Supporting document 1

Safety Assessment Report – Application A1114

Food derived from High Yield Corn Line MON87403

Summary and conclusions

Background

A genetically modified (GM) corn line with OECD Unique Identifier MON-87403-1 (henceforth referred to as MON87403) has been developed by Monsanto Company.

The corn has been modified to have increased ear biomass at an early reproductive phase compared to conventional corn and thus is higher yielding than conventional corn. The modification is achieved through expression of a truncated ATHB17 transcription factor[1] encoded by the *ATHB17* gene from *Arabidopsis thaliana*.

In conducting a safety assessment of food derived from MON87403, a number of criteria have been addressed including: a characterisation of the transferred gene sequences, their origin, function and stability in the corn genome; the changes at the level of DNA, and protein in the whole food; compositional analyses; and evaluation of intended and unintended changes.

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

- environmental risks related to the environmental release of GM plants used in food production
- the safety of animal feed, or animals fed with feed, derived from GM plants
- the safety of food derived from the non-GM (conventional) plant.

History of Use

In terms of production, corn is the world's dominant cereal crop, ahead of wheat and rice and is grown in over 160 countries. It has a long history of safe use in the food supply. Sweet corn is consumed directly while corn-derived products are routinely used in a large number and diverse range of foods (e.g. cornflour, starch products, breakfast cereals and high fructose corn syrup). Corn is also widely used as a feed for domestic livestock.

---

[1] A transcription factor is a protein that binds to specific DNA sequences and regulates gene expression

Molecular Characterisation

MON87403 was generated through *Agrobacterium*-mediated transformation. Comprehensive molecular analyses of MON87403 indicate there is a single insertion site containing a single copy of the *ATHB17* gene plus regulatory elements. The introduced gene is stably inherited from one generation to the next. There are no antibiotic resistance marker genes present in the line and no plasmid backbone has been incorporated into the transgenic locus.

Characterisation and safety assessment of new substances

Newly expressed proteins

MON87403 contains one newly expressed DNA binding protein, ATHB17Δ113. Expression levels were low in all tissues analysed. The highest mean level was in leaf at 0.014 µg/g dw and the lowest mean level was in grain where it was below the limit of detection (<0.3 ng/g).

The identity of the MON87403-produced protein was confirmed by Western blot analysis, sequence analysis of the *ATHB17Δ113* mRNA transcript produced in MON87403, matrix assisted laser desorption/ionization time-of-flight mass spectrometry, and liquid chromatography-tandem mass spectrometry. Indirect evidence also indicated that the MON87403-produced ATHB17Δ113 is not N-glycosylated and that it has the expected functional activity.

Bioinformatic studies confirmed the lack of any significant amino acid sequence similarity to known protein toxins or allergens and digestibility studies have demonstrated ATHB17Δ113 would be completely digested before absorption in the gastrointestinal tract would occur. The protein also loses DNA binding activity with heating. Taken together, the evidence indicates the ATHB17Δ113 protein is unlikely to be toxic or allergenic to humans.

Compositional Analyses

Detailed compositional analyses were done to establish the nutritional adequacy of grain from MON87403 and to characterise any unintended compositional changes. Analyses were done of proximates, fibre, minerals, amino acids, fatty acids, vitamins, secondary metabolites and anti-nutrients. The levels were compared to levels in a) an appropriate non-GM hybrid line b) a tolerance interval compiled from results taken for a total of 17 non-GM hybrid lines grown in the same field trials and c) levels recorded in the literature. None of the 52 analytes that were statistically analysed deviated in level from the control in a statistically significant manner. It can therefore be concluded that grain from line MON87403 is compositionally equivalent to grain from conventional corn varieties.

Conclusion

No potential public health and safety concerns have been identified in the assessment of high yield corn line MON87403. On the basis of the data provided in the present Application, and other available information, food derived from MON87403 is considered to be as safe for human consumption as food derived from conventional corn varieties.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADF	acid detergent fibre
ATHB17	<u>A</u>rabidopsis <u>t</u>haliana <u>h</u>omeo<u>b</u>ox-leucine zipper protein <u>17</u>
BLAST	Basic Local Alignment Search Tool
BLASTP	Basic Local Alignment Search Tool Protein
bp	base pairs
DNA	deoxyribonucleic acid
T-DNA	transferred DNA
dw	dry weight
ELISA	enzyme linked immunosorbent assay
FAO	Food and Agriculture Organization of the United Nations
FARRP	Food Allergy Research and Resource Program
FASTA	Fast Alignment Search Tool - All
FSANZ	Food Standards Australia New Zealand
GM	genetically modified
HD-Zip	Homeodomain-leucine zipper
IgE	Immunoglobulin E
JSA	junction sequence analysis
kDa	kilo Dalton
LB	Left Border of T-DNA
LC-MS/MS	liquid chromatography-tandem mass spectrometry
LOD	Limit of detection
MALDI-TOF MS	matrix-assisted laser desorption/ionisation–time of flight mass spectrometry
NCBI	National Center for Biotechnology Information
NDF	neutral detergent fibre
NGS	next generation sequencing
OECD	Organisation for Economic Co-operation and Development
OGTR	Australian Government Office of the Gene Technology Regulator
ORF	open reading frame
PCR	polymerase chain reaction
RT-PCR	Real time PCR
P or P-value	probability value
RB	Right Border of T-DNA
RNA	ribonucleic acid
mRNA	messenger RNA
SAS	Statistical Analysis Software
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SE	standard error
U.S.	United States of America

1 Introduction

Monsanto Australia Limited has submitted an application to FSANZ to vary Schedule 26 in the *Australia New Zealand Food Standards Code* (the Code) to include food from a new genetically modified (GM) corn line with OECD Unique Identifier MON-87403-1 (referred to as MON87403). The corn has been modified to have increased ear biomass at an early reproductive phase compared to conventional corn.

This modification is achieved through expression of a truncated ATHB17 (Arabidopsis thaliana homeobox-leucine zipper protein 17) transcription factor[2] encoded by the *ATHB17* gene from *Arabidopsis thaliana*. The truncated form (designated ATHB17Δ113) is missing the first 113 N-terminal amino acids of the wild-type protein and modulates certain pathways in the corn ear leading to increased partitioning of photosynthate and hence increased growth at an early stage and therefore increased grain yield at harvest (Rice et al. 2014).

The Applicant has stated that MON87403 is not intended to be a stand-alone product and will be crossed by conventional breeding with other approved GM corn lines (a process known as 'stacking').

The Applicant states the intention is that any lines containing the MON-87403-1 event will be grown in North America, and approval for cultivation in Australia or New Zealand is not being sought. Therefore, if approved, food derived from this line may enter the Australian and New Zealand food supply as imported food products.

2 History of use

2.1 Host organism

Mature corn (*Zea mays*) plants contain both female and male flowers and usually reproduce sexually by wind-pollination. This provides for both self-pollination and natural out-crossing between plants, both of which are undesirable since the random nature of the crossing leads to lower yields (CFIA 1994). The commercial production of corn now utilises controlled cross-pollination of two inbred lines (using conventional techniques) to combine desired genetic traits and produce hybrid varieties known to be superior to open-pollinated varieties in terms of their agronomic characteristics.

This inbred-hybrid concept and resulting yield response is the basis of the modern corn seed industry and hybrid corn varieties are used in most developed countries for consistency of performance and production.

In terms of production, corn is the world's dominant cereal crop (2015 forecast = 1,007 MT[3]) ahead of wheat (723 MT) and rice (499 MT) and is grown in over 160 countries (FAOSTAT3 2015). In 2013, worldwide production of corn was over 1 billion tonnes, with the United States and China being the major producers (~353 and 217 million tonnes, respectively) (FAOSTAT3 2015). Corn is not a major crop in Australia or New Zealand and in 2013, production was approximately 506,000 and 201,000 tonnes respectively (FAOSTAT3 2015). In the U.S. it is estimated that approximately 93% of all corn planted is GM[4] while in Canada, the estimate of GM corn is approximately 80% of the total corn[5].

---

[2] Transcription factors are proteins that bind to specific DNA sequences and regulate gene expression.
[3] FAO Cereal Supply & Demand Brief, http://www.fao.org/worldfoodsituation/csdb/en/
[4] USDA, Economic Research Service, - http://www.ers.usda.gov/data-products/adoption-of-genetically-engineered-crops-in-the-us.aspx
[5] USDA Gain Report, CA14062, 2014 -

No GM corn is currently grown commercially in Australia or New Zealand.

Domestic production is supplemented by the import of corn grain and corn-based products, the latter of which are used, for example, in breakfast cereals, baking products, extruded confectionery and food coatings. In 2011, Australia and New Zealand imported, respectively, 856 and 5,800 tonnes of corn grain, 10,600 and 306 tonnes of frozen sweet corn and 8,427 and 900 tonnes of otherwise-processed sweet corn (FAOSTAT3 2015). Corn product imports to Australia and New Zealand included 6,050 and 2,096 tonnes respectively of corn flour and 3,455 and 13 tonnes respectively of corn oil (FAOSTAT3 2015). Corn is a major source of crystalline fructose and high fructose corn syrup, both of which are processed from corn starch. Approximately 3,000 tonnes of crystalline fructose, but negligible high fructose corn syrup, were imported into Australia in 2011 (Green Pool 2012); neither Australia nor New Zealand currently produce fructose (either crystalline or as high fructose corn syrup).

The majority of grain and forage derived from corn is used as animal feed, however corn also has a long history of safe use as food for human consumption. There are five main types of corn grown for food:

- Flour – *Zea mays* var. *amylacea*
- Flint – *Z. mays* var. *indurata*
- Dent – *Z. mays* var. *indentata*
- Sweet – *Z. mays* var. *saccharata* & *Z. mays* var. *rugosa*
- Pop – *Z. mays* var. *everta*

Dent corn is the type most commonly grown for grain and silage and is the predominant type grown in the U.S. (OGTR 2008). The parent line that was transformed to give MON87403 is a conventional corn hybrid line (LH244) resulting from a cross between the inbred lines LH197 and LH199 followed by a backcross to LH197. LH244 is a patented corn line assigned to Holden's Foundation Seeds LLC in 2001 (Armstrong 2001). It is a medium season, yellow dent corn line that is adapted to the central regions of the U.S. corn-belt.

Two main grain processing routes are followed for dent corn (White and Pollak 1995):

- Dry milling that gives rise to food by-products such as flour and hominy grits.
- Wet milling (CRA 2006), that involves steeping the grain, coarse and fine grinding, centrifugation and evaporating the steep, to yield food by-products such as starch (for corn starch, corn syrup and individual sweeteners such as dextrose and fructose) and germ (for oil) – see Figure 1. Corn products are used widely in processed foods.

---

http://gain.fas.usda.gov/Recent%20GAIN%20Publications/Agricultural%20Biotechnology%20Annual_Ottawa_Canada_7-14-2014.pdf
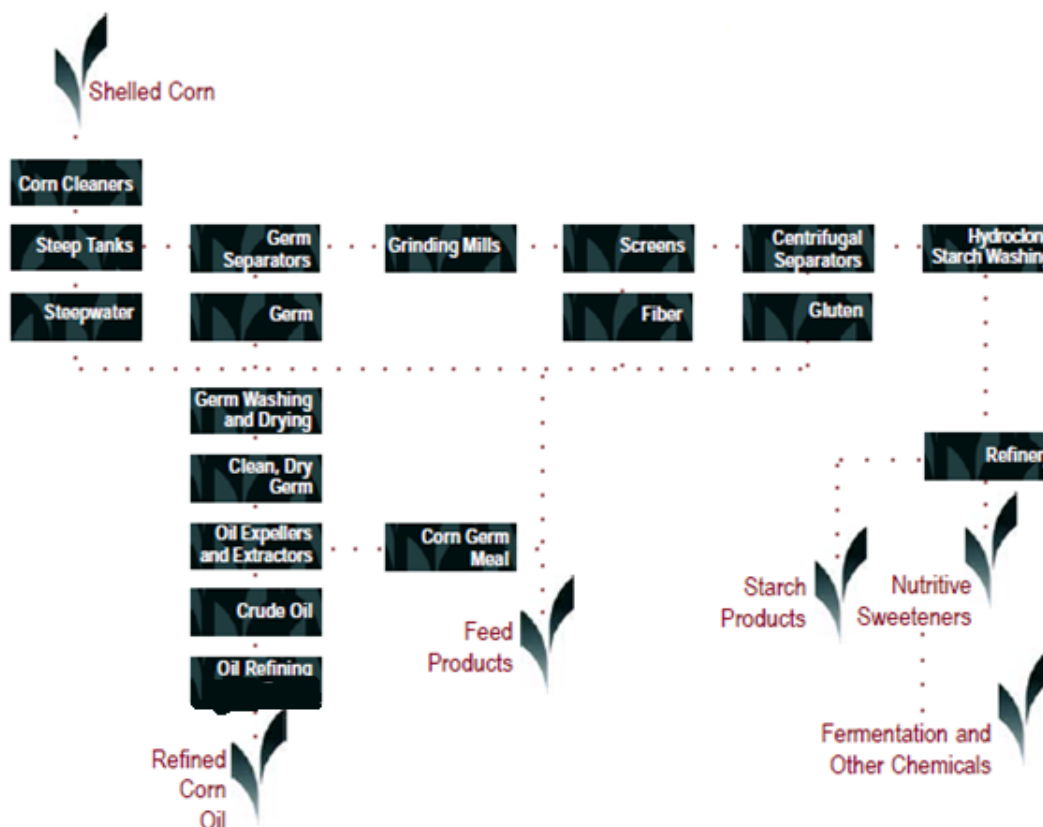
Figure 1: The corn wet milling process (diagram taken from CRA (2006))

2.2   Donor organisms

2.2.1   Arabidopsis thaliana

The donor organism for the *ATHB17* gene is *Arabidopsis thaliana* (common names: thale cress, mouse ear cress). *A. thaliana* is a small flowering plant belonging to the mustard (Brassicaceae) family, which includes cultivated species such as broccoli, cabbage, canola and radish. *A. thaliana* is widely used as a model organism in plant biology and genetics and its genome was the first plant genome to be fully sequenced.

2.2.2   Other organisms

Genetic elements from several other organisms have been used in the genetic modification of MON87403 (refer to Table 1). These non-coding sequences are used to drive, enhance or terminate expression of the novel genetic material. None of the sources of these genetic elements is associated with toxic or allergenic responses in humans. The genetic elements derived from the plant pathogen Cauliflower mosaic virus are not pathogenic in themselves and do not cause pathogenic symptoms in MON87403.

3   Molecular characterisation

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

- the transformation method together with a detailed description of the DNA sequences introduced to the host genome

- a characterisation of the inserted DNA including any rearrangements that may have occurred as a consequence of the transformation
- the genetic stability of the inserted DNA and any accompanying expressed traits.

Studies submitted:

2013. Amended Report for MSL0025316: Molecular Characterization of MON 87403. **MSL0025909.** Monsanto Company (unpublished)

2014. Bioinformatics evaluation of the transfer DNA insert in MON87403 utilizing the AD_2014, TOX_2014 and PRT_2014 databases. **MSL0025648**. Monsanto Company (unpublished)

2014. Bioinformatics Evaluation of DNA Sequences Flanking the 5' and 3' Junctions of Inserted DNA in MON 87403: Assessment of Putative Polypeptides. **MSL0025733**. Monsanto Company (unpublished)

2013. Segregation of the T-DNA Insert in MON 87403 Across Three Generations. **MSL0024676.** Monsanto Company (unpublished)

3.1 Method used in the genetic modification

Immature embryos from line LH244 were aseptically removed from 10 – 13 day post-pollination ears and transformed, using a disarmed strain (ABI) of *Agrobacterium tumefaciens,* with the T-DNA from plasmid vector PV-ZMAP5714 (see Figure 2) following the method of Sidorov and Duncan (2009). A tandem T-DNA approach (Huang et al. 2004) was used whereby a single Right Border (RB) and a single Left Border (LB) were used to achieve separate unlinked insertions of the T-DNA (containing the *ATHB17* gene) and a *cp4epsps* selectable marker gene located in the plasmid backbone.

After co-culturing with the *Agrobacterium* carrying the vector, the embryos were placed on selection medium containing glyphosate, and carbenicillin, to inhibit the growth of untransformed plant cells and excess *Agrobacterium*, and to permit the development of callus tissue. Resulting callus was then placed in a medium that supported shoot regeneration and root development. Rooted plants (generation R_0) with normal phenotypic characteristics and containing both the glyphosate (*cp4epsps*) expression cassette and the *ATHB17* expression cassette were self-pollinated to produce R1 seed. Plants that grew from this seed were then further selected by a polymerase chain reaction (PCR) method to eliminate those containing the *cp4epsps* cassette. Only those plants homozygous for the *ATHB17* expression cassette and not containing the *cp4esps* cassette were then selected for further assessment and development (see Section 3.3). From many hundreds of events MON87403 was ultimately chosen as the lead event based on superior agronomic, phenotypic and molecular characteristics.

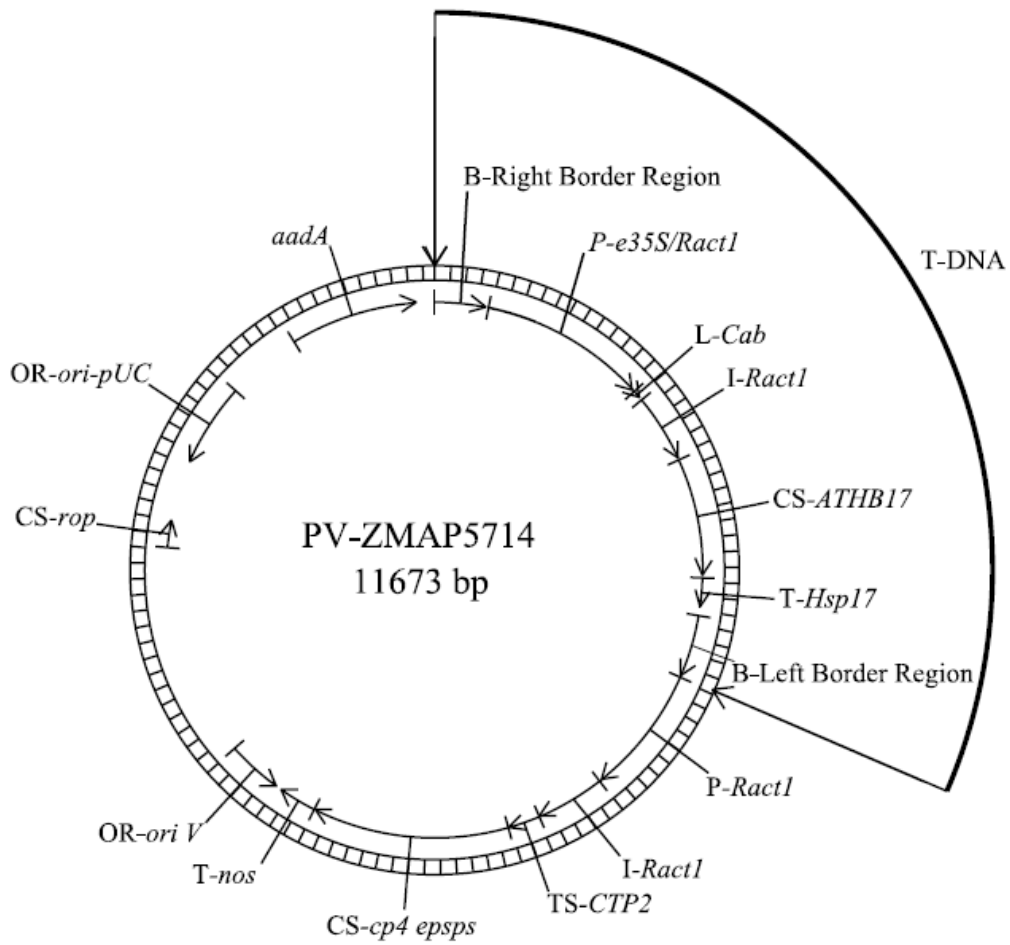*Figure 2: Genes and regulatory elements contained in plasmid PV-ZMAP5714*

3.2   Function and regulation of introduced genes

Information on the genetic elements in the T-DNA used for transformation is summarised in

Table 1. There is a single cassette comprising a total of 2,877 bp located between a 442 bp LB and a 357 bp RB. The complete plasmid is 11,673 bp in size (i.e. the vector backbone comprises 7,997 bp).

Table 1: Description of the genetic elements contained in the T-DNA of PV-ZMAP5714

Genetic element	Relative nt location on plasmid	Size (bp)	Source	Orient.	Description & Function	References
RIGHT BORDER	1 - 357	357				
ATHB17 cassette						
Intervening sequence	358 - 375	18				
e35S/Ract1	376 - 1556	1,181	Cauliflower mosaic virus (CaMV) and *Oryza sativa* (rice)	Clockwise	Chimeric promoter consisting of the duplicated enhancer region from the CaMV 35S RNA promoter combined with the promoter of the *actin1* gene from rice. • Directs transcription of the *ATHB17* gene	Kay *et al.* (1987); McElroy *et al.* (1990)
Intervening sequence	1557 - 1561	5				
Cab	1562 - 1622	61	*Triticum aestivum* (wheat)	clockwise	• 5' untranslated region leader sequence from chlorophyll a/b binding protein • Regulates expression of the *ATHB17* gene	Lamppa *et al.*(1985)
Intervening sequence	1623 - 1638	16				
Ract1	1639 - 2118	480	*Oryza sativa* (rice)	Clockwise	• Intron and flanking untranslated region sequence of the *actin 1* gene • Regulates expression of the *ATHB17* gene	McElroy *et al.* (1990)
Intervening sequence	2119 - 2130	12			•	
ATHB17	2131 - 2958	828	*Arabidopsis thaliana*	Clockwise	• Coding sequence of the *ATHB17* gene	Ariel *et al.* (2007)
Intervening sequence	2959 - 2971	13				
Hsp17	2972 - 3181	210	*Triticum aestivum* (wheat)	Clockwise	• 3' untranslated region from a heat shock protein • Terminates mRNA transcription and induces polyadenylation of the *ATHB17* mRNA	McElwain & Spiker (1989)
Intervening sequence	3182 - 3234	53				
LEFT BORDER	3235 - 3676	442				

3.2.1 *ATHB17* expression cassette

The *ATHB17* coding region is 828 bp in length and is driven by a constitutive chimeric promoter (*e35S/Ract 1*) made up of elements of the CaMV 35S RNA promoter and the rice actin 1 promoter. Expression of *ATHB17* is enhanced by the use of 5' untranslated regions taken from the wheat chlorophyll a/b binding protein gene (*Cab*) and the rice actin 1 gene (*Ract1*). A sequence from the 3' untranslated region of a wheat heat shock protein (*Hsp17*) gene terminates transcription.

3.3 Breeding of MON87403

The breeding pedigree for the various generations is given in Figure 3.

From a single R_0 plant, several rounds of self-pollination and seed bulking proceeded in order to produce specific generations that were used in characterisation and analysis (as indicated in Table 2). At the R_4 and R_5 generations, plants were crossed with conventional proprietary hybrid lines and the progeny were used to generate information on insert

stability.
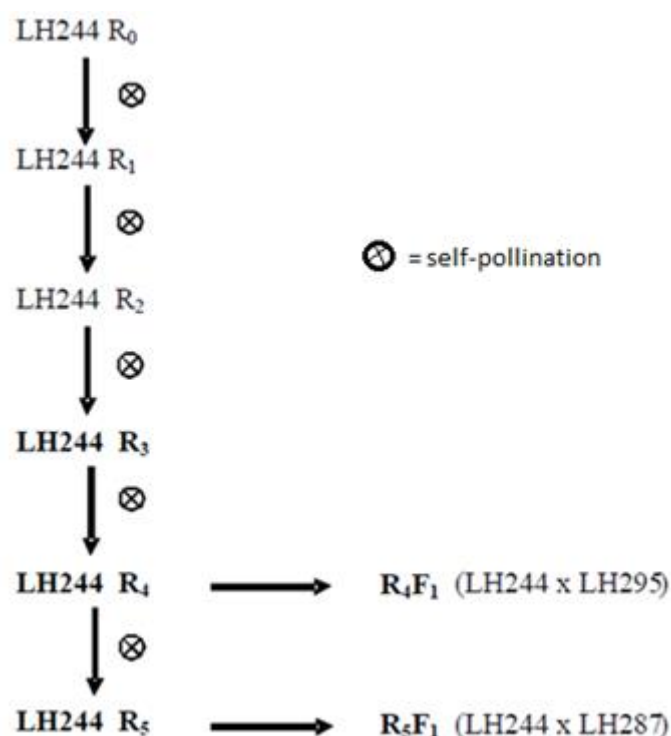
Figure 3: Breeding diagram for MON87403

Table 2: MON87403 generations used for various analyses

Analysis	MON87403 generation used	Control(s) used	Reference comparators
Molecular characterisation (Section 3.4)	R_3	LH244	
Genetic stability (Section 3.5.1)	$R_3, R_4, R_4F_1, R_5, R_5F_1$	LH244, LH244 x LH295; LH244 x LH287	
Mendelian inheritance (Section 3.5.2)	$BC_1F_1, BC_2F_1, BC_3F_1$ (see Fig 6)	N/A	
Protein characterisation (Section 4.1.3)	Protein from R_5F_1	Protein from *E. coli*	
Protein expression levels in plant parts (Section 4.1.4)	R_5F_1	LH244 x LH287	
Compositional analysis (Section 5)	R_5F_1	LH244 x LH287	13 non-GM commercial reference lines

3.4   Characterisation of the genetic modification in the plant

A range of analyses were undertaken to characterise the genetic modification in MON87403. These analyses focussed on the nature of the insertion of the introduced genetic elements and whether any unintended genetic re-arrangements may have occurred as a consequence of the transformation procedure.

The molecular characterisation of MON87403 incorporated an approach (Kovalic et al. 2012; DuBose et al. 2013) that applies Next Generation Sequencing (NGS) and Junction Sequence Analysis (JSA) together with bioinformatics to determine the number of inserts. In the past, this has been determined by Southern blot analysis. The organisation and sequence of the insert and adjacent flanking DNA, and the sequencing of the insertion site were all determined by methods employing directed sequencing.

The rationale for junction sequence analysis is that, since junctions are characteristic of DNA insertion, it follows that each insertion will produce two (i.e. 5' and 3') unique junction sequences. By evaluating the number of unique junctions detected, the number of insertion sites can be determined. In addition to this, information can be obtained about the presence or absence of any backbone sequences.

The Applicant supplied the flow-diagram shown in Figure 4 to illustrate this approach to molecular characterisation.

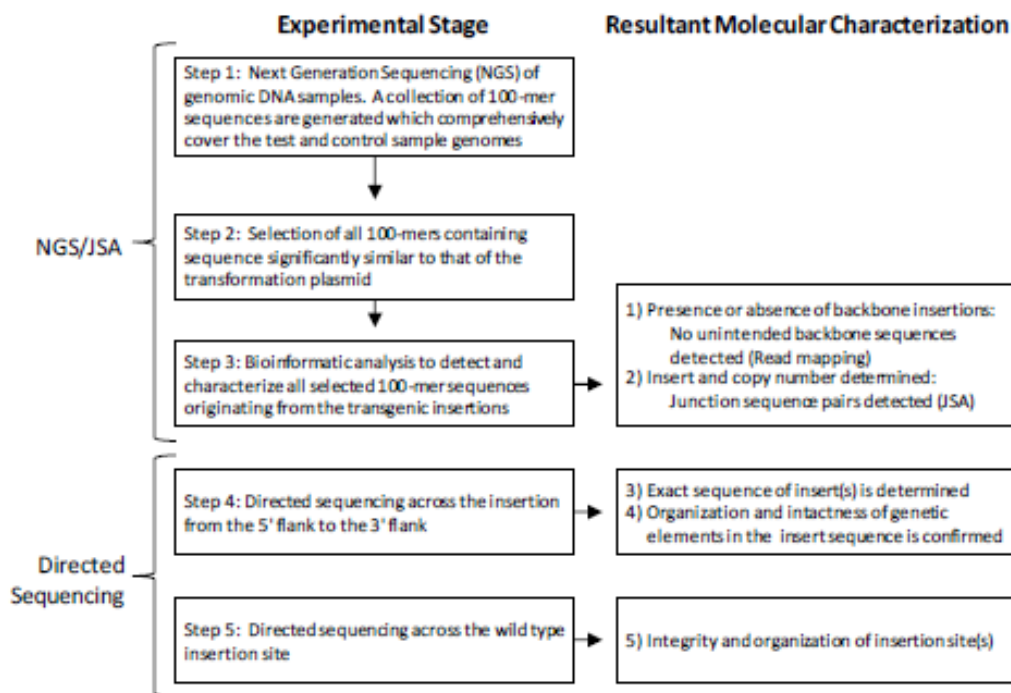Genomic DNA from the MON 87403 test and the conventional control was sequenced using technology that produces a large set of short, randomly distributed sequence reads (each approximately 100 bp long) that comprehensively cover both genomes (Step 1). Utilising these genomic sequence reads, bioinformatics search tools were used to select all sequence reads that were significantly similar to the transformation plasmid (Step 2) for use in read mapping to determine the presence/absence of backbone sequences and Junction Sequence Analysis (JSA) bioinformatics to determine the insert and copy number (Step 3). Overlapping PCR products are also produced which span any inserts and their wild type locus (Step 4 and Step 5 respectively); these overlapping PCR products are sequenced to allow for detailed characterisation of the inserted DNA and insertion site(s).

Figure 4: Steps in the molecular characterisation of MON87403

3.4.1 Insert number and backbone presence

Total genomic DNA from seed of verified MON87403 (generation R_3) and the untransformed parent (LH244) was sequenced using Illumina®[6] NGS technology. Reference DNA was also used from the plasmid vector PV-ZMAP5714. As a positive control, plasmid DNA was spiked into LH244 DNA at a single copy genome equivalent ratio and 1/10 copy genome equivalent ratio. It was noted from the subsequent positive control results, that any portion of the plasmid could be detected at both single copy and 1/10 copy; this indicated there was adequate sensitivity to be able to observe any inserted fragment.

The DNA was sheared into approximately 325 bp fragments with 5' or 3' overhangs, processed for deep sequencing (end-repaired, A-tailed and ligated to adapters), enriched through ten cycles of PCR and then sequenced using Illumina HiSeq® technology that produces short-sequence reads approximately 100 bp long. To confirm sufficient sequence coverage in the samples, the 100-mer sequence reads from all samples were analysed to determine the effective depth of coverage (i.e. the average number of times any base of the genome is expected to be independently sequenced) by mapping all reads to a known single-copy endogenous corn pyruvate decarboxylase gene (*pdc3*). The analysis showed that *pdc3* was covered by the 100-mers at ≥106x for each sample, a coverage that is greater than that of the 75x considered to be comprehensive (Kovalic et al. 2012), and established that the sequencing was adequate for analysis of the sample genomes.

An *in silico* analysis using the BLAST[7] algorithm then followed, in order to collect those 100-mer reads containing sequence similarity to the plasmid PV-ZMAP1574 i.e. this analysis found all 100-mer reads that were either fully matched to the insert plasmid sequences or contained both plasmid sequences and junction sequences.

Junction sequence analysis

The analysis collected all sequencing reads with an e-Value[8] of less than 1e-5 and at least 30 bases match of greater than 96.7% identity to the transformation plasmid (Kovalic et al. 2012) i.e. it collected all reads that were either fully matched to the insert or contained at least 30 bp of insert plus junction sequence. Using Bowtie[9] short sequence alignment software, non-duplicated reads were collected and were further processed *in silico* to remove adapters (Novoalign software[10]) and low quality read ends (Phred score[11] < 12).The remaining reads were then aligned to the whole plasmid PV-ZMAP5714 sequence in order to find junction region sequences (Figure 5). Reads were also aligned against the control genome in order to remove those reads sourced from endogenous homologues.

Figure 5 shows a map of the junction sequences (illustrated as stacked bars) that were detected. Each detected junction sequence read is shown trimmed to include only 30 bp of plasmid sequence. Only two unique junction sequence classes, both containing portions of T-DNA and flanking sequence were detected. This indicates that MON87403 contains a single DNA insert. No junction sequences were found in the DNA from LH244.

---

[6] http://www.illumina.com/technology/next-generation-sequencing/sequencing-technology.ilmn
[7] BLAST is the acronym for Basic Local Alignment Search Tool (Altschul et al. 1990), a computer algorithm that can rapidly align and compare a query DNA sequence with other DNA sequences..
[8] Comparisons between highly homologous proteins yield *E*-values approaching zero, indicating the very low probability that such matches would occur by chance. A larger *E*-value indicates a lower degree of similarity.
[9] http://bowtie-bio.sourceforge.net/index.shtml
[10] http://www.novocraft.com/main/index.php
[11] Phred is a base calling programme for DNA sequences. Phred quality scores have become widely accepted to characterise the quality of sequences (http://www.phrap.com/phred/) .
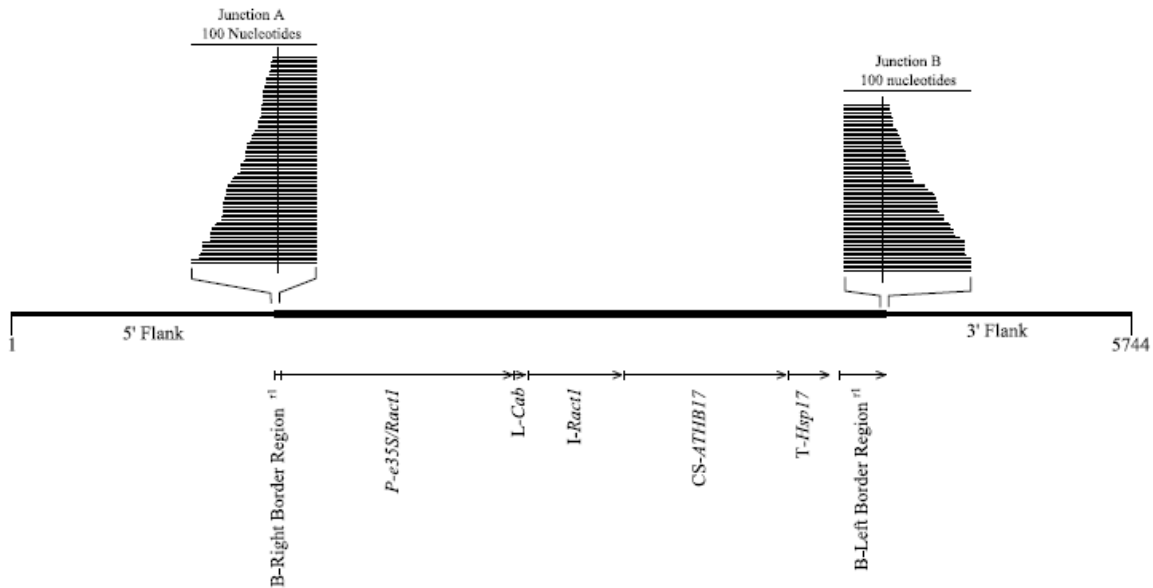
Figure 5: Schematic representation of the junction sequences detected in MON87403

Backbone analysis

Through mapping the sequence reads obtained from MON87403 and the control, to plasmid PV-ZMAP5714 , the presence/absence of backbone sequences was also determined. For a single insert at a single locus, few if any reads aligning with plasmid backbone would be expected. In fact, 1,111 of the over 3,500 reads in total, obtained from MON87403, and 2 reads from the conventional control did align with sequences in the backbone. However, of the 1,111 reads, 1,107 of these alignments were to *Ract1* sequences that were also present in the T-DNA (see Figure 2). The fact that a) full sequence analysis (Section 3.4.2) indicated no *Ract1* sequences were present in the junctions and b) over 3,500 reads in total aligned with the T-DNA (compared to the 2 reads from the control and 4 from MON87403 that were scattered across the length of the backbone) leads to the weight-of-evidence conclusion that MON87403 does not contain plasmid backbone sequences.

3.4.2   Insert organization and sequence

PCR primers were designed to amplify three overlapping regions of MON87403 (generation R_4) genomic DNA incorporating the insert and flanking regions [no products were obtained for DNA from LH244]. The products were used to determine the nucleotide sequence of the insert and flanking regions using BigDye® Terminator chemistry[12]. A consensus sequence was generated by compiling sequences from multiple sequencing reactions performed on the overlapping PCR products. The consensus sequence was then aligned to the PV-ZMAP5714 sequence to determine the integrity and organisation of the insert and flanking regions.

The results showed that the insert is 3,132 bp in length and comprises the identical 2,877 bp *ATHB17* expression cassette sequence found within the T-DNA I of plasmid PV-ZMAP5714 together with terminally truncated RB and LB regions (RB missing the last 333 bp and LB missing the first 211 bp – see Table 1 for details of full sequence length). This analysis also confirmed the conclusion from the NGS/JSA analysis that a single copy of T-DNA I has been inserted and no backbone sequence or sequence from T-DNA II is present.

---

[12] http://www.appliedbiosystems.com.au/

In addition to the insert, 1,345 bp flanking the 5' end of the insert and 1,267 bp flanking the 3' end of the insert were sequenced.

3.4.3 The insertion site

In order to identify any changes to the genomic DNA as a result of the insertion event, two primers (one specific to the 5' flanking sequence of the MON87403 and one specific to the 3' sequence) were used for PCR of genomic DNA isolated from the untransformed parent (LH244). The product (approximately 2,200 bp) was then sequenced and the sequence was compared with the sequences obtained for the 5' and 3' flanking regions of MON87403. The results showed that a 149 bp deletion had occurred during transformation. This deletion is not uncommon and most likely results from double-stranded break repair in the plant during the *Agrobacterium*-mediated transformation process (Salomon and Puchta 1998).

3.4.4 Open reading frame (ORF) analysis

An *in silico* analysis of the flanking regions was done to determine whether any novel ORFs had been created in MON87403. Sequences spanning the 5' and 3' junctions of the MON87403 insert were translated using DNAStar software[13] from stop codon to stop codon (TGA, TAG, TAA) in all six reading frames. A total of 10 ORFs (four in the 5' junction and six in the 3' junction) ≥ 8 amino acids in length were identified that encode putative polypeptides ranging in size from 14 – 119 amino acids. No analysis was done to determine whether any potential regulatory elements were associated with the polypeptides. The putative polypeptides encoded by the 10 identified ORFs were then analysed using a bioinformatic strategy to determine similarity to known protein toxins or allergens (refer to Section 4.1.7).

For the DNA in the T-DNA insert, the DNA sequences in the sense and anti-sense strands were translated to yield six reading frames and all sequences were then translated using DNAStar, EditSeq (Version 10.0.1). The resultant amino acid sequences were used for bioinformatic analyses described in Section 4.1.7.

3.5 Stability of the genetic changes in MON87403

The concept of stability encompasses both the genetic and phenotypic stability of the introduced trait over a number of generations. Genetic stability refers to maintenance of the modification (as produced in the initial transformation event) over successive generations. It is best assessed by molecular techniques, such as Southern blot analysis or NGS/JSA. Phenotypic stability refers to the expressed trait remaining unchanged over successive generations. It is often quantified by a trait inheritance analysis to determine Mendelian heritability via assay techniques (chemical, molecular, visual).

3.5.1 Genetic stability

The genetic stability of event MON87403 was evaluated by NGS/JSA (as described in Section 3.4.1) in verified genomic DNA isolated from the grain of plants of five generations (refer to Figure 3) as given in Table 3. Reads for the generations additional to R_3 were compared to the fully characterised R_3 generation. The depth of coverage for all samples was ≥106. Control genomic DNA was isolated from the non-GM parental line and two other conventional hybrid lines as indicated in Table 3.

---

[13] http://www.dnastar.com/

Table 3: Source of genomic DNA used for genetic stability analysis

DNA source	Generation
MON87403	R_3
	R_4
	R_4F_1
	R_5
	R_5F_1
LH244	
LH244 x LH295	
LH244 x LH287	

No junction sequences were detected in DNA obtained from the control lines (LH244, LH244 x LH295; LH244 x LH287). Analysis of the MON87403 DNA from all generations showed the presence of the same two junction sequences identified in Section 3.4.1. No other junction sequences were present. The consistency of this junction sequence data across all generations tested, demonstrates that the single insert is stably maintained in MON87403.

3.5.2   Phenotypic stability

Since it was demonstrated that the insert resides at a single locus within the MON87403 genome, the expectation would be that the genetic material within it would be inherited according to Mendelian principles.

Chi-square (X^2) analysis was undertaken over several generations to confirm the segregation and stability of the complete T-DNA sequence within the insert. The breeding path followed for this analysis was different from that represented in Figure 3 and is shown in Figure 6. The inheritance of the MON 87403 T-DNA was assessed in the BC1F1, BC2F1, and BC3F1 generations using End-Point TaqMan PCR.
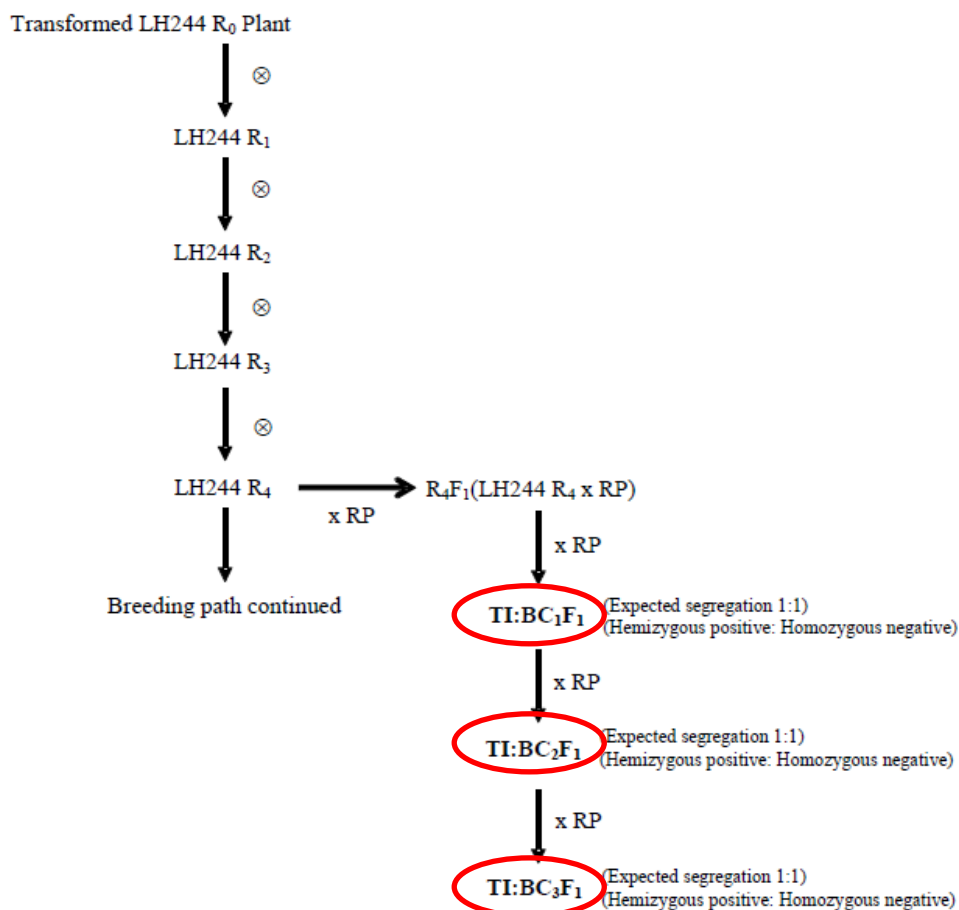
TI: Trait Integration: Replacement of genetic background of MON 87403 by recurrent background except inserted gene.
RP: Recurring parent
BC: Back-Cross.
⊗: Self- Pollinated.

Figure 6: Breeding path for generating segregation data for MON87403

The results (Table 4) indicated there were no significant differences between the observed and expected segregation ratios in any of the generations. This supported the conclusion that the T-DNA resides at a single locus and showed that the T-DNA is inherited according to Mendelian principles.

Table 4: Segregation of the MON87403 T-DNA sequences over three generations

Generation	Total plants	Ratio[1]		X^2	Probability (P)[2]
		Observed	Expected		
BC_1F_1	180	1:1.04	1:1	0.09	0.766
BC_2F_1	178	1:1.17	1:1	1.10	0.294
BC_3F_1	181	1:0.79	1:1	2.44	0.119

[1] Ratio is hemizygous positive:homozygous negative for each of observed and expected
[2] Statistical significance is when P≤0.05

3.6   Antibiotic resistance marker genes

No antibiotic marker genes are present in corn line MON87403. The insert sequence analysis (Section 3.4.2) showed no plasmid backbone has been integrated into the MON87403 genome during transformation, i.e. the *aadA* gene, which was used as a bacterial selectable marker gene, is not present in MON87403.

3.7   Conclusion

Comprehensive molecular analyses of MON87403 indicate there is a single insertion site containing a single copy of the *ATHB17* gene plus regulatory elements. The introduced gene is stably inherited from one generation to the next. There are no antibiotic resistance marker genes present in the line and no plasmid backbone has been incorporated into the transgenic locus.

4   Characterisation and safety assessment of new substances

In considering the safety of newly expressed proteins it is important to note that a large and diverse range of proteins are ingested as part of the normal human diet without any adverse effects, although a small number have the potential to impair health, e.g., because they are allergens or anti-nutrients (Delaney et al. 2008). As proteins perform a wide variety of functions, different possible effects have to be considered during the safety assessment including potential toxic, anti-nutritional and allergenic effects. To effectively identify any potential hazards requires knowledge of the characteristics, concentration and localisation of all newly expressed proteins in the organism as well as a detailed understanding of their biochemical function and phenotypic effects. It is also important to determine if the newly expressed protein is expressed as expected, including whether any post-translational modifications have occurred.

Two types of proteins were considered:

- The protein expected to be directly produced as a result of the translation of the introduced gene. A number of different analyses were done to characterise this protein and determine *in planta* expression.
- Those that may be potentially translated as a result of the creation of ORFs during the transformation process (see Section 3.4.4).

4.1   Newly expressed protein

4.1.1   The ATHB17Δ113 protein

Transcription factors work by binding to DNA at enhancer sites and/or to other proteins in the initiation complex that must assemble in order for mRNA to be transcribed from DNA. They are either activators or repressors of transcription. Transcription factors are encoded by genes referred to as homeobox genes and they contain a highly conserved 60 – 61 amino acid homeodomain (HD) that is involved in DNA binding. Homeobox genes are present in many animal and plant systems and their products control a vast array of developmental processes by directly modulating gene expression.

The ATHB17 protein is a member of the homeodomain-leucine zipper (HD-Zip) family of transcription factors which is unique to plants (Schena and Davis 1992; Ariel et al. 2007; Elhiti and Stasolla 2009; Harris et al. 2011).

The term HD-Zip indicates that there is a leucine zipper[14] adjacent to the C-terminus of the HD (Mukherjee et al. 2009; Elhiti and Stasolla 2009). This Zip domain is responsible for dimerization and thus HD-Zip proteins function as homodimers (binding with themselves) or as heterodimers (binding with other HD-Zip proteins) both of which help to enhance high affinity DNA binding.

Some 48 of 110 homeobox genes in *Arabidopsis* have been defined as encoding HD-Zip proteins (Mukherjee et al. 2009). The HD-Zip family comprises four sub-families (Mukherjee et al. 2009) and ATHB17 belongs to class II (HD-ZIP II). The precise role of HD-Zip II proteins is still not fully known but they appear to be involved in the developmental regulation of apical meristems and lateral organ polarity (see references in Rice et al. 2014). Fourteen endogenous HD-Zip II genes have been identified in corn of which all are expressed in at least reproductive tissues, with eight being predominantly expressed in reproductive tissues (Rice et al. 2014).

In Arabidopsis, the ATHB17 protein functions as a transcriptional repressor and is implicated in abscisic acid and water-stress responses (Park et al. 2013). All transcription factors are modular proteins composed of distinct functional domains. The major domains of the ATHB17 protein comprise:

- the HD, which as in all HD-Zip II proteins, recognises, and therefore binds to, a pseudopalindromic 9 bp DNA sequence CAAT(C/G)ATTG (Sessa et al. 1993).
- the Zip domain which permits the formation of homo- and hetero-dimers.
- a repressor domain which is key to the protein's function as a repressor. A repressor domain is a polypeptide sequence that represses transcription when it is fused to a DNA-binding domain. In ATHB17 it contains a unique 73 amino acid N-terminal extension rich in tyrosine and cysteine that has not been identified in any other HD-Zip II proteins.

The ATHB17 protein normally expressed in *A. thaliana* by the *ATHB17* gene is 275 amino acids in length (see Figure 7). However, during the characterisation of the mRNA of *ATHB17* expressed in MON87403, by Real Time PCR (RT-PCR) followed by Sanger sequencing (Rice et al. 2014), it was found that pre-mRNA splicing in the corn had resulted in the removal of the *Ract1* intron (refer to Table 1) along with a portion of the *ATHB17* gene. Further characterisation using a BLASTX[15] search to compare the *ATHB17* mRNA expressed in MON87403 with that expressed in the wild type *Arabidopsis* predicted that the MON87403 ATHB17 protein would lack the N-terminal 113 amino acids (see Figure 7), which contains a significant portion of the repression domain, and would therefore be 162 amino acids long and have a predicted molecular weight of approximately 22 kDa. This predicted protein was given the designation ATHB17Δ113. Since a large part of the repression domain is lacking, this means ATHB17Δ113 cannot repress the transcription of target genes.

---

[14] Leucine zipper is a motif comprising a periodic repetition of a leucine residue at every seventh position and forms an α-helical conformation (Landschulz et al. 1988). The Zip motif is also found in proteins other than transcription factors and is thought to be one of the general modules for protein-protein interactions.
[15] BLASTX = Basic Local Alignment Search Tool X, a bioinformatic programme that is used to search a protein database using a translated nucleotide query

```
  1 MIKLLFTYIC TYTYKLYALY HMDYACVCMY KYKGIVTLQV CLFYIKLRVF
 51 LSNFTFSSSI LALKNPNNSL IKIMAILPEN SSNLDLTISV PGFSSSPLSD
101 EGSGGGRDQL RLDMNRLPSS EDGDDEEFSH DDGSAPPRKK LRLTREQSRL
151 LEDSFRQNHT LNPKQKEVLA KHLMLRPRQI EVWFQNRRAR SKLKQTEMEC
201 EYLKRWFGSL TEENHRLHRE VEELRAMKVG PTTVNSASSL TMCPRCERVT
251 PAASPSRAVV PVPAKKTFPP QERDR
```

Figure 7: Amino acid sequence of the ATHB17 protein (shaded + unshaded) and the truncated ATHB17Δ113 protein (unshaded)

Experimental work by Rice et al (2014) has suggested a mechanism by which ATHB17Δ113 functions in MON87403 (see Figure 8). The truncated protein, although losing its repression domain, does retain both the HD domain (i.e. ability to bind to the CAAT(C/G)ATTG DNA sequence) and Zip domain (i.e. ability to form dimers). As a result, it is able to interact with endogenous corn HD-Zip II proteins through a dominant-negative mechanism that results in the heterodimer either being unable to bind at all to the enhancer DNA site, or being able to bind but be less active and therefore have reduced repressor capability. In addition, because ATHB17Δ113 homodimers can still bind to the target sequence, they can at times 'outcompete' the endogenous HD-ZIP II dimers at the enhancer DNA site.

The consequences of the presence of ATHB17Δ113 is that the endogenous HD-Zip II proteins have their repression activity either nullified or reduced and this, in turn, leads to positive changes in ear inflorescence growth. The changes are small, would not be expected to have an effect on global transcription in corn and do not cause undesirable phenotypic off-types.
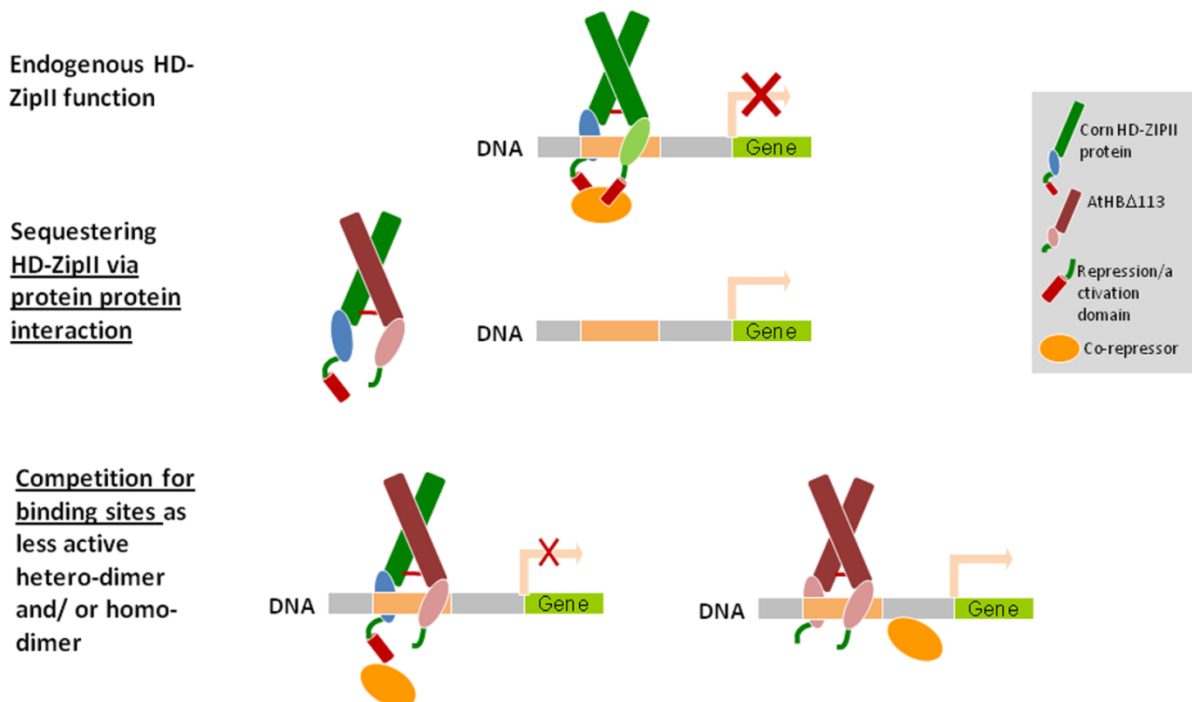

Figure 8: Proposed mechanism of action of the ATHB17Δ113 protein (diagram taken from Figure S2 of Rice et al (2014)

4.1.2   Assessment of the effectiveness of the genetic modification

The Applicant provided a study assessing the effect of the expression of ATHB17Δ113 in MON87403 on ear biomass.

> **Study submitted:**
>
> 2014. Assessment of R1 ear biomass of maize MON 87403 in 2012 U.S. field trials. **MSL0026473**, Monsanto Company (unpublished).

Plants of MON87403 (generation R_5F_1) and the control (LH244 x LH287) were grown from verified seed lots at 13 field sites in the U.S. during the 2012 growing season. These plantings overlapped with the five plantings used for the expression analysis described in Section 4.1.4 and with eight plantings used for the compositional analysis described in Section 5.2. At each site, ear biomass was measured at the R1 growth stage. In a combined-site analysis, the mean ear biomass of MON87403 (144.5 g) was significantly ($P<0.05$) higher than the mean ear biomass of the control (129.3 g). This supported the product concept for MON87403.

4.1.3 ATHB17Δ113 characterisation, and equivalence of the protein produced *in planta* and in a bacterial expression system

> **Studies submitted:**
>
> 2014. Amended report for MSL0025447: Characterization of the ATHB17Δ113 protein purified from the maize leaf of MON87403 and comparison of the physicochemical properties of the plant-produced and *Escherichia coli* (*E. coli*)-produced ATHB17Δ113 proteins. **MSL0025829**, Monsanto Company (unpublished).
> 2015. Characterization of the MON87403-produced ATHB17Δ113 protein by mass spectrometry. **MSL0026621**, Monsanto Company (unpublished).
> 2015. Characterization of the DNA binding activity of the *Escherichia coli*-produced ATHB17Δ113 protein. **MSL0026645**, Monsanto Company (unpublished).

The amount of ATHB17Δ113 protein produced in MON87403 was insufficient for safety evaluations. Therefore, the ATHB17Δ113 protein for these evaluations was produced in *Escherichia coli.* In order to confirm that the *E. coli*-produced ATHB17Δ113 is equivalent to that expressed in MON87403, two analytical techniques were directly employed – a) determination of the molecular weight, and b) immunoreactivity.

The *E. coli*-derived protein was purified from transgenic *E. coli* containing the confirmed coding sequence for expression of ATHB17Δ113 (see Figure 7). The MON87403-derived protein was purified from lyophilised leaf material of generation R_5F_1. It is noted that due to the very low amount of ATHB17Δ113 present in leaves, the protein could not be isolated to a high level of purity and, in fact, was only approximately 3% pure. The low amount and low purity identify ATHB17Δ113 as an 'intractable' protein (Bushey et al. 2014) that is therefore difficult to assess for safety by more routine methods.

The identification of the MON87403-produced protein was confirmed by a) sequence analysis of the *ATHB17Δ113* mRNA transcript produced in MON87403 (see discussion Section 4.1.1) b) matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) of the protein, and c) liquid chromatography-tandem mass spectrometry (LC-MS/MS) – which replaced the more traditional Edman degradation N-terminal sequencing approach. Due to the low quantity of ATHB17Δ113 produced in MON87403, it was not possible to undertake a direct glycosylation analysis or functional activity analysis of the plant-produced MON87403 protein. However, the LC-MS/MS analysis and an *in silico* search provided an indirect indication of glycosylation, and characterisation of the DNA binding activity of the *E.coli* produced protein was used to infer functional activity.

4.1.3.1 Molecular weight and immunoreactivity

The molecular weights of ATHB17Δ113 protein from the two sources were estimated from analysis of SDS-PAGE gels stained with silver. Immunoreactivity was detected on the Western blots using a polyclonal rabbit anti-ATHB17Δ113 primary antibody and a horseradish peroxidase-conjugated goat anti-rabbit IgG secondary antibody.

On the SDS-PAGE gel, the protein from *E. coli* showed a single band at approximately 22 kDa, the same weight predicted for the ATHB17Δ113 protein (see Section 4.1.1). The plant-derived protein showed numerous bands (as expected from the low purity of the sample), one of which had an average apparent molecular weight of approximately 22 kDa.

Western blot analysis showed a single immunoreactive band, increasing in intensity with protein load, that had co-migrated at approximately 20 kDa in separate extracts from both MON87403 and *E. coli.* This co-migrating band can be taken as evidence of extensive immunological cross reactivity between the proteins from the two sources. Taken together with the SDS-PAGE analysis, the results confirm that the samples from both *E. coli* and MON87403 contain a protein with a molecular weight of approximately 20 kDa and immunoreactivity to an ATHB17Δ113 antibody.

4.1.3.2 MALDI-TOF tryptic mass fingerprint

A protein identification made by peptide mass fingerprinting is considered to be reliable if the measured coverage of the sequence is 15% or higher with a minimum of five peptide matches (Jensen et al. 1997).

Purified ATHB17Δ113 protein from MON87403 was denatured on SDS-PAGE and the corresponding ≈22 kDa protein band was excised and digested with trypsin followed by matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) to determine the peptide mass fingerprint coverage. The masses of the detected peptides were searched *in silico* against those deduced from potential trypsin cleavage sites within the ATHB17Δ113 amino acid sequence, using the Mascot® search engine and ProteinPilot™ software (Applied Biosystems).

A total of eight peptide matches were obtained for the MON87403 protein and it was estimated that there was 50% coverage (81 out of 162 amino acids). This was adequate to provide convincing evidence of the identity of the protein from MON87403.

4.1.3.3 LC-MS/MS

Amino acid sequencing of trypsin-digested plant-produced ATHB17Δ113 was undertaken on a state-of-the-art Thermo Scientific™ Orbitrap Fusion™ Tribrid™ mass spectrometer[16]. The MS/MS dataset contained 49 unique peptides that corresponded to the expected peptides for trypsin-digested ATHB17Δ113 and the assembled sequences of which yielded coverage of 99% (160 out of 162 amino acids) of the expected sequence (see Figure 7) of the ATHB17Δ113 protein. This coverage included the N- and C-termini and a potential N-glycosylation sequence (NHT, starting at amino acid position 45) of the MON 87403-produced ATHB17Δ113 protein. The masses of two peptide residues containing the putative glycosylation site indicated that they were the same as those calculated for the unmodified amino acid sequences of each residue i.e. that there had been no addition of any sugar moieties that would be associated with glycosylation.

---

[16] For details of the instrument see http://www.thermoscientific.com/content/tfs/en/product/orbitrap-fusion-tribrid-mass-spectrometer.html

The same potential glycosylation site was also predicted by a basic search of the ATHB17Δ113 amino acid sequence using NetNGlyc[17]; this search further indicated that the absence of a signal peptide targeting the protein to the site of N-glycosylation in the endoplasmic reticulum means the protein would be unlikely to be N-glycosylated.

4.1.3.4 Functional activity

The DNA binding activity of the *E. coli*-produced ATHB17Δ113 protein was assessed using a qualitative ELISA-based protein-DNA binding end-point assay that evaluates the ability of ATHB17Δ113 to bind to the pseudo-palindromic DNA sequence CAAT(C/G)ATTG (see discussion in Section 4.1.1).

The principle of the analysis is that biotinylated oligonucleotide containing the target DNA sequence is placed in a reaction mixture containing the ATHB17Δ113 protein within a streptavidin-coated well in a micro-plate. The ATHB17Δ113 should bind to the DNA sequence and the oligonucleotide will bind strongly to the well because of the biotin/streptavidin interaction. The well is then washed out and any bound ATHB17Δ113 is detected using a monoclonal mouse anti-ATHB17Δ113 primary antibody followed by an anti-mouse horseradish peroxidase-conjugated secondary antibody; addition of tetramethylbenzidine substrate leads to a blue colour reaction that is detected spectrophometrically (460 nm) and can be directly related to the binding activity of the ATHB17Δ113. The negative control is a non-biotinylated oligonucleotide containing the target DNA sequence.

The results indicated that the ATHB17Δ113 protein did bind to the pseudo-palindromic sequence and therefore confirmed the functional activity of the protein.

4.1.3.5 Conclusion

ATHB17Δ113 produced in MON87403 and in *E. coli* were compared. The proteins from both sources were found to have the same apparent molecular weight in SDS-PAGE, and to be recognised by an anti-ATHB17Δ113 antibody in Western blotting. Based on this, it can be said that the *E. coli*-derived ATHB17Δ113 protein is a suitable surrogate for MON87403-derived ATHB17Δ113 in safety assessment studies.

The amino acid sequences of tryptic digests of MON87403-derived ATHB17Δ113 was established by, MALDI-TOF analysis and LC-MS/MS peptide sequencing. These amino acid sequences showed good agreement with the expressed ATHB17Δ113 protein deduced from DNA *in silico*. Indirect evidence also indicated that the MON87403-produced ATHB17Δ113 is unlikely to be N-glycosylated and that it has functional DNA-binding activity.

4.1.4   ATHB17Δ113 expression in the tissues of MON87403

Study submitted:

2015. Assessment of ATHB17Δ113 protein levels in leaf, root, forage and grain tissues collected from Maize MON87403 produced in United States field trials during 2012. **MSL0026598**. Monsanto Company (unpublished)

Plants of MON87403 (generation R_5F_1) were grown from verified seed lots at five field sites in the U.S.[18] during the 2012 growing season.

---

[17] http://www.cbs.dtu.dk/services/NetNGlyc/
[18] Jackson County, Arkansas; Story County, Iowa; Jefferson County, Iowa; Pawnee County, Kansas; Lehigh County, Pennsylvania

These plantings overlapped with the eight plantings used for the compositional analysis described in Section 5.2. There were four replicated plots at each site planted in a randomised complete-block design.

Samples from different plants were taken at various stages of growth (Table 5) to give a total of 20 samples for each tissue and time point (except for the roots for which there were inconclusive results for four of the samples). Levels of ATHB17Δ113 were determined for each sample type using a validated enzyme linked immunosorbent assay (ELISA). Detection of the protein utilised a monoclonal mouse anti-ATHB17Δ113 capture antibody and a biotinylated polyclonal goat anti- ATHB17Δ113 detection antibody followed by addition of .a commercial streptavidin-horseradish peroxidase conjugate and then horseradish peroxidase substrate. Plates were analysed on a microplate spectrophotometer, and commercial software (SoftMax Pro GxP, Molecular Devices) was used to convert optical density values to protein concentration

The results, averaged over all sites, are given in Table 5.

Table 5: ATHB17Δ113 protein content in MON87403 parts at different growth stages (averaged across 5 sites)

Tissue/Growth stage[1]	ATHB17Δ113 µg/g dw			
	n	Mean	Range	LOD[2]
Leaf V3 - V4	20	0.014	0.0096 – 0.017	0.00049
Root V3 - V4	16	0.0023	0.00083 – 0.0058	0.00065
Forage[3] R5	20	0.0018	0.0011 – 0.0035	0.00063
Grain R6	20	<LOD	-	0.00028

[1]For information on corn growth stages see e.g. Ransom & Endres (2014)
[2]Limit of Detection
[3]Forage is the above-ground plant parts used for animal feed

Expression levels were low in all tissues analysed. This was expected given the 'intractable' nature of the protein (see Section 4.1.4). The highest mean level was in leaf at 0.014 µg/g dw and the lowest mean level was in grain where it was below the limit of detection.

4.1.5   Potential toxicity of the ATHB17Δ113 protein

While the vast majority of proteins ingested as part of the diet are not typically associated with toxic effects, a small number may be harmful to health. Therefore, if a GM food differs from its conventional counterpart by the presence of one or more novel proteins, these proteins should be assessed for their potential toxicity. The main purpose of an assessment of potential toxicity is to establish, using a weight of evidence approach, that the novel protein will be digested like most other dietary proteins.

The assessment focuses on: whether the novel protein has a prior history of safe human consumption, or is sufficiently similar to proteins that have been safely consumed in food; amino acid sequence similarity with known protein toxins and anti-nutrients; structural properties of the novel protein including whether it is resistant to heat or processing and/or digestion. Appropriate oral toxicity studies in animals may also be considered, particularly where results from the biochemical, bioinformatic, digestibility or stability studies indicate a concern.

4.1.5.1 History of human consumption

As outlined in Section 4.1.1, HD-Zip proteins are found in higher plants, and have been specifically characterised in the crop plants rice, corn and soybean (Zhao et al. 2011; Belamkar et al. 2014); in particular, endogenous HD-Zip II genes are all expressed in the developing ears of corn.

Several large databases, containing thousands of protein sequences, are now available and can be searched using algorithms, to compare a particular protein with the proteins in the database(s). A BLASTP (Basic Local Alignment Search Tool Protein) algorithm (Pearson and Lipman 1988) was used to conduct a bioinformatics search of the National Center for Biotechnology Information (NCBI) Protein database[19], using the ATHB17Δ113 protein 162 amino acid sequence as the query, identified many homologous sequences. These were found in diverse plant species including the food crops soybean, rice, corn, tomato, potato, apple, grape, and cruciferous vegetables. Within the first 100 hits the protein sequence identity of ATHB17Δ113 to homologs in food species ranged from ~56-91%, with the highest identity to the homologs in *Camelina sativa* (a close relative of *A. thaliana*), *Brassica napus* (canola), and *Brassica rapa* (turnip). Since ATHB17Δ113 shares sequence identity and structural similarity with proteins present in crop plants this indicates the protein is sufficiently similar to proteins that have been safely consumed in food.

4.1.5.2 Similarity of ATHB17Δ113 with known toxins

Study submitted:

2013. Bioinformatics evaluation of the ATHB17Δ113 protein utilizing the AD_2013, TOX_2013 and PRT_2013 databases. **MSL0025242,** Monsanto Company (unpublished).

A subset of sequences derived from the Genbank[20] protein database (release 193) was compiled to contain only toxin proteins (TOX_13). The ATHB17Δ113 sequence was compared for structural similarities with the 8,881 sequences present in TOX-13. The Fast Alignment Search Tool - All (FASTA) algorithm (Pearson and Lipman 1988), version 3.4t 26, was used to search the database using the BLOSUM50[21] scoring matrix (Henikoff and Henikoff 1992).

A search generates a parameter known as the *E* value (see eg Baxevanis 2005). Comparisons between highly homologous proteins yield *E*-values approaching zero, indicating the very low probability that such matches would occur by chance. For this search an E-value of 1×10^{-5} was set as an initial high cut-off value for alignment significance.

There were no significant matches with any sequences from known protein toxins.

4.1.5.3 Thermolability

The thermolability of a protein provides an indication of its stability under processing or cooking conditions.

Study submitted:

2014. Effect of heat treatment on the functional activity of *Escherichia coli* (*E. coli*)-produced ATHB17Δ113 protein. **MSL0025364**, Monsanto Company (unpublished).

---

[19] NCBI Protein database - http://www.ncbi.nlm.nih.gov/guide/proteins/
[20] Genbank - http://www.ncbi.nlm.nih.gov/genbank/
[21] The BLOSUM series of matrices tabulate the frequency with which different substitutions occur in conserved blocks of protein sequences and are effective in identifying distant relationships.

The temperature stability of microbially-derived ATHB17Δ113 protein was evaluated by examining the functional activity of the protein during 15 min and 30 min incubations at various temperatures (25°, 37°, 55°, 75° and 95° C). The higher temperatures represent those within a range that is applicable to industrial processing conditions for corn. Functional activity was measured using the method described in Section 4.1.3. SDS-PAGE analysis of the treatments was also done.

Results from the functional activity analysis showed there was no reduction in DNA-binding at 25°, 37°, or 55°. Reduced relative DNA binding values, (equating to relative activity reducing to 59% at 95° C after 15 min and 38% after 30 min) were observed for the 75 and 95 °C temperature treatments for both incubation periods. This indicated the ATHB17Δ113 protein is partially labile when heated for 15 or 30 minutes at temperatures ≥75 °C. The SDS-PAGE analysis showed the appearance of lower molecular weight species only when ATHB17Δ113 was heated to 95° C; these may be due to slight degradation of the protein when exposed to high temperatures. Overall, the results show that ATHB17Δ113 has a tendency towards loss of functional activity at elevated temperatures.

4.1.5.4 *Acute toxicity study*

Although not required, since no toxicity concerns were raised in the data considered in the sub-sections above, the Applicant supplied an acute oral toxicity study.

Study submitted:

2014. An acute toxicity study of *E.coli*-produced ATHB17Δ113 protein by oral gavage in mice. **CRO-2013-0121**, Monsanto Company (unpublished).

The ATHB17Δ113 protein derived from *E. coli* was administered to CD-1 mice by oral gavage twice in one day. The mice were observed for 14 days then euthanased for necropsy examination. No adverse effects were observed at a dose of 1,335 mg protein/kg body weight.

4.1.6   Potential allergenicity of the ATHB17Δ113 protein

The potential allergenicity of the ATHB17Δ113 protein was evaluated using an integrated, step-wise, case-by-case approach relying on various criteria used in combination. This is because no single criterion is sufficiently predictive of either allergenicity or non-allergenicity (see eg Thomas et al. 2009). The assessment focuses on:

- the source of the novel protein;
- any significant amino acid sequence similarity between the novel protein and known allergens;
- the structural properties of the novel protein, including susceptibility to digestion, heat stability and/or enzymatic treatment; and
- specific serum screening if the novel protein is derived from a source known to be allergenic or has amino acid sequence similarity to a known allergen, additional in vitro and in vivo immunological testing may be warranted.

Applying this approach systematically provides reasonable evidence about the potential of the novel protein to act as an allergen.

The allergenic potential of the ATHB17Δ113 protein was assessed by:

- consideration of the source of the *ATHB17* gene and history of use or exposure

- bioinformatic comparison of the amino acid sequence of the ATHB17Δ113 protein with known protein allergen sequences
- evaluation of the digestibility of the microbially-produced ATHB17Δ113 using an *in vitro* gastric digestion model.

4.1.6.1 Source of the ATHB17Δ113 protein

Although a member of the mustard family, *A. thaliana* is not commonly cultivated or harvested for food due to its small size and therefore does not have a history of significant human consumption. The species *Camelina sativa* (false flax) which has been grown as an oilseed crop for hundreds of years is the cultivated species most closely related to Arabidopsis (Flannery et al. 2006). The oil has been used in both edible and industrial products (Fleenor 2011). No Arabidopsis proteins are reported in databases of known allergens. One case of occupational asthma due to exposure to Arabidopsis pollen has been reported in a laboratory worker (Yates et al. 2008).

4.1.6.2 Similarity to known allergens

Study submitted:

2013. Bioinformatics evaluation of the ATHB17Δ113 protein utilizing the AD_2013, TOX_2013 and PRT_2013 databases. **MSL0025242,** Monsanto Company (unpublished).

Bioinformatic analysis provides part of a "weight of evidence" approach for assessing potential allergenicity of novel proteins introduced to GM plants. It is a method for comparing the amino acid sequence of the introduced protein with sequences of known allergens in order to indicate potential cross-reactivity between allergenic proteins and the introduced protein. As with the bioinformatic analysis that looked at similarities of the ATHB17Δ113 protein with known protein toxins (see Section 4.1.5), the generation of a small *E*-value provides an important indicator of significance of matches (Pearson 2000; Baxevanis 2005).

To evaluate the similarity to known allergens of the ATHB17Δ113 protein an epitope search was carried out to identify any short sequences of amino acids that might represent an isolated shared allergenic epitope. This search compared the sequences with known allergens in the Allergen, Gliadin and Glutenin sequence database (AD_2013), containing 1,630 sequences and residing in the 2013 FARRP (Food Allergy Research and Resource Program) dataset within AllergenOnline[22]. As for the toxin search, the FASTA algorithm was used.

No alignments generated an E-score of ≤1e^{-5}, no alignment met or exceeded the Codex Alimentarius (Codex 2009) FASTA alignment threshold for potential allergenicity (35% identity over 80 amino acids) and no alignments of eight or more consecutive identical amino acids (Metcalfe et al. 1996) were found. It was concluded that ATHB17Δ113 does not contain any cross-reactive Immunoglobulin E (IgE) binding epitopes with known allergens.

4.1.6.3 In vitro *digestibility*

Typically, food proteins that are allergenic tend to be stable to enzymes such as pepsin and the acidic conditions of the digestive system, exposing them to the intestinal mucosa and leading to an allergic response (Astwood and Fuchs 1996; Metcalfe et al. 1996; Kimber et al. 1999). Therefore a correlation exists between resistance to digestion by pepsin and potential allergenicity although this may be inconsistent because not all resistant proteins are allergens (Fu et al. 2002; Thomas et al. 2004; Herman et al. 2007).

---

[22] University of Nebraska; http:www.allergenonline.org/

As a consequence, one of the criteria for assessing potential allergenicity is to examine the stability of novel proteins in conditions mimicking human digestion. Proteins that are rapidly degraded in such conditions are considered less likely to be involved in eliciting an allergic response. However, evidence of slow or limited protein digestibility does not necessarily indicate that the protein is allergenic.

Study submitted:

2014. Assessment of the *in vitro* digestibility of ATHBΔ113 protein in simulated gastric and simulated intestinal fluids. **MSL0025516**, Monsanto Company (unpublished).

Analyses of microbially-derived ATHB17Δ113 in simulated gastric fluid (SGF), containing pepsin, and simulated intestinal fluid (SIF) containing pancreatin (a mixture of enzymes including amylase, trypsin, lipase, ribonuclease and protease) were done. The SIF study by itself may not be entirely informative because ordinarily an ingested protein would first be exposed to pepsin-mediated hydrolysis in the acidic environment of the stomach before being subject to further digestion in the small intestine.

The SGF assay was done, using a commercially available pepsin, at 37° C with incubation times of 0, 0.5, 2, 5, 10, 20, 30 and 60 minutes. The equivalent of a zero time point was prepared by quenching the SGF solution with sodium carbonate then heating to approximately 100° C prior to adding the ATHB17Δ113 protein sample.

SIF was prepared using the method described in The United States Pharmacopeia (U.S.Pharmacopeia 2000) and the assay was run at 37° C with incubation times of 0, 5 min, 15 min, 30 min, 1 h, 2 h, 4 h, 8 h, and 24 h.

Following incubation, the SGF and SIF samples were run on SDS-PAGE. Proteins from the SGF assay were visualised by Brilliant Blue G-Colloidal staining of the resulting gels. Western blotting of SGF and SIF SDS-PAGE gels was also performed using a rabbit anti-ATHB17Δ113 primary antibody and an HRP-conjugated goat anti-rabbit secondary antibody. Immunoreactive bands were visualized using an enhanced chemiluminescence detection system (Lorimier et al. 1993) and quantified via reading of chemiluminescence film on a densitometer.

In the SDS-PAGE gel for ATHB17Δ113, it was estimated that >97.5% of the full-length protein was digested within 0.5 min of incubation in SGF, while there was no significant digestion in the absence of pepsin (or with pepsin at time zero). A peptide fragment of ≈ 4 kDa was present at 0.5 min but had disappeared by 2 min. The Western blot for ATHB17Δ113 indicated that >93.5% of the protein was digested within 0.5 min; no peptide fragments were detected at any time point.

The Western blot analysis of the digestibility of ATHB17Δ113 in SIF demonstrated that >93.5% of the protein was digested within 5 min.

4.1.7 Bioinformatic analysis of additional ORFs created by the transformation procedure

---

Study submitted:

2014. Bioinformatics evaluation of the transfer DNA insert in MON87403 utilizing the AD_2014, TOX_2014 and PRT_2014 databases. MSL0025648. Monsanto Company (unpublished)
2014. Bioinformatics Evaluation of DNA Sequences Flanking the 5' and 3' Junctions of Inserted DNA in MON 87403: Assessment of Putative Polypeptides. MSL0025733. Monsanto Company (unpublished)

A bioinformatics analysis was performed to assess the similarity to known allergens and toxins of the putative polypeptides encoded by the 10 identified ORFs in the flanking regions of the MON87403 insert and of the six reading frames in the translated sequences from within the T-DNA insert itself (see Section 3.4.4).

The bioinformatics analyses of toxins and allergens were carried out as already described in the relevant bioinformatics parts of Sections 4.1.5 and 4.1.6 respectively, i.e. the ORF and insert sequences were compared with sequences present in a toxin (TOX_14, 10,419 sequences) and allergen (AD_2014, 1706 sequences) database using the FASTA search algorithm. The allergen analyses considered both ≥35% identity over 80 amino acids and an eight amino acid sliding window search.

No significant homologies were found in comparisons of the 10 putative polypeptides in the flanking regions with sequences in either database, i.e. in the unlikely event any of the potential ORFs were to be expressed in MON87403 there is no significant similarity between the encoded sequence and any known protein toxins or allergens.

Similarly, no significant similarities of any T-DNA reading frame sequence to any sequences in either of the databases were found, i.e. in the event that an unexpected translation product was derived from reading frames 1-6 of the T-DNA insert, these putative polypeptides are not expected to possess functional cross-reactivity with known allergenic proteins or be toxic.

4.1.8 Conclusion

MON87403 expresses one novel DNA binding protein, ATHB17Δ113. Expression levels were low in all tissues analysed. The highest mean level was in leaf at 0.014 µg/g dw and the lowest mean level was in grain where it was below the limit of detection.

The identity of the MON87403-produced protein was confirmed by Western blot analysis, sequence analysis of the *ATHB17Δ113* mRNA transcript produced in MON87403, matrix assisted laser desorption/ionization time-of-flight mass spectrometry, and liquid chromatography-tandem mass spectrometry. Indirect evidence indicated that the MON87403-produced ATHB17Δ113 is not N-glycosylated and that it has the expected functional activity.

Bioinformatic studies have confirmed the lack of any significant amino acid sequence similarity to known protein toxins or allergens and digestibility studies have demonstrated ATHB17Δ113 would be rapidly and completely digested in the gastrointestinal tract. The protein also loses DNA-binding activity with heating. Taken together, the evidence indicates the ATHB17Δ113 protein is unlikely to be toxic or allergenic to humans.

5 Compositional analysis

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

The classic approach to the compositional analysis of GM food is a targeted one. Rather than analysing every single constituent, which would be impractical, the aim is to analyse only those constituents most relevant to the safety of the food or that may have an impact on the whole diet. Important analytes therefore include the key nutrients, toxicants and anti-

nutrients for the food in question. The key nutrients and anti-nutrients are those components in a particular food that may have a substantial impact in the overall diet. They may be major constituents (fats, proteins, carbohydrates or enzyme inhibitors such as anti-nutrients) or minor constituents (minerals, vitamins). Key toxicants are those toxicologically significant compounds known to be inherently present in an organism, such as compounds whose toxic potency and level may be significant to health (e.g. solanine in potatoes).

5.1   Key components

For corn there are a number of components that are considered to be important for compositional analysis (OECD 2002). As a minimum, the key nutrients of corn grain appropriate for a comparative study include the proximates (crude protein, fat, ash, acid detergent fibre and neutral detergent fibre), amino acids and fatty acids. In addition, mineral and vitamin levels may be considered and international guidance also suggests that levels of the key anti-nutrients and secondary metabolites could be determined for new varieties of corn.

Corn contains a number of anti-nutrients: phytic acid, raffinose, 2,4-dihydroxy-7-methoxy-2H-1,4-benzoxazin-3(4H)-one (DIMBOA) and trypsin and chymotrypsin inhibitor. Only phytic acid and raffinose are considered to be biologically relevant (OECD 2002). DIMBOA is present at highly variable levels in corn hybrids and little evidence is available on either its toxicity or anti-nutritional effects, and corn contains only low levels of trypsin and chymotrypsin inhibitor, neither of which is considered nutritionally significant.

Phytic acid is considered an important anti-nutrient for animals, especially non-ruminants, since it can significantly reduce the bioavailability of phosphorus in the plant. Feed formulators add the enzyme phytase to pig and poultry diets to improve the utilisation of phosphorus. Raffinose is a non-digestible oligosaccharide and is considered an anti-nutrient because of its gastrointestinal effects (flatulence).

Secondary metabolites are defined as those natural products which do not function directly in the primary biochemical activities which support the growth, development and reproduction of the organism in which they occur. Secondary plant metabolites are neither nutrients nor anti-nutrients but are sometimes analysed as further indicators of the absence of unintended effects of the genetic modification on metabolism (OECD 2002). Characteristic metabolites in corn are furfural and the phenolic acids, ferulic acid and p-coumaric acid.

There are no generally recognised anti-nutrients in corn at levels considered to be harmful, but for the purposes of comparative assessment, the OECD has recommended considering analytical data for the content of the anti-nutrients phytic acid and raffinose, and the secondary metabolites furfural, ferulic acid, p-coumaric acid.

5.2   Study design and conduct for key components

Study submitted:

2015. Amended report for MSL0025076: Composition Analyses of Maize Forage and Grain from MON 87403 Grown in the United States during 2012. **MSL0025787**. Monsanto Company (unpublished).

Verified (event-specific PCR) MON87403 generation R_5F_1 (refer to Figure 3) was used for compositional analysis. Ideally, the comparator in compositional analyses should be the near isogenic parental line grown under identical conditions (OECD 2002). In the case of MON87403, the control was the hybrid line LH244 x LH287 since this represents the closest non-GM genetic line for the purposes of comparison.

The test and control lines were grown from verified seed lots at eight field sites across the US corn belt[23] during the 2012 growing season. Five of the plantings were the same as those used for the protein expression analysis (Section 4.1.4). There were four replicated blocks at each site planted in a randomised complete-block design. Maintenance fertilizer and pesticides were applied as needed in order to maintain a relatively weed-free and insect-free environment. Additionally, a total of 17 non-GM hybrid lines were also grown as reference lines with four different lines being grown at each site in order to generate tolerance ranges for each analyte and hence to aid in the determination of the normal variation found in corn analyte levels.

Grain samples were analysed for proximates, fibre (acid detergent fibre – ADF; neutral detergent fibre – NDF; total dietary fibre), fatty acids, amino acids, minerals, vitamins, anti-nutrients and secondary metabolites.

Key analyte levels (proximates, fibre, calcium and phosphorus) for forage (harvested at R5) were also obtained but are not reported here; it is noted, however, that in the combined site analysis none of the analyte levels in MON87403 differed significantly from those of the control.

Methods of composition analysis were based on internationally recognised procedures (e.g. those of the Association of Official Analytical Chemists), methods specified by the manufacturer of the equipment used for analysis, or other published methods.

5.3 Analyses of key components in grain

In total, 68 analyte levels were measured and carbohydrate was calculated rather than being measured i.e. there was a total of 69 analytes considered. Moisture was measured for conversion of components to dry weight but was not statistically analysed. Sixteen analytes had more than 50% of the observations below the assay limit of quantitation (LOQ) and were excluded from the statistical analysis. The data for 52 analytes were therefore analysed. This analysis used a mixed model analysis of variance. Data were transformed into Statistical Analysis Software[24] (SAS) data sets and analysed using SAS® software (SAS MIXED, version 9.2). The SAS GLM procedure was applied to all data (test, control and reference) to detect potential outliers in the dataset by screening studentised PRESS residuals[25].

The replicated sites were analysed both separately and combined across all sites (combined-site analysis). Descriptive statistics (mean and standard error (SE) were generated and are presented in Tables 7 – 13 which represent results from combined-site analyses. In assessing the significance of any difference between means, a P-value of 0.05 was used (i.e. a P-value of ≥0.05 was not significant).

Any statistically significant differences between MON87403 and the LH244 x LH287 control have been compared to the 95% tolerance interval (i.e. 95% confidence that the interval contains 99% of the values expressed in the commercial lines) compiled from the results of the 17 commercial reference lines combined across all sites, to assess whether the differences are likely to be biologically meaningful. Additionally, the results for MON87403 and the hybrid control have been compared to a combined literature range for each analyte, compiled from published literature for commercially available corn[26].

---

[23] Jackson, Arkansas; Story, Iowa; Jefferson, Iowa; Warren, Illinois; Boone, Indiana; Pawnee, Kansas; Polk, Nebraska; Lehigh, Pennsylvania.
[24] SAS website - http://www.sas.com/technologies/analytics/statistics/stat/index.html
[25] A PRESS (predicted residual sum of squares) statistic provides a comparison of the predicted marginal mean and the observed mean when the predicted value is calculated without the deleted observation in question (Schabenberger 2004).
[26] Published literature for corn incorporates references used to compile listings in the ILSI Crop Composition

It is noted, however, that information in the published literature is limited and is unlikely to provide a broad reflection of the natural diversity that occurs within corn (Harrigan et al. 2010; Zhou et al. 2011; Ridley et al. 2011). Therefore, even if means fall outside the published range, this is unlikely to raise a concern.

5.3.1   Proximates and fibre

Results of the proximate and fibre analysis are shown in Table 7. There was no significant difference between the mean level in MON87403 and the control for any analyte. All means were also within both the tolerance interval and the literature range.

Table 6: Mean (±SE) percentage dry weight (%dw) of proximates and fibre in grain from MON87403 and the hybrid control

Analyte	MON87403 (%dw)	Control (%dw)	P-value	Tolerance interval (%dw)	Combined literature range (%dw)
Ash	1.31 ± 0.026	1.33 ± 0.026	0.371	1.08, 1.60	0.62 – 6.28
Protein	10.13 ± 0.34	10.15 ± 0.34	0.911	7.72, 12.67	6.15 – 17.26
Total Fat	3.56 ± 0.072	3.54 ± 0.072	0.682	1.93, 5.49	1.74 – 5.82
Carbohydrate[1]	84.98 ± 0.34	84.98 ± 0.34	0.986	81.8, 87.71	77.4 – 89.5
ADF	3.63 ± 0.068	3.62 ± 0.068	0.893	2.36, 4.43	1.82 – 11.34
NDF	9.47 ± 0.19	9.41 ± 0.19	0.696	5.32, 21.85	5.59 – 22.64
Total dietary fibre	13.04 ± 0.15	12.95 ± 0.15	0.657	10.05, 15.51	9.0 – 35.3

[1] Carbohydrate calculated as 100% - (protein %dw+ fat %dw + ash %dw)

5.3.2   Fatty Acids

The levels of 22 fatty acids were measured. Of these, the following had ≥50% of observations below the LOQ and were therefore excluded from analysis - C8:0 caprylic, C10:0 capric, C12:0 lauric, C14:0 myristic, C14:1 myristoleic, C15:0 pentadecanoic, C15:1 pentadecenoic, C16:0 palmitoleic, C17:0 heptadecanoic, C17:1 heptadecenoic, C18:3 gamma linolenic, C20:2 eicosadienoic, C20:3 eicosatrienoic, and C20:4 arachidonic.. Results for the remaining eight fatty acids are given in Table 8. There was no significant difference between the mean level in MON87403 and the control for any fatty acid. All means were also within both the tolerance interval and the literature range.

Table 7: Mean (±SE) percentage composition, relative to total fat, of major fatty acids in grain from MON87403 and the hybrid control

Analyte	MON87403[1] (%total)	Control (%total)	P-value	Tolerance interval (%total)	Combined literature range (%total)
Palmitic acid (C16:0)	12.77 ± 0.14	12.68 ± 0.14	0.219	7.90, 14.94	7.94 – 20.71
Stearic acid (C18:0)	2.04 ± 0.022	2.06 ± 0.022	0.206	1.05, 2.72	1.02 – 3.40
Oleic acid (C18:1)	21.84 ± 0.30	21.77 ± 0.30	0.716	14.77, 38.25	17.4 – 40.2
Linoleic acid (C18:2)	61.23 ± 0.30	61.39 ± 0.30	0.372	45.88, 70.60	36.2 – 66.5

Database Version 4 (ILSI 2014).

Analyte	MON87403[1] (%total)	Control (%total)	P-value	Tolerance interval (%total)	Combined literature range (%total)
Linolenic acid (C18:3)	1.31 ± 0.014	1.30 ± 0.014	0.267	0.69, 1.67	0.57 – 2.25
Arachidic acid (C20:0)	0.45 ± 0.012	0.45 ± 0.012	0.634	0.26, 0.55	0.28 – 0.965
Eicosenoic acid (C20:1)	0.20 ± 0.0030	0.20 ± 0.0030	0.243	0.14, 0.33	0.17 – 1.917
Behenic acid (C22:0)	0.16 ± 0.0081	0.16 ± 0.0081	0.544	0, 0.26	0.11 – 0.29

5.3.3 Amino acids

Levels of 18 amino acids were measured. Since asparagine and glutamine are converted to aspartate and glutamate respectively during the analysis, levels for aspartate include both aspartate and asparagine, while glutamate levels include both glutamate and glutamine.

The results in Table 9 show there was no significant difference between the control and MON87403 for any of the amino acid means. All means were also within both the tolerance interval and the literature range.

Table 8: Mean (±SE) percentage dry weight of amino acids in grain from line MON87403 and the hybrid control

Analyte	MON87403[1] (%dw)	Control (%dw)	P-value	Tolerance interval (%dw)	Combined literature range (%dw)
Alanine	0.77 ± 0.030	0.77 ± 0.030	0.975	0.55, 1.01	0.439 – 1.393
Arginine	0.48 ± 0.0093	0.48 ± 0.0093	0.988	0.41, 0.59	0.119 – 0.639
Aspartate	0.63 ± 0.018	0.63 ± 0.018	0.943	0.52, 0.78	0.335 – 1.208
Cystine	0.21 ± 0.0041	0.21 ± 0.0041	0.589	0.16, 0.25	0.125 – 0.514
Glutamate	1.85 ± 0.077	1.86 ± 0.077	0.870	1.26, 2.52	0.965 – 3.536
Glycine	0.37 ± 0.0065	0.37 ± 0.0065	0.966	0.32, 0.44	0.184 – 0.539
Histidine	0.28 ± 0.0068	0.28 ± 0.0068	0.648	0.20, 0.35	0.137 – 0.434
Isoleucine	0.36 ± 0.014	0.36 ± 0.014	0.944	0.26, 0.46	0.179 – 0.692
Leucine	1.27 ± 0.059	1.27 ± 0.059	0.956	0.81, 1.73	0.642 – 2.492
Lysine	0.27 ± 0.0048	0.27 ± 0.0048	0.625	0.24, 0.31	0.172 – 0.668
Methionine	0.21 ± 0.0063	0.20 ± 0.0063	0.803	0.15, 0.26	0.124 – 0.468
Phenylalanine	0.52 ± 0.022	0.52 ± 0.022	0.801	0.36, 0.68	0.244 – 0.930
Proline	0.95 ± 0.030	0.95 ± 0.030	0.830	0.64, 1.17	0.462 – 1.632
Serine	0.44 ± 0.015	0.45 ± 0.015	0.356	0.34, 0.57	0.235 – 0.769
Threonine	0.36 ± 0.0098	0.35 ± 0.0098	0.570	0.28, 0.43	0.224 – 0.666
Tryptophan	0.077 ± 0.0015	0.077 ± 0.0015	0.827	0.064, 0.093	0.027 – 0.215
Tyrosine	0.41 ± 0.016	0.41 ± 0.016	0.959	0.30, 0.53	0.103 – 0.642
Valine	0.46 ± 0.013	0.46 ± 0.013	0.880	0.35, 0.57	0.266 – 0.855

5.3.4 Minerals

The levels of eight minerals in grain from MON87403 and the hybrid control were measured. Results for the analytes are given in Table 10 and show there was no significant difference between the control and MON87403 for any of the mineral means. All means were also within both the tolerance interval and the literature range.

Table 9: Mean (±SE) levels of minerals in the grain of MON87403 and the hybrid control

Analyte	Unit	MON87403[1]	Control	P-value	Tolerance interval	Combined literature range
Calcium	% dw	0.0037 ± 0.00021	0.0037 ± 0.00021	0.321	0.0011, 0.0059	0.0012– 0.02
Copper	mg/kg dw	1.57 ± 0.11	1.55 ± 0.11	0.852	0.29, 3.17	0.73 – 18.5
Iron	mg/kg dw	19.33 ± 0.82	19.60 ± 0.82	0.240	10.87, 27.03	10.4 – 49.1
Magnesium	%dw	0.12 ± 0.0033	0.12 ± 0.0033	0.939	0.092, 0.15	0.059 – 0.194
Manganese	mg/kg dw	6.17 ± 0.31	6.14 ± 0.31	0.787	2.59, 10.23	1.69 – 14.30
Phosphorus	%dw	0.30 ± 0.0068	0.30 ± 0.0068	0.683	0.24, 0.34	0.147 – 0.533
Potassium	%dw	0.33 ± 0.0076	0.33 ± 0.0076	0.611	0.23, 0.42	0.181 – 0.603
Zinc	mg/kg dw	20.52 ± 0.83	20.98 ± 0.83	0.165	9.09, 32.95	6.5 – 37.2

5.3.5 Vitamins

Levels of seven vitamins were measured. Results are given in Table 11 and show there was no significant difference between the control and MON87403 for any of the vitamin means. All means were also within both the tolerance interval and the literature range.

Table 10: Mean (±SE) weight (mg/k g dry weight) of vitamins in grain from MON87403 and the hybrid control

Analyte	MON87403[1] (mg/kg dw)	Control (mg/kg dw)	P-value	Tolerance interval (mg/kg dw)	Combined literature range (mg/kg dw)
Vitamin A (β-Carotene)	1.16 ± 0.039	1.14 ± 0.039	0.319	0, 3.10	0.19 – 46.81
Vitamin B_1 (Thiamine HCl)	3.48 ± 0.13	3.52 ± 0.13	0.277	1.73, 5.12	1.26 – 40.00
Vitamin B_2 (Riboflavin)	1.83 ± 0.059	1.71 ± 0.059	0.057	1.25, 2.22	0.50 – 2.36
Vitamin B_3 (Niacin)	16.47 ± 0.78	16.59 ± 0.78	0.788	7.36, 30.18	10.37 – 46.94
Vitamin B_6 (Pyridoxine HCl)	7.11 ± 0.17	6.89 ± 0.17	0.255	4.51, 8.98	3.68 – 11.32
Vitamin B9 (Folic acid)	0.39 ± 0.016	0.39 ± 0.016	0.940	0.038, 0.69	0.147 – 1.464
Vitamin E (α-Tocopherol	17.87 ± 0.65	18.33 ± 0.65	0.095	2.50, 27.12	1.537 – 68.672

5.3.6 Anti-nutrients

Levels of two key anti-nutrients were measured. Results in Table 12 show that none of the means differed significantly between MON87403 and the control. Both means also fell within both the tolerance interval and the literature range.

Table 11: Mean (±SE) of anti-nutrients in grain from MON87403 and the hybrid control

Analyte	MON87403	Control	P-value	Tolerance interval	Combined literature range
Phytic acid (%dw)	0.89 ± 0.017	0.87 ± 0.017	0.164	0.68 – 1.18	0.111 – 1.570
Raffinose %dw)	0.22 ± 0.012	0.23 ± 0.012	0.487	0.00088, 0.40	0.020 – 0.320

5.3.7   Secondary metabolites

The levels of two secondary metabolites were measured (see Table 13), There was no significant difference between the control and MON87403 for either metabolite. Both means

Table 12: Mean weight (±SE) of two secondary metabolites in grain from MON87403 and the hybrid control

Analyte	MON87403[1] (µg/g dw)	Control (µg/g dw)	P-value	Reference range (mg/kg dw)	Combined literature range (mg/kg dw)
p-coumaric acid	216.73 ± 5.25	212.58 ± 5.25	0.441	6.62, 433.65	53.4 – 576.2
Ferulic acid	2262.60 ± 61.27	2213.54 ± 61.27	0.341	827.07, 3473.40	291.9 – 3885.8

5.4   Conclusion from compositional analyses

Detailed compositional analyses were done to establish the nutritional adequacy of grain from MON87403 and to characterise any unintended compositional changes. Analyses were done of proximates, fibre, minerals, amino acids, fatty acids, vitamins, secondary metabolites and anti-nutrients. The levels were compared to levels in a) an appropriate non-GM hybrid line b) a tolerance interval compiled from results taken for a total of 17 non-GM hybrid lines grown in the same field trials and c) levels recorded in the literature. None of the 52 analytes reported in Tables 7 - 13 deviated in level from the control in a statistically significant manner. It can therefore be concluded that grain from line MON87403 is compositionally equivalent to grain from conventional corn varieties.

6   Nutritional impact

In assessing the safety of a GM food, a key factor is the need to establish that the food is nutritionally adequate and will support typical growth and well-being. In most cases, this can be achieved through an understanding of the genetic modification and its consequences, together with an extensive compositional analysis of the food.

If the compositional analysis indicates biologically significant changes to the levels of certain nutrients in the GM food, additional nutritional assessment should be undertaken to assess the consequences of the changes and determine whether nutrient intakes are likely to be altered by the introduction of such foods into the food supply.

Where a GM food has been shown to be compositionally equivalent to conventional varieties, the evidence to date indicates that feeding studies using target livestock species will add little to the safety assessment and generally are not warranted (see e.g. OECD 2003; Bartholomaeus et al. 2013; Herman and Ekmay 2014).

MON87403 is the result of a genetic modification designed to increase grain yield with no intention to significantly alter nutritional parameters in the food. In addition, the extensive compositional analyses of grain that have been undertaken to demonstrate the nutritional adequacy of MON87403 indicate it is equivalent in composition to conventional corn cultivars. The introduction of food from MON87403 into the food supply is therefore expected to have little nutritional impact and as such no additional studies, including animal feeding studies, are required.

References [27]

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215(3):403–410

Ariel FD, Manavella PA, Dezar CA, Chan RL (2007) The true story of the HD-Zip family. Trends in Plant Science 12(9):419–426

Armstrong, M. F. (2001) Inbred corn line LH244. Patent US 6,252,148 B1 Patent Application 09/455,440. http://www.google.com/patents/US6252148

Astwood JD, Fuchs RL (1996) Allergenicity of foods derived from transgenic plants. Monographs in Allergy 32 (Highlights in Food Allergy):105–120

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

Baxevanis AD (2005) Assessing Pairwise Sequence Similarity: BLAST and FASTA. Ch 11 In: Baxevanis AD, Ouellette BFF (eds) Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. John Wiley & Sons, Inc., p. 295–324

Belamkar V, Weeks NT, Bharti AK, Farmer AD, Graham MA, Cannon SB (2014) Comprehensive characterization and RNA-Seq profiling of the HD-Zip transcription factor family in soybean (*Glycine max*) during dehydration and salt stress. BMC Genomics 15:950

Bushey DF, Bannon GA, Delaney BF, Graser G, Hefford M, Jiang X, Lee TC, Madduri KM, Pariza M, Privalle LS, Ranjan R, Saab-Rincon G, Schafer BW, Thelen JJ, Zhang JXQ, Harper MS (2014) Characteristics and safety assessment of intractable proteins in genetically modified crops. Regulatory Toxicology and Pharmacology 69:154–170

CFIA (1994) The biology of *Zea mays* L. (corn/maize). BIO1994-11. Canadian Food Inspection Agency, Ottawa. http://www.inspection.gc.ca/plants/plants-with-novel-traits/applicants/directive-94-08/biology-documents/zea-mays-l-/eng/1330985739405/1330985818367

Codex (2009) Foods derived from modern biotechnology, 2nd edition. Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme, Food and Agriculture Organization of the United Nations, Rome. http://www.codexalimentarius.org/standards/thematic-publications/

CRA (2006) Corn oil. 5 ed, Corn Refiners Association, Washington D.C. http://www.corn.org/wp-content/uploads/2009/12/CornOil.pdf.

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

DuBose AJ, Lichtenstein ST, Narisu N, Bonnycastle LL, Swift AJ, Chines PS, Collins FS (2013) Use of microarray hybrid capture and next-generation sequencing to identify the anatomy of a transgene.

---

[27] All website references were current as at 25 August 2015

Nucleic Acids Research 41(6):e70

Elhiti M, Stasolla C (2009) Structure and function of homeodomain-leucine zipper (HD-Zip) proteins. Plant Signaling & Behavior 4(2):86–88

FAOSTAT3 (2015) Online database of the Food and Agriculture Organization of the United Nations. http://faostat3.fao.org/browse/T/TP/E

Flannery ML, Mitchell FJG, Coyne S, Kavanagh TA, Burke JI, Salamin N, Dowding P, Hodkinson TR (2006) Plastid genome chracterisation in *Brassica* and Brassicaceae using a new set of nine SSRs. Theoretical and Applied Genetics 113(7):1221–1231

Fleenor RA (2011) Plant Guide for Camelina (*Camelina sativa*)., USDA-Natural Resources Conservation Service, Spokane, Washington. http://plants.usda.gov/plantguide/pdf/pg_casa2.pdf

Fu T-J, Abbott UR, Hatzos C (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

Green Pool (2012) Sugar consumption in Australia. A statistical update. Green Pool Commodity Specialists, Brisbane. http://greenpoolcommodities.com/files/8113/4932/3223/121004_Sugar_Consumption_in_Australia_-_A_Statistical_Update_-_Public_Release_Document.pdf

Harrigan GG, Glenn KC, Ridley WP (2010) Assessing the natural variability in crop composition. Regulatory Toxicology and Pharmacology 58:S13–S20; doi:10.1016/j.yrtph.2010.08.023

Harris JC, Hrmova M, Lopato S, Langridge P (2011) Modulation of plant growth by HD-Zip class I and II transcription factors in response to environmental stimuli. New Phytologist 190:823–837

Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences 89:10915–10919

Herman RA, Ekmay R (2014) Do whole-food animal feeding studies have any value in the safety assessment of GM crops? Regulatory Toxicology and Pharmacology 68:171–174

Herman RA, Woolhiser MM, Ladics GS, Korjagin VA, Schafer BW, Storer NP, Green SB, Kan L (2007) Stability of a set of allergens and non-allergens in simulated gastric fluid. International Journal of Food Sciences and Nutrition 58:125–141

Huang S, Gilbertson L, Adams T, Malloy K, Reisenbigler E, Birr D, Snyder M, Zhang Q, Luethy M (2004) Generation of marker-free transgenic maize by regular two-border *Agrobacterium* transformation vectors. Transgenic Research 13(5):451–461

ILSI (2014) International Life Sciences Institute Crop Composition Database Version 4.0. http://www.cropcomposition.org/query/index.html

Jensen ON, Podtelejnikov AV, Mann M (1997) Identification of the components of simple protein mixtures by high-accuracy peptide mass mapping and database searching. Annals of Chemistry 69(23):4741–4750

Kay R, Chan A, Daly M, McPherson J (1987) Duplication of CaMV 35S promoter sequences creates a strong enhancer for plant genes. Science 236(4806):1299–1302

Kimber I, Kerkvliet NI, Taylor SL, Astwood JD, Sarlo K, Dearman RJ (1999) Toxicology of protein allergenicity: prediction and characterization. Toxicological Sciences 48(2):157–162

Kovalic D, Garnaat C, Guo L, Yan Y, Groat J, Silvanovich A, Ralston L, Huang M, Tian Q, Christian A, Cheikh N, Hjelle J, Padgette S, Bannon G (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(3):149–163

Lamppa GK, Morelli G, Chua N-H (1985) Structure and developmental regulation of a wheat gene encoding the major chlorophyll a/b-binding polypeptide. Molecular and Cellular Biology 5(6):1370–1378

Landschulz WH, Johnson PF, McKnight SL (1988) The leucine zipper: A hypotherical structure common to a new class of DNA binding proteins. Science 240(4860):1759–1764

Lorimier P, Lamarcq L, Labat-Moleur F, Guillermet C, Bethier R, Stoebner P (1993) Enhanced chemiluminescence: a high-sensitivity detection system for *in situ* hybridization and immunohistochemistry. The Journal of Histochemistry and Cytochemistry 41(11):1591–1597

McElroy D, Zhang W, Cao J, Wu R (1990) Isolation of an efficient actin promoter for use in rice transformation. Plant Cell 2(2):163–171

McElwain EF, Spiker S (1989) A wheat cDNA clone which is homologous to the 17 kd heat-shock protein gene family of soybean. Nucleic Acids Research 17(4):1764

Metcalfe DD, Astwood JD, Townsend R, Sampson HA, Taylor SL, Fuchs RL (1996) Assessment of the allergenic potential of foods derived from genetically engineered crop plants. Critical Reviews in Food Science and Nutrition 36 Suppl:S165–S186

Mukherjee K, Brocchieri L, Bürglin TR (2009) A comprehensive classification and evolutionary analysis of plant homeobox genes. Mol Biol Evol 26(12):2775–2794

OECD (2002) Consensus document on compositional considerations for new varieties of maize (*Zea mays*): Key food and feed nutrients, anti-nutrients and secondary plant metabolites. Organisation for Economic Co-operation and Development, Paris. http://www.oecd.org/officialdocuments/displaydocumentpdf?cote=env/jm/mono(2002)25&doclanguage=en

OECD (2003) Considerations for the safety assessment of animal feedstuffs derived from genetically modified plants. ENV/JM/MONO(2003)10. Organisation for Economic Co-operation and Development, Paris. http://www.oecd.org/officialdocuments/displaydocumentpdf?cote=ENV/JM/MONO(2003)10&doclanguage=en

OGTR (2008) The biology of *Zea mays* L. ssp *mays* (maize or corn). Prepared by the Office of the Gene Technology Regulator, Australia. http://www.ogtr.gov.au/internet/ogtr/publishing.nsf/Content/riskassessments-1

Park MY, Kim S, Lee S, Kim SY (2013) ATHB17 is a positive regulator of abscisic acid response during early seedling growth. Molecules and Cells 35(2):125–133

Pearson WR (2000) Flexible Sequence Similarity Searching with the FASTA3 Program Package. Ch 10 In: Misener S, Krawetz SA (eds) Methods in Molecular Biology, Volume 132: Bioinformatics Methods and Protocols. Human Press Inc., Totowa, NJ, p. 185–219

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. Proceedings of the National Academy of Sciences 85(8):2444–2448

Ransom J, Endres GJ (2014) Corn growth and management quick guide. A-1173, North Dakota State University Extension Service. http://www.ag.ndsu.edu/publications/landing-pages/crops/corn-growth-and-management-quick-guide-a-1173

Rice EA, Khandelwal A, Creelman RA, Griffith C, et al. (2014) Expression of a truncated ATHB17 protein in maize increases ear weight at silking. PLoS ONE (open access) 9(4):e94238

Ridley WP, Harrigan GG, Breeze ML, Nemeth MA, Sidhu RS, Glenn KC (2011) Evaluation of compositional equivalence for multitrait biotechnology crops. Journal of Agricultural and Food Chemistry 59:5865–5876

Salomon S, Puchta H (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. The EMBO Journal 17(20):6086–6095

Schabenberger O (2004) Mixed model influence diagnostics. Paper 189-29. Proceedings of the 29th Annual SAS Users Group International Conference (SUGI 29), May 9 - 12, 2004, Montreal. http://www2.sas.com/proceedings/sugi29/189-29.pdf

Schena M, Davis RW (1992) HD-Zip proteins: members of an *Arabidopsis* homeodomain protein superfamily. Proceedings of the National Academy of Sciences 89:3894–3898

Sessa G, Morelli G, Rubert I (1993) The Athb-1 and -2 HD-Zip domains homodimerize forming complexes of different DNA binding specificities. The EMBO Journal 12(9):3507–3517

Sidorov V, Duncan D (2009) *Agrobacterium*-mediated maize transformation: Immature embryos versus callus. Ch 4 In: Scott MP (ed) Methods in molecular biology: Transgenic maize. Humana Press Inc, Totowa, NJ, p. 47–58

Thomas K, MacIntosh S, Bannon G, Herouet-Guicheney C, Holsapple M, Ladics G, McClain S, Vieths S, Woolhiser M, Privalle L (2009) Scientific advancement of novel protein allergenicity evaluation: An overview of work from the HESI Protein Allergenicity Technical Committee (2000 - 2008). Food and Chemical Toxicology 47:1041–1050

Thomas K, Aalbers M, Bannon GA, Bartels M, Dearman RJ, Esdaile DJ, Fu T-J, Glatt CM, Hadfield N, Hatzos C, Hefle SL, Heylings JR, Goodman RE, Henry B, Herouet C, Holsapple M, Ladics GS, Landry TD, MacIntosh SC, Rice EA, Privalle LS, Steiner HY, Teshima R, Van Ree R, Woolhiser M, Zawodny J (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

U.S.Pharmacopeia (2000) The United States Pharmacopeia 24. The National Formulary 19th ed. United States Pharmacopeia Convention, Inc, Rockville, MD

White PJ, Pollak LM (1995) Corn as a food source in the United States: Part II. Processes, products, composition, and nutritive values. In: Cereal Foods World 40. p. 756–762

Yates B, De Soyza A, Harkawat R, Stenton C (2008) Occupational asthma caused by *Arabidopsis thaliana*: a case of laboratory plant allergy. European Respiratory Journal 32:1111–1112

Zhao Y, Zhou Y, Jiang H, Li X, Gan D, Peng X, Zhu S, Cheng B (2011) Systematic analysis of sequences and expression patterns of drought-responsive members of the HD-Zip gene family in maize. PLoS ONE (open access) 6(12):e28488

Zhou J, Harrigan GG, Berman KH, Webb EG, Klusmeyer TH, Nemeth MA (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