 SPECIFIC DATA REQUIREMENTS FOR SAFETY ASSESSMENT

A. TECHNICAL INFORMATION ON THE FOOD PRODUCED USING GENE TECHNOLOGY

A1 Nature and Identity of the Genetically Modified

A1(a) A description of the new GM organism

Monsanto Company has developed insect-protected cotton MON 88702, which produces a mCry51Aa2 insecticidal crystal (Cry) protein derived from *Bacillus thuringiensis* (*Bt*) that protects against feeding damage caused by targeted hemipteran and thysanopteran insect pests.

MON 88702 will be combined, through traditional breeding methods, with other biotechnology-derived traits that have been deregulated by USDA, completed consultations with FDA and, where applicable, registered by U.S. EPA. These next generation combined-trait cotton products will increase grower pest and weed control choices, to help sustainably improve production efficiency.

A1(b) Name, line number and OECD unique identifier of each new lines or strains of GM organism from which the food is derived

In accordance with OECD's "Guidance for the Designation of a Unique Identifier for Transgenic Plants" MON 88702 has been assigned the unique identifier MON-887Ø2-4.

A1(c) The name the food will be marketed under (if known)

A commercial trade name for the product has not been determined at the time of this submission and will be available prior to commercial launch of the product.

A1(d) The types of products likely to include the food or food ingredient

Other than the introduction of the insect protection trait, MON 88702 is not materially different from conventional cotton and can be processed into four major products: oil, meal, hulls and linters. Only cottonseed oil and linters are utilised as food sources. For a further description of food uses and processing of oil and linters, refer to Section A2(b)(iii) and Section A2(b)(iv).

A2 History of Use of the Host and Donor Organisms

A2(a) Description of all donor organism(s)

A2(a)(i) Common and scientific names and taxonomic classification

Wild-type Cry51Aa2 was initially identified from *B. thuringiensis* (*Bt*), and observed to have activity against two species of cotton pests, *L. hesperus* and *L. lineolaris* (Baum *et al.*, 2012). *Bt* is used commercially in the U.S. to produce microbial-derived products with insecticidal activity, and applications of sporulated *Bt* have a long history of safe use for pest control in

agriculture, especially in organic farming (U.S. EPA, 1988; Cannon, 1993; WHO, 1999). The mCry51Aa2 expressed by MON 88702 is highly homologous (approximately 96% sequence similarity) to the amino acid sequence of wild-type Cry51Aa2 from *Bt*, with 8 amino acid substitutions (F46S, Y54H, S95A, F147S, Q149E, S167R, P219R, R273W) and a deletion of 3 amino acids (Δ196-198) (Gowda *et al.*, 2016).

Kingdom: Eubacteria

Phylum: Firmicutes

Class: Bacilli

Order: Bacillales

Family: Bacillaceae

Genus: *Bacillus*

A2(a)(ii) Information on pathogenicity, toxicity, or allergenicity of relevance to the food

Wild-type Cry51Aa2 was initially identified from *B. thuringiensis* (*Bt*), and observed to have activity against two species of cotton pests, *L. hesperus* and *L. lineolaris* (Baum *et al.*, 2012). *Bt* is used commercially in the U.S. to produce microbial-derived products with insecticidal activity, and applications of sporulated *Bt* have a long history of safe use for pest control in agriculture, especially in organic farming (U.S. EPA, 1988; Cannon, 1993; WHO, 1999).

There is no evidence of human or animal pathogenicity for any of the donor organisms of the coding and non-coding DNA sequences present in MON 88702.

A2(a)(iii) History of use of the organism in the food supply or human exposure to the organism through other than intended food use

The *B. thuringiensis* (*Bt*) donor organism from which the *mCry51Aa2* coding sequence was derived, has a long history of safe use, and is not commonly known for human or animal pathogenicity, or allergenicity.

A history of safe use for *Bt* microbial biopesticide applications has been established through the documented use of these products for over 50 years (Hammond, 2004; OECD, 2010b). There are at least 180 registered microbial *Bt* products in the U.S. and over 120 microbial products approved in the European Union (Hammond, 2004). Applications of sporulated *Bt* have a long history of safe use for pest control in agriculture, especially in organic farming (Cannon, 1993; U.S. EPA, 1988). They have been safely and directly applied to consumed agriculture commodities including berry crops, cabbage, grapes, tomatoes, celery, lettuce, and spinach (U.S. EPA, 1988). Furthermore, extensive toxicity testing of commercial *Bt* microbial biopesticides that contain numerous Cry proteins (e.g., Cry2A, Cry1Aa, Cry1Ab, Cry1Ac, Cry1C, and Cry1F) has resulted in no evidence for adverse effects to human or animal health (McClintock *et al.*, 1995; Koch *et al.*, 2015; Moar *et al.*, 2017; OECD, 2010a). Thus, *Bt* microbial formulations and the insecticidal proteins present in *Bt* microbial biopesticide formulations have been safely consumed by humans and animals for over 50 years.

A2(b) Description of the host organism

A2(b)(i) Phenotypic information

Family – Malvaceae

Tribe – Gossypieae

Genus – *Gossypium* L.

Primary Cultivated Species –*G. hirsutum* L.

Cotton is of the genus *Gossypium* of the tribe Gossypieae of the family Malvaceae (OECD, 2008). Approximately 50 *Gossypium* species are currently recognized, only four of which are commercially cultivated (OECD, 2008; Percival *et al.*, 1999). Phylogenetic classifications of the *Gossypium* genus have expanded in recent years. The major diploid *Gossypium* species lineages are: Australia (C, G, and K genomes); the American continents (D genome); and Africa/Arabia (A, B, E, and F genomes) (Percival *et al.*, 1999). The tetraploid species (2n=4x=52), including *G. hirsutum*, *G. barbadense*, and *G. tomentosum* (in Hawaii), are comprised of the A and D nuclear genomes (AADD) with A genome cytoplasm (Wendel *et al.*, 2009; OECD, 2008). These species originated in the Americas (OECD, 2008). The diploid species, AA, BB, *etc*. (2n=2x=26), are distributed among tropical and subtropical regions worldwide.

A2(b)(ii) How the organism is propagated for food use

Cotton is primarily a self-pollinated species (Niles and Feaster, 1984) and is propagated by seed. The structure of the cotton flower consists of five mostly fused sepals that form the calyx, enclosing the five petals of the sympetalous corolla (Oosterhuis and Jernstedt, 1999). A fused staminal column surrounds the style. The pistil is composed of the ovary, containing 3 to 5 carpels; the style; and the stigma (Oosterhuis and Jernstedt, 1999). Each carpel defines one locule and contains multiple ovules (Oosterhuis and Jernstedt, 1999). The cotton flower is receptive to pollen on the day of opening and stays open for one day (Oosterhuis and Jernstedt, 1999). Shortly after the petals open, the anthers dehisce (Oosterhuis and Jernstedt, 1999). Cotton pollen grains are heavy and have a sticky coating, a biological factor that limits dispersal by wind (McGregor, 1976). Cotton pollen grains are also relatively short-lived. Under highly humid laboratory conditions, Richards et al. (2005) found that most control cotton pollen was viable after 8 hours, with substantial reductions in viability occurring by 16 and 32 hours. Pollen sampled from *Helicoverpa armigera* moth probosces was already mostly non-viable by 8 hours (Richards *et al.*, 2005). Although pollen-mediated gene flow can be carried out by insect pollinators, it is impacted by the effective foraging range of the insect pollinators under local environmental conditions (Llewellyn *et al.*, 2007). The extent of spontaneous (unaided) or natural outcrossing thus depends greatly upon local insect populations, including introduced and native species (Llewellyn *et al.*, 2007; Van Deynze *et al.*, 2005). Pollen-mediated gene flow declines steeply with distance from the source; for example, in solid planted cotton, Van Deynze (2005) found that pollen-mediated gene flow was less than 1% beyond 9 meters from the source; Kairichi (2008) found 0% beyond 8 meters.

A2(b)(iii) What part of the organism is used for food

After ginning to remove fibres for textile manufacturing, cottonseed is processed into four major products: oil, meal, hulls, and linters. Processing of cottonseed typically yields (by weight): 16% oil, 45% meal, 26% hulls, and 9% linters, with 4% lost during processing (Cherry, 1983). Only cottonseed oil and linters are utilised as food sources, both are further discussed below.

A2(b)(iv) Whether special processing is required to render food derived from the organism safe to eat

Cottonseed is highly processed during the production of oil and meal. After hulling, the cottonseed is flaked by a rolling process to facilitate oil removal. Prior to oil extraction, the flakes are heated at temperatures of 88 ºC to greater than 130 ºC to break down the cell walls, reduce the viscosity of oil, inactivate proteins, and detoxify gossypol (Harris, 1981; NCPA, 1993). After heating, oil is typically removed from the meal by direct solvent extraction with hexane. Crude cottonseed oil is further processed with refining, bleaching and deodorisation steps to produce high purity vegetable oil. Temperatures up to 230 ºC are used in the deodorisation process (Harris, 1981; NCPA, 1993). Further processing (refining) for all the uses of cottonseed oil includes deodorisation and bleaching. Deodorisation greatly reduces the cyclopropenoid fatty acid content of the oil due to extreme pH and temperature conditions (NCPA, 1993). A winterisation step is added to produce cooking oil, whereas for solid shortening a hydrogenation step is added to transform the liquid oil into a solid fat. Previous studies have shown that the resulting oil contains no detectable protein (Reeves and Weihrauch, 1979). Cottonseed oil is traded as premium quality oil that is used for a variety of food uses, including frying oil, salad and cooking oil, mayonnaise, salad dressing, shortening, margarine, and packing oil. The material left after the extraction of the crude cottonseed oil is the cottonseed meal. The gossypol levels in the meal after extraction are reduced by approximately half. Linters are the short fibres on American upland cottonseed that remain after the long fibres have been removed at the ginning process for textile manufacturing. Linters consist of nearly pure (*i.e.*, >99%) cellulose (NCPA, 2002; Nida *et al.*, 1996) and after extensive processing at alkaline pH and temperatures >100 ºC (AOCS, 1991), the linters can be used as a high fibre dietary product. Food uses include fibre supplement, casings for processed meats, binder for solids in the pharmaceutical industry, and to improve viscosity in products such as toothpaste, ice cream, and salad dressings (NCPA, 2002). The highest grade linters can also be used in the manufacturing of absorbent cotton, medical pads, and gauze (NCPA, 2002), however as mentioned earlier these would consist of nearly pure cellulose, with negligible amounts of protein.

A2(b)(v) The significance to the diet in Australia and New Zealand of the host organism

Cottonseed has been a staple of the human diet for centuries, and its processed fractions are consumed in a multitude of food. Estimates of cotton consumption are available from the WHO Global Environmental Monitoring System - Food Contamination Monitoring and Assessment Programme (GEMS/Food) (www.who.int/foodsafety/chem/gems). The GEMS/Food programme has developed 13 Cluster Diets which are considered to be

representative of the major food consumption patterns exhibited by regional and cultural groups around the world. Australia is included in Cluster M, along with the United States and Canada and several other countries.

A3 The Nature of the Genetic Modification

A3(a) Method used to transform the host organism

MON 88702 was developed through *Agrobacterium*-mediated transformation of conventional cotton, based on the method described by (Chen *et al.*, 2014) which allows for the generation of transformed plants without the utilization of callus. Briefly, meristem tissues were excised from the embryos of germinated conventional seed. After co-culturing with *Agrobacterium* carrying the transformation construct, the meristems were placed on selection medium containing spectinomycin, carbenicillin disodium salt and cefotaxime sodium salt, to inhibit the growth of untransformed plant cells and excess *Agrobacterium*. The meristems were then placed in media conducive to shoot development followed by a transfer to a Jiffy Carefree® propagation plug for root development. Rooted plants with normal phenotypic characteristics were selected and transferred to soil for growth and further assessment.

The R_0 plants generated through this transformation process were self-pollinated to produce R_1 seed, and the unlinked insertions of T-DNA I and T-DNA II were segregated. R_1 plants that were positive for the *mCry51Aa2* expression cassette (T-DNA I) and did not contain the *aadA* expression cassette (T-DNA II) were identified by a quantitative polymerase chain reaction (PCR)-based analysis. Subsequently, R_1 plants homozygous for T-DNA I were selected for further development and their progenies were subjected to further molecular analysis, insect efficacy and phenotypic assessments. As is typical of a commercial event production and selection process, hundreds of different transformation events (regenerants) were generated in the laboratory using PV-GHIR508523. After careful selection and evaluation of these events in the laboratory, greenhouse and field, MON 88702 was selected as the lead event based on superior agronomic, phenotypic, and molecular characteristics. Studies on MON 88702 were initiated to further characterize the genetic insertion and the expressed products, and to establish the food and feed safety and unaltered environmental risk compared to conventional cotton. The major development steps of MON 88702 are depicted in Figure 1.

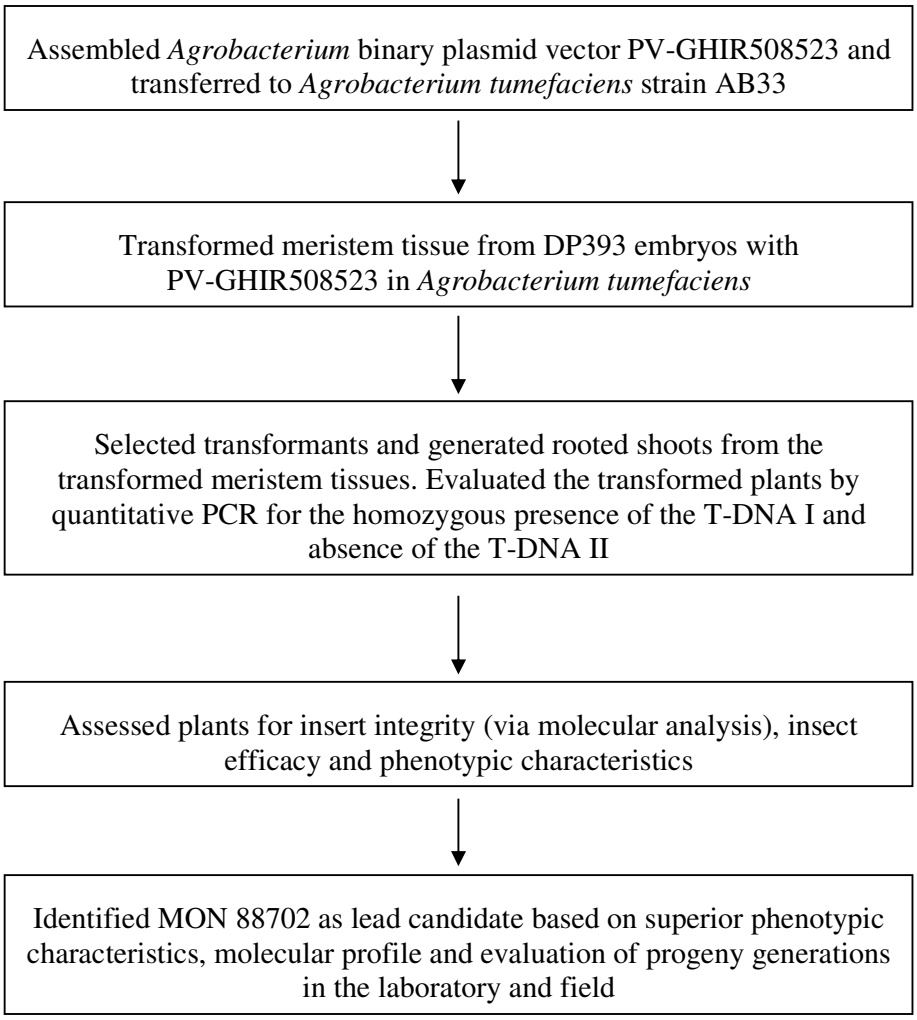

Figure 1. Schematic of the Development of MON 88702

A3(b) Intermediate hosts (*e.g.* bacteria)

A disarmed strain of *Agrobacterium tumefaciens* was the intermediate host used to transfer the plasmid PV-GHIR508523 into cotton cells. PV-GHIR508523 contains one T-DNA containing the *mCry51Aa2* expression cassette. Following transformation, self-pollination, breeding, and segregation methods were used to produce MON 88702.

A3(c) Gene contruct including size, source and function of all elements

A3(c)(i) The size, source and function of all the genetic components including marker genes, regulatory and other elements

Plasmid PV-GHIR508523 was used for the transformation of conventional cotton to produce MON 88702. A map of the plasmid is shown in Figure 3. The elements included in this plasmid vector are described in Table 1. PV-GHIR508523 is approximately 14.6 kb and contains two separate T-DNAs, each delineated by Left and Right Border regions. The first T-DNA, designated as T-DNA I, contains the *mCry51Aa2* expression cassette. The second T-DNA, designated as T-DNA II, contains the *aadA* expression cassette; *aadA* encodes an aminoglycoside-modifying enzyme that confers spectinomycin and streptomycin resistance (Fling *et al.*, 1985), acting as a selectable marker to allow selection of transformed tissue.

The backbone region of PV-GHIR508523, located outside both of the T-DNAs, contains two origins of replication for maintenance of the plasmid vector in bacteria (*ori-pRi*, *ori-pBR322*), a bacterial selectable marker gene (*nptII*), and a coding sequence for repressor of primer (ROP) protein for the maintenance of the plasmid vector copy number in *Escherichia coli*. A description of the genetic elements and their prefixes (e.g., B, E, P, TS, CS, T, and OR) in PV-GHIR508523 is provided in Table 1.

During transformation, both T-DNAs were inserted into the cotton genome (Section A3(a)). Subsequently, traditional breeding, segregation, selection, and screening were used to isolate those plants that contained the *mCry51Aa2* expression cassette (T-DNA I) but did not contain the *aadA* expression cassette (T-DNA II) or plasmid backbone.

The *mCry51Aa2* Coding Sequence and mCry51Aa2 Protein

The *mCry51Aa2* expression cassette in MON 88702 encodes the ~34 kDa mCry51Aa2 protein, which consists of a single polypeptide of 306 amino acids (Figure 14). The mCry51Aa2 protein expressed by MON 88702 is highly homologous (approximately 96% sequence similarity) to the amino acid sequence of wild-type Cry51Aa2 from *Bt*, with 8 amino acid substitutions (F46S, Y54H, S95A, F147S, Q149E, S167R, P219R, R273W) and a deletion of 3 amino acids (Δ196-198) (Gowda *et al.*, 2016).

```
1     MAILDLKSLV LNAINYWGPK NNNGIQGGDF GYPISEKQID TSIITSTHPR LIPHDLTIPQ
61    NLETIFTTTQ VLTNNTDLQQ SQTVSFAKKT TTTTATSTTN GWTEGGKISD TLEEKVSVSI
121   PFIGEGGGKN STTIEANFAH NSSTTTSQEA STDIEWNISQ PVLVPPRKQV VATLVIMGGN
181   FTIPMDLMTT IDSTEHYSGY PILTWISSPD NSYSGRFMSW YFANWPNLPS GFGPLNSDNT
241   VTYTGSVVSQ VSAGVYATVR FDQYDIHNLW TIEKTWYARH ATLHNGKKIS INNVTEMAPT
301   SPIKTN
```

Figure 2. Deduced Amino Acid Sequence of the mCry51Aa2 Protein
The amino acid sequence of the mCry51Aa2 protein sequence was deduced from the full-length coding nucleotide sequence present in MON 88702.

Regulatory Sequences

The *mCry51Aa2* coding sequence in T-DNA I is under the regulation of the *FMV* enhancer, *Hsp81-2* promoter, and the *35S CaMV* 3' untranslated region. The *FMV* sequence is a genetic element of the 35S RNA of figwort mosaic virus (FMV) (Richins *et al.*, 1987), which enhances transcription in most plant cells (Rogers, 2000). The *Hsp81-2* sequence is the promoter for the heat shock protein *81-2* gene of *A. thaliana* (Yabe *et al.*, 1994), which functions to direct transcription in plant cells. The *35S* 3' sequence is the 3' untranslated region of the 35S RNA of cauliflower mosaic virus (CaMV), which directs polyadenylation of mRNA in plant cells (Mogen *et al.*, 1990).

T-DNA II contains the *aadA* coding sequence under the regulation of the *FMV* enhancer (described above), the *EF-1α* promoter, the *CTP2* targeting sequence, and the *T-E9* untranslated region. The *EF-1α* promoter consists of the leader, promoter, and intron sequences from *A. thaliana* encoding elongation factor *EF-1α* (Axelos *et al.*, 1989), which functions to direct transcription in plant cells. The *CTP2* targeting sequence is from the *ShkG* gene encoding the EPSPS transit peptide region in *A. thaliana* (Klee *et al.*, 1987; Herrmann, 1995), which functions to direct transport of the protein to the chloroplast. The *E9* 3' sequence is the 3' untranslated region from *P. sativum* (pea) *rbcS* gene family encoding the small subunit of ribulose bisphosphate carboxylase protein (Coruzzi *et al.*, 1984), which functions to direct polyadenylation of the mRNA.

T-DNA Border Regions

PV-GHIR508523 contains Left and Right Border regions (Figure 3 and Table 1) that were derived from *A. tumefaciens* plasmids. The border regions each contain a nick site that is the site of DNA exchange during transformation (Barker *et al.*, 1983; Depicker *et al.*, 1982; Zambryski *et al.*, 1982). The border regions separate the T-DNA from the plasmid backbone region and are involved in the efficient transfer of T-DNA I into the cotton genome. Because PV-GHIR508523 is a two T-DNA vector, it contains two Left Border regions and two Right Border regions, where one border region set flanks T-DNA I and the other border region set flanks T-DNA II.

Genetic Elements Outside the T-DNA Border Regions

Genetic elements that exist outside of the T-DNA border regions are those that are essential for the maintenance or selection of PV-GHIR508523 in bacteria and are referred to as

plasmid backbone. The origin of replication, *ori-pRi*, is required for the maintenance of the plasmid in *Agrobacterium* and is derived from the broad host plasmid *pRi* (Ye *et al.*, 2011). The origin of replication, *ori-pBR322*, is required for the maintenance of the plasmid in *E. coli* and is derived from the plasmid vector pBR322 (Sutcliffe, 1979). Coding sequence *rop* encodes the repressor of primer (ROP) protein which is necessary for the maintenance of plasmid vector copy number in *E. coli* (Giza and Huang, 1989). The *rrn* promoter is the promoter for the ribosomal RNA operon from *A. tumefacien*s (Bautista-Zapanta *et al.*, 2002). The selectable marker *nptII* is the coding sequence for an enzyme from transposon Tn5 that confers neomycin and kanamycin resistance (Fraley *et al.*, 1983) in *E. coli* and *Agrobacterium* during molecular cloning. Because these elements are outside the border regions, they are not expected to be transferred into the cotton genome. The absence of the backbone and other unintended plasmid sequence in MON 88702 was confirmed by sequencing and bioinformatic analyses (described in Section A3(d)(ii)).

Table 1. Summary of Genetic Elements in PV-GHIR508523

Genetic Element	Location in Plasmid Vector	Function (Reference)
T–DNA-I		
B[1]-Right Border Region	1-285	DNA region from *Agrobacterium tumefaciens* containing the right border sequence used for transfer of the T–DNA (Depicker *et al.*, 1982; Zambryski *et al.*, 1982)
Intervening Sequence	286-338	Sequence used in DNA cloning
E[2]-*FMV*	339-745	Enhancer from the 35S RNA of figwort mosaic virus (FMV) (Richins *et al.*, 1987) that enhances transcription in most plant cells (Rogers, 2000)
Intervening Sequence	746-820	Sequence used in DNA cloning
P[3]-*Hsp81-2*	821-1828	Promoter and 5′ UTR leader sequence for the heat shock protein 81-2 (Hsp81-2) from *Arabidopsis thaliana* that directs transcription in plant cells (Yabe *et al.*, 1994)
Intervening Sequence	1829-1865	Sequence used in DNA cloning
CS[4]-*Cry51Aa2.834_16*	1866-2786	Coding sequence of the modified Cry51Aa2 protein of *Bacillus thuringiensis* that provides insect resistance (Baum *et al.*, 2012; Anderson *et al.*, 2015; Gowda *et al.*, 2016)
Intervening Sequence	2787-2818	Sequence used in DNA cloning
T[5]-*35S*	2819-3018	3' UTR sequence of the 35S RNA of cauliflower mosaic virus (CaMV) (Mogen *et al.*, 1990) that directs polyadenylation in plant cells
Intervening Sequence	3019-3156	Sequence used in DNA cloning
B-Left Border Region	3157-3598	DNA region from *Agrobacterium tumefaciens* containing the left border sequence used for transfer of the T–DNA (Barker *et al.*, 1983)
Vector Backbone		
Intervening Sequence	3599-3807	Sequence used in DNA cloning
CS-*nptII*	3808-4602	Coding sequence of the *neo* gene from transposon Tn*5* of *Escherichia coli* encoding neomycin phosphotransferase II (NPT II) (Beck *et al.*, 1982) that confers neomycin and kanamycin resistance (Fraley *et al.*, 1983)

Table 1. Summary of Genetic Elements in PV-GHIR508523 (continued)

Genetic Element	Location in Plasmid Vector	Function (Reference)
P-*rrn*	4603-4827	Promoter of the ribosomal RNA operon from *Agrobacterium tumefaciens* (Bautista-Zapanta *et al.*, 2002) that directs transcription in bacteria
Intervening Sequence	4828-4903	Sequence used in DNA cloning
OR[6]-*ori-pBR322*	4904-5492	Origin of replication from plasmid pBR322 for maintenance of plasmid in *E. coli* (Sutcliffe, 1979)
Intervening Sequence	5493-5919	Sequence used in DNA cloning
CS-*rop*	5920-6111	Coding sequence for repressor of primer protein from the ColE1 plasmid for maintenance of plasmid copy number in *E. coli* (Giza and Huang, 1989)
Intervening Sequence	6112-6299	Sequence used in DNA cloning
OR-*ori-pRi*	6300-10413	Origin of replication from plasmid pRi for maintenance of plasmid in *Agrobacterium* (Ye *et al.*, 2011)
Intervening Sequence	10414-10420	Sequence used in DNA cloning
T–DNA-II		
B-Left Border Region	10421-10739	DNA region from *Agrobacterium tumefaciens* containing the left border sequence used for transfer of the T–DNA (Barker *et al.*, 1983)
Intervening Sequence	10740-10803	Sequence used in DNA cloning
T-*E9*	10804-11446	3' UTR sequence from *Pisum sativum* (pea) *rbcS* gene family encoding the small subunit of ribulose bisphosphate carboxylase protein (Coruzzi *et al.*, 1984) that directs polyadenylation of the mRNA
Intervening Sequence	11447-11461	Sequence used in DNA cloning
aadA	11462-12253	Bacterial promoter, coding sequence, and 3' UTR for an aminoglycoside-modifying enzyme, 3"(9)-*O*-nucleotidyltransferase from the transposon Tn*7* (Fling *et al.*, 1985) that confers spectinomycin and streptomycin resistance

Table 1. Summary of Genetic Elements in PV-GHIR508523 (continued)

Genetic Element	Location in Plasmid Vector	Function (Reference)
TS[7]-*CTP2*	12254-12481	Targeting sequence of the *ShkG* gene from *Arabidopsis thaliana* encoding the EPSPS transit peptide region that directs transport of the protein to the chloroplast (Klee *et al.*, 1987; Herrmann, 1995)
Intervening Sequence	12482-12490	Sequence used in DNA cloning
P-*EF-1α*	12491-13638	Promoter, leader, and intron sequences of the *EF–1α* gene from *Arabidopsis thaliana* encoding elongation factor *EF–1α* that directs transcription in plant cells (Axelos *et al.*, 1989)
Intervening Sequence	13639-13661	Sequence used in DNA cloning
E-*FMV*	13662-14198	Enhancer from the 35S RNA of figwort mosaic virus (FMV) (Richins *et al.*, 1987) that enhances transcription in most plant cells (Rogers, 2000)
Intervening Sequence	14199-14248	Sequence used in DNA cloning
B-Right Border Region	14249-14605	DNA region from *Agrobacterium tumefaciens* containing the right border sequence used for transfer of the T–DNA (Depicker *et al.*, 1982; Zambryski *et al.*, 1982)
Vector Backbone		
Intervening Sequence	14606-14620	Sequence used in DNA cloning

[1] B, Border

[2] E, Enhancer

[3] P, Promoter

[4] CS, Coding Sequence

[5] T, Transcription Termination Sequence

[6] OR, Origin of Replication

[7] TS, Targeting Sequence

A3(c)(ii) A detailed map of the location and orientation of all the genetic components contained within the construct and vector, including the location of relevant restriction sites

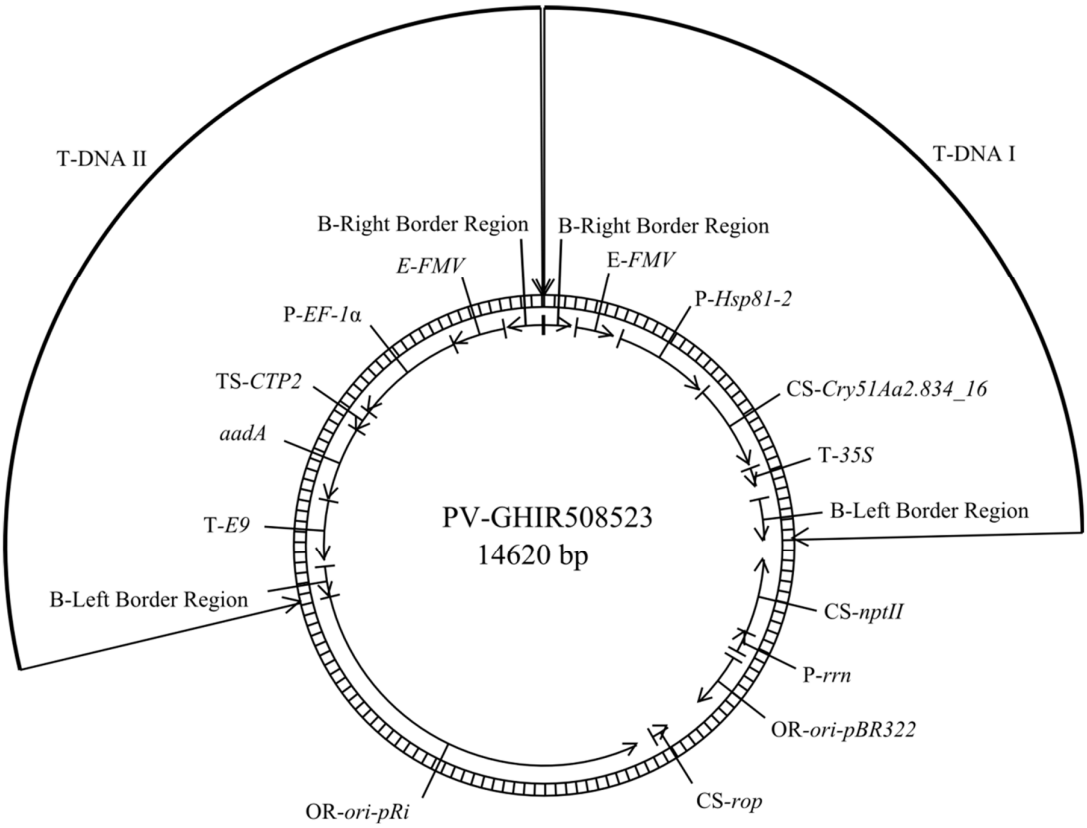

Figure 3. Circular Map of PV-GHIR508523

A circular map of PV-GHIR508523 used to develop MON 88702 is shown. PV-GHIR508523 contains two T–DNAs. Genetic elements are shown on the exterior of the map.

A3(d) Full characterisation of the genetic modification in the new organism, including:

A3(d)(i) Identification of all transferred genetic material and whether it has undergone any rearrangements

This section contains a comprehensive molecular characterization of the genetic modification present in MON 88702. It provides information on the DNA insertion into the plant genome of MON 88702, and additional information relative to the arrangement and stability of the introduced genetic material.

A schematic representation of the Next Generation Sequencing (NGS) methodology and the basis of the characterization using NGS and PCR sequencing is illustrated in Figure 4 below.

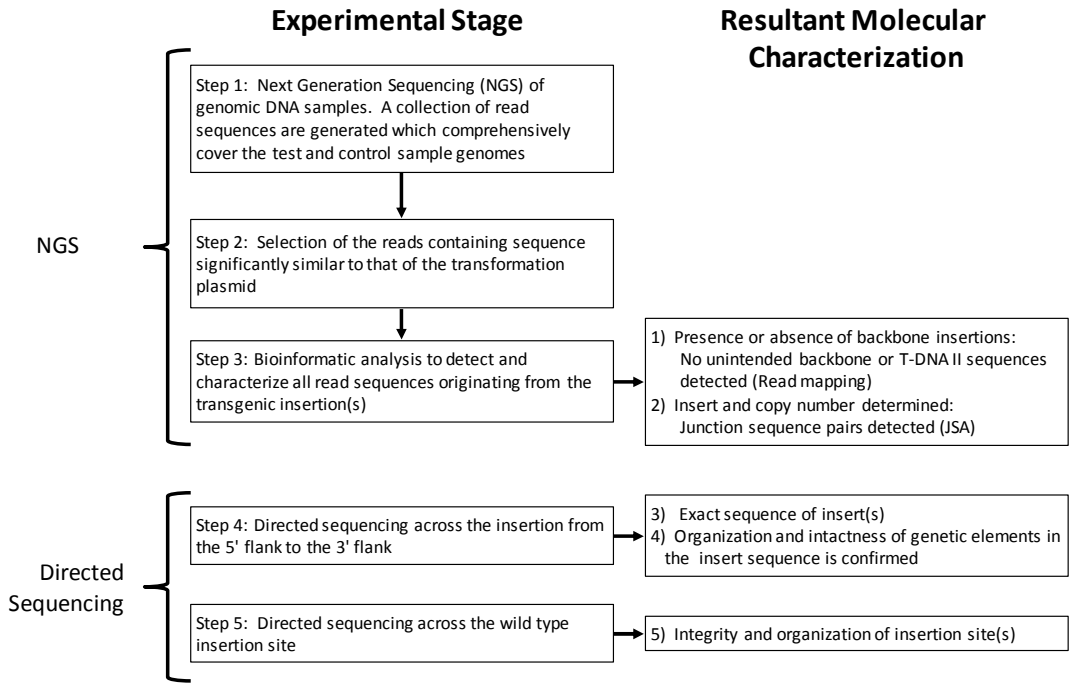

Figure 4. Molecular Characterization using Sequencing and Bioinformatics
Genomic DNA from MON 88702 and the conventional control was sequenced using technology that produces a set of short, randomly distributed sequence reads that comprehensively cover test and control genomes (Step 1). Utilizing these genomic sequence reads, bioinformatics searches are conducted to identify all sequence reads that are significantly similar to the transformation plasmid (Step 2). These captured reads are then mapped and analyzed to determine the presence/absence of transformation plasmid backbone and/or T-DNA II sequences, identify insert junctions, and to determine the insert and copy number (Step 3). Overlapping PCR products are also produced which span any insert and their wild type locus (Step 4 and 5 respectively); these overlapping PCR products are sequenced to allow for detailed characterization of the inserted DNA and insertion site.

Genomic DNA from five breeding generations of MON 88702 (Figure 6) and the conventional control was isolated from seed and prepared for sequencing using the KAPA Hyper Prep Kit (Kapa Biosystems). These genomic DNA libraries were used to generate short (~125 bp) randomly distributed sequence fragments (sequencing reads) of the cotton genome (see Figure 4).

The NGS method was used to characterize the genomic DNA from MON 88702 and the conventional control using sequencing reads generated in sufficient numbers to ensure comprehensive coverage of the sample genomes. It has been previously demonstrated that 75× coverage of the soybean genome is adequate to provide comprehensive coverage and ensure detection of inserted DNA (Kovalic *et al.*, 2012). Similarly, it is expected that 75× will provide comprehensive coverage of the cotton genome. To confirm sufficient sequence coverage of the genome, the 125-mer sequence reads are analyzed to determine the coverage of a known single-copy endogenous gene, demonstrating the depth of coverage (the median number of times each base of the genome is independently sequenced). The level of sensitivity of this method was demonstrated by detection of a positive control spike that is randomly sampled at 1 and 1/10th copy-per-genome equivalent; this confirms the method's ability to detect any sequences derived from the transformation plasmid that may be present in MON 88702. Bioinformatics analysis was then used to select sequencing reads that contained sequences significantly similar to the transformation plasmid, and these were analysed in depth to determine the number of DNA inserts. NGS was run on all five generations of MON 88702 samples and the conventional controls. The NGS results are shown in Sections A3(d)(ii) and A3(f)(i).

To demonstrate sufficient sequence coverage the 125-mer sequence reads were analyzed by mapping all reads in each of the five breeding generations to a single copy locus for acyl carrier protein (*acp1)* selected from the *Gossypium hirsutum* genome in each of the five breeding generations. The analysis of sequence coverage plots showed that the depth of coverage was 80× or greater for the five generations of MON 88702 (R3 through R7) and the conventional control.

To demonstrate the method's ability to detect any sequences derived from the PV-GHIR508523 transformation plasmid, a sample of conventional control genomic DNA spiked with PV-GHIR508523 DNA was analyzed by NGS and bioinformatics. The level of sensitivity of this method was demonstrated to a level of one genome equivalent and 1/10th genome equivalent, 100% nucleotide identity was observed over 100% of PV-GHIR508523. This result demonstrates that all nucleotides of PV-GHIR508523 are observed by the sequencing and bioinformatic assessments performed and that a detection level of at least 1/10th genome equivalent was achieved for the plasmid DNA sequence assessment.

The DNA inserts of MON 88702 were determined by mapping of sequencing reads relative to the transformation plasmid and identifying junctions and unpaired read mappings adjacent to the junctions. The junctions of the DNA insert and the flanking DNA are unique for each insertion (Kovalic *et al.*, 2012). Therefore, insertion sites can be recognized by analyzing for sequence reads containing such junctions.

Directed sequencing (locus-specific PCR and DNA sequencing analyses, Figure 4, Step 4) complements the NGS method. Sequencing of the insert and flanking genomic DNA determined the complete sequence of the insert and flanks by evaluating if the sequence of the insert was identical to the corresponding sequence from the T-DNA I in PV-GHIR508523, and if each genetic element in the insert was intact. It also characterizes the flank sequence beyond the insert corresponding to the genomic DNA of the transformed cotton. Results are described in Sections A3(d)(ii) and A3(d)(iii). For details, please also refer to [redacted], 2017 (MSL0028391).

The stability of the T-DNA I present in MON 88702 across multiple breeding generations was evaluated by NGS as described above. This information was used to determine the number and identity of the DNA inserts in each generation. For the single copy T-DNA I insert, two junction sequence classes are expected. In the case of an event where a single locus is stably inherited over multiple breeding generations, two identical junction sequence classes would be detected in all the breeding generations tested. Results are described in Section A3(f)(i). For details, please also refer to [redacted] 2017 (MSL0028391).

Segregation analysis of the T-DNA I was conducted to determine the inheritance and stability of the insert in MON 88702. Segregation analysis corroborates the insert stability demonstrated by NGS and independently establishes the genetic behavior of the T-DNA I. Results are described in Section A3(f)(i). For details, please also refer to [redacted] 2016 (MSL0027485).

A3(d)(ii) Determination of the number of insertion sites, and the number of copies at each insertion site

The number of insertion sites of PV-GHIR508523 DNA in MON 88702 was assessed by performing NGS on MON 88702 genomic DNA using the R4 generation (Figure 6). A plasmid map of PV-GHIR508523 is shown in Figure 3.

A schematic representation of the insert and flanking sequences in MON 88702 is shown in Figure 5. Table 2 provides a description of the genetic elements present in MON 88702.

Table 2. Summary of Genetic Elements in MON 88702

Genetic Element	Location in Sequence	Function (Reference)
Flanking DNA	1-1642	Flanking DNA
B[1]-Right Border Region[r1]	1643-1710	DNA region from *Agrobacterium tumefaciens* containing the right border sequence used for transfer of the T–DNA (Depicker *et al.*, 1982; Zambryski *et al.*, 1982)
Intervening Sequence	1711-1763	Sequence used in DNA cloning
E[2]-*FMV*	1764-2170	Enhancer from the 35S RNA of figwort mosaic virus (FMV) (Richins *et al.*, 1987) that enhances transcription in most plant cells (Rogers, 2000)
Intervening Sequence	2171-2245	Sequence used in DNA cloning
P[3]-*Hsp81-2*	2246-3253	Promoter and 5′ UTR leader sequence for the heat shock protein 81-2 (Hsp81-2) from *Arabidopsis thaliana* that directs transcription in plant cells (Yabe *et al.*, 1994)
Intervening Sequence	3254-3290	Sequence used in DNA cloning
CS[4]-*Cry51Aa2.834_16*	3291-4211	Coding sequence of the modified Cry51Aa2 protein of *Bacillus thuringiensis* that provides insect resistance (Baum *et al.*, 2012; Anderson *et al.*, 2015; Gowda *et al.*, 2016)
Intervening Sequence	4212-4243	Sequence used in DNA cloning
T[5]-*35S*	4244-4443	3' UTR sequence of the 35S RNA of cauliflower mosaic virus (CaMV) (Mogen *et al.*, 1990) that directs polyadenylation in plant cells
Intervening Sequence	4444-4581	Sequence used in DNA cloning
B-Left Border Region[r1]	4582-4785	DNA region from *Agrobacterium tumefaciens* containing the left border sequence used for transfer of the T–DNA (Barker *et al.*, 1983)
Flanking DNA	4786-6748	Flanking DNA

[1] B, Border
[2] E, Enhancer
[3] P, Promoter
[4] CS, Coding Sequence
[5] T, Transcription Termination Sequence
[r1] Superscript in Left and Right Border Regions indicate that the sequence in MON 88702 was truncated compared to the sequences in PV-GHIR508523

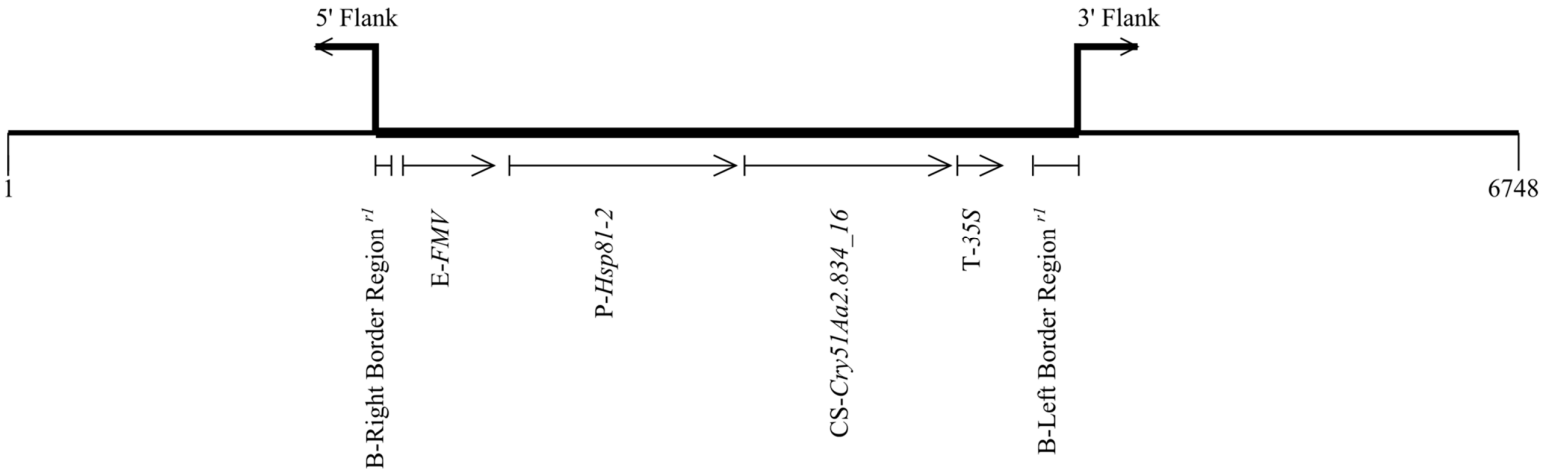

Figure 5. Schematic Representation of the Insert and Flanking Sequences in MON 88702

DNA derived from T-DNA I of PV-GHIR508523 integrated in MON 88702. Right-angled arrows indicate the ends of the integrated T-DNA and the beginning of the flanking sequence. Identified on the map are genetic elements within the insert. This schematic diagram may not be drawn to scale.

[r1] Superscripts in Left and Right Border Regions indicate that the sequence in MON 88702 was truncated compared to the sequences in PV-GHIR508523.

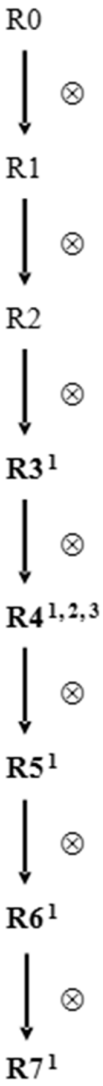

Figure 6. Breeding History of MON 88702

The generations used for molecular characterization and insert stability analyses are indicated in bold text. R0 corresponds to the transformed plant, ⊗ designates self-pollination.

[1]Generations used to confirm insert stability

[2] Generation used for molecular characterization and compositional assessment

[3] Generation used for breeding commercial varieties of MON 88702

As T-DNA from plasmid PV-GHIR508523 was transformed into the parental variety DP393 to produce MON 88702, any DNA inserted into MON 88702 will consist of sequences that are similar to the PV-GHIR508523 DNA sequence. Complete analysis of only the sequence reads that have similarity to plasmid PV-GHIR508523 (Figure 4, Step 2) is sufficient to characterize the DNA from PV-GHIR508523 inserted in MON 88702. Any inserted transformation plasmid vector sequence, regardless of origin, either T-DNA I, T-DNA II, or backbone, can be identified by mapping sequence reads to the transformation plasmid vector sequence, while the number of inserted DNA molecules can be determined by identifying the number of junction sequences. Therefore, the NGS method described above used the entire plasmid vector sequence as a query to determine the DNA insertion site number.

Unlike the traditional Southern blot analysis that separately hybridizes T-DNA or backbone probes, NGS uses identification of sequence reads that match PV-GHIR508523 to determine T-DNA I presence and insert number, as well as the absence of backbone, T-DNA II, or unintended sequences. This alternative method can be used to reach the same conclusions regarding the number of inserts, and presence or absence of backbone or T-DNA II that can be determined using traditional Southern blots (Kovalic *et al.*, 2012).

Using established criteria, sequence reads similar to PV-GHIR508523 were selected from MON 88702 sequence datasets and mapped relative to the transformation plasmid sequence in order to identify junction sequences (Figure 8). PV-GHIR508523 sequences were also compared against the conventional control sequence dataset.

No reads from the conventional control dataset were found to map with T-DNA I, T-DNA II, OR-*ori-pRi*, CS-*nptII* or P-*rrn*. However, a small number of alignments were found to align with OR-*ori-pBR322* or CS-*rop* sequences (Figure 7). The sporadic low level detection of plasmid sequences such as OR-*ori-pBR322* has previously been described (Zastrow-Hayes *et al.*, 2015), and reported (see Supplemental Figure S1 in (Yang *et al.*, 2013)) and is due to the presence of environmental bacteria in tissue samples used in the preparation of genomic DNA used for library construction. Despite, the low level presence of sequence from environmental bacteria, altogether these results indicate the expected absence of inserted transformation plasmid DNA in the control.

When reads from the MON 88702 (R4) dataset were aligned with the transformation plasmid sequence, large numbers of reads mapped to T-DNA I, E-*FMV* contained in T-DNA II, and a significantly smaller number to OR-*ori-pBR322* and CS-*rop* contained in the transformation plasmid backbone (Figure 8).

The mapping of large numbers of reads (> 2000) from the MON 88702 (R4) dataset to T-DNA I is expected and fully consistent with the presence of the inserted DNA. Likewise, the mapping to E-*FMV* in T-DNA- II is not unexpected as the E-*FMV* sequence is contained within both T-DNA I and T-DNA II. Since these sequences are identical, the mapped read mate pairs were used to distinguish their true mapping location. Since all pairs which were not fully contained within the E-*FMV* region had their mates map to T-DNA I and not T-DNA II, these reads were uniquely assigned to T-DNA I. Furthermore, those paired reads fully contained in the E-*FMV* region that could map to either T-DNA I or T-DNA II were also uniquely assigned to T-DNA I. Consequently, no reads in the MON 88702 R4

generation dataset were identified that uniquely aligned with the plasmid backbone or T-DNA II. The small number of reads mapping with OR-*ori-pBR322* and CS-*rop* are comparable to those previously described in the conventional control dataset. As a result, it is concluded that MON 88702 (R4) does not contain inserted sequence from the transformation plasmid backbone or T-DNA II.

To determine the insert number in MON 88702 (R4), selected reads mapping to T-DNA I were analyzed to identify junctions. This bioinformatic analysis is used to identify and partially matched reads at the ends of insertions. The number of unique junctions determined by this analysis are shown in Table 3.

Table 3. Unique Junction Results

Sample	Junctions Detected
MON 88702 (R4)	2
DP393	0

The location and orientation of the flanking DNA relative to the T-DNA I insert determined for MON 88702 (as described in Section A3(d)(iii)) are illustrated in Figure 5. There are two junctions identified in MON 88702, both contain the T-DNA I border sequence joined to adjacent flanking sequence. As such they represent the sequences at the junctions of the intended T-DNA I insert and the adjacent flanking sequence.

Complete alignment to the full flank/insert sequence confirms that both of these junctions originate from the same locus of the MON 88702 genome and are linked by contiguous, DNA that makes up the single insert.

For details, please also refer to [redacted] 2017 (MSL0028391).

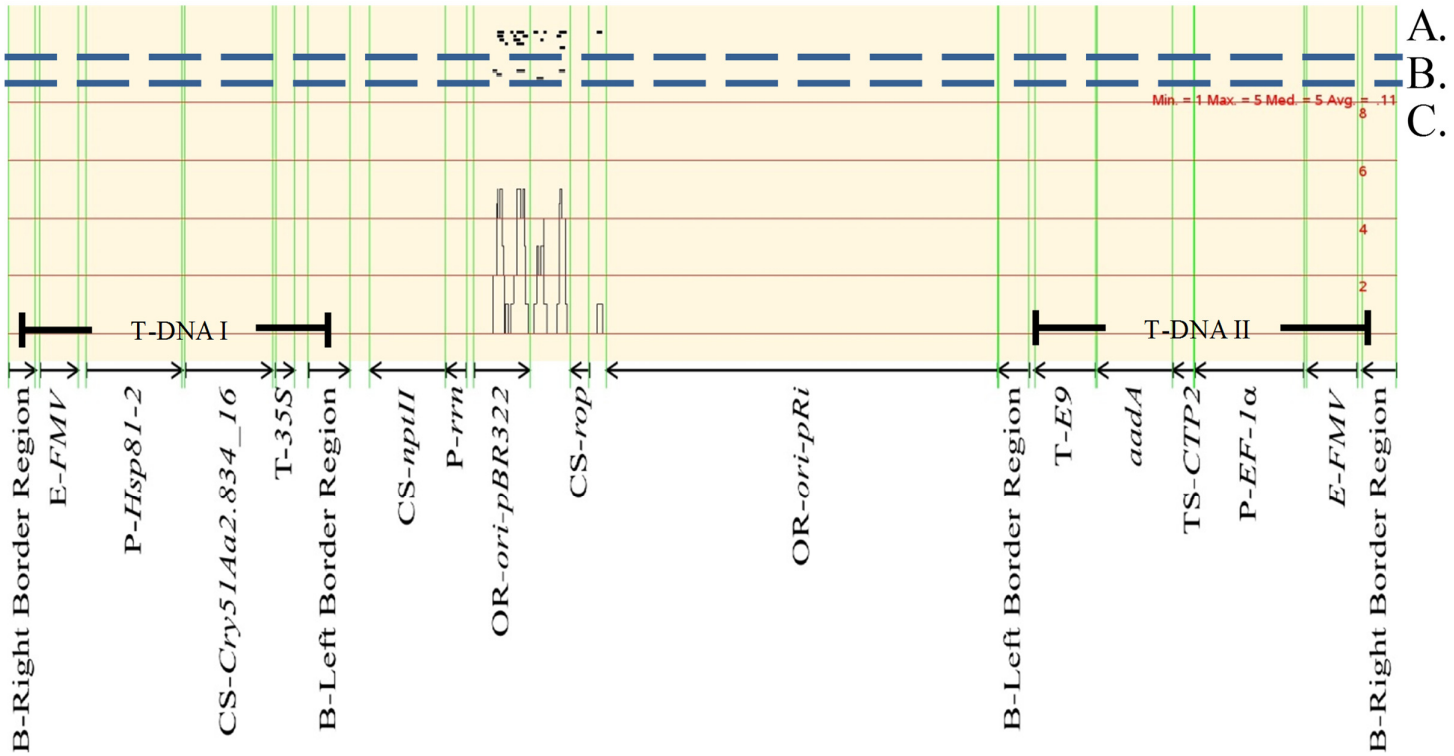

Figure 7. Read Mapping of Conventional Cotton Versus PV−GHIR508523

Panel A shows the location of unpaired mapped reads, Panel B shows paired mapped reads, and Panel C shows a representation of combined read depth for unpaired and paired reads.

Vertical lines show genetic element boundaries.

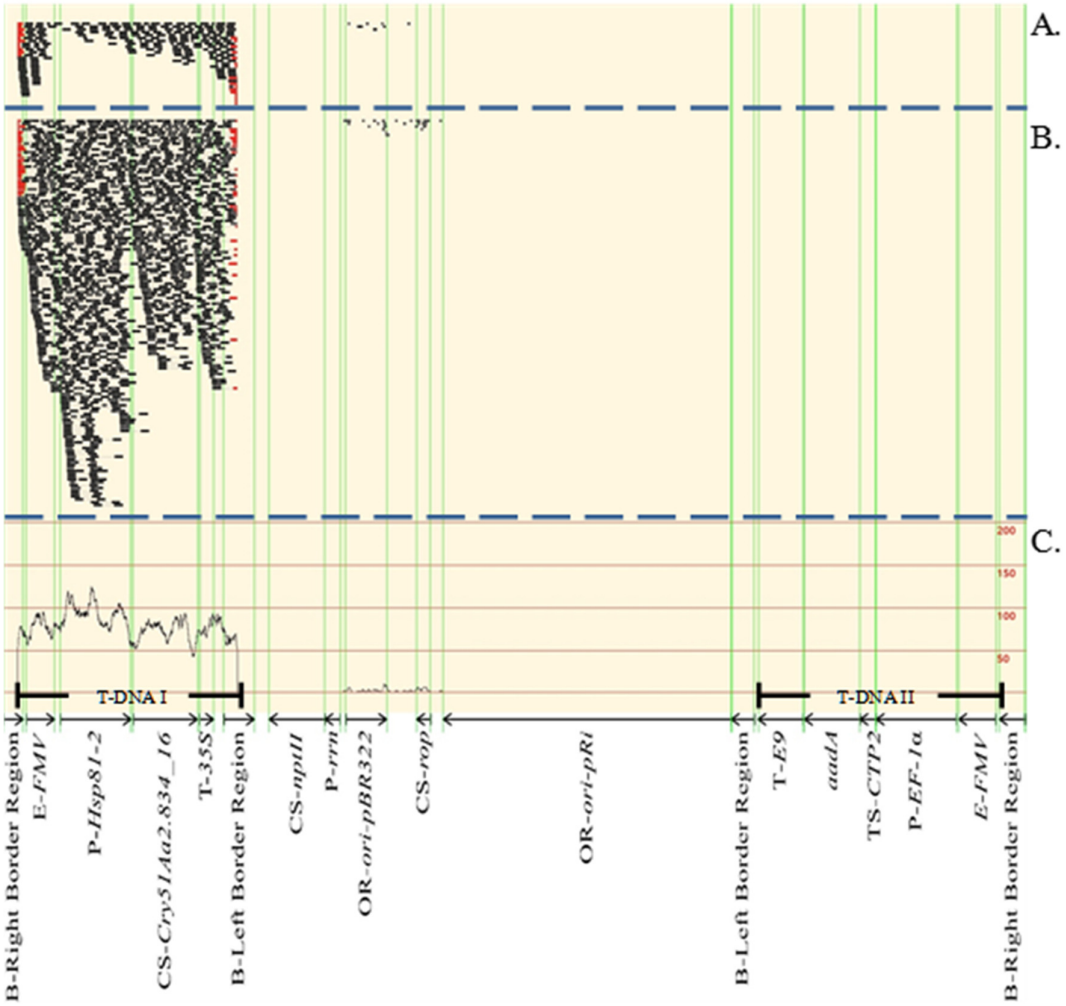

Figure 8. Read Mapping of MON 88702 (R4) Versus PV−GHIR508523

Panel A shows the location of unpaired mapped reads. Panel B shows paired mapped reads and Panel C shows a representation of combined read depth for unpaired and paired reads. Vertical lines show genetic element boundaries. The region of flank junction sequences that aligns with transformation plasmid is shown in red.

A3(d)(iii) Full DNA sequence, including junction regions

Organization and Sequence of Insert and Adjacent DNA

The intactness and organization of the elements within the DNA insert in the R4 generation of MON 88702 was confirmed by using PCR to amplify and subsequently sequence two overlapping DNA amplicons that span the entire insert and the associated flanking DNA sequence. The positions of the PCR products relative to the insert, as well as the results of the PCR analyses, are shown in Figure 9. The amplified PCR products were subjected to DNA sequencing analyses. The results of this analysis confirm that the MON 88702 insert is 3,143 bp, that each genetic element (except for the T-DNA border regions) in the insert is intact, and the sequence of the insert is identical to the corresponding sequence in PV-GHIR508523. This analysis also shows that only T-DNA I elements (described in Table 2) were present. Moreover, the result, together with the conclusion of single DNA insert detected by NGS, demonstrated that no PV-GHIR508523 backbone or T-DNA II elements are present in MON 88702.

For details, please also refer to [redacted], 2017 (MSL0028391).

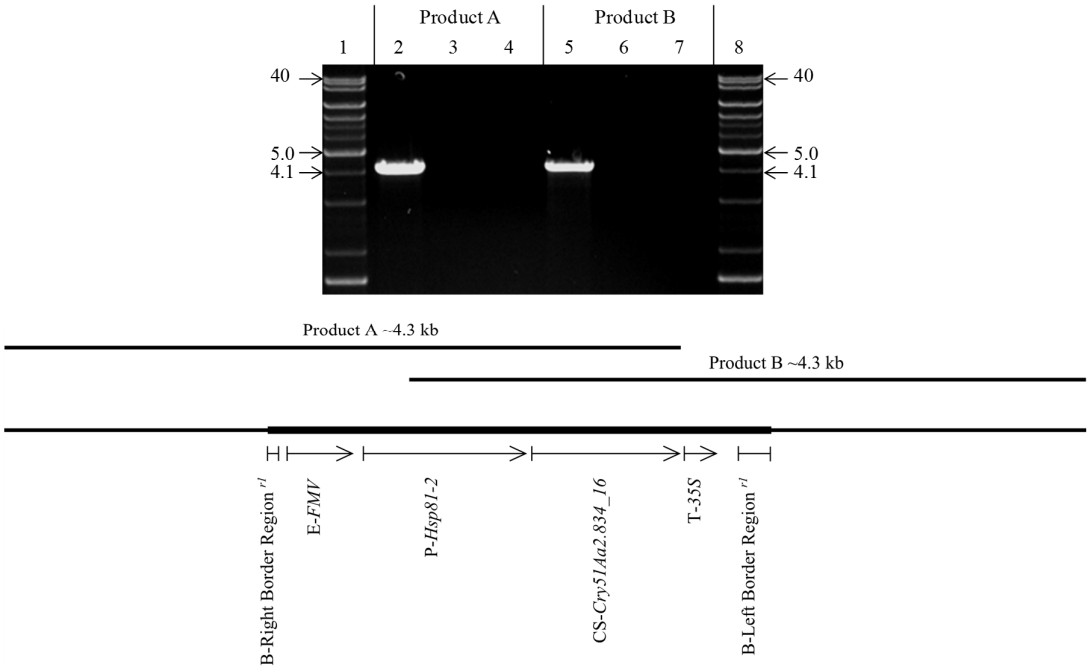

Figure 9. Overlapping PCR Analysis across the Insert in MON 88702

PCR was performed on both conventional control genomic DNA and genomic DNA of the R4 generation of MON 88702 using two pairs of primers to generate overlapping PCR fragments from MON 88702 for sequencing analysis. To verify the PCR products, 2µl of each of the PCR reactions was loaded on the gel. The expected product size for each amplicon is provided in the illustration of the insert in MON 88702 that appears at the bottom of the figure. This figure is a representative of the data generated in the study. Lane designations are as follows:

Lane	
1	1Kb DNA Extension Ladder
2	MON 88702
3	Conventional Control
4	No template control
5	MON 88702
6	Conventional Control
7	No template control
8	1Kb DNA Extension Ladder

Arrows on the agarose gel photograph denote the size of the DNA, in kilobase pairs, obtained from the 1Kb DNA Extension Ladder (Invitrogen) on the ethidium bromide stained gel.
r1 Superscript in Left and Right Border Regions indicate that the sequence in MON 88702 was truncated compared to the sequences in PV-GHIR508523

Sequence of the Insertion Site

PCR and sequence analysis were performed on genomic DNA extracted from the conventional control to examine the insertion site in conventional cotton (see Figure 4, Step 5). The PCR was performed with one primer specific to the genomic DNA sequence flanking the 5' end of the MON 88702 insert paired with a second primer specific to the genomic DNA sequence flanking the 3' end of the insert (Figure 10). A sequence comparison between the PCR product generated from the conventional control and the sequence generated from the 5' and 3' flanking sequences of MON 88702 indicates that 244 bases of cotton genomic DNA were deleted and 4 bases inserted in the MON 88702 3' flanking sequence during integration of the T-DNA I. The remainder of the flanks in MON 88702 are identical to the conventional control. Such changes are common during plant transformation (Anderson *et al.*, 2016) and these changes presumably resulted from double stranded break repair mechanisms in the plant during *Agrobacterium*-mediated transformation process (Salomon and Puchta, 1998).

For details, please also refer to [redacted]2017 (MSL0028391).

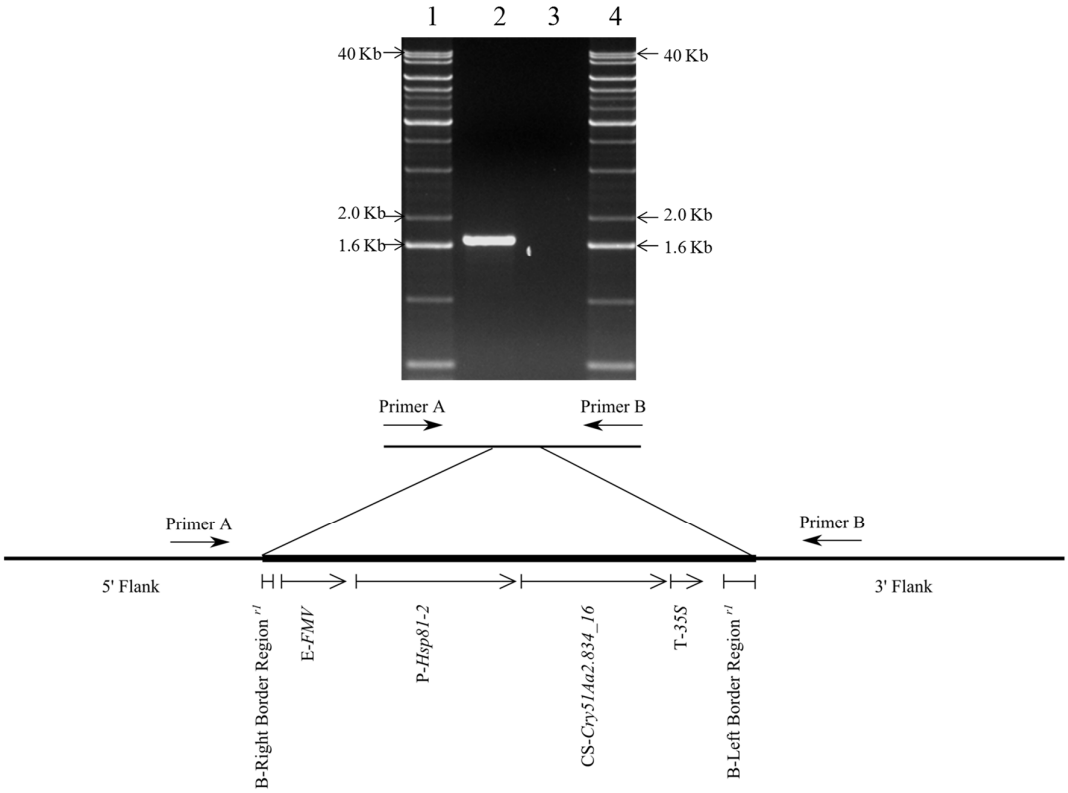

Figure 10. PCR Amplification of the MON 88702 Insertion Site

PCR Analysis was performed to evaluate the insertion site. PCR was performed on conventional control DNA using Primer A, specific to the 5′ flanking sequence, and Primer B, specific to the 3′ flanking sequence of the insert in MON 88702. The amplicon generated from PCR with the conventional control genomic DNA was used for sequencing analysis. This illustration depicts the MON 88702 insertion site in the conventional control (upper panel) and the MON 88702 inset (lower panel). Approximately 2 µl of each of the PCR reactions was loaded on the gel. This figure is representative of the data generated in the study. Lane designations are as follows:

Lane	
1	1Kb DNA Extension Ladder
2	Conventional Control
3	No template control
4	1Kb DNA Extension Ladder

Arrows on the agarose gel photograph denote the size of the DNA, in kilobase pairs, obtained from the 1Kb DNA Extension Ladder (Invitrogen) on the ethidium bromide stained gel.

r1 Superscript in Left and Right Border Regions indicate that the sequence in MON 88702 was truncated compared to the sequences in PV-GHIR508523.

A3(d)(iv) Map of the organization of the inserted DNA (each site)

PCR and DNA sequence analyses performed on MON 88702 and the conventional control determined the organisation of the genetic elements within the insert as given in Figure 9.

A3(d)(v) Identification and characterisation of unexpected ORFs

Bioinformatic Assessment of Putative Open Reading Frames (ORFs) of MON 88702 Insert and Flanking Sequences

The 2009 Codex Alimentarius Commission guidelines for the safety assessment of food derived from biotechnology crops (Codex Alimentarius, 2009) includes an assessment element on the identification and evaluation of "open reading frames within the inserted DNA or created by the insertion with contiguous plant genomic DNA." These assessments examine the potential homology of any putative polypeptides or proteins that could be produced from open reading frames (ORFs) in the insert or at the plant-insert junction to known toxins or allergens. These analyses are conducted even if there is no evidence that such alternative reading frames in the insert or such ORFs at the plant-insert junction are capable of being transcribed or translated into a protein. Results from these bioinformatics analyses demonstrate that any putative polypeptides in MON 88702 are unlikely to exhibit allergenic, toxic or otherwise biologically adverse properties.

Bioinformatic analyses were performed on the MON 88702 insert and flanking genomic DNA sequences to assess the potential for allergenicity, toxicity, or biological activity of putative polypeptides encoded by all six reading frames present in the MON 88702 insert DNA, as well as ORFs spanning the 5' and 3' insert DNA-flanking sequence junctions. ORFs spanning the 5' and 3' cotton genomic DNA-inserted DNA junctions were translated from stop codon to stop codon in all six reading frames (three forward reading frames and three reading frames in reverse orientation)[3]. Polypeptides of eight amino acids or greater from each reading frame were then compared to toxin, allergen and all proteins databases using bioinformatic tools. Similarly, the entire T-DNA I sequence was translated in all six reading frames and the resulting deduced amino acid sequence was subjected to bioinformatic analyses.

Translated sequences were compared to allergen (AD_2017), toxin (TOX_2017) and all protein (PRT_2017) databases. The FASTA sequence alignment tool was used to assess the relatedness between the query sequences and any protein sequence in the AD_2017, TOX_2017, and PRT_2017 databases. Similarities shared between the sequence with each

[3] An evaluation of sequence translated from stop codon to stop codon represents the most conservative approach possible for flank junction analysis as it does not assume that a start codon is necessary for the production of a protein sequence.

sequence in the database were examined. The extent of relatedness was evaluated by detailed visual inspection of the alignment, the calculated percent identity, and the *E*-score. Alignments having *E*-scores of ≤1e-5 (1×10^{-5}) are deemed significant because they may reflect shared structure and function among sequences. In addition to sequence similarity, sequences were screened for short polypeptide matches using a pair-wise comparison algorithm. In these analyses, eight contiguous and identical amino acids were defined as immunologically relevant, where eight represents the typical minimum sequence length likely to represent an immunological epitope (Silvanovich *et al.*, 2006).

There are no analytical data that indicate that any putative polypeptides or proteins other than mCry51Aa2 are produced by the MON 88702 T-DNA insert. Moreover, the data generated from these analyses confirm that even in the highly unlikely occurrence that a translation product other than the mCry51Aa2 protein were derived from frames one to six of the insert DNA or the ORFs spanning the insert junctions, they would not share a sufficient degree of sequence similarity with other proteins to indicate they would be potentially allergenic, toxic, or have other safety implications. Therefore, there is no evidence for concern regarding the relatedness of putative polypeptides from MON 88702 to known toxins and allergens, or biologically active putative peptides.

Bioinformatics Assessment of Insert DNA Reading Frames

Bioinformatic analyses were performed to assess the potential of toxicity, allergenicity or biological activity of any putative peptides encoded by translation of reading frames 1 through 6 of the inserted DNA in MON 88702 (Figure 11).

The FASTA sequence alignment tool was used to assess structural relatedness between the query sequences and any protein sequences in the AD_2017, TOX_2017, and PRT_2017 databases. Structural similarities shared between each putative polypeptide with each sequence in the database were examined. The extent of structural relatedness was evaluated by detailed visual inspection of the alignment, the calculated percent identity and alignment length to ascertain if alignments exceeded Codex (Codex Alimentarius, 2009) thresholds for FASTA searches of the AD_2017 database, and the *E*-score. Alignments having an *E*-score less than 1×10^{-5} are deemed significant because they may reflect shared structure and function among sequences. In addition to structural similarity, each putative polypeptide was screened for short polypeptide matches using a pair-wise comparison algorithm. In these analyses, eight contiguous and identical amino acids were defined as immunologically relevant, where eight represents the typical minimum sequence length likely to represent an immunological epitope (Silvanovich *et al.*, 2006) and evaluated against the AD_2017 database.

The results of the search comparisons showed that no relevant structural similarity to known allergens were observed for any of the putative polypeptides when compared to proteins in the allergen (AD_2017) database. Furthermore, no short (eight amino acid) polypeptide matches were shared between any of the putative polypeptides and proteins in the allergen database and no FASTA alignments met or exceeded the Codex Alimentarius (2009) threshold of >35% identity in ≥80 contiguous amino acids with the query sequence.

From the FASTA searches using the TOX_2017 database, a single FASTA alignment that yielded *E*-score of 2.6e-06, less than the *E*-score threshold of 1e-5, with low linear sequence alignment identity of approximately 29%, was obtained with TOX_2017 database entry GI-1102943401. The aligning region corresponds to the mCry51Aa2 coding region and the relevance of the alignment has been discussed in detail in the Section B4(a). The small differences in *E*-score are related to the translated protein from the TDNA+4_2 sequence being different (longer) from that of the mCry51Aa2 coding sequence. In Section B4(a), it has been concluded that GI-1102943401 and mCry51Aa2 share structural similarities in the pore-forming region, while they are more dissimilar in the receptor-binding regions that confer specificity for their target hosts. Therefore, this alignment does not provide any additional information to indicate that mCry51Aa2 would be toxic outside of the target insect species.

When used to query the PRT_2017 database, translations of frames 1 and 2 yielded alignments with *E*-scores less than or equal to 1e-5. One of the top alignments from frame 1 was to an "unnamed protein product" from figwort mosaic virus (GI-58814), with an *E*-score of 1.3e-58 and >95% identity over a 142 amino acid overlap. Not surprisingly, the region of overlap (query 35-176 : subject 368-509) demarks the enhancer region from figwort mosaic virus, the E-*FMV* element, used to enhance transcription within the MON 88702 T-DNA. Several other alignments were observed to this region of the frame 1 query sequence, including GI-325369, protein "VI" from figwort mosaic virus, with an *E*-score of 1.2e-18. No other alignments were observed with an *E*-score of ≤1e-5. Positive identification of the E-*FMV* enhancer element in the MON 88702 T-DNA was expected and does not suggest potential adverse biological activity.

The top scoring hit to reading frame 2 was to "Sequence 34 from U.S. patent 9322033" (GI-1115466244), with an *E*-score of 3.9e-142 and 100% identity over a 306 amino acid overlap. Not surprisingly, the region of overlap (query 550-855 : subject 1-306) demarks the coding region of the mCry51Aa2 protein within the MON 88702 T-DNA. Several other alignments were observed to this region of the frame 2 query sequence and displayed an *E*-score of ≤1e-5; most of which were annotated as patent sequences or crystal protein sequences from *Bacillus thuringiensis*. Positive identification of Cry family proteins was expected (Crickmore *et al.*, 2016), as this region of the MON 88702 frame 2 T-DNA encodes the mCry51Aa2 protein. These observations do not suggest the potential for adverse biological activity. A second region of alignments to reading frame 2 was observed and is represented by GI-482549750, with an *E*-score of 2.6e-9 and >56% identity over an 82 amino acid overlap. This region (query 223-297) corresponds the P-*Hsp81-2* promoter and 5′ UTR leader sequence for the heat shock protein 81-2 (Hsp81-2) from *Arabidopsis thaliana*. GI-482549750 corresponds to "hypothetical protein CARUB_v10027062mg" from *Capsella rubella*, a plant species and close relative to *Arabidopsis thaliana*. The anchoring residues for this second region of alignments were all derived from query amino acid positions 262-297, displaying approximately 100% amino acid identity. This region of the T-DNA is not expected to be contained in a transcript and therefore a translation product is not expected. These analyses correctly identified the *Arabidopsis thaliana* promotor element of the MON 88702 T-DNA through conceptually translated database entries. This positive

identification does not suggest potential adverse biological activity. Other individual alignments were observed with numerically greater *E*-scores, and ≤1e-5, but all observations were represented by the two regions described above.

The remaining translated frames yielded alignments with poor *E*-score values that did not suggest relevant sequence similarity.

For details, please also refer to [redacted] 2017 (MSL0028694).

Insert Junction Open Reading Frame Bioinformatics Analysis

Analyses of putative polypeptides encoded by DNA spanning the 5' and 3' genomic junctions of the MON 88702 inserted DNA were performed using a bioinformatic comparison strategy (Figure 11).

The purpose of the assessment is to evaluate the potential for novel ORFs that may have homology to known allergens, toxins, or proteins that display adverse biological activity. Sequences spanning the 5' and 3' cotton genomic DNA-inserted DNA junctions (Figure 11) were translated from stop codon (TGA, TAG, TAA) to stop codon in all six reading frames. Putative polypeptides from each reading frame of eight amino acids or greater in length, were compared to AD_2017, TOX_2017, and PRT_2017 databases using FASTA and to the AD_2017 database using an eight amino acid sliding window search.

The FASTA sequence alignment tool was used to assess structural relatedness between the query sequences and protein sequences in the AD_2017, TOX_2017, and PRT_2017 databases. Structural similarities shared between 10 putative polypeptides (Figure 11) with each sequence in the database were examined. The extent of structural relatedness was evaluated by detailed visual inspection of the alignment, the calculated percent identity and the alignment length to ascertain if alignments exceeded Codex (Codex Alimentarius, 2009) thresholds for FASTA searches of the AD_2017 database, and the *E*-score. In addition to structural similarity, each putative polypeptide was screened for short polypeptide matches using a pair-wise comparison algorithm. In these analyses, eight contiguous and identical amino acids were defined as immunologically relevant, where eight represents the typical minimum sequence length likely to represent an immunological epitope, and evaluated against the AD_2017 database (Silvanovich *et al.*, 2006).

No FASTA alignments were observed with an *E*-score of ≤1e-5 using the AD_2017 database or a sliding window 8 amino acid search and no alignments were observed to meet or exceed the Codex Alimentarius (2009) threshold of >35% identity in 80 contiguous amino acids with the query sequence. Additionally, no FASTA alignments were observed with an *E*-score of ≤1e-5 using the TOX_2017 or the PRT_2017 database Consequentially, there is no reason to suspect allergenic cross reactivity, adverse toxic effects, or to suspect adverse biological activity from putative flanking peptides encoded by any of the T-DNA frames from MON 88702.

For details, please also refer to [redacted] 2017 (MSL0028798).

Bioinformatic Assessment of Allergenicity, Toxicity, and Adverse Biological Activity Potential of MON 88702 Polypeptides Putatively Encoded by the Insert and Flanking Sequences Summary and Conclusions

A conservative bioinformatic assessment of allergenicity, toxicity and adverse biological activity for putative polypeptides that are encoded on all reading frames and spanning the 5' and 3' junctions of the insert DNA in MON 88702 was conducted. The data generated from these analyses confirm that even in the highly unlikely occurrence that a translation product other than the mCry51Aa2 protein sequence was derived from frames 1 to 6 for the insert DNA, or the insert junctions, they would not share a sufficient degree of sequence similarity with other proteins to indicate they would be potentially allergenic, toxic, or have other safety implications. A single FASTA alignment in the TOX_2017 database exceeded the *E*-score threshold when using the reading frame that contains the mCry51Aa2 protein. It was concluded that the alignment did not provide any additional information to indicate mCry51Aa2 to be toxic (see Section B4(a)). Furthermore, no short (eight amino acid) polypeptide matches were shared between any of the putative polypeptides and proteins in the allergen database. Therefore, there is no evidence for concern regarding health implications of putative polypeptides for MON 88702.

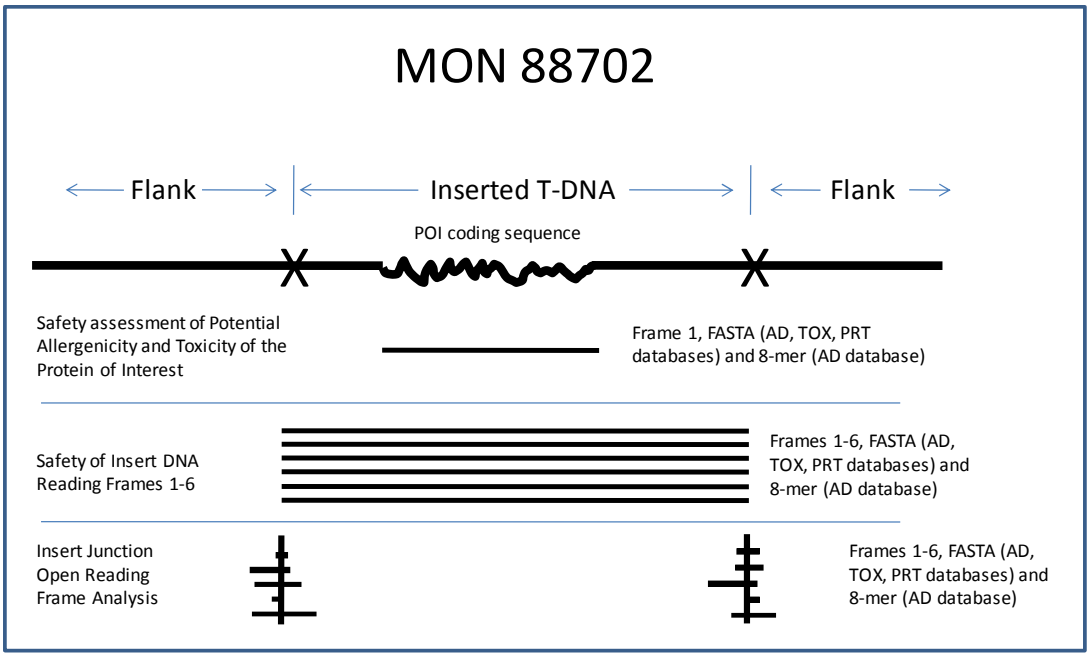

Figure 11. Schematic Summary of MON 88702 Bioinformatic Analyses

AD= AD_2017; TOX= TOX_2017 and PRT= PRT_2017 (GenBank release #217): 8-mer = the eight amino acid sliding window search.

A3(e) Family tree or breeding process

The cotton variety used as the recipient for the DNA insertion to create MON 88702 was DP393, a non-transgenic conventional upland variety developed by Delta and Pineland Technology Holding Company, LLC (Bridge *et al.*, 2005), which was released in 2005 in the U.S. DP393 was used as the conventional parental cotton comparator (referred to in this consultation document as the conventional control) in the safety assessment of MON 88702. MON 88702 and the conventional control have similar genetic backgrounds with the exception of the T-DNA; thus, the effect of the T-DNA and the expressed mCry51Aa2 protein can be assessed in an objective manner. Please also refer to Section A3(f)(i) and Figure 12.

A3(f) Evidence of the stability of the genetic changes

A3(f)(i) Pattern of inheritance of insert and number of generations monitored

Determination of Insert Stability over Multiple Generations

In order to demonstrate the stability of the T-DNA I present in MON 88702 through multiple breeding generations, NGS was performed using DNA obtained from five breeding generations of MON 88702. The breeding history of MON 88702 is presented in Figure 6 and the specific generations tested are indicated in the figure legend. The MON 88702 R4 generation was used for the molecular characterization analyses discussed in Sections A3(d)(ii)-A3(d)(iii). To assess stability, four additional generations were evaluated by the NGS method as previously described in Section A3(d)(ii), and compared to the fully characterized R4 generation. The conventional controls used for the generational stability analysis was DP393, a conventional variety with similar background genetics. Genomic DNA isolated from each of the selected generations of MON 88702 and conventional control was used for NGS.

To determine the insert number in the MON 88702 generations, the sequences were analyzed using NGS (Kovalic *et al.*, 2012). Table 4 shows the number of unique junctions containing PV-GHIR508523 DNA sequence determined by this analysis.

Table 4. Junction Sequence Classes Detected

Sample	Junction Sequence Classes Detected
MON 88702 R3	2
MON 88702 R4	2
MON 88702 R5	2
MON 88702 R6	2
MON 88702 R7	2
DP393	0

Alignment of the junction sequences from each of the assessed MON 88702 generations to the full flank/insert sequence determined for the MON 88702 R4 generation, confirms that the pair of junction sequences originates from the same region of the MON 88702 genome and is linked by contiguous, known and expected DNA sequence. This single identical pair of junction sequences is observed as a result of the insertion of PV-GHIR508523 T-DNA I at a single locus in the genome of MON 88702. The consistency of the data across all generations tested demonstrates that this single locus was stably maintained throughout the MON 88702 breeding process, thereby confirming the stability of the insert. Based on this comprehensive sequence data and bioinformatic analysis (NGS), it is concluded that MON 88702 contains a single and stable T-DNA I insertion.

For details, please also refer to [redacted]2017 (MSL0028391).

Inheritance of the Genetic Insert

The MON 88702 breeding path, from which segregation data were generated, is described in Figure 12. The transformed R0 plant was self-pollinated to generate R1 seed. An individual homozygous positive plant was identified in the R1 segregating population via an End Point TaqMan® PCR assay.

The homozygous positive R1 plant was self-pollinated to give rise to R2 seed. The R2 plants were self-pollinated to produce R3 seed. The R3 plants were self-pollinated to produce R4 seed. Homozygous positive R4 plants were crossed via traditional breeding techniques to a Monsanto proprietary recurrent parent that does not contain the *mCry51Aa2* coding sequence to produce hemizygous R4F1 seed. The R4F1 plants were crossed with the recurrent parent to produce BC1F1 seed. The BC1F1 generation was tested for the presence of the T-DNA I by End Point TaqMan® PCR assay. The inheritance of the MON 88702 T-DNA I was assessed in the BC1F1 generation. At the BC1F1 generation, the MON 88702 T-DNA I was predicted to segregate at a 1:1 ratio (positive: negative) according to Mendelian inheritance principles.

The BC1F1 plants hemizygous for MON 88702 T-DNA I were crossed with the recurrent parent to produce the BC2F1 plants. The BC2F1 plants hemizygous for MON 88702 were crossed with the recurrent parent to produce the BC3F1. The inheritance of the MON 88702 T-DNA I was assessed in the BC3F1 generation. At the BC3F1 generation, the MON 88702 T-DNA I was predicted to segregate at a 1:1 ratio (positive: negative) according to Mendelian inheritance principles.

The BC2F1 plants hemizygous for MON 88702 T-DNA I were also self-pollinated to produce the BC2F2 plants. The BC2F2 generation was tested for the presence of MON 88702 T-DNA I by End Point TaqMan® PCR assay. The inheritance of the MON 88702 T-DNA I was assessed in the BC2F2 generation. At the BC2F2 generation, the MON 88702 T-DNA I was predicted to segregate at a 1:2:1 ratio (homozygous positive: hemizygous positive: homozygous negative) according to Mendelian inheritance principles.

A Pearson's chi-square (χ^2) analysis was used to compare the observed segregation ratios of the MON 88702 T-DNA I to the expected ratios. The χ^2 analysis was performed using the statistical program R Version 3.2.2 (2015-08-14).

The Chi-square was calculated as:

$\chi^2 = \sum [(| o - e |)^2 / e]$

where o = observed frequency of the genotype or phenotype and e = expected frequency of the genotype or phenotype. The level of statistical significance was predetermined to be 5% ($\alpha = 0.05$).

The results of the χ^2 analysis of the segregating progeny of MON 88702 are presented in Table 5. The χ^2 value in the BC1F1, BC2F2, and BC3F1 generations indicated no statistically significant difference between the observed and expected segregation ratios of MON 88702 T-DNA I. These results support the conclusion that the MON 88702 T-DNA I resides at a single locus within the cotton genome and is inherited according to Mendelian principles of inheritance. These results are also consistent with the molecular characterization data indicating that MON 88702 contains a single intact copy of the T-DNA I inserted at a single locus in the cotton genome.

For details, please also refer to [redacted] 2016 (MSL0027485).

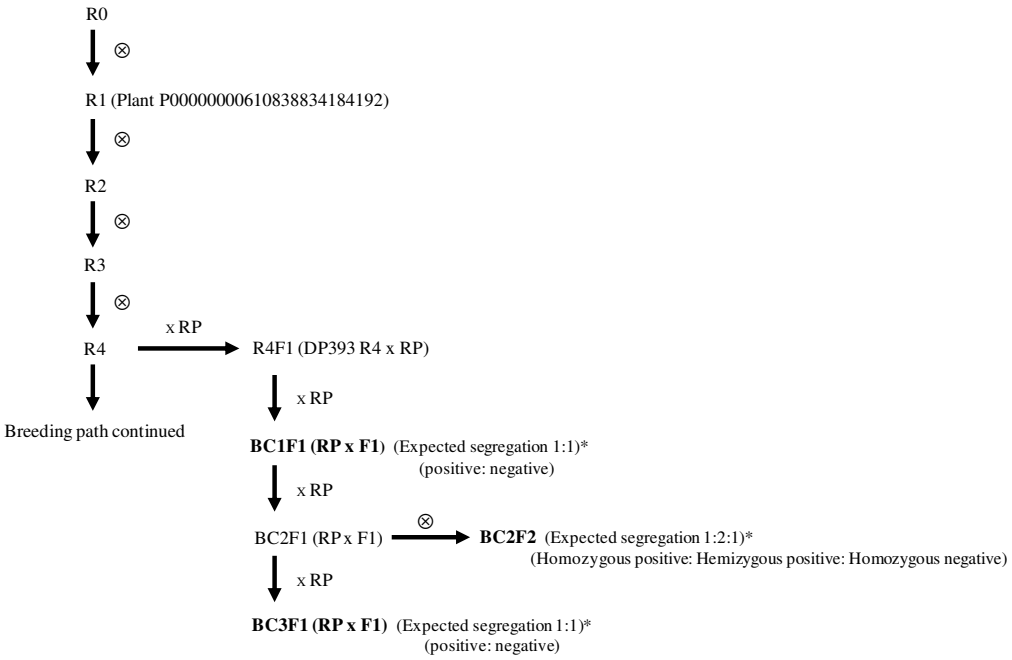

Figure 12. Breeding Path for Generating Segregation Data for MON 88702

*Chi-square analysis was conducted on segregation data from BC1F1, BC2F2, and BC3F1 generations (bolded text).

RP: Recurrent parent (12R241).
⊗: Self- Pollinated
BC: Back-Cross

Table 5. Segregation Results for MON 88702 from the BC1F1, BC2F2, and BC3F1

				1:1 Segregation			
Generation	**Total Plants**	**Observed # Plant Positive**	**Observed # Plant Negative**	**Expected # Plant Positive**	**Expected # Plant Negative**	**χ^2**	**Probability**
BC1F1	267	138	129	133.50	133.50	0.30	0.582
BC3F1	176	86	90	88.00	88.00	0.09	0.763

					1:2:1 Segregation				
Generation	**Total Plants**	**Observed # Plant Homozygous Positive**	**Observed # Plant Hemizygous Positive**	**Observed # Plant Homozygous Negative**	**Expected # Plant Homozygous Positive**	**Expected # Plant Hemizygous Positive**	**Expected # Plant Homozygous Negative**	**χ^2**	**Probability**
BC2F2	155	38	75	42	38.75	77.50	38.75	0.37	0.832

Characterization of the Genetic Modification Summary and Conclusion

As described above, characterization of the genetic modification in MON 88702 was conducted using a combination of sequencing, PCR, and bioinformatics. The results of this characterization demonstrate that MON 88702 contains a single copy of the intended T-DNA containing the *mCry51Aa2* expression cassette that is stably integrated at a single locus and is inherited according to Mendelian principles over multiple generations. These conclusions are based on the following:

- Molecular characterization of MON 88702 by NGS demonstrated that MON 88702 contained a single DNA insert. These whole-genome sequence analyses provided a comprehensive assessment of MON 88702 to determine the presence of sequences derived from PV-GHIR508523 (DuBose *et al.*, 2013; Kovalic *et al.*, 2012) and demonstrated that MON 88702 contained a single DNA insert with no detectable backbone sequences.
- Directed sequencing (locus-specific PCR, DNA sequencing and analyses) of MON 88702 which determined the complete sequence of the single DNA insert from PV-GHIR508523, the adjacent flanking DNA, and the 5' and 3' insert-to-flank junctions. This analysis confirmed that the sequence and organization of the DNA is identical to the corresponding region in the PV-GHIR508523 T-DNA. Furthermore, the genomic organization at the insertion site was assessed by comparing the sequences flanking the T-DNA I insert in MON 88702 to the sequence of the insertion site in conventional cotton. This analysis determined that no major DNA rearrangement occurred at the insertion site in MON 88702 upon DNA integration.
- Generational stability analysis by NGS demonstrated that the single PV-GHIR508523 T-DNA I insert in MON 88702 has been maintained through five breeding generations, thereby confirming the stability of the T-DNA I in MON 88702.
- Segregation data confirm that the inserted T-DNA I segregated following Mendelian inheritance patterns, which corroborates the insert stability demonstrated by NGS and independently establishes the nature of the T-DNA I at a single chromosomal locus.

Taken together, the characterization of the genetic modification in MON 88702 demonstrates that a single copy of the intended T-DNA was stably integrated at a single locus of the cotton genome and that no plasmid backbone sequences or T-DNA II are present in MON 88702.

A3(f)(ii) Pattern of expression of phenotype over several generations

Generational Stability of mCry51Aa2 Protein Expression in MON 88702

In order to demonstrate the presence of the mCry51Aa2 protein in MON 88702 across multiple breeding generations (refer to breeding history in Figure 6), western blot analysis of MON 88702 was conducted on leaf tissue collected from breeding generations R3 through R7 (Figure 6), and from a conventional control (DP393).

The presence of the mCry51Aa2 protein was demonstrated in five breeding generations of MON 88702 using Western blot analysis (Figure 13). A *Bt*-produced mCry51Aa2 standard

was used as a positive control for the identification of the mCry51Aa2 protein in MON 88702. The presence of mCry51Aa2 protein in leaf tissues of MON 88702 was determined by visual comparison of the bands produced in multiple breeding generations (Figure 13, lanes 5-9) to the mCry51Aa2 reference standard (Figure 13, lane 2). As expected, the mCry51Aa2 protein was present in all five breeding generations of MON 88702 and migrated with a mobility indistinguishable from that of the *Bt*-produced mCry51Aa2 protein standard analyzed on the same western blot. Also, as expected, the mCry51Aa2 protein was not detected in the conventional control extract (Figure 13, lane 4).

For details, please also refer to [redacted], 2016 (MSL0027352).

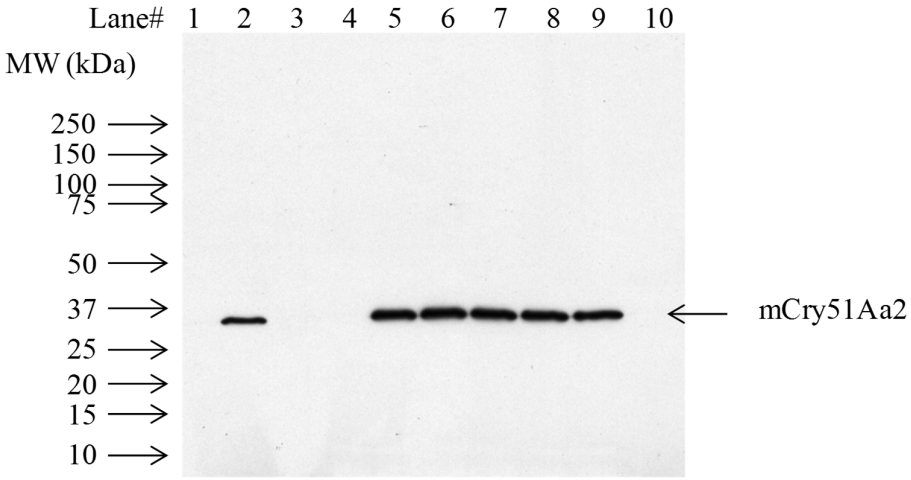

Figure 13. Presence of mCry51Aa2 Protein in Multiple Generations of MON 88702

Extracts from five generations of MON 88702 leaf tissues, conventional control leaf tissue, *Bt*-produced mCry51Aa2 protein standard, and molecular weight markers were subjected to SDS-PAGE and electrotransferred to a nitrocellulose membrane. The membrane was incubated with anti-mCry51Aa2 monoclonal antibody and immunoreactive bands visualized through the use of chemiluminescent reagents. Exposure time was five minutes. The molecular weights (in kDa) of the standards are shown on the left. Lane designations are as follows:

Lane	Description	Amount Loaded on Gel
1	Precision Plus Protein Dual Color MWM	5 µl
2	*Bt*-produced mCry51Aa2 protein (2 ng)	10 µl
3	Empty	N/A
4	Conventional Control,	10 µl
5	MON 88702 , R3,	10 µl
6	MON 88702, R4,	10 µl
7	MON 88702, R5,	10 µl
8	MON 88702, R6,	10 µl
9	MON 88702, R7,	10 µl
10	Empty	N/A

A4 Analytical Method for Detection

The event-specific DNA-based detection methods such as PCR can be used as the monitoring tool to determine the presence of MON 88702 in a collected sample.

B. INFORMATION RELATED TO THE SAFETY OF THE GM FOOD

B1 Equivalence Studies

B1(a) mCry51Aa2 protein identity and equivalence

The safety assessment of crops derived through biotechnology includes characterization of the physicochemical and functional properties of and confirmation of the safety of the introduced protein(s). The introduced gene is usually expressed at very low level (mg protein per kg of cottonseed). A heterologous system, *Bt* in this case, allows sufficient amount of mCry51Aa2 protein to be produced. For the safety data generated using mCry51Aa2 expressed in a heterologous system to be applied to mCry51Aa2 protein produced in MON 88702, the equivalence of the plant- and *Bt*-produced proteins must be assessed. To assess the equivalence between MON 88702-produced and *Bt*-produced mCry51Aa2 proteins, a small quantity of the mCry51Aa2 protein was purified from MON 88702 cottonseed with multiple protein enrichment and purification steps and a final affinity chromatography with a mCry51Aa2-spesific monoclonal antibody. The mCry51Aa2 protein was characterized and the equivalence of the physicochemical characteristics and functional activity between the MON 88702-produced mCry51Aa2 and the *Bt*-produced mCry51Aa2 proteins was assessed using a panel of six analytical tests as shown in Table 6 and sections below. Taken together, these data provide a detailed characterization of the mCry51Aa2 protein and establish the equivalence of MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 proteins.

For details, please also refer to [redacted], 2016 (MSL0027791).

Table 6. Summary of MON 88702 mCry51Aa2 Protein Identity and Equivalence

Analytical Test Assessment	Section Cross Reference	Analytical Test Outcome
1. N-terminal sequence analysis of the MON 88702-produced mCry51Aa2 protein to assess the N-terminal sequence	B1(b)	• The N-terminal sequence of mCry51Aa2 proteins was identified
2. MALDI-TOF MS[1] analysis of peptides derived from tryptic digested MON 88702-produced mCry51Aa2 protein to assess identity	B1(c)	• MALDI-TOF MS[1] analysis yielded peptide masses consistent with the peptide masses from the theoretical trypsin digest of the MON 88702-produced mCry51Aa2 sequence
3. Western blot analysis using anti- mCry51Aa2 monoclonal antibodies to assess identity and immunoreactive equivalence between MON 88702-produced mCry51Aa2 and the *Bt*-produced MON 88702 mCry51Aa2 proteins	B1(d)	• MON 88702-produced mCry51Aa2 protein identity was confirmed using a Western blot probed with an antibody specific for mCry51Aa2 proteins • Immunoreactive properties of the MON 88702-produced mCry51Aa2 and the *Bt*-produced mCry51Aa2 proteins were shown to be equivalent
4. SDS-PAGE[2] to assess equivalence of the apparent molecular weight between MON 88702-produced mCry51Aa2 and the *Bt*-produced mCry51Aa2 proteins	B1(e)	• Electrophoretic mobility and apparent molecular weight of the MON 88702-produced mCry51Aa2 and the *Bt*-produced mCry51Aa2 proteins were shown to be equivalent
5. Glycosylation analysis of the MON 88702 mCry51Aa2 protein to assess equivalence between the MON 88702-produced mCry51Aa2 and the *Bt*-produced mCry51Aa2 proteins	B1(f)	• MON 88702-produced mCry51Aa2 and the *Bt*-produced mCry51Aa2 proteins were both shown to not be glycosylated
6. mCry51Aa2activity analysis to assess functional equivalence between the MON 88702-produced mCry51Aa2 and the *Bt*-produced mCry51Aa2 proteins	B1(g)	• Functional activity of the MON 88702-produced mCry51Aa2 and the *Bt*-produced mCry51Aa2 proteins were shown to be equivalent

[1] MALDI-TOF MS = Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry

[2] SDS-PAGE = sodium dodecyl sulfate-polyacrylamide gel electrophoresis

The details of the materials and methods for the panel of analytical tests used to evaluate and compare the properties of the MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 proteins are described in [redacted] 2016 (MSL0027791). A summary of the data obtained to support the characterization of the MON 88702-produced mCry51Aa2 and a conclusion of protein equivalence is below.

B1(b) Results of the N-terminal sequencing analysis

N-terminal sequencing was performed on the MON 88702-produced mCry51Aa2 protein. The expected sequence for the mCry51Aa2 protein deduced from the *mCry51Aa2* gene present in MON 88702 was observed with the exception of the N-terminal methionine, which was not detected. This result is expected, as removal of the N-terminal methionine, catalyzed by methionine aminopeptidase, is a common modification that occurs co-translationally before completion of the nascent protein chain (Giglione and Meinnel, 2001). The data obtained correspond to the deduced mCry51Aa2 protein beginning at amino acid position 2 and to the *Bt*-produced mCry51Aa2 (Figure 14). Hence, the sequence information confirms the identity equivelance of the mCry51Aa2 protein isolated from both MON 88702 cottonseed and *Bt*.

For details, please also refer to [redacted]2016 (MSL0027791).

Amino acid residue # from the N-terminus	→	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Bt produced mCry51Aa2	→	X	A	I	L	D	L	K	S	L	V	L	N	A	I	N	Y
			|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|
Predicted mCry51Aa2 sequence	→	M	A	I	L	D	L	K	S	L	V	L	N	A	I	N	Y
			|	|	|	|	|	|	|	|	|	|	|	|	|	|	|	|
MON 88702 mCry51Aa2	→	X	A	I	L	D	L	K	S	L	V	L	N	A	I	N	Y

Figure 14. N-terminal Sequence of the MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2

The experimental sequence obtained from the MON 88702-produced mCry51Aa2 was compared to the expected sequence deduced from the *mCry51Aa2* gene present in MON 88702. The experimentally determined sequence corresponds to the deduced mCry51Aa2 protein beginning at amino acid position 2. The single letter International Union of Pure and Applied Chemistry - International Union of Biochemistry (IUPAC-IUB) amino acid code is; M, methionine; A, alanine; I, isoleucine; L, leucine; D, aspartic acid: K, Lysine; S, serine; V, valine; N, asparagine; Y, tyrosine. 'X' indicates the residue was not observed. The amino acid sequence of *Bt*-produced mCry51Aa2 was also listed here, and it was determined during the protein characterization.

B1(c) Results of MALDI-TOF Tryptic Mass Map analysis

Peptide mass fingerprint analysis is a standard technique used for confirming the identity of proteins. The identity of the MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 proteins were confirmed by MALDI-TOF MS analysis of peptide fragments produced by the trypsin digestion of this protein. The ability to identify a protein using this method is dependent upon matching a sufficient number of observed tryptic peptide fragment masses with predicted tryptic peptide fragment masses. In general, protein identification made by peptide mapping is considered to be reliable if ≥ 40% of the protein sequence was identified by matching experimental masses observed for the tryptic peptide fragments to the expected masses for the fragments (Biron *et al.*, 2006; Krause *et al.*, 1999).

There were 16 unique peptides identified from MON 88702-produced mCry51Aa2 protein that corresponded to the expected masses (Table 7). The identified masses were used to assemble a peptide map of the MON 88702-produced mCry51Aa2 protein. The experimentally determined coverage of the MON 88702-produced mCry51Aa2 protein was 55% (168 out of 305 amino acids, Figure 15A). This analysis further confirms the identity of MON 88702-produced mCry51Aa2 protein.

There were 9 unique peptides identified from *Bt*-produced mCry51Aa2 protein that corresponded to the expected masses (Table 8) during the protein characterization. The identified masses were used to assemble a peptide map of the *Bt*-produced mCry51Aa2 protein. The experimentally determined coverage of the *Bt*-produced mCry51Aa2 protein was 55% (169 out of 305 amino acids, Figure 15B). This analysis further confirms the identity of *Bt*-produced mCry51Aa2 protein.

The peptide maps confirm the equivalence in protein sequences between MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2.

For details, please also refer to [redacted] 2016 (MSL0027791).

Table 7. Summary of the Tryptic Masses Identified for the MON 88702-produced mCry51Aa2 Using MALDI-TOF MS

Experimental Mass[1]	Calculated Mass[2]	Diff.[3]	Fragment[4]	Sequence[5]
2126.9851	2127.2092	-0.2240	1 -19	AILD…WGPK
1473.8692	1473.7980	0.0712	7 - 19	SLVL…WGPK
1808.9677	1808.8329	0.1348	20 - 36	NNNG…ISEK
1467.8823	1467.7682	0.1142	37 - 49	QIDT…THPR
1942.0335	1941.9280	0.1055	88 - 106	KTTT…EGGK
1813.9540	1813.8330	0.1210	89 - 106	TTTT…EGGK
933.5212	933.4655	0.0557	107 - 114	ISDTLEEK
2260.9258	2261.1791	-0.2533	107 - 128	ISDT…GGGK
1345.8117	1345.7242	0.0875	115 - 128	VSVS…GGGK
4142.1988	4141.9828	0.2160	129 - 166	NSTT…VPPR
4269.7393	4270.0778	-0.3385	129 - 167	NSTT…PPRK
1821.0084	1820.8733	0.1351	260 - 273	FDQY…TIEK
876.5146	876.4566	0.0580	279 - 286	HATLHNGK
1004.6169	1004.5515	0.0654	279 - 287	HATLHNGKK
1842.1272	1841.9921	0.1352	287 - 303	KISI…SPIK
1714.0188	1713.8971	0.1217	288 - 303	ISIN…SPIK

[1] Only experimental masses that matched calculated masses with the smallest differences are listed in the table.

[2] The calculated mass is the relative molecular mass calculated from the matched peptide sequence.

[3] The calculated difference between the experimental mass and the calculated mass.

[4] Fragment numbering is based on the N-terminus of the protein.

[5] For peptide matches greater than nine amino acids in length, the first 4 residues and last 4 residues are show separated by dots (…).

Table 8. Summary of the Tryptic Masses Identified for the *Bt*-produced mCry51Aa2 Using MALDI-TOF MS

Experimental Mass[1]	Calculated Mass[2]	Diff.[3]	Fragment[4]	Sequence[5]
2127.2793	2127.2092	0.0702	1 -19	AILD…WGPK
1473.8527	1473.7980	0.0548	7 - 19	SLVL…WGPK
1808.9051	1808.8329	0.0722	20 - 36	NNNG…ISEK
1467.8180	1467.7682	0.0499	37 - 49	QIDT…THPR
4269.4009	4269.2169	0.1840	50 - 87	LIPH…SFAK
1345.7759	1345.7242	0.0517	115 - 128	VSVS…GGGK
4142.0503	4141.9828	0.0675	129 - 166	NSTT…VPPR
1820.9409	1820.8733	0.0676	260 - 273	FDQY…TIEK
1713.9567	1713.8971	0.0596	288 - 303	ISIN…SPIK

[1] Only experimental masses that matched calculated masses with the smallest differences are listed in the table.

[2] The calculated mass is the relative molecular mass calculated from the matched peptide sequence.

[3] The calculated difference between the experimental mass and the calculated mass.

[4] Fragment numbering is based on the N-terminus of the protein.

[5] For peptide matches greater than nine amino acids in length, the first 4 residues and last 4 residues are show separated by dots (…).

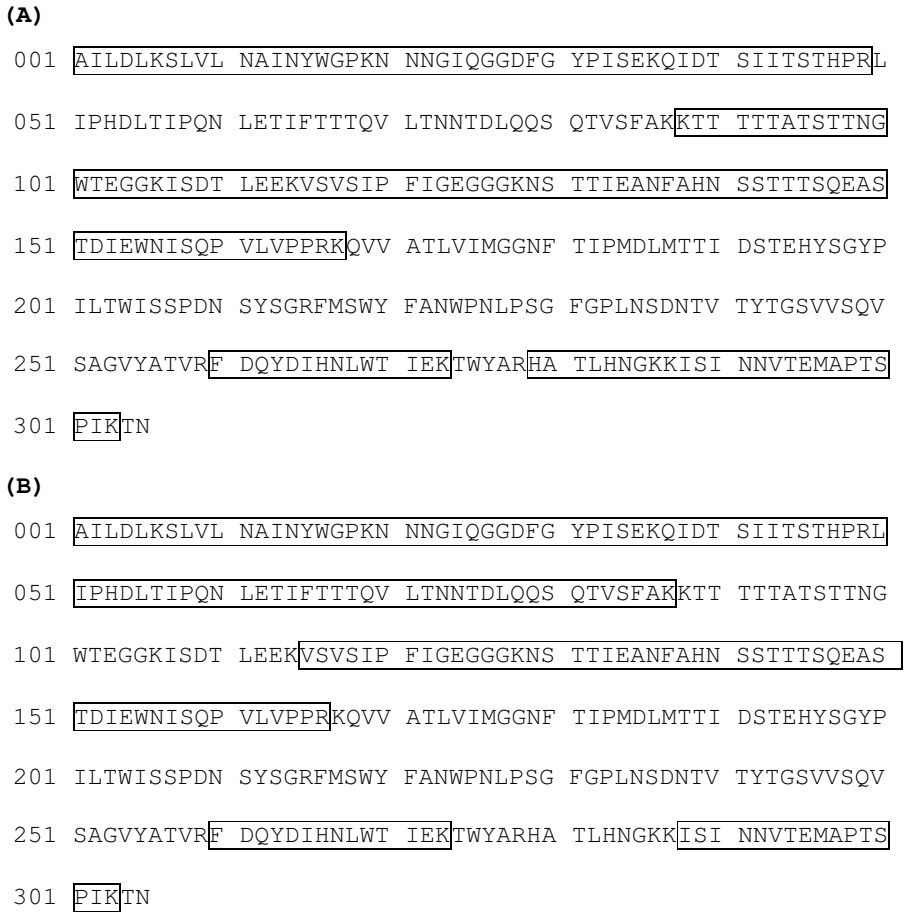

Figure 15. Peptide Maps of the MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2

Figure 15A. The amino acid sequence of the MON 88702-produced mCry51Aa2 protein was deduced from the *mCry51Aa2* gene present in MON 88702. Boxed regions correspond to peptides that were identified from the MON 88702-produced mCry51Aa2 protein sample using MALDI-TOF MS. In total, 55% (168 out of 305 amino acids) of the expected protein sequence was identified.

Figure 15B. The amino acid sequence of the *Bt*-produced mCry51Aa2 protein was deduced from the DNA sequence encoding the mCry51Aa2 protein that is contained in an expression plasmid for the heterologous expression. Boxed regions correspond to peptides that were identified from the *Bt*-produced mCry51Aa2 protein sample using MALDI-TOF MS. In total, 55% (169 out of 305 amino acids) of the expected protein sequence was identified.

Note that both calculations were performed based on the N-terminal methionine being processed (as determined by N-terminal sequence data) away from the mCry51Aa2 protein.

B1(d) Results of western blot analysis of the MON 88702 mCry51Aa2 protein isolated from the grain of MON 88702 and immunoreactivity comparison to *Bt*-produced mCry51Aa2

Western blot analysis was conducted using anti-mCry51Aa2 monoclonal antibody as an additional method to confirm the identity of the mCry51Aa2 protein isolated from the cottonseed of MON 88702, and to assess the equivalence of the immunoreactivity of the MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 proteins.

The results showed that immunoreactive bands migrating with the same electrophoretic mobility were present in all lanes loaded with the MON 88702-produced mCry51Aa2 or *Bt*-produced mCry51Aa2 proteins (Figure 16). For each amount loaded, comparable signal intensity was observed between the MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 protein bands (Figure 16). Increasing the amount of sample loaded onto the gels results in a corresponding increase in the intensity of the immunoreactive band at the correct molecular weight by the anti-mCry51Aa2 monoclonal antibody, supporting the conclusion that the band being detected in MON88702 and *Bt*-produced samples is the same mCry51Aa2 protein.

To compare the immunoreactivity of the MON 88702-produced mCry51Aa2 and the *Bt*-produced mCry51Aa2 proteins, densitometric analysis was conducted on bands that migrated to the expected apparent MW for mCry51Aa2 proteins (~ 34 kDa). The average signal intensity (reported in OD × mm^2, Table 9) of the band of interest in lanes loaded with MON 88702-produced mCry51Aa2 and the *Bt*-produced mCry51Aa2 proteins was reported. Because the mean signal intensity of the MON 88702-produced mCry51Aa2 protein band was within 35% (typical variability for a manually done western blot) of the mean signal of the *Bt*-produced mCry51Aa2 protein (Table 9), the MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 proteins were determined to have equivalent immunoreactivity.

For details, please also refer to [redacted]2016 (MSL0027791).

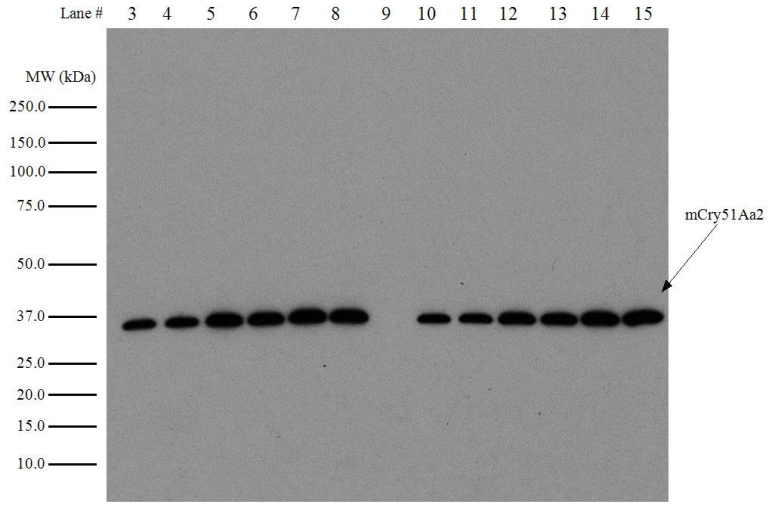

Figure 16. Western Blot Analysis and Immunoreactivity of MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 Proteins

Aliquots of the MON 88702-produced mCry51Aa2 and the *Bt*-produced mCry51Aa2 proteins were subjected to SDS-PAGE and electrotransferred to a nitrocellulose membrane. Proteins were detected using anti-mCry51Aa2 monoclonal antibodies as the primary antibodies. Immunoreactive bands were visualized using HRP-conjugated secondary antibodies and an ECL system. The approximate MWs (kDa) of the standards are shown on the left. The Precision Plus protein standards lane (lane 1) and empty lane (lane 2) were cropped from the image for easier viewing. Three amounts of ng proteins were loaded at ng level. Lane designations are as follows:

Lane	Sample	Load (ng)
1	Precision Plus Protein™ Standards	
2	Blank	
3	*Bt.*-produced mCry51Aa2	1
4	*Bt.*-produced mCry51Aa2	1
5	*Bt.*-produced mCry51Aa2	2
6	*Bt.*-produced mCry51Aa2	2
7	*Bt.*-produced mCry51Aa2	3
8	*Bt.*-produced mCry51Aa2	3
9	Blank	
10	MON 88702-produced mCry51Aa2	1
11	MON 88702-produced mCry51Aa2	1
12	MON 88702-produced mCry51Aa2	2
13	MON 88702-produced mCry51Aa2	2
14	MON 88702-produced mCry51Aa2	3
15	MON 88702-produced mCry51Aa2	3

Table 9. Comparison of Immunoreactive Signals between MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 Proteins

Mean Signal intensity from MON 88702-produced mCry51Aa2 [1] (OD x mm^2)	Mean Signal intensity from *Bt*-produced mCry51Aa2 [1] (OD x mm^2)	Acceptance limits[2] (OD x mm^2)
9.176	9.040	5.876-12.204

[1]Values refer to means of the sum of blank-subtracted density, calculated on n = 6. Values are rounded to three decimal places.

[2]The acceptance limits are based on the interval representing the average ± 35 % of the *Bt*-produced mCry51Aa2 signal intensity at all loads (n=6).

B1(e) Results of mCry51Aa2 protein molecular weight analysis

For apparent MW and purity determination, the MON 88702-produced mCry51Aa2 protein was subjected to SDS-PAGE. Following electrophoresis, the gel was stained with Brilliant Blue G-Colloidal stain and analyzed by densitometry. The MON 88702-produced mCry51Aa2 protein appears to migrate to approximately the same position on the gel as the *Bt*-produced mCry51Aa2 protein. The apparent MW of MON 88702-produced mCry51Aa2 protein was calculated to be 34.2 kDa based on molecular weight standards (Figure 17, Table 10). Because the experimentally determined apparent MW of the MON 88702-produced mCry51Aa2 protein was within the acceptance limits for equivalence (Table 11), the MON 88702-produced and *Bt*-produced mCry51Aa2 proteins were determined to have equivalent apparent molecular weights.

For details, please also refer to [redacted]2016 (MSL0027791).

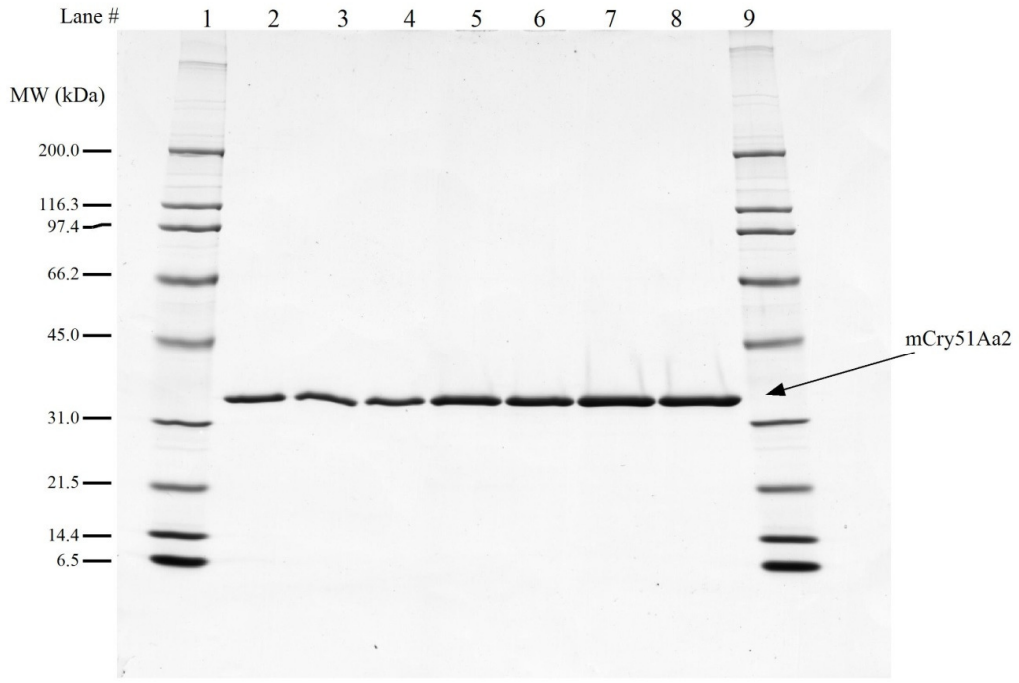

Figure 17. Purity and Apparent Molecular Weight Analysis of the MON 88702-produced mCry51Aa2 Protein

Aliquots of the MON 88702-produced mCry51Aa2 and the *Bt*-produced mCry51Aa2 proteins were subjected to SDS-PAGE. The gel was stained with Brilliant Blue G-Colloidal stain. The MWs (kDa) are shown on the left and correspond to the standards loaded in lanes 1 and 9. Empty lane (lane 10) was cropped from the image. Lane designations are as follows:

Lane	Sample	Amount (µg)
1	Broad Range Standards	~ 5
2	*Bt*-produced mCry51Aa2	1
3	MON 88702-produced mCry51Aa2	1
4	MON 88702-produced mCry51Aa2	1
5	MON 88702-produced mCry51Aa2	2
6	MON 88702-produced mCry51Aa2	2
7	MON 88702-produced mCry51Aa2	3
8	MON 88702-produced mCry51Aa2	3
9	Broad Range Standards	~ 5

Table 10. Apparent Molecular Weight Comparison between the MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 Proteins

Apparent MW of MON 88702-produced mCry51Aa2 Protein (kDa)	*Apparent MW of Bt-produced mCry51Aa2 Protein (kDa)*	*Acceptance Limits[1] (kDa)*
34.2	34.7	33.3 – 36.0

[1]Data obtained for the *Bt*-produced mCry51Aa2 protein was used to generate the prediction interval for setting the acceptance limits.

Table 11. Apparent Molecular Weight and Purity Analysis of the MON 88702-produced mCry51Aa2 Protein

	Apparent MW[1] (kDa)	*Purity[2] (%)*
Average (n=6)	34.2	100

[1]Final MW was rounded to one decimal place.

[2]Average % purity was rounded to the nearest whole number.

B1(f) MON 88702-produced mCry51Aa2 glycosylation analysis

Eukaryotic proteins can be post-translationally modified with carbohydrate moieties (Rademacher *et al.*, 1988). To test whether mCry51Aa2 protein was glycosylated when expressed in the cottonseeds of MON 88702, the MON 88702-produced mCry51Aa2 protein was analyzed using an ECL ™ Glycoprotein Detection Method (GE Healthcare). To assess equivalence of the MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 proteins, the *Bt*-produced mCry51Aa2 protein was also analyzed.

A clear glycosylation signal was observed at the expected size (~ 80 kDa) in the lanes containing the positive control (transferrin), and the band intensity increased with increasing concentration (Figure 18A). In contrast, no glycosylation signal was observed in the lanes containing the MON 88702-produced mCry51Aa2 or *Bt*-produced mCry51Aa2 protein (Figure 17).

To confirm that MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 proteins were appropriately loaded for glycosylation analysis, a second membrane with identical loadings and transfer time was stained with Coomassie Blue R250 for protein detection. Both the MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 proteins were detected (Figure 18B). These data indicate that the glycosylation status of MON 88702-produced mCry51Aa2 protein is equivalent to that of the *Bt*-produced mCry51Aa2 protein and neither protein is glycosylated.

For details, please also refer to [redacted], 2016 (MSL0027791).

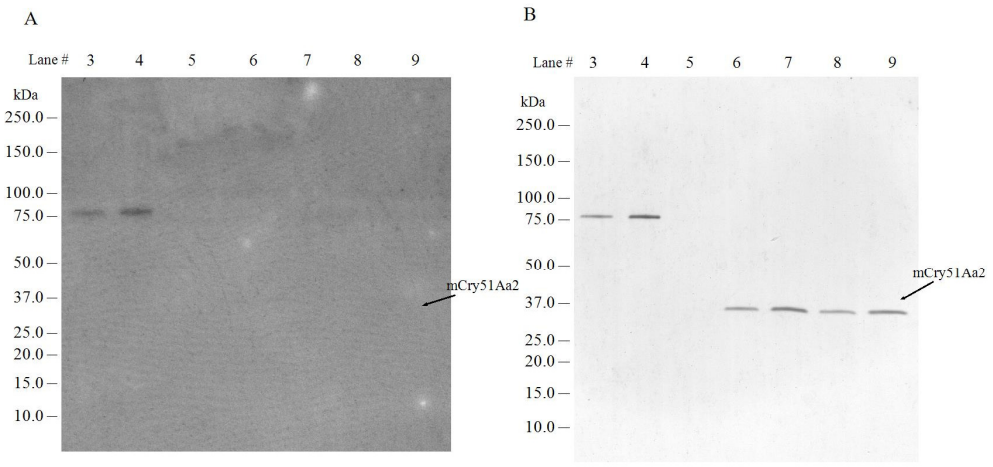

Figure 18. Glycosylation Analysis of the MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 Proteins

Aliquots of the transferrin (positive control), *Bt*-produced mCry51Aa2 and MON 88702-produced mCry51Aa2 were subjected to SDS-PAGE and electrotransferred to a PVDF membrane. The MWs (kDa) correspond to the Precision Plus Protein™ Standards. Lanes loaded with MW standards and blank are cropped. The arrows show the expected migration of the MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 proteins. (A) Where present, the labeled carbohydrate moieties were detected by addition of streptavidin conjugated to HRP followed by a luminol-based detection using ECL reagents and exposure to Hyperfilm®. The 5-minute exposure is shown. An equivalent blot was stained with Coomassie Blue R250 to confirm the presence of proteins (B). The Precision Plus protein standards lane (lane 1) and empty lanes (lanes 2 & 10) and were cropped from both images. Lane designations are as follows:

Lane	Sample	Amount (ng)
1	Precision Plus Protein™ Standards	
2	Blank	
3	Transferrin (positive control)	50
4	Transferrin (positive control)	100
5	Blank	
6	*Bt*-produced mCry51Aa2	50
7	*Bt*-produced mCry51Aa2	100
8	MON 88702-produced mCry51Aa2	50
9	MON 88702-produced mCry51Aa2	100
10	Blank	

B1(g) MON 88702 mCry51Aa2 functional activity

The functional activity of the MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 protein was determined by a Western tarnished plant bug (WTP) diet-incorporation insect bioassay. The MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 proteins were considered functionally equivalent if the LC_{50} of both were within acceptance limits of 1.266 to 3.692 µg mCry51Aa2/ml diet.

The LC_{50} of the MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 proteins were determined to be 2.114 and 1.951 µg mCry51Aa2/ml diet, respectively (Table 12). Because the LC_{50} values of MON 88702-produced and *Bt*-produced mCry51Aa2 proteins were within the acceptance limits calculated from previous studies with mCry51Aa2 protein produced in *Bt* bacteria (Table 12), the proteins were determined to have equivalent functional activity.

For details, please also refer to [redacted], 2016 (MSL0027791).

Table 12. mCry51Aa2 Functional Assay

MON 88702-produced mCry51Aa2[1] ***(µg mCry51Aa2/ml diet)***	***Bt-produced mCry51Aa2[1]*** ***(µg mCry51Aa2/ml diet)***	***Acceptance Limits[2]*** ***(µg mCry51Aa2/ml diet)***
2.114	1.951	1.266 – 3.692

[1]Value refers to mean calculated based on n = 3. Values are rounded to three decimal places.

[2]Data obtained for the *Bt*-produced mCry51Aa2 protein was used to generate a prediction interval for setting the acceptance limits.

B1(h) MON 88702 mCry51Aa2 protein identity and equivalence - Conclusion

The mCry51Aa2 protein purified from cottonseed of MON 88702 was characterized and the equivalence of the physicochemical and functional properties between the MON 88702-produced mCry51Aa2 and the *Bt*-produced mCry51Aa2 proteins was established using a panel of analytical tests: 1) the identity was confirmed by N-terminal sequence analysis; 2) MALDI-TOF MS analysis yielded peptide masses consistent with the expected peptide masses from the theoretical trypsin digest of the mCry51Aa2 sequence; 3) mCry51Aa2 protein was detected on a western blot probed with antibodies specific for mCry51Aa2 protein and the immunoreactive and physiochemical properties of the MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 proteins were shown to be equivalent; 4) the electrophoretic mobility and apparent molecular weight of the MON 88702-produced mCry51Aa2 and *Bt*-produced mCry51Aa2 proteins were shown to be equivalent; 5) glycosylation status of MON 88702-produced mCry51Aa2 and *Bt*-produced Cry51Aa2 proteins were determined to be equivalent; and 6) functional activity of the MON 88701-produced mCry51Aa2 and the *Bt*-produced mCry51Aa2 proteins were demonstrated to be equivalent.

Taken together, these data provide a detailed characterization of the mCry51Aa2 protein in MON 88702 and establish the equivalence of the MON 88702-produced mCry51Aa2 and the *Bt*-produced mCry51Aa2 protein. This equivalence justifies the use of the Bt-produced mCry51Aa2 as a test subtance in the protein safety studies.

B2 Antibiotic Resistance Marker Genes

MON 88702 does not contain genes that encode resistance to antibiotic markers. Molecular characterisation data presented in Section A demonstrate the absence of antibiotic resistance marker gene in MON 88702.

B2(a) Clinical importance of antibiotic that GM is resistant to (if any)

Not applicable.

B2(b) Presence in food of antibiotic resistance protein (if any)

Not applicable.

B2(c) Safety of antibiotic protein

Not applicable.

B2(d) If GM organism is micro-organism, is it viable in final food?

Not applicable.

B3 Characterisation of Novel Proteins or Other Novel Substances

B3(a) Biochemical function and phenotypic effects of novel substances

Description of the mCry51Aa2 Protein and Mode-of-Action

The mCry51Aa2 expressed by MON 88702 was developed through limited modification to the Cry51Aa2 protein identified from *Bt*, and has 96% amino acid sequence identity to wild-type Cry51Aa2 (Baum *et al.*, 2012; Jerga *et al.*, 2016). Analysis of cottonseed extracts from MON 88702 determined that the expressed protein had an apparent molecular weight of 34.2 kDa and corresponded to the mCry51Aa2 protein, composed of 305 amino acids without the first methionine (Figure 3), which was removed during co-translational processing in MON 88702. N-terminal methionine cleavage is common and naturally occurs in the vast majority of proteins (Meinnel and Giglione, 2008).

The insecticidal mode-of-action of *Bt* proteins in general requires insect uptake (ingestion), proteolytic activation (which converts the inactive protoxin form of the *Bt* protein to the active toxin form), receptor binding in the insect midgut, oligomerization at the membrane interface, and membrane pore formation in the midgut cells, which in turn leads to insect death (Pigott and Ellar, 2007; Gill *et al.*, 1992; Schnepf *et al.*, 1998; Vachon *et al.*, 2012; OECD, 2007).

The mode-of-action for mCry51Aa2 has been well assessed, and follows the same general steps as other *Bt* insecticidal proteins currently in commercial use for insect crop protection. Biophysical characterization has confirmed that the full-length mCry51Aa2 is a stable dimer in solution. Activation of mCry51Aa2 occurs through exposure to *Lygus* saliva, which results in proteolytic cleavage at the C-terminal end of each mCry51Aa2 protein in the dimer. This C-terminal proteolytic cleavage results in the removal of amino acids 280 to 306 (see deduced amino acid sequence, Figure 2) and the dissociation of the dimer into two separate monomers. mCry51Aa2 ligand binding immunoblotting assays indicate the activated monomeric form displays binding to a single band of the *Lygus* brush border membrane proteins on a SDS-PAGE, and forms a membrane-associated oligomeric complex both *in vitro* and *in vivo* (Jerga *et al.*, 2016). Immunohistochemistry analysis further demonstrated that upon mCry51Aa2 exposure, midgut epithelium and cellular sloughing occurs (Jerga *et al.*, 2016), which is consistent with observations of other insecticidal *Bt* Cry proteins. Finally, chemical cross-linking of the mCry51Aa2 dimer was shown to render the protein inactive, but still competent to compete for binding sites with the native mCry51Aa2 protein *in vivo*. Thus, disassociation of the mCry51Aa2 dimer into sterically unhindered monomers is required for brush border membrane binding, oligomerization, and the subsequent steps to culminate in insect mortality (Jerga *et al.*, 2016).

B3(a)(i) Description, mode-of-action, and specificity of mCry51Aa2 protein expressed in MON 88702

The insecticidal mode-of-action of *Bt* proteins in general requires insect uptake (ingestion), proteolytic activation (which converts the inactive protoxin form of the *Bt* protein to the active toxin form), receptor binding in the insect midgut, oligomerization at the membrane interface, and membrane pore formation in the midgut cells, which in turn leads to insect death (Pigott and Ellar, 2007; Gill *et al.*, 1992; Schnepf *et al.*, 1998; Vachon *et al.*, 2012; OECD, 2007).

The mode-of-action for mCry51Aa2 has been well assessed, and follows the same general steps as other *Bt* insecticidal proteins currently in commercial use for insect crop protection (see Section B3(a) for further details).

B3(b) Identification of novel substances (e.g. metabolites), levels and site

mCry51Aa2 protein levels in various tissues of MON 88702 relevant to the risk assessment were determined by a validated enzyme-linked immunosorbent assay (ELISA). Tissues of MON 88702 were collected from four replicate plots planted in a randomized complete block field design during the 2015 growing season from five field sites in the United States: Graham County, Arizona, (AZSA); Rapides County, Louisiana (LACH); Washington County, Mississippi (MSLE); Perquimans County, North Carolina (NCBD) and San Patricio County, Texas (TXPO). The field sites were representative of U.S. cotton-producing regions suitable for commercial production. Leaf, root, pollen, and seed tissue samples were collected from each replicated plot at all field sites.

The ELISA results obtained for each sample were averaged across the five sites and are summarized in Table 13. The mCry51Aa2 protein levels in MON 88702 across all samples analyzed from all sites ranged from 2.0 to 1700 µg/g dw. The mean mCry51Aa2 protein level among all tissue types was highest in OSL1 at 1200 µg/g dw and lowest in pollen at 2.6 µg/g dw.

For details, please refer to [redacted] 2016 (MSL0027766).

Table 13. Summary of mCry51Aa2 Protein Levels in Cotton Tissues Collected from MON 88702-produced in United States Field Trials during 2015

Tissue Type[1]	Growth Stage[2]	Mean (SD) Range (µg/g dw)[3]	LOQ/LOD (µg/g dw)[4]
OSL1	2 to 6-Leaf	1200 (380) 550-1700	0.078/0.010
OSL4	Cut out	1000 (160) 700-1300	0.078/0.010
Root	Peak Bloom	190 (41) 150-290	0.078/0.028
Pollen	Peak Bloom	2.6 (0.41) 2.0-2.9	0.078/0.016
Seed	Maturity	130 (17) 91-170	0.078/0.021

[1]OSL = over season leaf
OSR = over season root
[2]The crop development stage at which each tissue was collected.
[3]Protein levels are expressed as the arithmetic mean and standard deviation (SD) as microgram (µg) of protein per gram (g) of tissue on a dry weight basis (dw). The means, SD, and ranges (minimum and maximum values) were calculated for each tissue across all sites (n=20 except in pollen where n=5 as four replicates per site were pooled)
[4]LOQ=limit of quantitation; LOD=limit of detection.

B3(c) Site of expression of all novel substances and levels

Please refer to Section B3(b).

B3(d) Post-translational modifications to the novel protein(s)

Not applicable.

B3(e) Evidence of silencing, if silencing is the method of modification

Not applicable.

B3(f) History of human consumption of novel substances or similarity to substances previously consumed in food

Evaluation of history of safe use (HOSU) is only one aspect of the overall safety assessment, but can help to determine whether an initial presumption for safety can be made for the protein of interest, or if additional testing may be needed (Moar *et al.*, 2017).

The mCry51Aa2 protein shares 96% sequence identity to the Cry51Aa2 wild-type protein identified from *Bt*. As previously described, the mode-of-action of mCry51Aa2 follows the same commonly recognized steps as for other *Bt* insecticidal proteins, including a requirement for insect uptake and activation, specific binding to receptors in the insect midgut, oligomerization, and pore formation in the midgut cell membrane (Jerga *et al.*, 2016).

A history of safe use for *Bt* microbial biopesticide applications has been established through the documented use of these products for over 50 years (Hammond, 2004; OECD, 2010b). There are at least 180 registered microbial *Bt* products in the U.S. and over 120 microbial products approved in the European Union (Hammond, 2004). Applications of sporulated *Bt* have a long history of safe use for pest control in agriculture, especially in organic farming (Cannon, 1993; U.S. EPA, 1988). They have been safely and directly applied to consumed agriculture commodities including berry crops, cabbage, grapes, tomatoes, celery, lettuce, and spinach (U.S. EPA, 1988). Furthermore, extensive toxicity testing of commercial *Bt* microbial biopesticides that contain numerous Cry proteins (e.g., Cry2A, Cry1Aa, Cry1Ab, Cry1Ac, Cry1C, and Cry1F) has resulted in no evidence for adverse effects to human or animal health (McClintock *et al.*, 1995; Koch *et al.*, 2015; Moar *et al.*, 2017; OECD, 2010a). Thus, *Bt* microbial formulations and the insecticidal proteins present in *Bt* microbial biopesticide formulations have been safely consumed by humans and animals for over 50 years.

The mCry51Aa2 protein, and the parent Cry51Aa2 protein, are β-pore forming proteins (β-PFPs) that belong to the Aerolysin-like/ ETX_MTX2 family of proteins. ETX_MTX2 family members are found in a broad range of plant, animal and bacterial species (Moar *et al.*, 2017) that have documented safe exposure to humans and animals (Moar *et al.*, 2017). The molecular steps involved in the mode-of-action (MoA) of β-PFPs are generally known (Narva *et al.*, 2017), and the MoA of the mCry51Aa2 protein has been well-characterized (Jerga *et al.*, 2016). Such β-PFPs are produced by numerous species including bacteria, cnidaria, fungi,

plants, and in mammals and other vertebrates (Iacovache *et al.*, 2008; Iacovache *et al.*, 2016).

A specific example of an Aerolysin-like protein with an established history of safe consumption by humans and animals is the Cry35Ab1 protein, which is expressed in some HERCULEX® (is a registered trademark of Dow AgroSciences LLC), SmartStax®, (is a registered trademark of Monsanto Technology LLC) and AcreMax® (a registered trademark of Pioneer Hi-Bred International, Inc.) biotechnology-derived corn hybrids. These corn varieties have been safely grown on millions of acres annually in the U.S. since 2006 (Moar *et al.*, 2017; U.S. EPA, 2010). Additional sources of Aerolysin-like/ ETX_MTX2 family members include numerous genes, transcripts or proteins that have been identified in food species, including food that are directly consumed by humans such as fish, common crop plants and vegetables (Moar *et al.*, 2017), that display structural and sequence homology to the ETX_MTX2 protein family. Thus, human consumption of ETX_MTX2 proteins in the diet is common, and no adverse effects have been reported.

The global use of biopesticides derived from the *Bt* subspecies *israelensis* (*Bti*) and *Lysinibacillus sphaericus*, which are used for the control mosquito and black flies and express the ETX_MTX2 proteins MTX2 and MTX3 (Hu *et al.*, 2008; Liu *et al.*, 1996; Thanabalu and Porter, 1996), is further evidence for safe human and animal exposure to ETX_MTX2 proteins. No adverse effects to humans or the environment have been documented to result from the use of *Bti* and *L. sphaericus* biopesticide formulations, which are often applied in highly-populated urban areas and aquatic environments, including drinking water reservoirs (Moar *et al.*, 2017; Berry, 2012).

A HOSU can therefore be established for many Aerolysin-like/ ETX_MTX2 protein family members through the extensive human and animal exposure described above. Importantly, key differences, inclulding sequence and structural differences in the receptor binding head region (Moar *et al.*, 2017) distinguish these proteins, including mCry51Aa2, from a limited number of Aerolysin-like/ ETX_MTX2 proteins associated with toxicity in humans and other mammals (e.g., epsilon toxin from *Clostridium perfringens*, for which the protein family partially receives its namesake).

The MoA of mCry51Aa2 and other proteins from the Aerolysin-like/ ETX_MTX2 family, as noted previously and detailed elsewhere (Moar *et al.*, 2017; Jerga *et al.*, 2016) is an important determinant of conserved or divergent sequence and structural domains between family members. Classification within the Aerolysin-like/ ETX_MTX2 protein family is determined primarily by conserved sequence and structural relatedness in the “tail” domain of the proteins, which contains the regions responsible for the pore-formation and oligomerization functions. Pore-formation and oligomerization are required steps in the overall general MoA common to all family members, and are separate from target receptor-binding (Moar *et al.*, 2017; Berry and Crickmore, 2017).

In contrast, significant sequence and structural diversity is observed between the “head” domains of Aerolysin-like/ ETX_MTX2 proteins, which is the region responsible for target receptor binding (Moar *et al.*, 2017). The interaction between the head domain of a given Aerolysin-like/ ETX_MTX2 protein and its target receptor requires involvement of specific

amino acid motifs, and results in a high degree of selectivity and specificity in target binding. Importantly, the diversity in sequence and structure in the head domain determines the diversity in target receptor binding, and is thus a key characteristic relevant for evaluating safety and distinguishing potential to exert human or animal toxicity between family members (Moar *et al.*, 2017). The specific protein-receptor interaction required for Cry protein binding and activity are not present in mammals; thus, Cry proteins have not been observed to have adverse effects on mammals, nor on birds, amphibians, or reptiles (OECD, 2007; Schnepf *et al.*, 1998), and there is a presumption of safety associated with these proteins.

B4 Assessment of Potential Toxicity

The assessment of the potential toxicity of an introduced protein is based on comparing the biochemical characteristics of the introduced protein to characteristics of known toxins. A protein is unlikely to be associated with toxicity if: 1) the protein lacks any structural similarity to known toxins or other biologically-active proteins that could cause adverse effects in humans or animals; and there is negligible chance for exposure to to the protein because: 2) the protein is rapidly degraded by pepsin and pancreatin; and 3) the protein is unstable to heat treatment.

B4(a) Bioinformatic comparison (aa) of novel protein(s) to toxins

Bioinformatic analyses was performed as part of the assessment for potential toxicity of the MON 88702 encoded mCry51Aa2 protein sequence. The translated sequence was compared to the toxin (TOX_2017) and all protein (PRT_2017) databases using bioinformatic tools. The GenBank protein database, release 217, was downloaded from NCBI and was used to create the PRT_2017 database which contains 61,983,812 sequences.

Potential structural similarities shared between the mCry51Aa2 protein with sequences in a protein database were evaluated using the FASTA sequence alignment tool. The FASTA program directly compares amino acid sequences (*i.e.*, primary, linear protein structure) and the alignment data may be used to infer shared higher order structural similarities between two sequences (*i.e.*, secondary and tertiary protein structures). Proteins that share a high degree of similarity throughout the entire sequence often contain some degree of homology. Homologous proteins often have some common secondary structures, common three-dimensional configuration, and, consequently, may share similar functions (Caetano-Anollés *et al.*, 2009; Illergård *et al.*, 2009).

The FASTA sequence alignment tool was used to assess the relatedness between the mCry51Aa2 protein query sequence and any protein sequences in the TOX_2017 database. The toxin database, TOX_2017, is a subset of sequences derived from the PRT_2017 database, that was selected using a keyword search and filtered to remove likely non-toxin proteins and proteins that are not relevant to human or animal health. The TOX_2017 database contains 22,760 sequences.

Similarities shared between the mCry51Aa2 protein sequence with each sequence in the database were examined. The extent of relatedness was evaluated by detailed visual

inspection of the alignment, the calculated percent identity, and the *E*-score. As described above, the *E*-score is a statistical measure of the likelihood that the observed similarity score could have occurred by chance in a search. A larger *E*-score indicates a lower degree of similarity between the query sequence and the sequence from the database. Typically, alignments between two sequences require an *E*-score of 1×10^{-5} or less to be considered to have sufficient sequence similarity to infer some degree of homology and are deemed significant because they may reflect shared structure and function among sequences.

When compared to the TOX_2017 and PRT_2017 databases, mCry51Aa2 resulted in *E*-score threshold exceeding alignments with both the TOX_2017 and PRT_2017 databases. When compared to the PRT_2017 database, mCry51Aa2 was positively identified through alignment with Cry51Aa1 that displayed 97.7% identity over the full length of mCry51Aa2. Moreover, inspection of the *E*-score threshold exceeding alignments with the TOX_2017 and PRT_2017 databases confirmed that mCry51Aa2 is a β-pore forming protein. Inspection of the GenBank records of these threshold-exceeding aligning proteins revealed that β-pore forming proteins share a broad range of descriptive synonyms for structure or function including: Parasporal (parasporin) crystal protein, crystal protein, pesticidal protein, insecticidal protein, epsilon toxin (ETX), mosquitocidal toxin (MTX or MTX2), natterin-like protein, aerolysin, hemolysin and leucocidin.

When the results of the bioinformatic analysis are viewed in there entirety, the MON 88702 encoded mCry51Aa2 protein is not expected to possess functional cross-reactivity with known allergenic proteins, or be toxic, or display adverse biological activity in organisms other than the intended hemipteran and thysanopteran insect target pests.

Assessment of Potential Toxicity

A single FASTA alignment that yielded *E*-score of $9.2e^{-06}$, less than the *E*-score threshold of $1e^{-5}$, was obtained with TOX_2017 database entry GI-1102943401. Inspection of the GenBank record for GI-1102943401 shows this database entry is the product of an automated genome sequencing and assembly for a specific bacterial strain, *Francisella* sp. CA97-1460. Based upon the automated nature of how these results are obtained and annotated, bioinformatics was used to assign the structure and function identity, “aerolysin toxin family protein”, to GI-1102943401. While contained in a GenBank record, no additional information related to the existence or toxicity of GI-1102943401 has been published. Therefore, this alignment that exceeds the established threshold does not impact the safety of the mCry51Aa2 protein, as determined through the weight of evidence evaluation, and futher analysis of this alignment indicates that it has homology mainly in domains common to this family of proteins.

Assessment of Potential Adverse Biological Activity

The FASTA search against the PRT_2017 database resulted in 222 alignments displaying an *E*-score of $\leq1e^{-5}$. The top alignment was with GI-1115466244, a patent sequence whose alignment displays 100% identity over the full length, 306 amino acids, of mCry51Aa2 and yields an *E*-score of $1.3e^{-134}$. The first named non-patent database entry resulting in an alignment was with GI-112253719 a Cry51Aa1 protein that yielded the 85th best alignment

having an *E*-score of $7e^{-131}$. This alignment positively confirmed the identity of the query as being a member of the Cry51Aa family, as it displayed 97.7% identity across 303 of 306 amino acids of mCry51Aa2. mCry51Aa2 is a β-pore forming toxin (Moar *et al.*, 2017) and member of the ETX/MTX2 family of proteins which includes proteins with a broad range of descriptive synonyms for structure or function including: Parasporal (parasporin) crystal protein, crystal protein, pesticidal protein, insecticidal protein, epsilon toxin (ETX), mosquitocidal toxin (MTX or MTX2), natterin-like protein, aerolysin, hemolysin and leucocidin. Although not exceeding the threshold for a search of the PRT_2017 database, GI-1102943401, which was identified in the search of the TOX_2017 database, resulted in the 256th of 305 alignments. It yielded an alignment displaying 28.6% identity in 245 of 306 amino acids with an *E*-score of 0.033. Given the positive identification of mCry51Aa2, 97.7% identity across 303 of 306 amino acids with Cry51Aa1, there is no data contained in the results of the FASTA search of PRT_2017 that would indicate potential toxicity towards organisms other than the intended hemipteran and thysanopteran target insect pests.

Further Evaluation of the Alignment with GI-1102943401

Whole genome sequencing bioinformatic methods used to assign structure, function, and naming to entries such as GI-1102943401 involve an automated pipeline that includes high throughput sequencing, genome assembly and openreading frame identification. In this context and depending upon the methodology that is used, a range of similar terms are frequently used to describe entries for β-pore forming proteins, which includes description of GI-1102943401 as being an "aerolysin toxin family protein". Classification within this protein family is based on observation of conserved amino acid sequence in specific regions of the protein, which are shared across family members, including known toxins, as well as proteins with an extensive history of safe use and consumption (Moar *et al.*, 2017).

Thus, it is not surprising that mCry51Aa2 yielded an alignment with a protein described as an "aerolysin toxin family protein", as mCry51Aa2, being an ETX/MTX2 protein, is also classified within this larger family. Although exceeding the *E*-score threshold for a search of the TOX_2017 database, the alignment itself was far less significant than the alignment with Cry51Aa from PRT_2017, as it included only 245 of 306 amino acids.

Further, upon closer inspection of the alignment from the domain-based perspective (Moar *et al.*, 2017) revealed that the similarity between GI-1102943401 and mCry51Aa2 is observed primarily within the portions of the proteins corresponding to the oligomerization (aka "tail") and pore forming domains (aka "loop") in mCry51Aa2, which are responsible for the shared pore-formation function across family members.

Most importantly, in the regions of the proteins most relevant for assessing safety, the receptor-binding domain (aka "head") of mCry51Aa2, only three of 48 amino acids (two similar, one identical) in head domain 1 of mCry51Aa2 align with the corresponding 96 amino acid region of GI-1102943401. When viewed in the context of domains, it is clear that mCry51Aa2 and GI-1102943401 should both be classified as aerolysin-like ETX/MTX2 β-pore forming proteins. And while they exhibit a small degree of expected sequence conservation in the regions of the proteins required for oligomerization and pore-formation, they display lack of similarity in the receptor-binding domain, which is key for target

receptor identity within the overall mode-of-action (Jerga *et al.*, 2016), and thus specificity and safety.

To further understand the relationship between mCry51Aa2 and GI-1102943401, GI-1102943401 sequence was used as a query to search the PRT_2017 database. In addition to self-identifying, GI-1102943401 yielded 19 additional alignments with *E*-scores of $\leq 1e^{-5}$. Inspection of the GenBank records for each of the 19 aligning sequences revealed that 16 of the records contained the words pesticidal, larvicidal, insect, or insecticidal. None of the 19 records made mention of, or were sequences from the pathogenic bacteria *Aeromonas hydrophila*, the source of aerolysin protein. Moreover, when GenBank records for all 290 alignments having an *E*-score of ≤1, which is an even more conservative threshold meant to capture even distantly related proteins, were inspected, none of the alignments were with sequences from *Aeromonas hydrophila* and other than the self identified GI-1102943401, only one other record contained the word "aerolysin". When combined, the results of the search of PRT_2017 using GI-1102943401 as the query revealed that the description "aerolysin" was used as a synonym rather than an exact description for GI-1102943401. Based upon the observation that 16 of the top 19 sequences aligning with GI-1102943401 were annotated with words such as pesticidal, larvicidal, insect or insecticidal, GI-1102943401 could have just as easily been described as a β-pore forming protein or by using one of the other synonyms such as ETX/MTX protein or the other synonyms listed in section above (Assessment of Potential Adverse Biological Activity).

Conclusions for the Structural Similarities of mCry51Aa2 to Known Toxins

The results of these analyses indicate that no relevant sequence similarities are observed between the MON 88702 encoded mCry51Aa2 protein and allergens contained in AD_2017. When used as a query to search TOX_2017, mCry51Aa2 resulted in a single alignment more significant (lower numerical value) than the *E*-score threshold of $\leq 1e^{-5}$, GenBank record GI-1102943401, an "aerolysin toxin family protein".

When this alignment was evaluated using a domain-based approach, the annotation of GI-1102943401 more closely reflected the use of "aerolysin" as a synonym rather than an exact description, and GI-1102943401 could have just as easily been described as being a β-pore forming protein or by using one of the other synonyms such as ETX/MTX protein. Further, alignment identities between mCry51Aa2 and GI-1102943401 were ~28%, with similarity observed primarily within the conserved regions of ETX/MTX protein family members associated with oligomerization and pore-formation functions (Moar *et al.*, 2017).

Therefore, when viewed as a whole, the bioinformatic evaluation of mCry51Aa2 did not provide any indication of potential allergenicity or toxicity. Ultimately, the alignment of mCry51Aa2 with GI-1102943401 did not provide any information to indicate mCry51Aa2 to be toxic, or display adverse biological activity toxicity towards organisms other than the intended hemipteran and thysanopteran insect pests.

For details, please also refer to [redacted] 2017 (MSL0028423).

B4(b) Stability to heat or processing and/or degradation in gastric model

B4(b)(i) Digestive fate of the mCry51Aa2 protein

Proteins introduced into crops using biotechnology are evaluated for their safety for human and animal consumption. The majority of ingested dietary proteins undergo hydrolytic degradation and/or proteolytic degradation to their constituent amino acids or small peptides, which are then absorbed and used for synthesis of proteins or other glucogenic or ketogenic metabolites by the body (Delaney *et al.*, 2008). Therefore, evaluating a protein's intrinsic sensitivity to proteolytic degradation with enzymes of the gastrointestinal tract is a key aspect to understanding the safety of any introduced proteins in GM crops. One characteristic of protein toxins and many allergens is their ability to withstand proteolytic degradation by enzymes present in the gastrointestinal tract (Astwood *et al.*, 1996; Moreno *et al.*, 2005; Vassilopoulou *et al.*, 2006; Vieths *et al.*, 1999). Allergenic proteins or their fragments, when presented to the intestinal immune system, can lead to a variety of gastrointestinal and systemic manifestations of immune-mediated allergy. The complete enzymatic degradation of an ingested protein by exposure to gastric pepsin and intestinal pancreatic proteases makes it highly unlikely that either the intact protein or protein fragment(s) will reach the absorptive epithelial cells of the small intestine where antigen processing cells reside (Moreno *et al.*, 2005). To reach these cells, protein or protein fragment(s) must first pass through the stomach where they are exposed to pepsin and then the duodenum where they are exposed to pancreatic fluid containing a mixture of enzymes called pancreatin. Therefore, the susceptibility of the mCry51Aa2 protein to degradation by pepsin and pancreatin was assessed.

A correlation between the resistance to protein degradation by pepsin and the likelihood of the protein being an allergen has been previously assessed with a group of proteins consisting of both allergens and non-allergens (Astwood *et al.*, 1996; Codex Alimentarius, 2009), but this correlation is not absolute (Fu *et al.*, 2002). A standardized protocol to compare the relative resistance of proteins to degradation by pepsin has been established based on results obtained from an international, multi-laboratory study (Thomas *et al.*, 2004). The multi-laboratory study showed that the results of *in vitro* pepsin degradation assays were reproducible when a standard protocol was followed. Using this standardized *in vitro* pepsin degradation protocol, the susceptibility of the mCry51Aa2 protein to pepsin degradation was assessed.

Incubation of test protein with pancreatin is also used to assess the susceptibility of the protein to proteolytic degradation (Yagami *et al.*, 2000; Okunuki *et al.*, 2002). The relationship between protein allergenicity and susceptibility to pancreatin degradation is limited for several reasons. Namely, the protein has not been first exposed to the acidic, and proteolytic denaturing condition of the stomach, as would be the case *in vivo* (Helm, 2001). Using an established protocol, the susceptibility of the mCry51Aa2 to pancreatin degradation was assessed.

Assessing the Susceptibility of mCry51Aa2 to Pepsin Degradation

The degradation of mCry51Aa2 by pepsin was assessed using two methods: visual analysis of a Brilliant Blue G-Colloidal stained SDS-PAGE gel and visual analysis of a western blot probed with an anti-mCry51Aa2 polyclonal antibody. For this assessment, a separate SDS-PAGE gel containing dilutions of the pre-reaction test substance was run concurrently to estimate the limit of detection (LOD) of the *Bt*-produced mCry51Aa2 protein.

Degradation of mCry51Aa2 by pepsin in solution was evaluated over time by analyzing reaction mixtures incubated for targeted time intervals. For SDS-PAGE analysis, approximately 1 µg of total protein was analyzed for each time point (Figure 19). The 0 min No Test Protein Control and 60 min No Test Protein Control (Figure 19A, lanes 2 and 13), demonstrated that pepsin stable throughout the experimental phase.

No change in the mCry51Aa2 protein band intensity was observed in the absence of pepsin in the 0 minutes No Pepsin Control and 60 minutes No Pepsin Control (Figure 19A, lanes 3 and 12). This indicates that the degradation of the mCry51Aa2 protein was due to the proteolytic activity of pepsin and not due to instability of the protein while incubated in the assay buffer for 60 minutes.

Visual examination of SDS-PAGE data showed that intact mCry51Aa2 was completely degrade within 0.5 min of incubation in the presence of pepsin (Figure 19A, lane 5). For the SDS-PAGE analysis, the LOD of the mCry51Aa2 protein was visually estimated to be approximately 6.3 ng (Figure 19B, lane 7). This LOD used to calculate the maximum amount of intact mCry51Aa2 protein that could remain visually undetected after degradation, which corresponded to approximately 0.6% of the total protein loaded. Based on that LOD, more than 99.6% (100% - 0.6% = 99.4%) of the intact mCry51Aa2 protein was degraded within 0.5 min of incubation in the presence of pepsin. A peptide fragment of ~31 kDa was observed at the 0.5 min time point but was not observed at 2 min (Figure 19A, lanes 5), and a peptide fragment of ~4 kDa at 0.5 to 10 min time points but was not observed at the 20 min time point (Figure 19A, lanes 5-8), which is likely a result of a partially digested product.

For details, please also refer to [redacted]2016 (MSL0027977).

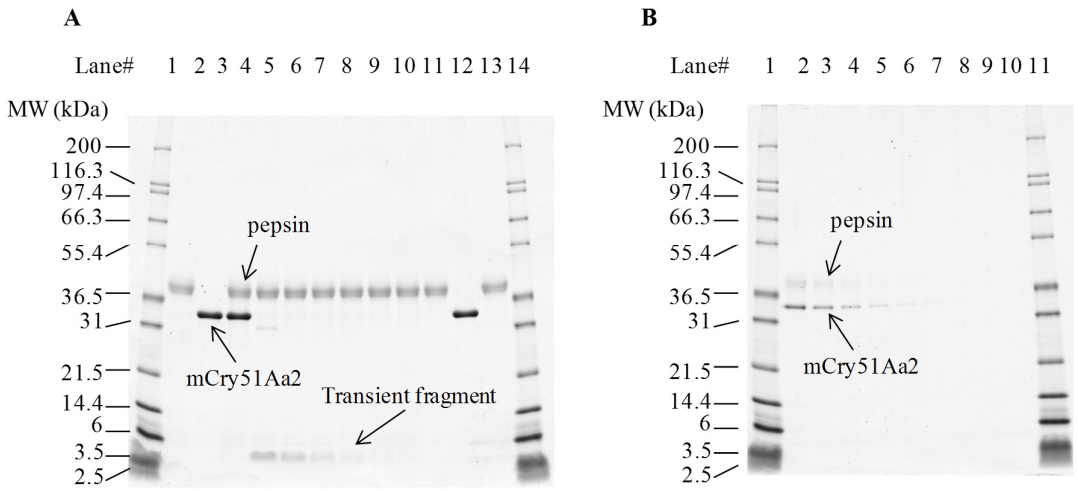

Figure 19. SDS-PAGE Analysis of the Susceptibility of mCry51Aa2 Protein to Pepsin Degradation

Colloidal Brilliant Blue G stained SDS-PAGE gels were used to assess the degradation of mCry51Aa2 protein by pepsin. Molecular weights (kDa) are shown on the left of each gel, and correspond to the markers loaded. In each gel, mCry51Aa2 protein migrated to approximately 34 kDa and pepsin to approximately 38 kDa. Blank lanes were cropped from the images.

A: mCry51Aa2 protein degradation in the presence of pepsin. Based on pre-reaction protein concentrations, 1 µg of mCry51Aa2 protein was loaded in each lane.

B: LOD determination. Indicated amounts of the mCry51Aa2 protein from the Pepsin Treated T0 sample were loaded to estimate the LOD of the mCry51Aa2 protein.

A

Lane	Sample	Incubation Time (min)
1	Mark 12 MWM	-
2	0 min No Test Protein Control	0
3	0 min No Pepsin Control	0
4	Pepsin Treated T0	0
5	Pepsin Treated T1	0.5
6	Pepsin Treated T2	2
7	Pepsin Treated T3	5
8	Pepsin Treated T4	10
9	Pepsin Treated T5	20
10	Pepsin Treated T6	30
11	Pepsin Treated T7	60
12	60 min No Pepsin Control	60
13	60 min No Test Protein Control	60
14	Mark 12 MWM	-
15	Blank	-

B

Lane	Sample	Amount (ng)
1	Mark 12 MWM	-
2	Pepsin Treated T0	200
3	Pepsin Treated T0	100
4	Pepsin Treated T0	50
5	Pepsin Treated T0	25
6	Pepsin Treated T0	12.5
7	Pepsin Treated T0	6.3
8	Pepsin Treated T0	3.1
9	Pepsin Treated T0	1.6
10	Pepsin Treated T0	0.8
11	Mark 12 MWM	-
12	Blank	-
13	Blank	-
14	Blank	-
15	Blank	-

Two blots were run concurrently, one used for the western blot analysis of the susceptibility of mCry51Aa2 to degradation by pepsin (Figure 20A) and another used to estimate the LOD of the mCry51Aa2 protein (Figure 20B). Twenty ng of total protein was analyzed by western blot for each time point. No immunoreactive bands were observed in controls, 0 min No Test Protein Control and 60 min No Test Protein Control (Figure 20A, Lanes 2 and 13). This result indicates that non-specific interactions between the test system components and the mCry51Aa2-specific antibody did not occur under these experimental conditions.

No change in the intact mCry51Aa2 protein band intensity was observed in the absence of pepsin in the 0 minutes No Pepsin Control and 60 minutes No Pepsin Control (Figure 20A, lanes 3 and 12). This indicates that the degradation of the mCry51Aa2 protein was due to the proteolytic activity of pepsin and not due to instability of the protein while incubated in the assay buffer for 60 minutes.

Western blot analysis demonstrated that the intact mCry51Aa2 protein was degraded below the LOD within 0.5 min of incubation in the presence of pepsin (Figure 20A, Lane 5). The LOD of the mCry51Aa2 protein was visually estimated to be approximately 0.31 ng (Figure 20B, Lane 8). The LOD estimated for the mCry51Aa2 protein was used to calculate the maximum amount of mCry51Aa2 protein that could remain visually undetected after degradation, which corresponded to approximately 1.6% of the total protein loaded. Based on the western blot LOD for the mCry51Aa2 protein, the conclusion was that more than 98% (100% - 1.6% = 98.4%) of the intact mCry51Aa2 protein was degraded within 0.5 min. A immunoreactive peptide fragment at ~31 kDa was deteected by the Western blot analysis at 0.5 min time points but was not observed at 2 min.

In summary, the results from western blot analysis demonstrate that greater than 98% of the mCry51Aa2 protein was degraded in the presence of pepsin within 0.5 min and an immunoreactive band at ~31 kDa was detected at 0.5 min time points but was not observed at 2 min.

For details, please also refer to [redacted] 2016 (MSL0027977).

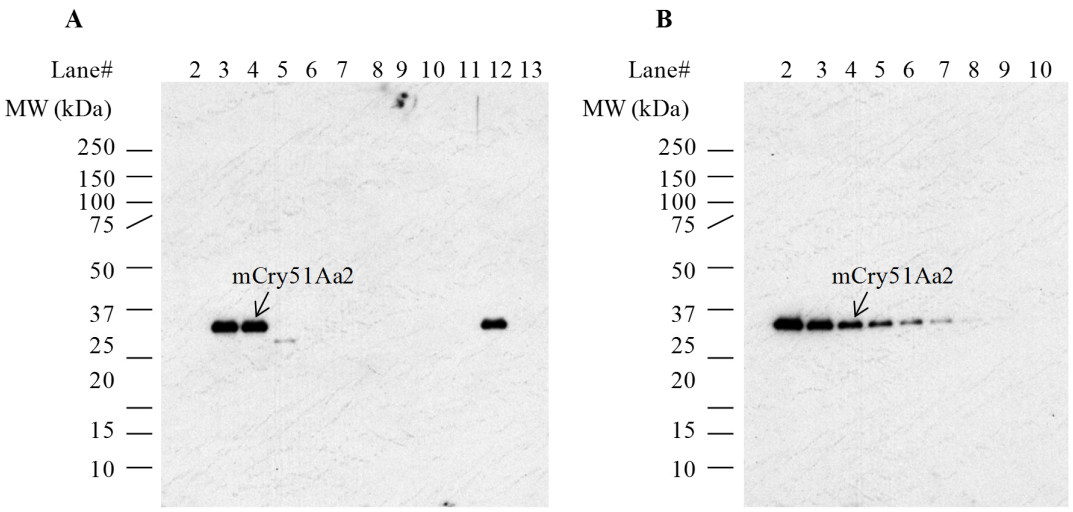

Figure 20. Western Blot Analysis of the Susceptibility of mCry51Aa2 Protein to Pepsin Degradation

Western blots probed with an anti-Cry51 antibody were used to assess the degradation of mCry51Aa2 by pepsin. Molecular weights (kDa) are shown on the left of each gel, and correspond to the markers loaded (cropped from images). Blank lanes were cropped from the images. A 10 sec exposure is shown.

A: mCry51Aa2 protein degradation by pepsin. Based on pre-reaction protein concentrations, 20 ng of mCry51Aa2 protein was loaded in each lane.

B: LOD determination. Indicated amounts of the test protein from the Pepsin Treated T0 sample were loaded to estimate the LOD of the mCry51Aa2 protein.

A

Lane	Sample	Incubation Time (min)
1	Precision Plus MWM	-
2	0 min No Test Protein Control	0
3	0 min No Pepsin Control	0
4	Pepsin Treated T0	0
5	Pepsin Treated T1	0.5
6	Pepsin Treated T2	2
7	Pepsin Treated T3	5
8	Pepsin Treated T4	10
9	Pepsin Treated T5	20
10	Pepsin Treated T6	30
11	Pepsin Treated T7	60
12	60 min No Pepsin Control	60
13	60 min No Test Protein Control	60
14	Precision Plus MWM	-
15	Blank	-

B

Lane	Sample	Amount (ng)
1	Precision Plus MWM	-
2	Pepsin Treated T0	20
3	Pepsin Treated T0	10
4	Pepsin Treated T0	5
5	Pepsin Treated T0	2.5
6	Pepsin Treated T0	1.25
7	Pepsin Treated T0	0.63
8	Pepsin Treated T0	0.31
9	Pepsin Treated T0	0.16
10	Pepsin Treated T0	0.08
11	Precision Plus MWM	-
12	Blank	-
13	Blank	-
14	Blank	-
15	Blank	-

Susceptibility of mCry51Aa2 to the Pancreatin Degradation

The degradation of the mCry51Aa2 protein by pancreatin was assessed by western blot (Figure 21). The western blot used to assess the *in vitro* degradation of the mCry51Aa2 protein by pancreatin (Figure 21A) was run concurrently with a western blot used to estimate the LOD of the intact mCry51Aa2 protein (Figure 21B) in this assay. The gel used to assess the degradation of the mCry51Aa2 protein by pancreatin by Western blot was loaded with 20 ng total protein (based on pre-reaction protein concentrations) for each of the incubation time points.

No immunoreactive bands were observed on the western blot in the 0 min No Test Protein Control and 24 hr No Test Protein Control, which represent the pancreatin test system without mCry51Aa2 protein (Figure 21A, Lanes 2 and 14). This result demonstrates the absence of non-specific antibody interactions with the pancreatin test system.

No change in mCry51Aa2 protein band intensity was observed in the controls, 0 min No Pancreatin Control and 24 hr No Pancreatin Control (Figure 21A, Lanes 3 and 13), which represent the test system without pancreatin. This result reaffirms that mCry51Aa2 was stable in the test system without pancreatin over the course of the experiment.

Western blot analysis demonstrated that a band that migrated to a position just below that intact mCry51Aa2 protein was present during incubation of pancreatin (Figure 21A, Lane 5-12). The LOD was visually estimated to be 0.63 ng (Figure 21B, Lane 7). The mCry51Aa2 had commonly observed characteristics that are present in many other Cry proteins, such as Cry3Bb1 and Cry1Ac, as part of activation step by trypsin prior to oligomerization and pore formation. Therefore, it is not unexpected that the protein may not be digested by the enzymes in pancreatin or digested to its active form similar to what occurs in the insect.

For details, please also refer to [redacted] 2016 (MSL0027977).

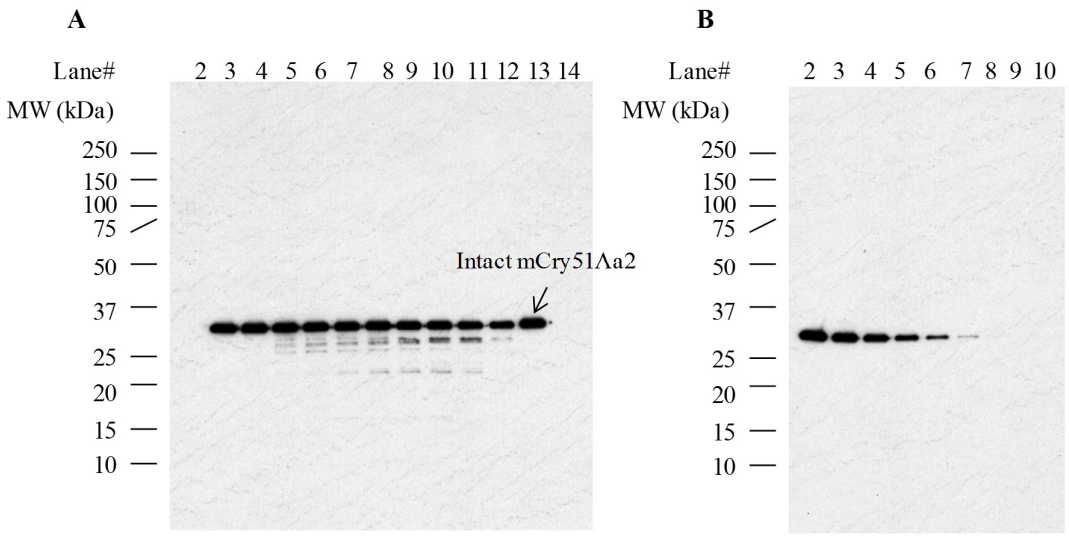

Figure 21. Western Blot Analysis of the Susceptibility of mCry51Aa2 Protein to Pancreatin Degradation

Western blots probed with an anti-Cry51 antibody were used to assess the degradation of mCry51Aa2 by pancreatin. Molecular weights (kDa) are shown on the left of each gel, and correspond to the markers loaded (cropped from images). Blank lanes were cropped from the images. A 10 sec exposure is shown.

A: mCry51Aa2 protein degradation by pancreatin. Based on pre-reaction protein concentrations, 20 ng of mCry51Aa2 protein was loaded in each lane.

B: LOD determination. Indicated amounts of the test protein from the Pancreatin Treated T0 sample were loaded to estimate the LOD of the mCry51Aa2 protein.

A

Lane	Sample	Incubation Time
1	Precision Plus MWM	-
2	0 min No Test Protein Control	0
3	0 min No Pancreatin Control	0
4	Pancreatin Treated T0	0
5	Pancreatin Treated T1	5 min
6	Pancreatin Treated T2	15 min
7	Pancreatin Treated T3	30 min
8	Pancreatin Treated T4	1 hr
9	Pancreatin Treated T5	2 hr
10	Pancreatin Treated T6	4 hr
11	Pancreatin Treated T7	8 hr
12	Pancreatin Treated T8	24 hr
13	24 hr No Pancreatin Control	24 hr
14	24 hr No Test Protein Control	24 hr
15	Precision Plus MWM	-

B

Lane	Sample	Amount (ng)
1	Precision Plus MWM	-
2	Pancreatin Treated T0	20
3	Pancreatin Treated T0	10
4	Pancreatin Treated T0	5
5	Pancreatin Treated T0	2.5
6	Pancreatin Treated T0	1.25
7	Pancreatin Treated T0	0.63
8	Pancreatin Treated T0	0.31
9	Pancreatin Treated T0	0.16
10	Pancreatin Treated T0	0.08
11	Precision Plus MWM	-
12	Blank	-
13	Blank	-
14	Blank	-
15	Blank	-

Degradation of mCry51Aa2 by Sequential Digestion

To further evaluate the susceptibility of the mCry51Aa2 protein by gastrointestinal digestion enzymes, the mCry51Aa2 protein was also assessed by sequential digestion, which consisted with a pepsin digestion for 2 min, followed by pancreatin digestion. This sequential degradation was assessed both by visual analysis of a Colloidal Brilliant Blue G stained SDS-PAGE gel, and visual analysis of a western blot probed with an anti-mCry51Aa2 polyclonal antibody.

Degradation of mCry51Aa2 by pepsin followed by pancreatin in solution was evaluated over time by analyzing reaction mixtures incubated for different time intervals. Approximately 1 µg of total protein was analyzed for each time point using SDS-PAGE Colloidal Brilliant Blue G staining (Figure 22A). Examination of SDS-PAGE data showed that the intact mCry51Aa2 protein was digested within 2 min of incubation in pepsin (Figure 22A, lane 3) as was observed previously (Figure 19). The small transient fragment of ~4 kDa with light intensity (also observed in Figure 19) was completely digested within 0.5 min of incubation in pancreatin (Figure 22A, lane 7). No change in the ~5 kDa fragment band intensity was observed in the absence of pancreatin in the 0 min No Pancreatin Control and 2 hr No Pancreatin Control (Figure 22A, lanes 5 and 14). This indicates that the digestion of the ~4 kDa fragment was due to the proteolytic activity of pancreatin present in pancreatin and not due to instability of the fragment while incubated at pH~7.5 at 37 ± 2ºC for 2 hr.

The SEQ 0 min No Test Protein Control and SEQ 2 hr No Test Protein Control (Figure 22A, lanes 4 and 15) demonstrated the integrity of the pancreatin over the course of the experiment. The intensity of some pancreatin bands decreased somewhat during the course of the experiment, most likely due to auto-digestion. This is not expected to adversely impact the pancreatin degradation results, as the ~4 kDa transiently stable fragment was digested within 0.5 minutes of exposure to pancreatin.

The sequential digestion of the mCry51Aa2 protein was also assessed by western blot (Figure 22B), with 16 ng of the test protein (based on total protein pre-digestion concentrations) loaded per lane. No bands were detected in the 2 min Pepsin Treated sample (Figure 22B, lane 3).

For details, please also refer to [redacted] 2016 (MSL0027977).

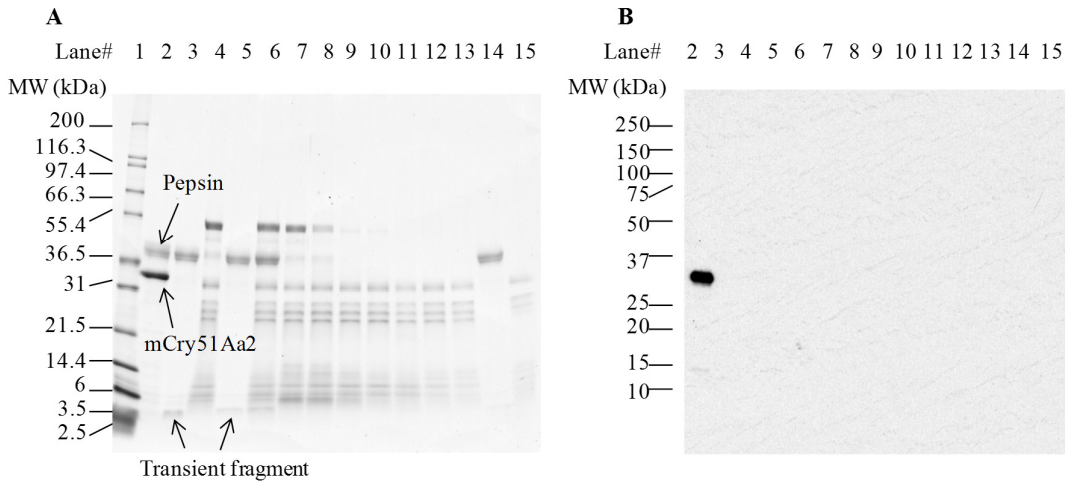

Figure 22. SDS-PAGE and Western Blot Analysis of the Degradation of mCry51Aa2 Protein by Sequential Digestion

SDS-PAGE and western blot analysis were used to assess the degradation of mCry51Aa2 in sequential digestion. Molecular weights (kDa) are shown on the left of each gel, and correspond to the markers loaded (cropped from image B).

A: Colloidal Brilliant Blue G stained SDS-PAGE gel analysis of mCry51Aa2 in sequential digestion. Based on pre-digestion protein concentrations, 1 µg of test protein was loaded in each lane containing mCry51Aa2 protein.

B: Western blot analysis of mCry51Aa2 in sequential digestion. Based on pre-digestion protein concentrations, 16 ng of test protein was loaded in each lane containing mCry51Aa2 protein. A 10 sec exposure is shown.

A

Lane	Sample	Incubation Time
1	Mark 12	-
	Pepsin Degradation	
2	0 min Pepsin Treated	0 min
3	2 min Pepsin Treated	2 min
	Pancreatin Degradation	
4	SEQ 0 min No Test Protein Control	0 min
5	SEQ 0 min No Pancreatin Control	0 min
6	SEQ T0	0 min
7	SEQ T1	0.5 min
8	SEQ T2	2 min
9	SEQ T3	5 min
10	SEQ T4	10 min
11	SEQ T5	30 min
12	SEQ T6	1 hr
13	SEQ T7	2 hr
14	SEQ 2 hr No Pancreatin Control	2 hr
15	SEQ 2 hr No Test Protein Control	2 hr

B

Lane	Sample	Incubation Time
1	Precision Plus MWM	-
	Pepsin Degradation	
2	0 min Pepsin Treated	0 min
3	2 min Pepsin Treated	2 min
	Pancreatin Degradation	
4	SEQ 0 min No Test Protein Control	0 min
5	SEQ 0 min No Pancreatin Control	0 min
6	SEQ T0	0 min
7	SEQ T1	0.5 min
8	SEQ T2	2 min
9	SEQ T3	5 min
10	SEQ T4	10 min
11	SEQ T5	30 min
12	SEQ T6	1 hr
13	SEQ T7	2 hr
14	SEQ 2 hr No Pancreatin Control	2 hr
15	SEQ 2 hr No Test Protein Control	2 hr

Conclusions on the Susceptibility of mCry51Aa2 to Digestive Enzymes

Suceptability of the mCry51Aa2 protein to degradation by pepsin and pancreatin was evaluated. The results of these studies demonstate that greater than 99% of the mCry51Aa2 protein was degraded within 0.5 minutes in the presence of pepsin. Two fragments were observed with one being degarded within 2 minutes and the other in 20 minutes. The latter was not stable for more that 0.5 minutes when a sequential digestion method, pepsin followed by pancreatin, was utilized. The Cry51Aa2 protein was largely stable to pancreatin which is common in many types of Cry proteins.

The results from this assessment demonstrate that the mCry51Aa2 is rapidly degraded in pepsin which would support the conclusion that the mCry51Aa2 protein is highly unlikely to pose a safety concern to human and animal health.

Please also refer to [redacted] 2016 (MSL0027977).

B4(b)(ii) Heat stability of the purified mCry51Aa2 protein

Temperature can have a profound effect on the intergrity and function of proteins. Heat treatment is widely used in the preparation of foods derived from cotton grain (Hammond and Jez, 2011). It is reasonable that such processing will have an effect on the functional activity and intergrity of mCry51Aa2 protein when consumed in different food products derived from MON 88702. Although the weight of evidence for mCry51Aa2 makes a strong case for this protein being safe, assessing the effect of heating on the activity and structure of this protein would further inform on its safety. Therefore, an assessment of the effect of heating was conducted as a surrogate for the conditions encountered during the preparation of food and feed from MON 88702 cottonseed.

The effect of heat treatment on the activity of MON 88702-produced mCry51Aa2 protein was evaluated using the *Bt*-produced mCry51Aa2 protein. Heat treated samples and an unheated control sample of mCry51Aa2 protein were analyzed: 1) using a functional assay to assess the impact of temperature on the biological activity of mCry51Aa2 protein; and 2) using SDS PAGE to assess the impact of temperature on protein integrity.

Aliquots of mCry51Aa2 were heated to 25, 37, 55, 75, and 95 °C for 15 minutes, while a separate aliquot of mCry51Aa2 was maintained on ice for the duration of the heat treatments to serve as a temperature control. The effect of heat treatment on the biological activity of mCry51Aa2 was evaluated using a functional activity assay namely the Western tarnished plant bug (WTP) diet-incorporation insect bioassay. The effect of heat treatment on the integrity of the mCry51Aa2 protein was evaluated using SDS PAGE analysis of the heated and temperature control mCry51Aa2 protein samples.

The effects of heating on the functional activity of mCry51Aa2 are presented in Table 14. The unheated control mCry51Aa2 protein had an LC_{50} of 2.557 µg mCry51Aa2/ml diet. The mCry51Aa2 protein incubated at 25ºC and at 37ºC was shown to retain functional activity with LC_{50} values of 1.164 and 2.557 µg mCry51Aa2/ml diet, respectively. However, when heated to temperatures of 55, 75 and 95ºC for 15 minutes, the LC_{50} values were >60.0 µg mCry51Aa2/ml diet which is a reduction in functional activity of >95% relative to the control mCry51Aa2 protein. Given the impact on protein functionality observed within 15 minutes it was not deemed necessary to evaluate longer incubation times as no further loss of activity would be observed.

The results of the SDS-PAGE analysis of the heat treated samples incubated for 15 minutes are illustrated in Figure 23. The intensity of the ~35 kDa band of the unheated control sample (Figure 23, lane 2) was equivalent to that of the 100% reference standard (Figure 23, lane 8), demonstrating that the mCry51Aa2 protein was stable on wet ice during the incubation period.

No apparent decrease in band intensity of the ~35 kDa mCry51Aa2 protein was observed when heated at temperatures of 25, 37 and 55ºC for 15 minutes (Figure 23, lanes 3-5). The mCry51Aa2 protein heated to 75 and 95ºC for 15 minutes (Figure 23, lanes 6 and 7) showed some appearance of lower molecular weight bands and a decrease in the intensity of the ~35 kDa band.

These data demonstrate that the mCry51Aa2 protein behaves with a predictable tendency toward protein denaturation and loss of functional activity at elevated temperatures. Heat treatment is widely used in the preparation of foods containing components derived from cottonseed. Given the impact on protein structure observed within 15 minutes it was not deemed necessary to evaluate longer incubation times as no further impact on structure would be observed by this method. Therefore, it is reasonable to conclude that mCry51Aa2 protein would not be consumed as an active protein in food or feed products derived from MON 88702 due to standard processing practices that include heat treatment.

For details, please refer to [redacted] 2016 (MSL0029144).

Table 14. LC_{50} Values and 95% Confidence Limits (CI) for the Heat Treated mCry51Aa2 Protein After 15 Minutes

Treatment	LC_{50} (µg mCry51Aa2/ml diet)	95% CI (µg mCry51Aa2/ml diet)
Control Treatment (wet ice)	2.557	1.289 – 3.799
25 ºC	1.164	0.556 – 1.905
37 ºC	2.557	1.646 – 3.675
55 ºC	>60.0 [1]	NA
75 ºC	>60.0 [1]	NA
95 ºC	>60.0 [1]	NA

[1] 60.0 µg mCry51Aa2/ml diet represents the highest concentration tested.

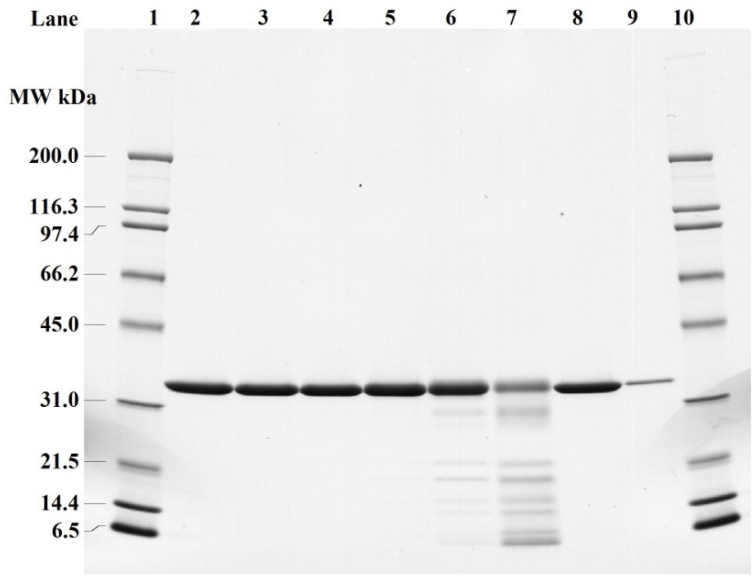

Figure 23. SDS-PAGE of mCry51Aa2 Protein Following Heat Treatement for 15 Minutes

Heat treated samples of mCry51Aa2 (3.0 µg total protein) separated on a SDS-PAGE. The gel was stained with Brilliant Blue G Colloidal. Approximate molecular weights (kDa) are shown on the left and correspond to molecular weight markers in lanes 1 and 10.

Lane	Description	Total Amount
1	Broad Range Molecular Weight Markers	5 µg
2	mCry51Aa2 Protein Control (wet ice)	3.0 µg
3	mCry51Aa2 Protein 25 ºC	3.0 µg
4	mCry51Aa2 Protein 37 ºC	3.0 µg
5	mCry51Aa2 Protein 55 ºC	3.0 µg
6	mCry51Aa2 Protein 75 ºC	3.0 µg
7	mCry51Aa2 Protein 95 ºC	3.0 µg
8	mCry51Aa2 Protein Reference 100 % Equivalence	3.0 µg
9	mCry51Aa2 Protein Reference 10 % Equivalence	0.3 µg
10	Broad Range Molecular Weight Markers	5 µg

B4(c) Acute oral toxicity study with the mCry51Aa2 protein

Acute Oral Toxicity Study with the mCry51Aa2 Protein

Most known protein toxins act through acute mechanisms to exert toxicity (Sjoblad *et al.*, 1992; Pariza and Johnson, 2001; Hammond and Fuchs, 1998). The primary exceptions to this rule consist of certain anti-nutritional proteins such as lectins and protease inhibitors, which manifest toxicity in a short-term (three-week) feeding study (Liener, 1994). The amino acid sequence of the mCry51Aa2 protein produced in MON 88702 is not similar to any of these anti-nutritional proteins or, when thoroughly evaluayed, to any other known protein toxin. Therefore, an acute oral mouse toxicity study was considered sufficient to evaluate the potential toxicity of the mCry51Aa2 protein.

Bt-produced mCry51Aa2 protein was administered by gavage to 10 male and 10 female CD-1 mice at dose of 5000 mg/kg body wt (bw). Additionally, 10 male and10 female mice were administered a comparable amount (mg/kg bw) of bovine serum albumin (BSA) to serve as a protein controls, respectively. Following dosing, all mice were observed daily for mortality or signs of toxicity. Food consumption was measured on days 0 to 7 and 7 to 14. Body weights were recorded on day 0 prior to dosing and on study days 7 and 14. All animals were sacrificed on day 14 and subjected to a gross necropsy. There were no treatment-related effects on survival, clinical observations, body weight gain, food consumption or gross pathology. Therefore, the No Observable Adverse Effect Level (NOAEL) for mCry51Aa2 is considered to be 5000 mg/kg bw.

For details, please refer to [redacted] 2017 (MSL0028578).

B5 Assessment of Potential Allergenicity

The potential allergenicity, toxicity and dietary safety of mCry51Aa2 protein is assessed in an integrated weight-of-evidence approach for the safety of MON 88702 as food and feed. History of safe use is a key consideration in the potential for allergenicity and has been previously addressed in Section B3(f).

According to the Codex Alimentarius Commission “if the protein(s) resulting from the inserted gene is present in the food, it should be assessed for potential allergenicity in all cases” (Codex Alimentarius, 2009). The only human food currently produced from cottonseed is refined, bleached, and deodorized (RBD) oil, and to a smaller extent, linters. RBD oil contains undetectable amounts of protein (Reeves and Weihrauch, 1979) and linters are a highly processed product composed of nearly pure (i.e., >99.9%) cellulose (NCPA, 2002; Nida *et al.*, 1996). Because RBD oil and linters are processed fractions that contain negligible amounts of total protein and mCry51Aa2 protein represents a very small portion of the total protein in the cottonseed of MON 88702 (Section B5(e)), an allergenicity, toxicity, and dietary safety assessment is primarily considered a theoretical assessment due to the extremely low exposure anticipated for food uses of MON 88702.

Nevertheless, following the guidelines adopted by the Codex Alimentarius Commission, an assessment of potential allergenicity of the introduced proteins has been conducted by comparing the characteristics of the introduced protein to characteristics of known allergens

(Codex Alimentarius, 2009). A protein is not likely to be associated with allergenicity if: 1) the protein is from a non-allergenic source; 2) the protein represents only a very small portion of the total plant protein; 3) the protein does not share structural similarities to known allergens based on the amino acid sequence; 4) the protein is rapidly digested in mammalian gastrointestinal systems; and 5) the protein is not stable to heat treatment. The mCry51Aa2 protein in MON 88702 has been assessed for potential allergenicity according to these safety assessment guidelines.

B5(a) Source of introduced protein

As described in Section B3(f). mCry51Aa2 is derived from *Bacillus thuringensis* which has an extensive history of safe use and has never been reported to be a souce of allegens. This extensive history of safe use establishes the safety of the donor organism.

B5(b) Bioinformatic comparison (aa) of novel protein(s) to allergens

Assessment of Structural Similarity of mCry51Aa2 to Known Allergens

The Codex guidelines for the evaluation of the allergenicity potential of introduced proteins (Codex Alimentarius, 2009) are based on the comparison of amino acid sequences between introduced proteins and allergens, where allergenic cross-reactivity may exist if the introduced protein is found to have at least 35% amino acid identity with an allergen over any segment of at least 80 amino acids. The Codex guideline also suggests that a sliding window search with a scientifically justified peptide size be used to identify immunologically relevant peptides in otherwise unrelated proteins. Therefore, the extent of sequence similarities between the mCry51Aa2 protein sequence and known allergens, gliadins, and glutenins was assessed using the FASTA sequence alignment tool along with an eight-amino acid sliding window search (Codex Alimentarius, 2009; Thomas *et al.*, 2005). No evidence was observed to indicate that the MON 88702 encoded mCry51Aa2 protein shares relevant similarities with known allergens, gliadins, or glutenins.

The FASTA program directly compares amino acid sequences (i.e. primary, linear protein structure). This alignment data may be used to infer shared higher order structural similarities between two sequences (i.e., secondary and tertiary protein structures). Proteins that share a high degree of similarity throughout the entire sequence are often homologous. By definition, homologous proteins have common secondary structures, and three-dimensional configuration, and, consequently, may share similar functions. The allergen, gliadin, and glutenin sequence database (AD_2017) was obtained as the "COMprehensive Protein Allergen REsource" (COMPARE) database from the Health and Environmental Sciences Institute (HESI) and was used for the evaluation of sequence similarities shared between the mCry51Aa2 protein and all proteins in the database. The AD_2017 database contains 1,970 sequences. When used to align the sequence of the introduced protein to each protein in the database, the FASTA algorithm produces an *E*-score (expectation score) for each alignment. The *E*-score is a statistical measure of the likelihood that the observed similarity score could have occurred by chance in a search. A larger *E*-score indicates a low degree of similarity between the query sequence and the sequence from the database. Typically, alignments

between two sequences which have an *E*-score of less than or equal to 1×10^{-5} are considered to have meaningful homology. Results indicate that the mCry51Aa2 protein sequence does not share meaningful similarity with sequences in the allergen database. No alignment met or exceeded the threshold of 35% identity over 80 amino acids recommended by Codex Alimentarius (2009) or had an *E*-score of less than or equal to 1×10^{-5}.

A second bioinformatic tool, an eight-amino acid sliding window search, was used to specifically identify short linear polypeptide matches to known allergens. It is possible that proteins structurally unrelated to allergens, gliadins, and glutenins may contain smaller immunologically meaningful epitopes. An amino acid sequence may have allergenic potential if it has an exact sequence identity of at least eight linearly contiguous amino acids with a potential allergen epitope (Hileman *et al.*, 2002; Metcalfe *et al.*, 1996). Using a sliding window of less than eight amino acids can produce matches containing considerable uncertainty depending on the length of the query sequence (Silvanovich *et al.*, 2006), and is not useful to the allergy assessment process (Thomas *et al.*, 2005). No eight contiguous amino acid identities were detected when the mCry51Aa2 protein sequence was compared to the proteins in the AD_2017 sequence database.

The bioinformatic results demonstrated there were no biologically relevant sequence similarities to allergens when the mCry51Aa2 protein sequence was used as a query for a FASTA search of the AD_2017 database. Furthermore, no short (eight amino acid) polypeptide matches were shared between the mCry51Aa2 protein sequence and proteins in the allergen database. These data show that mCry51Aa2 protein sequence lacks both structurally and immunologically relevant similarities to known allergens, gliadins, and glutenins.

For details, please refer to [redacted] 2017 (MSL0028423).

B5(c) Structural properties, including digestion by pepsin, heat treatment

B5(c)(i) Digestive fate of the mCry51Aa2 protein

Proteins are an essential dietary component for humans and animals, and most are rapidly degraded to the component amino acids for nutritional purposes and, therefore, are not toxic when ingested (Hammond and Jez, 2011). The degradation of mCry51Aa2 protein was evaluated by incubation with solutions containing pepsin and pancreatin, and the results show that mCry51Aa2 protein was rapidly degraded in pepsin which would support the conclusion that the mCry51Aa2 protein is highly unlikely to pose a safety concern to human and animal health.

B5(c)(ii) Heat stability of the Purified mCry51Aa2 protein

The effect of heat treatment on the activity of the mCry51Aa2 protein was evaluated using functional assays to assess the impact of temperature on activity, and using SDS-PAGE to assess the impact of temperature on protein integrity. The results show that the mCry51Aa2 protein behaves with a predictable tendency toward protein denaturation and loss of functional activity at elevated temperatures (Section B4(b)) and therefore it was not deemed

necessary to evaluate longer incubation times as no further impact on structure would be observed by this method. Therefore, it is anticipated that exposure to functionally active mCry51Aa2 protein from the consumption of food or feed products derived from MON 88702 due to standard processing practices that include heat treatment is unlikely.

B5(d) Specific serum screening if protein from allergenic source

Not applicable.

B5(e) Protein as a proportion of total protein

The Cry51Aa2 Protein as a Proportion of Total Protein

The mCry51Aa2 protein was detected in all plant tissues assayed (Table 13). Concerns for assessing potential allergenicity are less relevant to MON 88702 since the only human food currently produced from cottonseed is refined, bleached, and deodorized (RBD) oil, and to a smaller extent, linters. RBD oil contains undetectable amounts of protein (Reeves and Weihrauch, 1979) and linters are a highly processed product composed of nearly pure (*i.e.*, >99%) cellulose (NCPA, 2002; Nida *et al.*, 1996). Because RBD oil and linters are processed fractions that contain negligible amounts of total protein an allergenicity assessment is primarily considered a theoretical assessment. However, since cottonseed is the source of cottonseed oil and linters, cottonseed is the most appropriate tissue to assess the potential food allergenicity of MON 88702 The mean level of mCry51Aa2 protein in cottonseed of MON 88702 is 130 µg/g dw. The mean percent dry weight of total protein in seed of MON 88702 is 24.09% (or 240,900 µg/g; Table 15). The percentage of mCry51Aa2 protein in MON 88702 seed is calculated as follows:

(130 µg/g ÷ 240900 µg/g) × 100% ≈ 0.054% or 540 ppm of total cottonseed protein

Therefore, the mCry51Aa2 protein represents a very small portion of the total protein in the cottonseed of MON 88702 and due to the harsh conditions used in cottonseed processing is most likely absent in the oil and linters that are used for food production.

B6 Toxicity of Novel Herbicide Metabolites in GM Herbicide-Tolerant Plants

Not Applicable

B7 Compositional Assessment

Safety assessments of biotechnology-derived crops follow the comparative safety assessment process (Codex Alimentarius, 2009) in which the composition of grain and/or other raw agricultural commodities of the biotechnology-derived crop are compared to the appropriate conventional control that has a history of safe use. For cotton, assessments are performed using the general principles outlined in the OECD consensus document for cotton composition (OECD, 2009).

Compositional equivalence between biotechnology-derived and conventional crops supports an "equal or increased assurance of the safety of foods derived from genetically modified plants" (OECD, 2002). OECD consensus documents on compositional considerations for new crop varieties emphasize quantitative measurements of key nutrients and known anti-nutrients.

These quantitative measurements effectively discern compositional changes that could imply potential nutritional or safety (e.g., anti-nutritional) concerns. Levels of the components in grain and/or other raw agricultural commodities of the biotechnology-derived crop product are compared to: 1) corresponding levels in a conventional comparator, a genetically similar conventional line, grown concurrently under similar field conditions, and 2) natural ranges from data published in the scientific literature or in publically available databases(e.g. ILSI Crop Composition Database). This second comparison places any potential differences between the assessed crop and its comparator in the context of the well-documented variation in the concentrations of crop nutrients and anti-nutrients.

A review of compositional assessments, conducted according to OECD guidelines, that encompassed a total of seven biotechnology-derived crop varieties, nine countries, and eleven growing seasons, concluded that incorporation of biotechnology-derived agronomic traits has little impact on natural variation in crop composition (Harrigan *et al.*, 2010). Numerous scientific publications have further documented the extensive variability in the concentrations of crop components that reflect the influence of environmental and genetic factors and are not related to the insertion of agronomic traits through biotechnology (Harrigan *et al.*, 2010; Harrigan *et al.*, 2009; Ridley *et al.*, 2011; Zhou *et al.*, 2011; Venkatesh *et al.*, 2015; Bellaloui *et al.*, 2011).

This section provides analyses of concentrations of key nutrients and anti-nutrients of MON 88702 compared to that of a conventional control cotton variety grown and harvested under similar conditions. The production of materials for compositional analyses used a sufficient variety of field trial sites, robust field designs (randomized complete block design with four blocks), and sensitive analytical methods that allow accurate assessments of compositional characteristics over a range of environmental conditions under which MON 88702 is expected to be grown.

B7(a) Levels of key nutrients, toxicants and anti-nutrients

Compositional Equivalence of MON 88702 Cottonseed to Conventional Cotton

Samples for this study were harvested from five sites in the United States during the 2015 field season. Starting seeds were planted in a randomized complete block design with four replicates at each site. MON 88702 and the conventional control were grown under agronomic field conditions typical for the different growing regions.

The evaluation of MON 88702 followed considerations relevant to the compositional quality of cotton as defined by the OECD consensus document (OECD, 2009).

Compositional analyses of acid-delinted cottonseed samples were conducted for nutrients and anti-nutrients (56 components). Of these, 8 components had more than 50% of the observations below the assay limit of quantitation (LOQ), and were excluded from statistical analysis. Moisture values for cottonseed were measured for conversion of components to dry weight, but were not statistically analyzed. Therefore, 47 components in cottonseed samples were statistically assessed using a mixed model analysis of variance method.

The statistical comparison of MON 88702 and the conventional control was based on compositional data combined across all field sites. Statistically significant differences were evaluated at the 5% level ($\alpha = 0.05$). A statistically significant difference between MON 88702 and the conventional control does not necessarily imply biological relevance from a food and feed perspective. Therefore, statistically significant differences observed between MON 88702 and the conventional control were evaluated further to determine whether the detected difference indicated a biologically relevant compositional change or supported a conclusion of compositional equivalence, as follows:

Step 1 – Determination of the Magnitude of Difference between MON 88702 and Conventional Control Means

The difference in means between MON 88702 and the conventional control was determined for use in subsequent steps. For protein and amino acids only, the relative magnitude of the difference (percent change relative to the control) between MON 88702 and the conventional control was determined to allow an assessment of any observed difference in amino acids in relation to the difference in protein.

Step 2 – Assessment of the Difference in the Context of Natural Variation within the Conventional Control across Multiple Sites

The relative impact of MON 88702 was evaluated in the context of variation within the conventional control germplasm grown across multiple sites (i.e, variation due to environmental influence). This assesses the mean difference between MON 88702 and the conventional control in the context of the individual replicate values for the conventional control (maximum value minus the minimum value). When a mean difference is less than the variability seen due to natural environmental variation within the single, closely related germplasm, the difference is typically not a food or feed safety concern (Venkatesh *et al.*, 2014).

Step 3 – Assessment of the Difference in the Context of Natural Variation Due to Multiple Sources

The relative impact of MON 88702 on composition was evaluated in the context of sources of natural variation such as environmental and germplasm influences. This assessment determined whether the component mean value of MON 88702 was within the natural variability defined by the literature values or the ILSI Crop Composition Database (ILSI-CCDB) values. This naturally occurring variability is important in assessing the biological relevance of statistically significant differences in composition between MON 88702 and the conventional control.

These evaluations of natural variation are important as crop composition is known to be greatly influenced by environment and variety (Harrigan *et al.*, 2010). Although used in the comparative assessment process, detection of statistically significant differences between MON 88702 and the conventional control mean values does not necessarily imply a meaningful contribution by MON 88702 to compositional variability. Only if the impact of MON 88702 on levels of components was large relative to natural variation inherent to conventional cottonseed would further assessments be required to establish whether the

change in composition would have an impact from a food and feed safety and nutritional perspective. The steps reviewed in this assessment, therefore, describe the process for determining whether the differences between MON 88702 and the conventional control are meaningful from a food and feed perspective or whether they support a conclusion of compositional equivalence.

There were no statistically significant differences ($p<0.05$) for 38 of the 47 components analysed (Table 15 - Table 20). There were 9 components (protein, lauric acid, myristic acid, palmitic acid, stearic acid, oleic acid, arachidic acid, behenic acid and calcium in cottonseed) that showed a statistically significant difference ($p<0.05$) between MON 88702 and the conventional control.

For protein, the difference was -0.59% dw (Table 15). As shown in Table 15, the magnitude of the difference for protein between MON 88702 and the conventional control was less than the corresponding conventional control range value. This indicates that MON 88702 does not impact levels of this component more than the natural variation within the conventional control grown at multiple locations. The mean level of protein was also within the natural variability of this component as published in the scientific literature on cotton composition and/or the ILSI-CCDB (Table 21).

For the seven fatty acids, the statistically significant differences reflected differences in mean values between the MON 88702 and the conventional control with magnitudes ranging from 0.0020 to 1.57% Total FA (Table 16). As shown in Table 16, the magnitude of differences for the seven fatty acids between MON 88702 and the conventional control were less than the corresponding conventional control range values. This indicates that MON 88702 does not impact levels of these components more than the natural variation within the conventional control grown at multiple locations.

The mean levels of the seven fatty acids were within the natural variability of these components as published in the scientific literature on cotton composition and/or the ILSI-CCDB, except for lauric acid, a low abundance fatty acid for which literature and ILSI Crop Composition Database (ILSI-CCDB) values were unavailable (Table 21). However, the observed MON 88702 mean value for 12:0 lauric acid (0.015 % Total FA) was within the data range (ND – 0.2% Total FA) reported in cottonseed oil by the Codex as the international food standard (Codex Alimentarius, 1999). The data demonstrated that MON 88702 was not a major contributor to variation in total fat or fatty acid levels in cotton and confirmed the compositional equivalence of MON 88702 to the conventional control in levels of these components.

For calcium, the difference was 0.011% dw (Table 18). As shown in Table 18, the magnitude of the difference for calcium between MON 88702 and the conventional control was less than the corresponding conventional control range value. This indicates that MON 88702 does not impact levels of this component more than the natural variation within the conventional control grown at multiple locations. The mean level of calcium was within the natural variability of this component as published in the scientific literature on cotton composition and/or the ILSI-CCDB (Table 21).

These results support the overall conclusion that MON 88702 was not a major contributor to variation in component levels in cottonseed and confirmed the compositional equivalence of MON 88702 to the conventional control in levels of these components. These data indicated that the statistically significant difference observed was not compositionally meaningful from a food and feed safety perspective.

For details, please refer to [redacted] 2017 (MSL0029119).

Table 15. Summary of Cottonseed Protein and Amino Acids for MON 88702 and the Conventional Control

				Difference (MON 88702 minus Control)		
Component (% dw)[1]	MON 88702 Mean (S.E.)[2] Range	Control Mean (S.E.) Range	Control Range Value[3]	Mean (S.E.)	p-Value	% Relative[4]
Protein	24.09 (0.63) 21.55 - 26.87	24.67 (0.63) 22.67 - 27.39	4.72	-0.59 (0.19)	0.038	-2.39
Alanine	0.88 (0.024) 0.79 - 0.95	0.88 (0.024) 0.78 - 1.01	0.23	-0.0012 (0.026)	0.965	-0.14
Arginine	2.58 (0.11) 2.07 - 3.07	2.69 (0.11) 2.36 - 3.24	0.88	-0.11 (0.081)	0.245	-4.12
Aspartic acid	2.08 (0.076) 1.73 - 2.27	2.15 (0.076) 1.89 - 2.49	0.59	-0.065 (0.072)	0.417	-3.03
Cystine	0.50 (0.019) 0.36 - 0.58	0.50 (0.019) 0.40 - 0.61	0.21	0.0045 (0.013)	0.733	0.90
Glutamic acid	4.67 (0.20) 3.78 - 5.13	4.82 (0.20) 3.94 - 5.47	1.53	-0.14 (0.16)	0.414	-2.97

Table 15. Summary of Cottonseed Protein and Amino Acids for MON 88702 and the Conventional Control (continued)

				Difference (MON 88702 minus Control)		
Component (% dw)[1]	MON 88702 Mean (S.E.)[2] Range	Control Mean (S.E.) Range	Control Range Value[3]	Mean (S.E.)	p-Value	% Relative[4]
Glycine	0.95 (0.028) 0.78 - 1.07	0.95 (0.028) 0.85 - 1.13	0.27	-0.0061 (0.032)	0.857	-0.64
Histidine	0.67 (0.020) 0.58 - 0.79	0.68 (0.020) 0.60 - 0.81	0.21	-0.0064 (0.017)	0.727	-0.95
Isoleucine	0.74 (0.023) 0.63 - 0.81	0.76 (0.023) 0.66 - 0.88	0.21	-0.015 (0.018)	0.453	-1.96
Leucine	1.41 (0.043) 1.25 - 1.53	1.43 (0.043) 1.27 - 1.64	0.37	-0.024 (0.034)	0.522	-1.65
Lysine	1.12 (0.031) 1.00 - 1.21	1.14 (0.031) 0.98 - 1.26	0.27	-0.016 (0.030)	0.624	-1.40
Methionine	0.36 (0.013) 0.28 - 0.41	0.36 (0.013) 0.31 - 0.42	0.12	-0.0037 (0.0072)	0.607	-1.03

Table 15. Summary of Cottonseed Protein and Amino Acids for MON 88702 and the Conventional Control (continued)

				Difference (MON 88702 minus Control)		
Component (% dw)[1]	MON 88702 Mean (S.E.)[2] Range	Control Mean (S.E.) Range	Control Range Value[3]	Mean (S.E.)	p-Value	% Relative[4]
Phenylalanine	1.31 (0.048) 1.06 - 1.46	1.34 (0.048) 1.16 - 1.68	0.51	-0.035 (0.034)	0.363	-2.61
Proline	0.87 (0.028) 0.76 - 0.93	0.88 (0.028) 0.77 - 1.01	0.23	-0.012 (0.023)	0.623	-1.38
Serine	0.92 (0.037) 0.66 - 1.35	0.91 (0.037) 0.75 - 1.16	0.41	0.011 (0.045)	0.818	1.21
Threonine	0.78 (0.020) 0.66 - 0.83	0.78 (0.020) 0.71 - 0.88	0.18	-0.00066 (0.019)	0.974	-0.085
Tryptophan	0.27 (0.0079) 0.23 - 0.29	0.27 (0.0079) 0.22 - 0.30	0.086	-0.00028 (0.0044)	0.949	-0.11
Tyrosine	0.47 (0.018) 0.40 - 0.52	0.48 (0.018) 0.38 - 0.61	0.22	-0.0053 (0.019)	0.790	-1.10

Table 15. Summary of Cottonseed Protein and Amino Acids for MON 88702 and the Conventional Control (continued)

				Difference (MON 88702 minus Control)		
Component (% dw)[1]	MON 88702 Mean (S.E.)[2] Range	Control Mean (S.E.) Range	Control Range Value[3]	Mean (S.E.)	p-Value	% Relative[4]
Valine	1.03 (0.031) 0.88 - 1.13	1.05 (0.031) 0.92 - 1.21	0.29	-0.018 (0.026)	0.529	-1.69

[1]dw=dry weight
[2] Mean (S.E.) = least-square mean (standard error)
[3]Maximum value minus minimum value for the control cotton variety
[4]The relative magnitude of the difference in mean values between MON 88702 and the control, expressed as a percent of the control.

Table 16. Summary of Cottonseed Total Fat and Fatty Acids for MON 88702 and the Conventional Control

Component	MON 88702 Mean (S.E.)[1] Range	Control Mean (S.E.) Range	Control Range Value[2]	Difference (Test minus Control) Mean (S.E.)	Difference (Test minus Control) p-Value
Total fat (% dw)[3]	19.54 (1.64) 12.31 - 23.71	18.89 (1.64) 11.90 - 22.33	10.43	0.64 (0.30)	0.097
12:0 Lauric acid (% Total FA)[4]	0.015 (0.0024) 0.0088 – 0.022	0.017 (0.0024) 0.0095 – 0.026	0.017	-0.0020 (0.00042)	0.009
14:0 Myristic acid (% Total FA)	0.57 (0.047) 0.44 - 0.67	0.71 (0.047) 0.55 - 0.88	0.33	-0.14 (0.015)	<.001
16:0 Palmitic acid (% Total FA)	19.21 (0.60) 17.21 - 20.43	20.78 (0.60) 18.71 - 22.25	3.54	-1.57 (0.15)	<.001
16:1 Palmitoleic acid (% Total FA)	0.55 (0.018) 0.47 - 0.65	0.53 (0.018) 0.47 - 0.60	0.13	0.027 (0.0096)	0.050
17:0 Heptadecanoic acid (% Total FA)	0.077 (0.0028) 0.071 - 0.093	0.079 (0.0028) 0.066 - 0.089	0.023	-0.0022 (0.00094)	0.076
17:1 Heptadecenoic acid (% Total FA)	0.042 (0.00088) 0.035 - 0.048	0.040 (0.00087) 0.034 - 0.048	0.013	0.0016 (0.00089)	0.085

Table 16. Summary of Cottonseed Total Fat and Fatty Acids for MON 88702 and the Conventional Control (continued)

Component	MON 88702 Mean (S.E.)[1] Range	Control Mean (S.E.) Range	Control Range Value[2]	Difference (Test minus Control) Mean (S.E.)	Difference (Test minus Control) p-Value
18:0 Stearic acid (% Total FA)	2.66 (0.15) 2.14 - 2.98	2.50 (0.15) 2.03 - 2.82	0.79	0.16 (0.025)	0.003
18:1 Oleic acid (% Total FA)	18.45 (1.16) 14.74 - 22.92	17.34 (1.16) 14.37 - 20.60	6.23	1.11 (0.29)	0.020
18:2 Linoleic acid (% Total FA)	56.60 (1.90) 50.74 - 62.92	56.13 (1.90) 51.25 - 61.83	10.58	0.47 (0.26)	0.141
18:3 Linolenic acid (% Total FA)	0.22 (0.0094) 0.19 - 0.26	0.22 (0.0094) 0.19 - 0.27	0.079	0.0021 (0.0021)	0.326
20:0 Arachidic acid (% Total FA)	0.26 (0.028) 0.18 - 0.35	0.27 (0.028) 0.20 - 0.38	0.18	-0.016 (0.0016)	<.001
20:1 Eicosenoic acid (% Total FA)	0.065 (0.0031) 0.059 - 0.078	0.066 (0.0031) 0.060 - 0.081	0.022	-0.0014 (0.00075)	0.132

Table 16. Summary of Cottonseed Total Fat and Fatty Acids for MON 88702 and the Conventional Control (continued)

Component	MON 88702 Mean (S.E.)[1] Range	Control Mean (S.E.) Range	Control Range Value[2]	Difference (Test minus Control) Mean (S.E.)	Difference (Test minus Control) p-Value
20:2 Eicosadienoic acid (% Total FA)	0.018 (0.00085) 0.014 - 0.023	0.019 (0.00085) 0.016 - 0.024	0.0085	-0.00083 (0.00045)	0.074
22:0 Behenic acid (% Total FA)	0.12 (0.017) 0.094 - 0.19	0.13 (0.017) 0.10 - 0.21	0.11	-0.010 (0.0015)	0.002

[1] Mean (S.E.) = least-square mean (standard error)
[2]Maximum value minus minimum value for the control cotton variety
[3]dw=dry weight
[4]FA=Fatty Acid
The following components with more than 50% of observations below the assay LOQ were excluded from statistical analysis: caprylic acid, capric acid, myristoleic acid, pentadecanoic acid, pentadecenoic acid, gamma linolenic acid, eicosatrienoic acid and arachidonic acid.

Table 17. Summary of Cottonseed Carbohydrates by Calculation and Fiber for MON 88702 and the Conventional Control

				Difference (MON 88702 minus Control)	
Component (% dw)[1]	MON 88702 Mean (S.E.)[2] Range	Control Mean (S.E.) Range	Control Range Value[3]	Mean (S.E.)	p-Value
Carbohydrates by calculation	52.22 (1.51) 49.04 - 60.26	52.33 (1.51) 49.51 - 59.91	10.40	-0.11 (0.23)	0.635
ADF	35.40 (0.55) 32.95 - 37.80	34.75 (0.55) 31.87 - 38.52	6.65	0.65 (0.59)	0.334
NDF	43.03 (0.90) 40.00 - 46.41	42.56 (0.90) 38.79 - 48.02	9.23	0.47 (0.36)	0.195
TDF	40.95 (1.69) 26.61 - 48.33	40.05 (1.69) 28.02 - 47.25	19.24	0.91 (1.34)	0.501

[1]dw=dry weight
[2] Mean (S.E.) = least-square mean (standard error)
[3]Maximum value minus minimum value for the control cotton variety

Table 18. Summary of Cottonseed Ash and Minerals for MON 88702 and the Conventional Control

				Difference (MON 88702 minus Control)	
Component (% dw)[1]	MON 88702 Mean (S.E.)[2] Range	Control Mean (S.E.) Range	Control Range Value[3]	Mean (S.E.)	p-Value
Ash	4.17 (0.18) 3.55 - 4.62	4.11 (0.18) 3.47 - 4.56	1.09	0.055 (0.021)	0.061
Calcium	0.13 (0.0047) 0.11 - 0.15	0.11 (0.0047) 0.099 - 0.13	0.034	0.011 (0.0039)	0.049
Phosphorus	0.66 (0.046) 0.46 - 0.79	0.63 (0.046) 0.45 - 0.77	0.31	0.029 (0.018)	0.178

[1]dw=dry weight
[2] Mean (S.E.) = least-square mean (standard error)
[3]Maximum value minus minimum value for the control cotton variety

Table 19. Summary of Cottonseed Vitamins for MON 88702 and the Conventional Control

				Difference (MON 88702 minus Control)	
Component (mg/kg dw)[1]	MON 88702 Mean (S.E.)[2] Range	Control Mean (S.E.) Range	Control Range Value[3]	Mean (S.E.)	p-Value
Vitamin E	122.94 (8.09) 98.60 - 159.97	120.26 (8.09) 100.96 - 149.01	48.05	2.68 (1.50)	0.090

[1]dw=dry weight
[2] Mean (S.E.) = least-square mean (standard error)
[3]Maximum value minus minimum value for the control cotton variety

Table 20. Summary of Cottonseed Anti-Nutrients for MON 88702 and the Conventional Control

Component	MON 88702 Mean (S.E.)[1] Range	Control Mean (S.E.) Range	Control Range Value[2]	Difference (Test minus Control) Mean (S.E.)	Difference (Test minus Control) p-Value
Total gossypol (% dw)[3]	1.01 (0.12) 0.55 - 1.44	1.01 (0.12) 0.59 - 1.46	0.87	-0.0064 (0.032)	0.852
Free gossypol (% dw)	0.62 (0.093) 0.26 - 0.93	0.61 (0.093) 0.23 - 0.89	0.66	0.0074 (0.014)	0.614
Malvalic acid (% Total FA)[4]	0.52 (0.046) 0.41 - 0.73	0.55 (0.046) 0.38 - 0.74	0.36	-0.034 (0.025)	0.246
Sterculic acid (% Total FA)	0.24 (0.015) 0.19 - 0.29	0.24 (0.015) 0.20 - 0.31	0.10	0.00063 (0.0058)	0.919
Dihydrosterculic acid (% Total FA)	0.38 (0.038) 0.30 - 0.54	0.37 (0.038) 0.31 - 0.56	0.25	0.0068 (0.0078)	0.437

[1] Mean (S.E.) = least-square mean (standard error)
[2]Maximum value minus minimum value for the control cotton variety
[3]dw=dry weight
[4]FA=Fatty Acid

Table 21. Literature and ILSI Database Ranges for Components in Cottonseed

Tissue Components[1]	Literature Range[2]	ILSI Range[3]
Cottonseed Nutrients		
Proximates		
protein (% dw)	21.76-30.83[a]	19.19-32.97
total fat (% dw)	15.44-25.16[a]	15.05-27.90
ash (% dw)	3.67-5.29[a]	3.006-5.476
Amino Acids		
alanine (% dw)	0.83-1.22[b]	0.69-1.29
arginine (% dw)	2.30-3.77[b]	1.76-3.93
aspartic acid (% dw)	1.79-2.74[b]	1.51-3.21
cystine (% dw)	0.29-0.47[b]	0.288-0.557
glutamic acid (% dw)	3.39-5.70[b]	3.04-6.72
glycine (% dw)	0.85-1.23[b]	0.73-1.32
histidine (% dw)	0.57-0.84[b]	0.452-0.985
isoleucine (% dw)	0.72-1.03[b]	0.58-1.05
leucine (% dw)	1.20-1.72[b]	1.01-1.86
lysine (% dw)	0.99-1.44[b]	0.84-1.46
methionine (% dw)	0.29-0.49[b]	0.29-0.49
phenylalanine (% dw)	1.10-1.66[b]	0.88-1.76
proline (% dw)	0.79-1.25[b]	0.60-1.37
serine (% dw)	0.81-1.24[b]	0.74-1.39
threonine (% dw)	0.67-0.96[b]	0.55-1.06
tryptophan (% dw)	0.31-0.52[b]	0.162-0.519
tyrosine (% dw)	0.63-0.91[b]	0.47-1.00
valine (% dw)	0.97-1.40[b]	0.76-1.49
Fatty Acids		
lauric acid (% Total FA)	NA	NA
myristic acid (% Total FA)	0.55-2.40[a]	0.426-2.400
palmitic acid (% Total FA)	21.23-28.10[a]	15.11-27.90
palmitoleic acid (% Total FA)	0.43-1.74[a]	0.375-1.190
heptadecanoic acid (% Total FA)	NA	0.077-1.120
heptadecenoic acid (% Total FA)	NA	NA
stearic acid (% Total FA)	1.99-3.11[a]	0.20-3.54
oleic acid (% Total FA)	12.90-20.10[a]	12.8-25.4
linoleic acid (% Total FA)	46.00-57.10[a]	42.5-63.0
linolenic acid (% Total FA)	0.050-0.32[a]	0.10-0.64
arachidic acid (% Total FA)	0.24-0.34[a]	0.149-0.484
eicosenoic acid (% Total FA)	NA	0.095-0.100
eicosadienoic acid (% Total FA)	NA	NA
behenic acid (% Total FA)	0.12-0.24[a]	0.099-0.295
Carbohydrates By Calculation		
carbohydrates by calculation (% dw)	43.69-53.62[a]	39.0-59.2
Fiber		
ADF (% dw)	21.10-34.80[a]	19.74-38.95
NDF (% dw)	32.92-45.83[a]	25.56-51.87
TDF (% dw)	NA	33.69-53.50

Table 21. Literature and ILSI Database Ranges for Components in Cottonseed (continued)

Tissue Components[1]	Literature Range[2]	ILSI Range[3]
Minerals		
calcium (% dw)	0.10-0.33[a]	0.070-0.326
phosphorus (% dw)	0.56-0.86[a]	0.384-0.992
Vitamins		
vitamin E (mg/kg dw)	84.07-162.98[b]	26.57-197.24
Cottonseed Other		
Anti-Nutrients		
total gossypol (% dw)	0.57-1.42[a]	0.350-1.613
free gossypol (% dw)	0.53-1.20[a]	0.384-1.418
malvalic acid (% Total FA)	0.17-0.61[a]	0.112-0.854
sterculic acid (% Total FA)	0.13-0.56[a]	0.061-0.556
dihydrosterculic acid (% Total FA)	0.12-0.24[a]	0.031-0.325

[1]dw=dry weight; FA=Fatty Acid

[2]Literature range references: [a](Hamilton *et al.*, 2004);[b](Harrison *et al.*, 2013)

[3]ILSI range is from ILSI Crop Composition Database, 2014 (Accessed July 27, 2016).

Compositional Assessment Conclusion

Compositional analysis was conducted on cottonseed of MON 88702 and the conventional control grown at five sites in the United States during the 2015 field season. Of the 47 components statistically assessed, 38 showed no statistically significant differences ($p<0.05$) between MON 88702 and the conventional control. A total of 9 components (protein, lauric acid, myristic acid, palmitic acid, stearic acid, oleic acid, arachidic acid, behenic acid and calcium) showed a statistically significant difference ($p<0.05$) between MON 88702 and the conventional control. For these components, the mean difference in component values between MON 88702 and the conventional control was less than the range of the conventional control values. In addition, the MON 88702 mean component values were within the natural variability defined by the range of values observed in the literature, the ILSI Crop Composition Database (ILSI-CCDB) and/or (Codex Alimentarius, 1999) values, except for lauric acid, a low abundance fatty acid for which literature and ILSI Crop Composition Database (ILSI-CCDB) values were unavailable.

These results support the overall conclusion that MON 88702 cotton was not a major contributor to variation in component levels in cottonseed and confirmed the compositional equivalence of MON 88702 to the conventional control in levels of these components. The statistically significant differences observed were not compositionally meaningful from a food and feed safety perspective.

B7(b) Levels of other GM-influenced constituents

Not applicable.

B7(c) Levels of naturally-occurring allergenic proteins

Not applicable.

C. NUTRITIONAL IMPACT

C1 Information Related to the Nutritional Impact of the Food Produced Using Gene Technology

Not applicable. This product is developed to confer insect protection. It is not a nutritionally altered product.

D. Other Information

The data and information presented in this submission demonstrate that the food and feed derived from MON 88702 are as safe and nutritious as those derived from commercially-available, conventional cotton for which there is an established history of safe consumption. Therefore, animal feeding studies do not add value to the safety of MON 88702.

PART 3 STATUTORY DECLARATION – AUSTRALIA

I, [redacted], declare that the information provided in this application fully sets out the matters required and that the same are true to the best of my knowledge and belief, and that no information has been withheld that might prejudice this application.

Signature: [redacted]

Declared before me ……

This ……**16th**…….. day of ……**August**…….. **2017**.

PART 4 REFERENCES

Anderson, H.M., D.J. Bowen, C.A. Chay, S. Flasinski, U.R. Kesanapalli, J.S. Milligan, R.N. Slightom and Y. Yin. 2015. Pesticidal toxin active against Coleopteran and/or Hemipteran insects. Patent 20150047076, U.S. Patent Office, Washington, D.C.

Anderson, J.E., J.-M. Michno, T.J.Y. Kono, A.O. Stec, B.W. Campbell, S.J. Curtin and R.M. Stupar. 2016. Genomic variation and DNA repair associated with soybean transgenesis: A comparison to cultivars and mutagenized plants. BMC Biotechnology 16:41.

AOCS. 1991. Cellulose yield pressure-cook method. Method Bb 3-47, American Oil Chemists' Society, Champaign, Illinois.

Astwood, J.D., J.N. Leach and R.L. Fuchs. 1996. Stability of food allergens to digestion in vitro. Nature Biotechnology 14:1269-1273.

Axelos, M., C. Bardet, T. Liboz, A. Le Van Thai, C. Curie and B. Lescure. 1989. The gene family encoding the *Arabidopsis thaliana* translation elongation factor EF-1α molecular cloning characterization and expression. Molecular and General Genetics 219:106-112.

Barker, R.F., K.B. Idler, D.V. Thompson and J.D. Kemp. 1983. Nucleotide sequence of the T-DNA region from the *Agrobacterium tumefaciens* octopine Ti plasmid pTi15955. Plant Molecular Biology 2:335-350.

Baum, J.A., U.R. Sukuru, S.R. Penn, S.E. Meyer, S. Subbarao, X. Shi, S. Flasinski, G.R. Heck, R.S. Brown and T.L. Clark. 2012. Cotton plants expressing a hemipteran-active *Bacillus thuringiensis* crystal protein impact the development and survival of *Lygus hesperus* (Hemiptera: Miridae) nymphs. Journal of Economic Entomology 105:616-624.

Bautista-Zapanta, J.-n., K. Suzuki and K. Yoshida. 2002. Characterization of four ribosomal RNA operons in the genome of *Agrobacterium tumefaciens* MAFF301001. Nucleic Acids Research S2:91-92.

Beck, E., G. Ludwig, E.A. Auerswald, B. Reiss and H. Schaller. 1982. Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from transposon Tn*5*. Gene 19:327-336.

Bellaloui, N., J.R. Smith, A.M. Gillen and J.D. Ray. 2011. Effects of maturity, genotypic background, and temperature on seed mineral composition in near-isogenic soybean lines in the early soybean production system. Crop Science 51:1161-1171.

Berry, C. 2012. The bacterium, *Lysinibacillus sphaericus*, as an insect pathogen. Journal of Invertebrate Pathology 109:1-10.

Berry, C. and N. Crickmore. 2017. Structural classification of insecticidal proteins - Towards an *in silico* characterisation of novel toxins. Journal of Invertebrate Pathology 142:16-22.

Biron, D.G., C. Brun, T. Lefevre, C. Lebarbenchon, H.D. Loxdale, F. Chevenet, J.-P. Brizard and F. Thomas. 2006. The pitfalls of proteomics experiments without the correct use of bioinformatics tools. Proteomics 6:5577-5596.

Bridge, R., J.M. Bridge and R. McGowen. 2005. Cotton cultivar DP 393. Patent US 6,930,228 B2, U.S. Patent Office, Washington, D.C.

Caetano-Anollés, G., M. Wang, D. Caetano-Anollés and J.E. Mittenthal. 2009. The origin, evolution and structure of the protein world. Biochemical Journal 417:621-637.

Cannon, R.J.C. 1993. Prospects and progress for *Bacillus thuringiensis*-based pesticides. Pesticide Science 37:331-335.

Chen, Y.R., A. Rivlin, A. Lange, X.D. Ye, Z. Vaghchhipawala, E. Eisinger, E. Dersch, M. Paris, B. Martinell and Y.C. Wan. 2014. High throughput *Agrobacterium tumefaciens*-mediated germline transformation of mechanically isolated meristem explants of cotton (*Gossypium hirsutum* L.). Plant Cell Reports 33:153-164.

Cherry, J.P. 1983. Cottonseed oil. Journal of the American Oil Chemists' Society 60:360-367.

Codex Alimentarius. 1999. Codex standards for fats and oils from vegetable sources. Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme, Food and Agriculture Organization of the United Nations, Rome, Italy. http://www.fao.org/docrep/004/y2774e/y2774e04.htm [Accessed May 4, 2017].

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

Coruzzi, G., R. Broglie, C. Edwards and N.-H. Chua. 1984. Tissue-specific and light-regulated expression of a pea nuclear gene encoding the small subunit of ribulose-1, 5-bisphosphate carboxylase. EMBO Journal 3:1671-1679.

Crickmore, N., J. Baum, A. Bravo, D. Lereclus, K. Narva, K. Sampson, E. Schnepf, M. Sun and D.R. Zeigler. 2016. *Bacillus thuringiensis* toxin nomenclature. http://www.btnomenclature.info/ [Accessed 4-15-2016].

Delaney, B., J.D. Astwood, H. Cunny, R.E. Conn, C. Herouet-Guicheney, S. MacIntosh, L.S. Meyer, L. Privalle, Y. Gao, J. Mattsson and M. Levine. 2008. Evaluation of protein safety in the context of agricultural biotechnology. Food and Chemical Toxicology 46:S71-S97.

Depicker, A., S. Stachel, P. Dhaese, P. Zambryski and H.M. Goodman. 1982. Nopaline synthase: Transcript mapping and DNA sequence. Journal of Molecular and Applied Genetics 1:561-573.

DuBose, A.J., S.T. Lichtenstein, N. Narisu, L.L. Bonnycastle, A.J. Swift, P.S. Chines and F.S. Collins. 2013. Use of microarray hybrid capture and next-generation sequencing to identify the anatomy of a transgene. Nucleic Acids Research 41:e70.

Fling, M.E., J. Kopf and C. Richards. 1985. Nucleotide sequence of the transposon Tn*7* gene encoding an aminoglycoside-modifying enzyme, 3"(9)-*O*-nucleotidyltransferase. Nucleic Acids Research 13:7095-7106.

Fraley, R.T., S.G. Rogers, R.B. Horsch, P.R. Sanders, J.S. Flick, S.P. Adams, M.L. Bittner, L.A. Brand, C.L. Fink, J.S. Fry, G.R. Galluppi, S.B. Goldberg, N.L. Hoffmann and S.C. Woo. 1983. Expression of bacterial genes in plant cells. Proceedings of the National Academy of Sciences of the United States of America 80:4803-4807.

Fu, T.-J., U.R. Abbott and C. Hatzos. 2002. Digestibility of food allergens and nonallergenic proteins in simulated gastric fluid and simulated intestinal fluid - A comparative study. Journal of Agricultural and Food Chemistry 50:7154-7160.

Giglione, C. and T. Meinnel. 2001. Organellar peptide deformylases: Universality of the N-terminal methionine cleavage mechanism. Trends in Plant Science 6:566-572.

Gill, S.S., E.A. Cowles and P.V. Pietrantonio. 1992. The mode of action for *Bacillus thuringiensis* endotoxins. Annual Review of Entomology 37:615-636.

Giza, P.E. and R.C.C. Huang. 1989. A self-inducing runaway-replication plasmid expression system utilizing the Rop protein. Gene 78:73-84.

Gowda, A., T.J. Rydel, A.M. Wollacott, R.S. Brown, W. Akbar, T.L. Clark, S. Flasinski, J.R. Nageotte, A.C. Read, X. Shi, B.J. Werner, M.J. Pleau and J.A. Baum. 2016. A transgenic approach for controlling *Lygus* in cotton. Nature Communications 7:12213.

Hamilton, K.A., P.D. Pyla, M. Breeze, T. Olson, M.H. Li, E. Robinson, S.P. Gallagher, R. Sorbet and Y. Chen. 2004. Bollgard II cotton: Compositional analysis and feeding studies of cottonseed from insect-protected cotton (*Gossypium hirsutum* L.) producing the Cry1Ac and Cry2Ab2 proteins. Journal of Agricultural and Food Chemistry 52:6969-6976.

Hammond, B. 2004. A review of the food/feed safety and benefits of *Bacillus thuringiensis* protein containing insect-protected crops. Pages 103-123 in Agricultural Biotechnology: Challenges and Prospects. M.K. Bhalgat, W.P. Ridley, A.S. Felsot, and J.N. Seiber (eds.), American Chemical Society, Washington, D.C.

Hammond, B.G. and R.L. Fuchs. 1998. Safety evaluation for new varieties of food crops developed through biotechnology. Pages 61-79 in Biotechnology and Safety Assessment. Second Edition. J.A. Thomas (ed.). Taylor & Francis, Philadelphia, Pennsylvania.

Hammond, B.G. and J.M. Jez. 2011. Impact of food processing on the safety assessment for proteins introduced into biotechnology-derived soybean and corn crops. Food and Chemical Toxicology 49:711-721.

Harrigan, G.G., D. Lundry, S. Drury, K. Berman, S.G. Riordan, M.A. Nemeth, W.P. Ridley and K.C. Glenn. 2010. Natural variation in crop composition and the impact of transgenesis. Nature Biotechnology 28:402-404.

Harrigan, G.G., W.P. Ridley, K.D. Miller, R. Sorbet, S.G. Riordan, M.A. Nemeth, W. Reeves and T.A. Pester. 2009. The forage and grain of MON 87460, a drought-tolerant corn hybrid, are compositionally equivalent to that of conventional corn. Journal of Agricultural and Food Chemistry 57:9754-9763.

Harris, W.D. 1981. Cottonseed. Pages 375-391 in Encyclopedia of Chemical Processing and Design. Volume 12. J.J. McKetta and W.A. Cunningham (eds.). Marcel Dekker, Inc., New York, New York.

Harrison, J.M., D. Howard, M. Malven, S.C. Halls, A.H. Culler, G.G. Harrigan and R.D. Wolfinger. 2013. Principal variance component analysis of crop composition data: A case study on herbicide-tolerant cotton. Journal of Agricultural and Food Chemistry 61:6412-6422.

Helm, R.M. 2001. Topic 5: Stability of known allergens (digestive and heat stability). Joint FAO/WHO expert consultation on foods derived from biotechnology. Food and Agriculture Organization of the United Nations, Rome, Italy.

Herrmann, K.M. 1995. The shikimate pathway: Early steps in the biosynthesis of aromatic compounds. The Plant Cell 7:907-919.

Hileman, R.E., A. Silvanovich, R.E. Goodman, E.A. Rice, G. Holleschak, J.D. Astwood and S.L. Hefle. 2002. Bioinformatic methods for allergenicity assessment using a comprehensive allergen database. International Archives of Allergy and Immunology 128:280-291.

Hu, X., W. Fan, B. Han, H. Liu, D. Zheng, Q. Li, W. Dong, J. Yan, M. Gao, C. Berry and Z. Yuan. 2008. Complete genome sequence of the mosquitocidal bacterium *Bacillus sphaericus* C3-41 and comparison with those of closely related Bacillus species. Journal of Bacteriology 190:2892-2902.

Iacovache, I., S. De Carlo, N. Cirauqui, M. Dal Peraro, F.G. van der Goot and B. Zuber. 2016. Cryo-EM structure of aerolysin variants reveals a novel protein fold and the pore-formation process. Nature Communications 7:12062.

Iacovache, I., F.G. van der Goot and L. Pernot. 2008. Pore formation: An ancient yet complex form of attack. Biochimica et Biophysica Acta 1778:1611-1623.

Illergård, K., D.H. Ardell and A. Elofsson. 2009. Structure is three to ten times more conserved than sequence - A study of structural response in protein cores. Proteins 77:499-508.

Jerga, A., D. Chen, C. Zhang, J. Fu, J.-L.K. Kouadio, Y. Wang, S.M.G. Duff, J.E. Howard, T.J. Rydel, A.G. Evdokimov, P. Ramaseshadri, A. Evans, R. Bolognesi, Y. Park and J.A. Haas. 2016. Mechanistic insights into the first Lygus-active β-pore forming protein. Archives of Biochemistry and Biophysics 600:1-11.

Kairichi, M.N., B.W. Waswa, C.N. Waturu, W. Wiesel, R.G. Ngigi, G.K. Njenga and S.M. Njinju. 2008. Assessment of pollen-mediated gene flow of *Bt*-cotton to local commerical variety, Hart 89m in Kari-Mwea Station. Page 257 in Proceedings of the 1st All African Congress on Biotechnology, Nairobi, Kenya.

Klee, H.J., Y.M. Muskopf and C.S. Gasser. 1987. Cloning of an *Arabidopsis thaliana* gene encoding 5-enolpyruvylshikimate-3-phosphate synthase: Sequence analysis and manipulation to obtain glyphosate-tolerant plants. Molecular and General Genetics 210:437-442.

Koch, M.S., J.M. Ward, S.L. Levine, J.A. Baum, J.L. Vicini and B.G. Hammond. 2015. The food and environmental safety of *Bt* crops. Frontiers in Plant Science 6:283.

Kovalic, D., C. Garnaat, L. Guo, Y. Yan, J. Groat, A. Silvanovich, L. Ralston, M. Huang, Q. Tian, A. Christian, N. Cheikh, J. Hjelle, S. Padgette and G. Bannon. 2012. The use of next generation sequencing and junction sequence analysis bioinformatics to achieve molecular characterization of crops improved through modern biotechnology. The Plant Genome 5:149-163.

Krause, E., H. Wenschuh and P.R. Jungblut. 1999. The dominance of arginine-containing peptides in MALDI-derived tryptic mass fingerprints of proteins. Analytical Chemistry 71:4160-4165.

Liener, I.E. 1994. Implications of antinutritional components in soybean foods. Critical Reviews in Food Science and Nutrition 34:31-67.

Liu, J.-W., A.G. Porter, B.Y. Wee and T. Thanabalu. 1996. New gene from nine *Bacillus sphaericus* strains encoding highly conserved 35.8-kilodalton mosquitocidal toxins. Applied and Environmental Microbiology 62:2174-2176.

Llewellyn, D., C. Tyson, G. Constable, B. Duggan, S. Beale and P. Steel. 2007. Containment of regulated genetically modified cotton in the field. Agriculture, Ecosystems and Environment 121:419-429.

McClintock, J.T., C.R. Schaffer and R.D. Sjoblad. 1995. A comparative review of the mammalian toxicity of *Bacillus thuringiensis*-based pesticides. Pesticide Science 45:95-105.

McGregor, S.E. 1976. Insect pollination of cultivated crop plants. Agricultural Handbook No. 496. U.S. Department of Agriculture, Agricultural Research Service, Washington, D.C.

Meinnel, T. and C. Giglione. 2008. Tools for analyzing and predicting N-terminal protein modifications. Proteomics 8:626-649.

Metcalfe, D.D., J.D. Astwood, R. Townsend, H.A. Sampson, S.L. Taylor and R.L. Fuchs. 1996. Assessment of the allergenic potential of foods derived from genetically engineered crop plants. Critical Reviews in Food Science and Nutrition 36:S165-S186.

Moar, W.J., A.J. Evans, C.R. Kessenich, J.A. Baum, D.J. Bowen, T.C. Edrington, J.A. Haas, J.-L.K. Kouadio, J.K. Roberts, A. Silvanovich, Y. Yin, B.E. Weiner, K.C. Glenn and M.L. Odegaard. 2017. The sequence, structural, and functional diversity within a protein family and implications for specificity and safety: The case for ETX_MTX2 insecticidal proteins. Journal of Invertebrate Pathology 142:50-59.

Mogen, B.D., M.H. MacDonald, R. Graybosh and A.G. Hunt. 1990. Upstream sequences other than AAUAAA are required for efficient messenger RNA 3'-end formation in plant. The Plant Cell 2:1261-1272.

Moreno, F.J., F.A. Mellon, M.S.J. Wickham, A.R. Bottrill and E.N.C. Mills. 2005. Stability of the major allergen Brazil nut 2S albumin (Ber e 1) to physciologically relevant *in vitro* gastrointestinal digestion. FEBS Journal 272:341-352.

Narva, K.E., N.X. Wang and R. Herman. 2017. Safety considerations derived from Cry34Ab1/Cry35Ab1 structure and function. Journal of Invertebrate Pathology 142:27-33.

NCPA. 1993. Cottonseed oil. L.A. Jones and C.C. King (eds.). National Cottonseed Products Association, Inc. and The Cotton Foundation, Memphis, Tennessee.

NCPA. 2002. Cottonseed and its products. National Cottonseed Products Association, Cordova, Tennessee. http://www.cottonseed.com/publications/cottonseedanditsproducts.asp [Accessed October 17, 2011].

Nida, D.L., S. Patzer, P. Harvey, R. Stipanovic, R. Wood and R.L. Fuchs. 1996. Glyphosate-tolerant cotton: The composition of the cottonseed is equivalent to that of conventional cottonseed. Journal of Agricultural and Food Chemistry 44:1967-1974.

Niles, G.A. and C.V. Feaster. 1984. Breeding. Pages 201-231 in Cotton. R.J. Kohel and C.F. Lewis (eds.). American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc., Madison, Wisconsin.

OECD. 2002. Report of the OECD workshop on the toxicological and nutritional testing of novel foods. SG/ICGB(1998)1/FINAL. Organisation for Economic Co-operation and Development, Paris, France.

OECD. 2007. Consensus document on safety information on transgenic plants expressing *Bacillus thuringiensis*-derived insect control proteins. ENV/JM/MONO(2007)14. Organisation for Economic Co-operation and Development, Paris, France.

OECD. 2008. Consensus document on the biology of cotton (*Gossypium* spp.). ENV/JM/MONO(2008)33. Organisation for Economic Co-operation and Development, Paris, France.

OECD. 2009. Consensus document on the compositional considerations for new varities of cotton (*Gossypium hirsutum* and *Gossypium barbadense*): Key food and feed nutrients and anti-nutrients. ENV/JM/MONO(2004)16. Organisation for Economic Co-operation and Development, Paris, France.

OECD. 2010a. Section 1. Cotton (*Gossypium* spp.). Safety Assessment of Transgenic Organisms. Volume 4. Organisation for Economic Co-operation and Development, Paris, France.

OECD. 2010b. 6.2 Human health assessment. Pages 234-237 in Safety Assessment of Transgenic Organisms. Volume 4. Organisation for Economic Co-operation and Development, Paris, France.

Okunuki, H., R. Techima, T. Shigeta, J. Sakushima, H. Akiyama, Y. Goda, M. Toyoda and J. Sawada. 2002. Increased digestibility of two products in genetically modified food (CP4 EPSPS and Cry1Ab) after preheating. Journal of the Food Hygienic Society of Japan 43:68-73.

Oosterhuis, D.M. and J. Jernstedt. 1999. Morphology and anatomy of the cotton plant. Pages 175-206 in Cotton: Origin, History, Technology, and Production. C.W. Smith and J.T. Cothren (eds.). John Wiley and Sons, Inc., New York, New York.

Pariza, M.W. and E.A. Johnson. 2001. Evaluating the safety of microbial enzyme preparations used in food processing: Update for a new century. Regulatory Toxicology and Pharmacology 33:173-186.

Percival, A.E., J.F. Wendel and J.M. Stewart. 1999. Taxonomy and germplasm resources. Pages 33-63 in Cotton: Origin, History, Technology, and Production. W.C. Smith and J.T. Cothren (eds.). John Wiley & Sons, Inc., New York, New York.

Pigott, C.R. and D.J. Ellar. 2007. Role of receptors in *Bacillus thuringiensis* crystal toxin activity. Microbiology and Molecular Biology Reviews 71:255-281.

Rademacher, T.W., R.B. Parekh and R.A. Dwek. 1988. Glycobiology. Annual Review of Biochemistry 57:785-838.

Reeves, J.B. and J.L. Weihrauch. 1979. Composition of foods: Fats and oils. Agriculture Handbook 8-4. U.S. Department of Agriculture, Washington, D.C.

Richards, J.S., J.N. Stanley and P.C. Gregg. 2005. Viability of cotton and canola pollen on the proboscis of *Helicoverpa armigera*: Implications for spread of transgenes and pollination ecology. Ecological Entomology 30:327-333.

Richins, R.D., H.B. Scholthof and R.J. Shepherd. 1987. Sequence of figwort mosaic virus DNA (caulimovirus group). Nucleic Acids Research 15:8451-8466.

Ridley, W.P., G.G. Harrigan, M.L. Breeze, M.A. Nemeth, R.S. Sidhu and K.C. Glenn. 2011. Evaluation of compositional equivalence for multitrait biotechnology crops. Journal of Agricultural and Food Chemistry 59:5865-5876.

Rogers, S.G. 2000. Promoter for transgenic plants. Patent 6,018,100, U.S. Patent Office, Washington, D.C.

Salomon, S. and H. Puchta. 1998. Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. EMBO Journal 17:6086-6095.

Schnepf, E., N. Crickmore, J. van Rie, D. Lereclus, J. Baum, J. Feitelson, D.R. Zeigler and D.H. Dean. 1998. *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiology and Molecular Biology Reviews 62:775-806.

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

Sjoblad, R.D., J.T. McClintock and R. Engler. 1992. Toxicological considerations for protein components of biological pesticide products. Regulatory Toxicology and Pharmacology 15:3-9.

Sutcliffe, J.G. 1979. Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. Cold Spring Harbor Symposia on Quantitative Biology 43:77-90.

Thanabalu, T. and A.G. Porter. 1996. A *Bacillus sphaericus* gene encoding a novel type of mosquitocidal toxin of 31.8 kDa. Gene 170:85-89.

Thomas, K., M. Aalbers, G.A. Bannon, M. Bartels, R.J. Dearman, D.J. Esdaile, T.J. Fu, C.M. Glatt, N. Hadfield, C. Hatzos, S.L. Hefle, J.R. Heylings, R.E. Goodman, B. Henry, C. Herouet, M. Holsapple, G.S. Ladics, T.D. Landry, S.C. MacIntosh, E.A. Rice, L.S. Privalle, H.Y. Steiner, R. Teshima, R. van Ree, M. Woolhiser and J. Zawodny. 2004. A multi-laboratory evaluation of a common in vitro pepsin digestion assay protocol used in assessing the safety of novel proteins. Regulatory Toxicology and Pharmacology 39:87-98.

Thomas, K., G. Bannon, S. Hefle, C. Herouet, M. Holsapple, G. Ladics, S. MacIntosh and L. Privalle. 2005. In silico methods for evaluating human allergenicity to novel proteins: International Bioinformatics Workshop Meeting Report, 23-24 February 2005. Toxicological Sciences 88:307-310.

U.S. EPA. 1988. Guidance for the reregistration of pesticide products containing *Bacillus thuringiensis* as the active ingredient. 540/RS-89-023. U.S. Environmental Protection Agency, Washington, D.C.

U.S. EPA. 2010. Biopesticides registration action document: *Bacillus thuringiensis* Cry34Ab1 and Cry35Ab1 proteins and the genetic material necessary for their production (PHP17662 T-DNA) in event DAS-59122-7 corn (OECD unique identifier: DAS-59122-7). U.S. Environmental Protection Agency, Washington, D.C.

Vachon, V., R. Laprade and J.-L. Schwartz. 2012. Current models of the mode of action of *Bacillus thuringiensis* insecticidal crystal proteins: A critical review. Journal of Invertebrate Pathology 111:1-12.

Van Deynze, A.E., F.J. Sundstrom and K.J. Bradford. 2005. Pollen-mediated gene flow in California cotton depends on pollinator activity. Crop Science 45:1565-1570.

Vassilopoulou, E., N. Rigby, F.J. Moreno, L. Zuidmeer, J. Akkerdaas, I. Tassios, N.G. Papadopoulos, P. Saxoni-Papageorgiou, R. van Ree and C. Mills. 2006. Effect of *in vitro* gastric and duodenal digestion on the allergenicity of grape lipid transfer protein. Journal of Allergy and Clinical Immunology 118:473-480.

Venkatesh, T.V., M.L. Breeze, K. Liu, G.G. Harrigan and A.H. Culler. 2014. Compositional analysis of grain and forage from MON 87427, an inducible male sterile and tissue selective glyphosate-tolerant maize product for hybrid seed production. Journal of Agricultural and Food Chemistry 62:1964-1973.

Venkatesh, T.V., G.G. Harrigan, T. Perez and S. Flint-Garcia. 2015. Compositional assessments of key maize populations: B73 hybrids of the Nested Association Mapping founder lines and diverse landrace inbred lines. Journal of Agricultural and Food Chemistry 63:5282-5295.

Vieths, S., J. Reindl, U. Müller, A. Hoffmann and D. Haustein. 1999. Digestibility of peanut and hazelnut allergens investigated by a simple in vitro procedure. European Food Research and Technology 209:379-388.

Wendel, J.F., C. Brubaker, I. Alvarez, R. Cronn and J.M. Stewart. 2009. Evolution and natural history of the cotton genus. Plant Genetics and Genomics 3:3-22.

WHO. 1999. Microbial pest control agent: *Bacillus thuringiensis*. Environmental Health Criteria 217. World Health Organization, Geneva, Switzerland. http://www.inchem.org/documents/ehc/ehc/ehc217.htm [Accessed May 14, 2009].

Yabe, N., T. Takahashi and Y. Komeda. 1994. Analysis of tissue-specific expression of *Arabidopsis thaliana* HSP90-family gene *HSP81*. Plant and Cell Physiology 35:1207-1219.

Yagami, T., Y. Haishima, A. Nakamura, H. Osuna and Z. Ikezawa. 2000. Digestibility of allergens extracted from natural rubber latex and vegetable foods. Journal of Allergy and Clinical Immunology 106:752-762.

Yang, L., C. Wang, A. Holst-Jensen, D. Morisset, Y. Lin and D. Zhang. 2013. Characterization of GM events by insert knowledge adapted re-sequencing approaches. Scientific Reports 3:2839.

Ye, X., E.J. Williams, J. Shen, S. Johnson, B. Lowe, S. Radke, S. Strickland, J.A. Esser, M.W. Petersen and L.A. Gilbertson. 2011. Enhanced production of single copy backbone-free transgenic plants in multiple crop species using binary vectors with a pRi replication origin in *Agrobacterium tumefaciens*. Transgenic Research 20:773-786.

Zambryski, P., A. Depicker, K. Kruger and H.M. Goodman. 1982. Tumor induction by *Agrobacterium tumefaciens*: Analysis of the boundaries of T-DNA. Journal of Molecular and Applied Genetics 1:361-370.

Zastrow-Hayes, G.M., H. Lin, A.L. Sigmund, J.L. Hoffman, C.M. Alarcon, K.R. Hayes, T.A. Richmond, J.A. Jeddeloh, G.D. May and M.K. Beatty. 2015. Southern-by-sequencing: A robust screening approach for molecular characterization of genetically modified crops. The Plant Genome 8:1-15.

Zhou, J., G.G. Harrigan, K.H. Berman, E.G. Webb, T.H. Klusmeyer and M.A. Nemeth. 2011. Stability in the composition equivalence of grain from insect-protected maize and seed from glyphosate-tolerant soybean to conventional counterparts over multiple seasons, locations, and breeding germplasms. Journal of Agricultural and Food Chemistry 59:8822-8828.