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

21 April 2023
237-23

Supporting document 1

Safety assessment – **Application A1264**

Food derived from drought-tolerant and herbicide-tolerant soybean line IND-00410-5

Executive summary

Background

Application A1264 seeks approval for the sale and use of food derived from genetically modified (GM) soybean line IND-00410-5 that has tolerance to drought and the herbicide glufosinate.

Drought tolerance is conferred by the expression of the novel transcription factor HaHB4, encoded by the *HaHB4* gene from sunflower. This novel transcription factor regulates gene transcription in IND-00410-5 soybean in response to environmental stressors, such as drought. The HaHB4 protein has previously been assessed by FSANZ in Application A1232.

Tolerance to glufosinate is achieved through expression of the enzyme phosphinothricin acetyltransferase (PAT), encoded by the *bar* gene from the soil bacterium *Streptomyces hygroscopicus*. FSANZ has assessed the PAT protein in numerous previous applications.

This safety assessment addresses food safety and nutritional issues associated with GM food. It therefore does not address:

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

History of use

Soybean is one of the leading oilseed crops in the world. It is grown as a commercial food crop worldwide and has a long history of safe use in the food supply. Soybean oil is widely used as cooking oil and an ingredient in a wide range of manufactured products, including shortening, margarine and confectionery products. Soybean grains are also used to make soy milk, soy sauce, soy lecithin and meat substitutes such as tofu and tempeh.

Molecular characterisation

The genes encoding HaHB4 (*HaHB4*) and PAT (*bar*) were introduced into soybean line IND-00410-5 via *Agrobacterium*-mediated transformation. Detailed molecular analyses indicate a

single, fully functional insert is present at one locus in the genome of IND-00410-5. There are no extraneous plasmid sequences or antibiotic-resistance genes present in this line.

The introduced genetic elements were shown by molecular techniques and phenotypic analyses to be present within a single locus and stably inherited across multiple generations.

Characterisation and safety assessment of new substances

Soybean line IND-00410-5 expresses the novel protein HaHB4. There is a history of human exposure to this protein through the consumption of sunflower seed, and homologous proteins found in the plant kingdom and commonly consumed food. As a transcription factor, it is expressed at very low levels in plants. A range of characterisation analyses confirmed the identity of HaHB4 in IND-00410-5. The safety of this protein has been assessed by FSANZ in a previous application (see Application A1232). Updated bioinformatic analyses undertaken for this application confirmed the expressed protein is unlikely to be allergenic or toxic to humans.

PAT is a newly expressed protein present in IND-00410-5. It is expressed at a low level in leaf tissues and at a higher level in grains. A range of characterisation analyses confirmed the identity of PAT in IND-00410-5. The safety of this protein has been assessed by FSANZ in numerous previous applications. Updated bioinformatic analyses undertaken for this application confirmed the expressed protein is unlikely to be allergenic or toxic to humans.

Herbicide metabolites

For PAT, the metabolic profiles resulting from the novel protein/herbicide interaction have been established through a significant history of use. There are no concerns that the spraying of soybean line IND-00410-5 with glufosinate would result in the production of metabolites that are not also produced in non-GM crops sprayed with the same herbicide and already used in the food supply.

Compositional analyses

Detailed compositional analyses were performed on IND-00410-5. Statistically significant differences in mean values were found between grain from IND-00410-5 and the control for 7 of the 44 analytes evaluated, however these differences were within the range established for existing commercial non-GM soybean varieties. Overall, the compositional data support the conclusion that there are no biologically significant differences in the levels of key constituents in grain from IND-00410-5 compared to non-GM soybean varieties available on the market.

Conclusion

No potential public health and safety concerns have been identified in the assessment of drought-tolerant and herbicide-tolerant soybean line IND-00410-5. On the basis of the data provided in the present application and other available information, food derived from IND-00410-5 is considered to be as safe for human consumption as food derived from non-GM soybean varieties.

Table of contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Definition
AFSI	Agriculture and Food Systems Institute
AQUA	Absolute quantification
bar	Gene from *Streptomyces hygroscopicus*
BLAST	Basic Local Alignment Search Tool
bp	Base pair
CERA	Centre for Environmental Risk Assessment
COMPARE	COMprehensive Protein Allergen REsource
Chr	Chromosome
DIG	Digitonin
DNA	Deoxyribonucleic acid
DW/dw	Dry weight
ELISA	Enzyme Linked Immunosorbent Assay
FAO	Food and Agriculture Organization of the United Nations
FSANZ	Food Standards Australia New Zealand
FW	Fresh weight
Gb	Gigabase
gDNA	Genomic DNA
GM	Genetically modified
HaHB4	Transcription factor gene from sunflower (*Helianthus annuus*)
HaHB4	Protein encoded by the *HaHB4* gene
HD	Homeodomain
IND-ØØ41Ø-5	OECD unique identifier for the soybean event selected for commercial approval
kDa	Kilodalton
kg	Kilogram
LB	Left border
LC MS/MS	Liquid chromatography-tandem mass spectrometry
LLOQ	Lower limit of quantification
LOD	Limit of detection
LPF	Long fragment promoter
mg	milligram
MT	Million tonnes
NCBI	National Centre for Biotechnology Information

Abbreviation	Definition
NGS	Next Generation Sequencing
OECD	Organisation for Economic Cooperation and Development
ORF	Open reading frame
PAT	Phosphinothricin-N-acetyl transferase
PCR	Polymerase chain reaction
ppm	Parts per million
RB	Right border
SE	Standard error
T-DNA	Transfer DNA

1 Introduction

FSANZ has received an application from Bioceres Crop Solutions to vary Schedule 26 in the Australia New Zealand Food Standards Code (the Code). The variation is to include food derived from the genetically modified (GM) soybean line IND-00410-5, with the OECD Unique Identifier IND-ØØ41Ø-5. This soybean line is tolerant to drought and the herbicide glufosinate.

Drought tolerance is achieved by expression of a sunflower transcription factor HaHB4 that drives the expression of abiotic stress response genes. The response increases the plant's tolerance to environmental stresses such as water scarcity and salinity, preventing crop yield loss. Tolerance to the herbicide glufosinate is achieved by the expression of phosphinothricin acetyltransferase (PAT) enzyme. PAT is encoded by the *bar* gene, which is derived from the bacterium *Streptomyces hygroscopicus*. Both the PAT protein and the HaHB4 protein have been assessed previously by FSANZ[1].

If approved, food derived from soybean line IND-00410-5 may enter the Australian and New Zealand food supply as imported food products.

2 History of use

2.1 Host organism

The host organism is soybean (*Glycine Max* L.), belonging to the family *Leguminosae* and the variety used for the genetic modification is Williams 82. The Williams 82 host organism was used as the conventional control for the purposes of comparative assessment with IND-00410-5.

Soybean has a long history of cultivation and human consumption (Hymowitz 1970). The commodity is the leading oilseed crop in the world, with total global production reaching 371.7 MT[2] in 2021. Figure 1 shows the major soybean producing countries in the world.

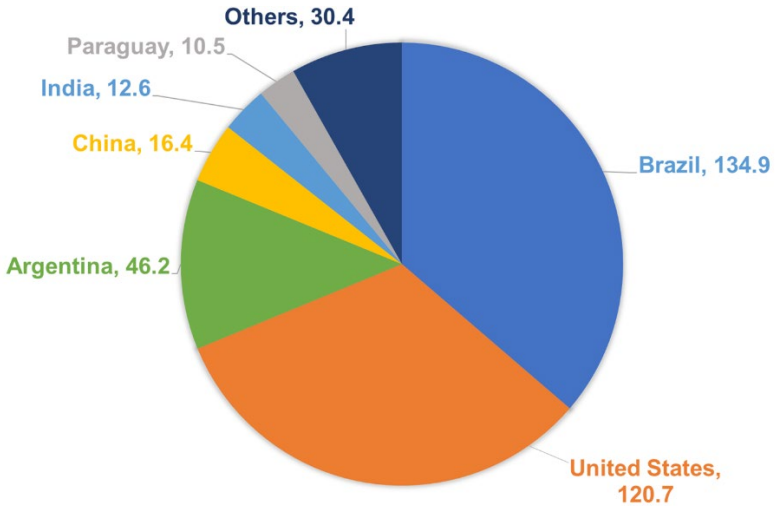


Figure 1: Major soybean producing countries in 2021 (in MT). Data obtained from FAOSTAT (2023)

[1] HaHB4 - Application A1232; PAT – 29 Applications
[2] Million tonnes

Australia remains a minor player in soybean production, producing 0.043 MT tonnes in 2021. According to FAOSTAT, New Zealand has no commercial soybean cultivation. Australia and New Zealand are net importers of soybeans, with 2325.03 tonnes and 3044.08 tonnes imported respectively in 2021 (FAOSTAT 2023).

Soybean production contributes significantly to Australia's $2.5 billion oilseed industry. Additionally, it is commonly used in farming systems as a rotational crop, either for green manure or forage and grain for animal feed. The Australian soybean growing region stretches from northern Queensland to southern New South Wales and Victoria (Australian Oilseeds Federation 2023).

Soybean grains are processed into two major products: oil and meal. The oil is the most important product for human consumption and has a long history of safe use. Soybean oil is used in a variety of manufactured foods, including cooking oil, shortening, margarine, frozen desserts and confectionery products. Soybean meal is a good source of protein and is primarily processed into livestock feed (pet and poultry food) and protein products such as soy flour, concentrates and isolates. Soybean is also used to make soy milk, soy sauce, soy lecithin and meat substitutes such as tofu and tempeh.

Soybeans may also be eaten with minimal processing, for example in the Japanese food edamame, in which immature soybeans are boiled whole in their pods and served with salt. Unprocessed (raw) soybean grain products are not suitable for food, due to presence of anti-nutrients, such as phytic acid and isoflavones (OECD 2012). The processing used with the soybean products inactivate these anti-nutrients, making them suitable for food use.

2.2 Donor organisms

2.2.1 Helianthus annuus

The *HaHB4* DNA sequence encoding the HaHB4 protein is derived from *Helianthus annuus*, also known as the common sunflower. The sunflower is native to North America and has been grown as a food crop for thousands of years (Lentz et al. 2008). Sunflower seeds are either eaten whole, milled for flour or meal to make bread and other baked goods, or the oil is extracted and used for cooking (Adeleke and Babalola 2020). Hence, the seeds have a history of safe human consumption.

2.2.2 Streptomyces hygroscopicus

The *bar* gene encodes the PAT protein and is derived from *Streptomyces hygroscopicus*, a non-pathogenic gram-positive spore-forming bacterium found in soil. FSANZ has previously assessed and approved 10 applications containing this gene sequence. The *bar* gene has been used to confer tolerance to glufosinate ammonium herbicides in food-producing crops around the world for over two decades (OECD 2006; CERA 2011).

2.2.3 Other organisms

Genetic elements from several other organisms have been used in the genetic modification of IND-00410-5 (refer to Table 1). These genetic elements are non-coding sequences that are used to regulate the expression of *HaHB4* and *bar*.

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.

3.1 Transformation method

To create the IND-00410-5 soybean line, the soybean variety Williams 82 was transformed using the plasmid pIND2-HB4 (Figure 2). Transformation of Williams 82 was achieved by incubating pre-germinated seedlings with *Agrobacterium tumefaciens* containing the pIND2-HB4 plasmid. The methodology is outlined in flowchart in Appendix 1 and summarised below.

The pre-germinated seedlings were then placed on shoot induction selective medium containing glufosinate and antibiotics. Glufosinate inhibits the growth of untransformed plant cells, while the antibiotics suppresses the growth of excess *Agrobacterium*. The seedlings were sub-cultured to a fresh medium every two weeks until true leaves[3] were visible.

Regenerated stems with leaves were excised and transferred to shoot elongation medium to promote shoot growth. Elongated shoots were then transferred to the rooting medium to promote root growth. Up to this point, the seedlings were maintained in glufosinate-containing selective medium.

Rooted plantlets were then transferred to soil, where they were tested for the presence of the *HaHB4* and *bar* genes using standard molecular biology techniques. Following the evaluation of insert integrity, gene expression, phenotypic characteristics and agronomic performance, soybean line IND-00410-5 was selected.

3.2 Detailed description of inserted DNA

Soybean line IND-00410-5 contains T-DNA from pIND2-HB4 plasmid (Figure 2) and includes the *HaHB4* and *bar* expression cassettes. Information on the genetic elements in the T-DNA used for transformation is summarised in Table 1. Additional detail, including the plasmid backbone and intervening sequences used to assist with cloning, mapping and sequence analysis, can be found in Appendix 2.

[3] The second set of leaves to grow, and are able to perform photosynthesis to start supplying the plant with food for its next growth stage.

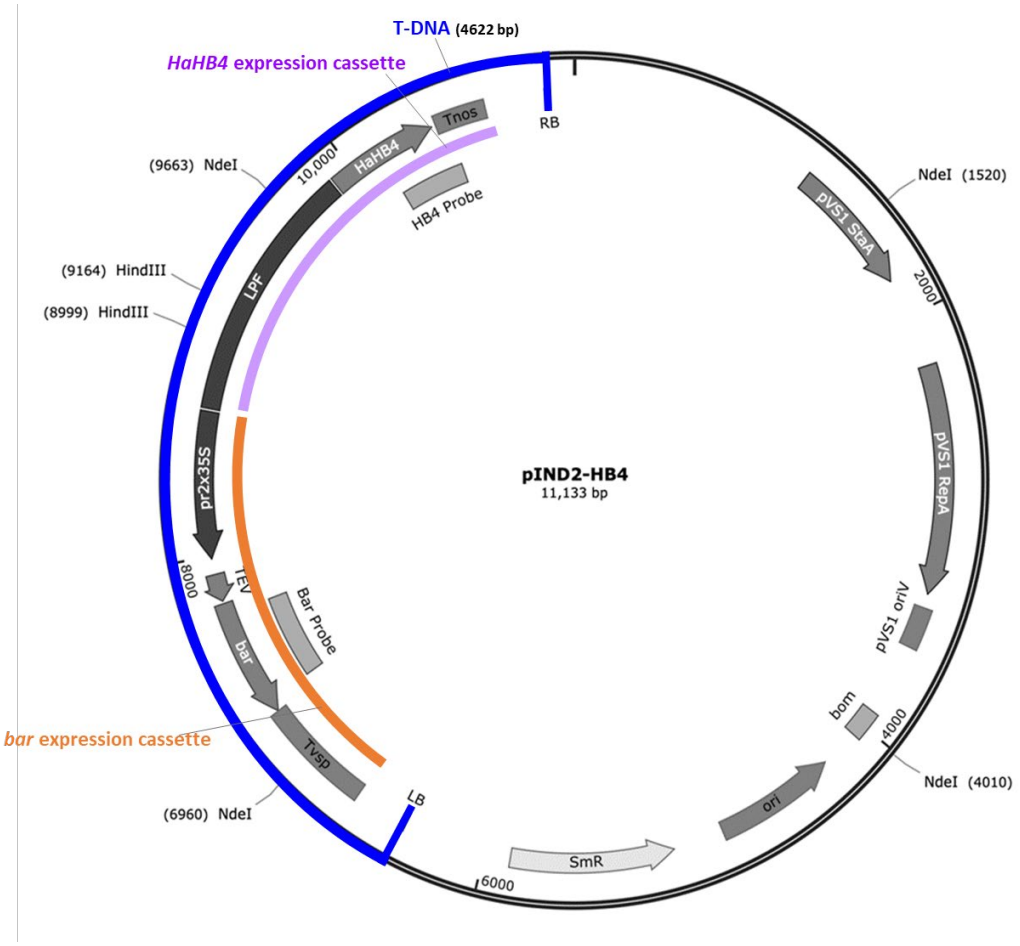


Figure 2: Plasmid map of pIND2-HB4. The T-DNA region comprising the HaHB4 and bar expression cassettes is highlighted using the blue bar. The HaHB4 expression cassette is highlighted using the purple bar. The bar expression cassette is highlighted using the orange bar.

Table 1: Expression cassettes contained in the T-DNA of pIND2-HB4

Expression cassette	Promoter (Drives expression)	Intron-containing 5'UTR (Expression enhancer)	Coding sequence	Terminator (Polyadenylation and termination of transcription)
HaHB4 expression cassette	Native allelic promoter of *HaHB4* gene, large promoter fragment (LPF) from *Helianthus annus*	–	*HaHB4* coding sequence from *Helianthus annus*	3'UTR derived from nopaline synthase (*NOS*) gene from *Agrobacterium tumefaciens*
bar expression cassette	2x35S Promoter from *Cauliflower Mosaic Virus*	5´ UTR derived from Tobacco Etch Virus (TEV)	*bar* coding sequence from *Streptomyces hygroscopicus*	3'UTR derived from vegetative storage protein (*VSP*) gene from *Glycine max*

3.3 Development of the soybean line from the original transformant

A breeding program was undertaken for the purpose of:

- obtaining generations suitable for analysing the characteristics of soybean line IND-00410-5; and
- ensuring that the IND-00410-5 event is incorporated into elite lines for commercialisation.

The breeding history of IND-00410-5, showing the generations used for characterisation studies, is shown in Figure 3. Table 2 indicates the specific generations and controls used in the characterisation of IND-00410-5.

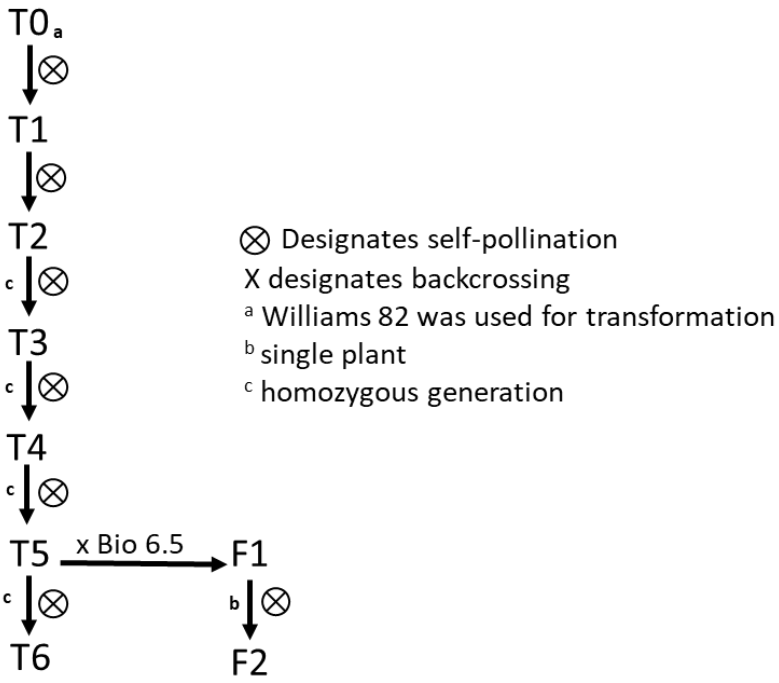


Figure 3: Breeding path used in the characterisation of IND-00410-5.

Table 2: IND-00410-5 generations used for various analyses

Analysis	Section	Generation(s) used	Comparators
Number of integration sites	Section 3.4.1	T5	*Glycine Max* L. Williams 82, *pIND-HB4*
Absence of backbone and other sequences	Section 3.4.2	T5	*Glycine Max* L. Williams 82, *pIND-HB4*
Insert integrity and site of integration	Section 3.4.3	T6	*Glycine Max* L. Williams 82
Genetic stability	Section 3.4.4.1	T1, T3, T5, T6	*Glycine Max* L. Williams 82
Mendelian inheritance	Section 3.4.4.2	F2	*Glycine Max* L. Williams 82 x *Bio 6.5*
Expression of phenotype over several generations	Section 3.4.4.2	T6	*Glycine Max* L. Williams 82
Compositional analysis	Section 5	T6	*Glycine Max* L. Williams 82

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

A range of analyses were undertaken to characterise the genetic modification in IND-00410-5. These analyses focused on the nature and stability of the insertion and whether any unintended re-arrangements or products may have occurred as a consequence of the transformation procedure.

3.4.1 Number of integration site(s)

Southern blot analysis was used to analyse the insertion site(s) and determine copy number. Genomic DNA (gDNA) from T5 homozygous IND-00410-5 plants was digested with restriction enzymes (*Hin*dIII or *Nde*I) and hybridised with DIG-labelled probes for either *HaHB4* or *bar*. gDNA from the conventional control (Williams 82) served as a negative control, while gDNA spiked with plasmid pIND2-HB4 served as a positive control. As expected, no hybridisation was observed in the non-spiked control samples while the data for IND-00410-5 revealed the presence of a single copy of T-DNA in the host genome.

To further confirm the Southern blot results, next-generation sequencing (NGS) was performed on gDNA isolated from the embryo axes of IND-00410-5 and control. In addition, plasmid pIND2-HB4 DNA was sequenced to serve as a reference. To assess the sensitivity of the NGS method, plasmid DNA was spiked and sequenced. Paired end reads (2x100 bp) with a total of 387.9 Gb of data was generated. Sufficient sequence reads were obtained to cover the inserted T-DNA and a selected soybean endogenous gene, with depth coverage of 30X and an adequate level of sensitivity[4].

Comparison of the sequence between IND-00410-5 and pIND2-HB4 detected two unique insert-flank junction sites, each comprised of the inserted T-DNA border sequence joined to a flanking sequence in the soybean genome. This indicates that a single copy of the intended DNA insert has been integrated into the genome of IND-00410-5 (Figure 4). As expected, no junction sites were detected in the control.

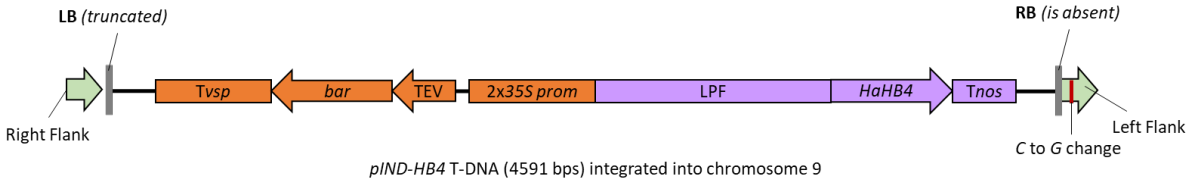


Figure 4: T-DNA insert present in IND-00410-5.

3.4.2 Absence of backbone and other sequences

NGS reads from IND-00410-5 (T5) and the pIND2-HB4 transformation plasmid were aligned. The results of this alignment confirmed there was no integration of pIND2-HB4 backbone sequences, including any antibiotic resistance genes, into the IND-00410-5 genome. Further confirmation was obtained using a Southern blot assay with probes for vector backbone sequences. The results were consistent with the NGS findings.

3.4.3 Insert integrity and site of integration

Locus-specific PCR and DNA sequence analysis of leaf-derived gDNA from IND-00410-5 showed that a single copy of the T-DNA from pIND2-HB4 was integrated into the host genome and the organisation of the genetic elements within the insert is as expected. No deletions, insertions, mutations or rearrangements of the expression cassettes were detected when the IND-00410-5 sequence was aligned with the T-DNA of the plasmid sequence. These results were fully consistent with the NGS dataset.
To examine the T-DNA insertion site, flanking sequences were obtained by aligning the NGS data of IND00410-5 and pIND2-HB4 plasmid sequence (Kovalic et al 2012). The identified soybean sequences flanking the insertion site were further subjected to homology searches

[4] The NGS method was sufficiently sensitive to detect 100% of the spiked plasmid when present at 1/10th of a copy per genome equivalent or greater.

against the reference genome sequence of the conventional control[5] (Altschul et al. 1990). These searches located the T-DNA insert at a single location in chromosome 9. A 142 bp deletion of the soybean genome at the T-DNA integration site was identified and this corresponded to an intergenic region. The insertion did not disrupt any genes or any other known annotated feature in the soybean genome.

The right border (RB) region of the inserted T-DNA was absent and there was some truncation of the left border (LB) region. In addition, a single nucleotide difference was identified when comparing the left flanking region in the transformed line and the reference genome (Figure 4). Such changes are common during *Agrobacterium*-mediated plant transformation due to double-strand break repair mechanisms (Gheysen et al. 1991; Mayerhofer et al. 1991; Gelvin 2021). These changes would not affect the expression of the *HaHB4* and *bar* genes.

3.4.4 Stability of the genetic changes in IND-00410-5

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 events) over successive generations. Phenotypic stability refers to the expressed trait remaining unchanged over successive generations.

3.4.4.1 Genetic stability

Genetic stability of the inserted DNA in IND-00410-5 was characterised through locus-specific PCR combined with Sanger sequencing analysis (T1, T3, T5) and NGS (T6). The analysis directly compared the inserted DNA and the adjacent flanking DNA for the four IND-00410-5 generations. This analysis showed the presence of the same two insert-flank junction sequences, as described in Section 3.4.1 in all generations. No other junction sequences were present. The consistency of these results across all generations tested demonstrates that the single inserted DNA is stably maintained in IND-00410-5.

3.4.4.2 Phenotypic stability

Mendelian inheritance

Since the inserted DNA resides at a single locus within the IND-00410-5 genome, it would be expected to be inherited according to Mendelian principles. Chi-square (X^2) analysis was undertaken in T5 F2 plants to confirm the segregation and stability of the inserted DNA.

A segregation analysis was performed on progeny derived from a cross between IND-00410-5 at T5 and a non-GM commercial variety Bio 6.5. The resulting F1 progeny were self-pollinated and 73 F2 plants were analysed by PCR for the presence insert-flanking DNA the native sequence. F2 plants were scored as:

- homozygous (I) when the amplicon for the insert-flanking DNA was present and the native sequence was absent.
- Hemizygous (H) when both the insert-flanking DNA and native sequence were present.
- Wildtype (W) when the amplicon for the insert-flanking DNA was absent and the amplicon for the native sequence was present.

According to Mendelian inheritance principles, the expected segregation ratio in the F2

[5] NCBI, Glycine max cultivar Williams 82 v4.0, Assembly Accession GCF_000004515.6.

generation is 1:2:1. The observed ratio presented in Table 3 matches the expected results, indicating the inserted DNA follows expected Mendelian inheritance rules.

Table 3: Segregation results of T-DNA in IND-00410-5 x Bio 6.5 at F2

Expected genotypes (number of plants)			Observed genotypes (number of plants)			χ^2	p-value
I	H	W	I	H	W	0.34246	0.8426
18.25	36.5	18.25	17	39	17		

I: IND-00410-5 homozygous; H: hemizygous; W: Wildtype.

Expressed phenotype over several generations

The expression of the HaHB4 protein in two generations of IND-00410-5 was examined. The drought-tolerant phenotype was assessed using T3 grain in 30 field trials in 16 different sites across six growing conditions. In comparison to the control, IND-00410-5 T3 plants conferred drought tolerance, particularly in warm and dry conditions (Ribichich et al. 2020). The LC-MS/MS protein quantification method was used with IND-00410-5 T6 generation to quantify HAHB4 protein expression. Grain and leaf tissues were collected from field trials in the United States and Argentina during the 2012-2013 growing seasons, with a conventional control serving as a negative control and *E.coli*-produced HaHB4 serving as an analytical reference standard. As expected for transcription factors, HaHB4 expression was found to be extremely low in IND-00410-5 grain and leaf tissues (Section 4.1.1).

The expression of the PAT protein was examined in two generations of IND-00410-5. T0 plants were grown in a glufosinate-containing medium, yielding 35 individual T1 lines with glufosinate tolerance (Ribichich et al. 2020). The level of PAT protein expression in IND-00410-5 T6 generation was determined using an ELISA kit. PAT protein expression was found to be comparable in IND-00410-5 to that previously reported in glufosinate-tolerant GM crops. (Block et al. 1987; CERA 2011). The PAT protein was not detected in the conventional control (Section 4.2.1).

Together this information supports the conclusion that HaHB4 and PAT proteins are stably expressed over several generations and multiple growing conditions.

3.4.5 Open reading frame (ORF) analysis

A bioinformatic analysis of the IND-00410-5 insert, as well as the flanking DNA regions, was undertaken to identify any novel open reading frames (ORFs) had been created in IND-00410-5 as a result of T-DNA insertion, a whether any putative peptides present in the insert have the potential for allergenicity or toxicity.

Sequence spanning the right and left insert-flank junction of IND-00410-5 were translated *in silico* from start-to-stop codon (TGA, TAG, TAA) in all six reading frames using a BioPython[6] script. A total of 74 ORFs were identified that correspond to putative peptides of eight amino acids or greater in length from the insert-flank junction sequences. If small putative proteins were contained within a larger sequence, the entire larger sequence was included in the ORF count and subsequent *in silico* analysis.

The 74 putative peptides were initially screened using the NCBI protein BLAST search tool[7]. The NCBI search showed 62 putative peptides did not align significantly (E score $<10^{-5}$) to any protein in this database. Two had significant homologies with the newly expressed proteins (HaHB4 and PAT), four are associated with proteins from the donor, the cauliflower

[6] http://www.biopython.org/
[7] blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch

mosaic virus, three are associated with known non-coding sequences (35S promoter, sunflower promoter) and three were similar to vectors and/or hypothetical proteins.

Putative peptides were used as query sequences in homology searches for known allergens and toxins in established databases. These analyses are theoretical only as there is no reason to expect that any of the identified ORFs would, in fact, be expressed.

3.4.5.1 Bioinformatic analysis for potential allergenicity

The 74 putative peptides were queried against known allergenic proteins listed in the Allergen Online database[8] (version 21). At the date of the search, there were 2,233 sequences in the allergen database. With an 80-mer sliding window, none of the putative peptides shared similarity ≥ 35% to any known allergen. Similar negative results were obtained using an 8-mer sliding window.

A conformational analysis of the putative peptides against known allergens was also performed using the Structural Database of Allergenic Proteins[9]. This database groups 1526 allergens, 1312 protein sequences, 92 crystallographic structures, 458 three-dimensional models and 29 IgE epitopes. No similarity was identified for any of the 74 putative peptides, when queried against all allergens and food allergens.

3.4.5.2 Bioinformatic analysis for potential toxicity

Putative peptides were examined for the presence of any known toxins found in the Toxin and Toxin Target Database[10] and the Toxin-antitoxin database[11]. Significant homology was determined based on a E score of $<10^{-5}$ and was detected with PAT-associated peptides (considered in Section 4.2.3). No other significant homology was found with the putative peptides and known toxins.

3.4.6 Conclusion

The data provided by the applicant showed that an integration event has occurred at a single locus in the soybean genome. The sequencing data confirmed a single, fully intact T-DNA with *HAHB4* and *bar* expression cassettes in the genome of IND-00410-5. No plasmid backbone sequences, including antibiotic resistance genes, from the transforming pIND2-HB4 plasmid were present. The introduced DNA was shown to be stably inherited from one generation to the next. No new ORFs that raise potential allergenicity or toxicity concerns were created by the insertion.

[8] www.allergenonline.org
[9] fermi.utmb.edu/
[10] www.t3db.ca/
[11] bioinfo-mml.sjtu.edu.cn/TADB2/tools.html

4 Characterisation and safety assessment of novel substances

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

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

Two novel substances are expressed in IND-00410-5, HaHB4 and PAT, and are assessed below.

4.1 HaHB4

The *Helianthus annuus* homeobox 4 (HaHB4) protein is a transcription factor. It is involved in regulating gene transcription in response to environmental stressors such as drought (Dezar et al. 2005b; Manavella et al. 2008b; Gonzalez et al. 2019).

Soybean line IND-00410-5 expressing HaHB4 protein shows increased grain yield compared to its non-GM control (Ribichich et al. 2020). The relative increase in grain yields in HaHB4 soybean is correlated to water use efficiency under conditions of water scarcity. Differences in grain yield are reflected by increased numbers of grain. This is associated with changes in plant architecture and an increase in photosynthetic rate during the critical period of grain filling, as well as an increase in the number of branches and grain pods per plant (Ribichich et al. 2020).

HaHB4 is a member of the homeodomain-leucine zipper (HD-Zip) gene family and was recently assessed by FSANZ in Application A1232. The gene family is unique to the plant kingdom and found in food crops with a history of safe use such as rice and wheat (Ariel et al. 2007; Yue et al. 2018). Homologous sequences to HaHB4 are found in commonly consumed food, such as artichoke, golden kiwifruit and citrus. Sunflower, the source of the gene encoding the HaHB4 protein, also has a long history of safe use as food (see Section 2.2.1). Hence there is a history of human exposure to this specific protein as well as proteins from the same gene family.

The *HaHB4* gene prepared by the applicant encodes a protein of 177 amino acids, with an expected mass of ~20.9 kDa and differs from sunflower HaHB4 by seven amino acids, which is equivalent to 96.1% similarity (Figure 5). These differences include the deletion of four amino acids at positions 7-10, and three amino acid substitutions: lysine to arginine at position 22, phenylalanine to leucine at position 159, and proline to leucine at position 175. These changes do not impact the conserved domains of the HD-Zip family. The proline to leucine substitution at position 175 is also found in a native sunflower gene variant[12].

[12] Accession number XP_022022563.1

```
HAHB4 in Sunflower  MSLQQVPTTETTTRKNRNEGRKRFTDKQISFLEYMFETQSRPELRMKHQL  50
HAHB4 in A1232      MSLQQV----TTTRKNRNEGRRRFTDKQISFLEYMFETQSRPELRMKHQL  46
HAHB4 in A1264      MSLQQV----TTTRKNRNEGRRRFTDKQISFLEYMFETQSRPELRMKHQL  46

HAHB4 in Sunflower  AHKLGLHPRQVAIWFQNKRARSKSRQIEQEYNALKHNYETLASKSESLKK 100
HAHB4 in A1232      AHKLGLHPRQVAIWFQNKRARSKSRQIEQEYNALKHNYETLASKSESLKK  96
HAHB4 in A1264      AHKLGLHPRQVAIWFQNKRARSKSRQIEQEYNALKHNYETLASKSESLKK  96

HAHB4 in Sunflower  ENQALLNQLEVLRNVAEKHQEKTSSSGSGEESDDRFTNSPDVMFGQEMNV 150
HAHB4 in A1232      ENQALLNQLEVLRNVAEKHQEKTSSSGSGEESDDRFTNSPDVMFGQEMNV 146
HAHB4 in A1264      ENQALLNQLEVLRNVAEKHQEKTSSSGSGEESDDRFTNSPDVMFGQEMNV 146

HAHB4 in Sunflower  PFCDGFAYFEEGNSLLEIEEQLPDPQKWWEF 181
HAHB4 in A1232      PFCDGFAYLEEGNSLLEIEEQLPDLQKWWEF 177
HAHB4 in A1264      PFCDGFAYLEEGNSLLEIEEQLPDLQKWWEF 177
```

Figure 5: Amino acid sequence alignment of HAHB4 in sunflower, A1232 and A1264 (IND-00410-5). Dashes represent amino acid deletions. Red text and yellow highlight represent amino acid substitutions.

4.1.1 Expression of HaHB4 in IND-00410-5 tissue

As a transcription factor, HaHB4 is most likely expressed at very low levels. An absolute quantification (AQUA) method of protein quantification by targeted liquid chromatography-tandem mass spectrometry (LC MS/MS) was used to detect HaHB4 in soybean line IND-00410-5 as standard molecular methodologies lack the required sensitivity (Gerber et al. 2003; Skinner et al. 2013).

The LC MS/MS analysis was performed on grain and leaf tissue samples from IND-00410-5 and the non-GM Williams 82 control. Samples were obtained from field trials in six different locations in Argentina during the 2012-2013 season and five different locations in the United States during the 2013 season. Steps were taken to maximise the recovery of proteins and sensitivity of the mass spectrometry detection. *E. coli*-derived HaHB4 was spiked in the Williams 82 control sample for an analytical reference standard. None of the grain samples produced a detectable HaHB4 signal[13]. HAHB4 protein was detected in two leaf tissue samples from two United States field trials of IND-00410-5 (5 and 4 ng/g dry weight (DW), respectively), though still below the lower limit of quantification (LLOQ)[14].

These results confirm that the levels of HaHB4 expression in grain and leaf are very low, consistent with the expression of native transcription factors.

4.1.2 Characterisation of HaHB4 expressed in IND-00410-5

The *HaHB4* gene prepared by the applicant encodes a protein of 177 amino acids. The protein sequence is perfectly matched to the expected sequence used in a previous application assessed and approved by FSANZ (Figure 5; Application A1232). It is therefore expected that the HaHB4 protein found in IND-00410-5 is structurally and biochemically similar to HaHB4 assessed in Application A1232.

In terms of function, the expression of HaHB4 protein in IND-00410-5 provides the soybean with tolerance to drought under field conditions (Ribichich et al. 2020).

[13] The limit of detection for HaHB4 was 0.007 µg/g and 0.003 µg/g DW for grain and leaf tissue, respectively.
[14] The lower limit of quantification was 0.027 µg/g and 0.041 µg/g DW for grain and leaf tissue, respectively.

4.1.3 Safety of the introduced HaHB4

The HaHB4 protein, encoded by the HaHB4 gene, has been considered by FSANZ in a previous assessment (Application A1232). In the previous FSANZ assessment, studies on potential allergenicity and toxicity were submitted and assessed. These previous assessments did not raise any safety concerns. Since the sequence of the protein expressed in IND-00410-5 is identical to the previous HaHB4 sequence assessed by FSANZ, no further safety evaluation is required other than the examination of updated bioinformatic searches[15].

Bioinformatic analyses of HaHB4

The applicant has submitted updated bioinformatic studies for HaHB4 that looked for amino acid sequence similarity to known protein allergens and toxins (March 2022). FSANZ has assessed the data submitted by the applicant and the results do not alter conclusions reached in previous assessments.

4.1.3 Conclusion

The data presented by the applicant confirms the HaHB4 expressed in IND-00410-5 is identical to previously assessed HaHB4 protein. HaHB4 in IND-00410-5 is functional, as it provides drought tolerance to the soybean (Ribichich et al. 2020). Expression studies confirmed very low expression levels of HaHB4 in IND-00410-5 tissue, similar to native transcription factors. Updated bioinformatic analyses confirm that the HaHB4 protein has no similarity with known toxins and allergens that is of significance or concern.

4.2 PAT

The *bar* gene from *S. hygroscopicus* encodes the PAT enzyme and confers tolerance to the antibiotic called bialaphos (Murakami et al. 1986; Thompson et al. 1987). This antibiotic is also produced by *S. hygroscopicus* i.e. the bacterium has evolved a mechanism to avoid the toxicity of its own product (Hara et al. 1991). Bialaphos, now also used as a non-selective herbicide, is a tripeptide comprising two L-alanine residues and an analogue of glutamate known as L-phosphinothricin (L-PPT) (Thompson et al. 1987) more recently known also as glufosinate ammonium. Free L-PPT released from bialaphos by peptidases (or applied directly as a synthetic herbicide) inhibits glutamine synthetase which in turn leads to rapid accumulation of ammonia and subsequent cell death.

The PAT protein encoded by the *bar* gene from *S. hygroscopicus* is homologous to the *pat* gene from *S. viridochromogenes*. They both are acetyl transferases with enzyme specificity for both L-PPT and demethylphosphinothricin (DMPT) in the acetylation reaction (Thompson et al. 1987). In the presence of acetyl-Coenzyme A (CoA), PAT catalyses the acetylation of the free amino group of L-PPT to N-acetyl-L-PPT, a herbicidally-inactive compound. The kinetics and substrate specificity of the PAT enzyme are well characterised; it has a high specificity for L-PPT and has been shown to have a very low affinity to related compounds and amino acids; even excess glutamate is unable to block the L-PPT-acetyltransferase reaction (Thompson et al. 1987). The proteins from the two different sources have a sequence identity of 85% (Wehrmann et al. 1996).

The commercialisation of plants engineered for glufosinate-tolerance using PAT began in the mid-1990s (CERA 2011). The history of use of the enzyme in crops therefore extends about 25 years, with FSANZ having assessed and approved 10 events across six commodities with *bar* encoded glufosinate-tolerance and 29 events in total for glufosinate-tolerance.

[15] The applicant did, however, provide additional data relevant to the assessment of the potential toxicity and allergenicity of HaHB4. These data do not change the conclusions reached in the previous assessment.

4.2.1 Expression of PAT in IND-00410-5 tissue

Protein expression in plant tissues was determined by ELISA. An analytical reference standard for plant-derived PAT was generated using a recombinant PAT protein.

In order to determine the sites of accumulation of the protein, samples were collected from IND-00410-5 grown in six field-trial sites in Argentina[16] during the 2012-2013 growing season and five field trials in United States[17] during the 2013 growing season. Leaf and grain tissues were examined from IND-00410-5 and the conventional control (Williams 82). Tissue samples were collected during the pod development stage. For each tissue sample analysed, four samples were processed from each field-trial site.

The results from the protein analysis showed the maximum levels found in grain were 69.05 µg/g of fresh weight (FW) and 12.68 µg/g FW in leaves (Table 4). These values are comparable to those previously reported for PAT protein in other GM crops, such as 127 µg/g FW (cotton seed) and 935 µg/g FW (corn leaf) (CERA 2011). Values differ across locations due to minor variations in weather conditions, as described for other PAT-expressing crop plants. There was no detection of PAT in the control. This result is as expected because the control does not contain the *bar* gene.

Table 4: Expression of PAT (µg/g FW) in leaf and grain tissues in IND-00410-5

		IND-00410-5	
		Leaf Mean ± SE	**Grain Mean ± SE**
Argentina	A	7.49 ± 1.39	69.05 ± 1.07
	D2	9.51 ± 0.56	34.49 ± 1.55
	G1	6.72 ± 1.00	30.33 ± 1.20
	Q1	5.44 ± 0.74	65.57 ± 1.54
	Q2	7.46 ± 1.61	68.70 ± 1.35
	W1	7.74 ± 0.65	23.00 ± 2.83
United States	IL3	10.19 ± 0.39	48.46 ± 1.82
	IN	12.14 ± 0.14	46.47 ± 6.25
	OH2	12.68 ± 0.93	58.68 ± 2.69
	IA	8.87 ± 1.09	58.31 ± 0.74
	KS	9.90 ± 1.69	50.80 ± 6.03

The field locations in Argentina were: Monte Buey, Cordoba (A); Corral de Bustos, Cordoba (D2); Carmen de Areco, Buenos Aires (G1); Hughes, Santa Fe (Q1); Hughes, Santa Fe (Q2); and Aranguren, Entre Rios (W1).
United States were: Effingham, IL (IL3); Ladoga, IN (IN); Pemberton, OH (OH2); Richland, IA (IA); and Troy KS (KS).

4.2.2 Characterisation of PAT expressed in IND-00410-5

The *bar* gene prepared by the applicant encodes a protein of 183 amino acids. The protein sequence is perfectly matched to the expected PAT protein sequence from *S. hygroscopicus* and is the same sequence used in previous applications assessed and approved by FSANZ. It is therefore expected that the PAT protein found in IND-00410-5 is structurally and biochemically similar to PAT found in other plants or from bacteria.

In terms of function, the expression of PAT protein in IND-00410-5 provides the soybean with tolerance to glufosinate. This was demonstrated during the transformation and selection

[16] The field locations in Argentina were: Monte Buey, Cordoba (A); Corral de Bustos, Cordoba (D2); Carmen de Areco, Buenos Aires (G1); Hughes, Santa Fe (Q1); Hughes, Santa Fe (Q2); and Aranguren, Entre Rios (W1).
[17] The field locations in the United States were Effingham, IL (IL3); Ladoga, IN (IN); Pemberton, OH (OH2); Richland, IA (IA); and Troy KS (KS).

process (Section 3.1).

4.2.3 Safety of the introduced PAT

The PAT protein, encoded by either the *pat* or *bar* genes (Wehrmann et al. 1996; Herouet et al. 2005), has now been considered in 29 FSANZ safety assessments[18]. These assessments, together with the published literature, firmly establish the safety of PAT and confirm that it does not raise toxicity or food allergenicity concerns in humans (Herouet et al. 2005; Delaney et al. 2008; Hammond and Jez 2011; ILSI 2016).

In previous FSANZ assessments, studies on potential allergenicity and toxicity were submitted and assessed. These previous assessments did not raise any safety concerns and there have been no credible reports of adverse health effects in humans.

The applicant has submitted updated bioinformatic studies for PAT that looked for amino acid sequence similarity to known protein allergens and toxins (March 2022). FSANZ has assessed the data submitted by the applicant and the results do not alter conclusions reached in previous assessments.

Since the sequence of the protein expressed in IND-00410-5 is identical to the previous PAT sequences assessed by FSANZ, no further safety evaluation is required.

4.2.4 Conclusion

The data presented by the applicant confirms the PAT expressed in IND-00410-5 is identical to previously assessed PAT proteins. IND-00410-5-derived PAT is immunoreactive to a PAT antibody and is functional i.e. provides glufosinate tolerance. The protein is expressed in various plant tissues, including grain. Updated bioinformatic analyses confirm that PAT has no similarity with known allergens or toxins that is of significance or concern.

4.3 Herbicide metabolites

FSANZ has assessed the novel herbicide metabolites for glufosinate in GM crops in multiple previous applications. These previous assessments indicate the spraying of IND-00410-5 with glufosinate ammonium would result in the same metabolites that are produced in non-GM soybean sprayed with the same herbicide. As no new glufosinate metabolites would be generated in soybean event IND-00410-5, further assessment is not required.

5 Compositional analysis

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

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

[18] A372, A375, A380, A385, A386, A446, A481, A518, A533, A543, A589, A1028, A1040, A1046, A1073, A1080, A1081, A1087, A1094, A1106, A1112, A1116, A1118, A1140, A1143, A1192, A1198, A1202, A1232.

constituents (fats, proteins, carbohydrates or enzyme inhibitors such as anti-nutrients) or minor constituents (minerals, vitamins). Key toxicants are those toxicologically significant compounds known to be inherently present in an organism, such as compounds whose toxic potency and level may be significant to health.

5.1 Key components

The key components to be analysed for the comparison of transgenic and conventional soybean are outlined in the OECD Consensus Document on Compositional Considerations for New Varieties of soybean (OECD 2012), and include: proximates, fibre, amino acids, fatty acids, minerals, vitamins (E and K), and anti-nutrients.

5.2 Study design

IND-00410-5 (T6 generation), the conventional control (Williams 82) and a set of commercial soybean reference varieties[19] were grown and harvested from six field trials in Argentina and five field trial sites in the United States during the 2012 and 2013 growing seasons. The sites were representative of environmentally diverse and major soybean production areas. The field sites were established in a randomised complete block design with four replicates per site. Plants were grown under agronomic field conditions typical for each of growing region.

IND-00410-5 plants grown during these field trials were not treated with glufosinate. FSANZ's previous safety assessments for glufosinate-tolerant soybean crops (A1081, A1073 and A1046) show that spraying had no effect on composition. Any significant differences that were found had no consistent trend, were within the range of natural variability and had no biological significance. This is in line with the scientific literature (Oberdoerfer et al. 2005; Harrison et al. 2013; Lepping et al. 2013; Bartholomaeus et al. 2015; Fast et al. 2016).

Compositional analyses were performed on grain samples was based on the OECD revised consensus document on compositional considerations for new varieties of soybean (OECD 2012). 44 different analytes were measured in grains (see Figure 6 for a complete list). Statistically analyses were performed using the SAS software[20]. Analytes were expressed as either percent dry weight (% dw), ppm dw or as mg/kg dw, as shown in Figure 7. For each analyte, 'descriptive statistics' (mean, standard error (SE), and range) were generated.

In assessing the significance of any difference between IND-00410-5 and the control, a *p*-value of 0.05 was used for all sites (combined-site analysis). Levels for each analyte in IND-ØØ41Ø-5 soybean were statistically compared to those measured in the control. Mean values from the commercial reference varieties were calculated to establish the reference range i.e. the natural variability of analytes in a plant grown under the same agronomical and environmental conditions (Chiozza et al. 2020).

The magnitude of difference in mean values between ND-00410-5 and the control were determined, and this difference was compared to the variation observed within the reference varieties. In addition, the natural variation of analytes from publicly available data was also considered. The applicant provided the AFSI[21] range cited within the OECD consensus document (OECD 2012). FSANZ has further compiled the ranges from the updated AFSI database (AFSI 2023) and the OECD consensus document (OECD 2012). This publicly available range takes into account variability present in non-GM soybean varieties due to a

[19] Argentinien varieties: Biosoja 4.6, DM 4670, DM 4210, SRM 3970, FN 3.85, A 3731 RG, NS 4009, and SPS 3900. United States varieties: Dow 32R280, Pioneer (93Y82, 93Y84, 93M94), Dupont 93Y82, Asgrow (AG) 3832, Asgrow AG393, Stine 39LD02, DynaGro 36RY38, Hoffman H38-12CR2, and NK S39-U2.

[20] SAS Institute, Cary, NC

[21] Formerly known as International Life Sciences Institute (ILSI)

wide range of agronomic and environmental conditions, as well as different genetic backgrounds. These data ranges assist with determining whether any statistically significant differences were likely to be biologically meaningful.

Key analyte levels were also analysed in forage but the results are not included in this report. It is noted however that, in the combined site analysis for forage, no statistically significant differences were observed in IND-00410-5 compared to the control.

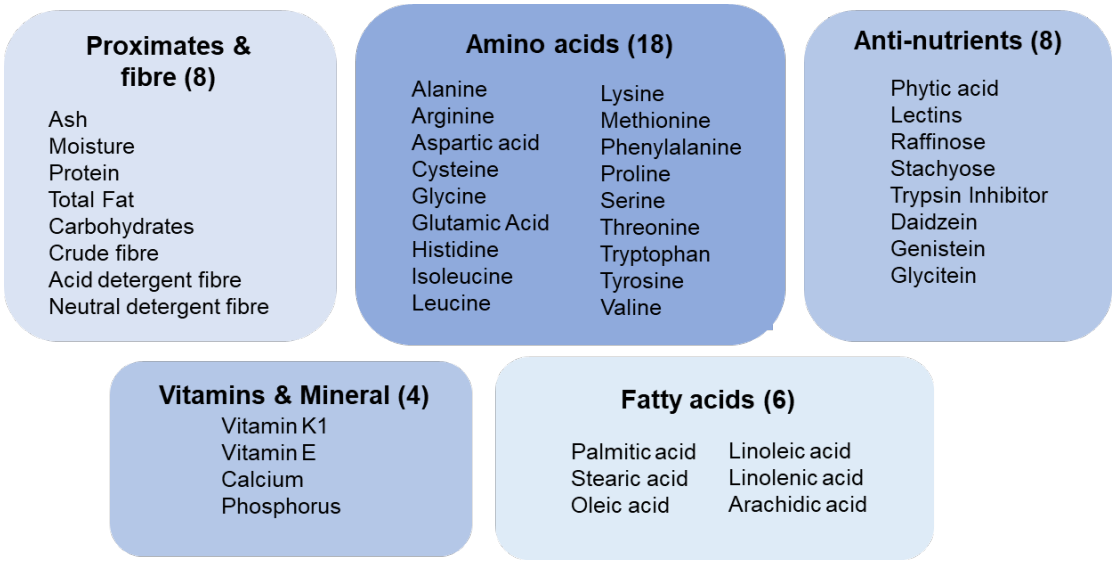


Figure 6: Analytes measured in grain samples.

5.3 Analyses of key components in grains

Of the 44 analytes measured in grain, there were 7 for which there was a statistically significant difference in mean values between soybean line IND-00410-5 and the control: vitamin K1, phytic acid, stachyose, daidzein, genistein and glycitein. A summary of these 7 analytes is provided in Figure 7. For the complete data set, including values for the 37 analytes for which no statistically significant differences were found, refer to the Application dossier (pages 55 – 60).

For all 7 analytes where a statistical significant difference was found, the deviation of the IND-00410-5 mean from the control mean ranged between 6.78% to 23.70% (Figure 7a). However, as can be observed in Figure 7b-h, the IND-00410-5 mean for each of these analytes was within the control range (orange bars), the reference mean range (light grey dots joined by lines) and/or the publicly available range (dark grey bars). The maximum range found in IND-00410-5 (blue bars) for two of the analytes, phytic acid and stachyose (Figure 7 d, e), was outside the control and publicly available ranges. However, given that the reported reference range is the mean range of the reference varieties, we expect the maximum/minimum values to be greater than the means i.e. we expect the range of natural variability to be broader. As such, the differences reported here are consistent with normal biological variability found in soybean.

Overall, the compositional data are consistent with the conclusion that there are no biologically significant differences in the levels of key constituents in IND-00410-5 when compared with conventional non-GM soybean varieties already available in agricultural markets. Grain from IND-00410-5 can therefore be regarded as equivalent in composition to grain from conventional non-GM soybean.

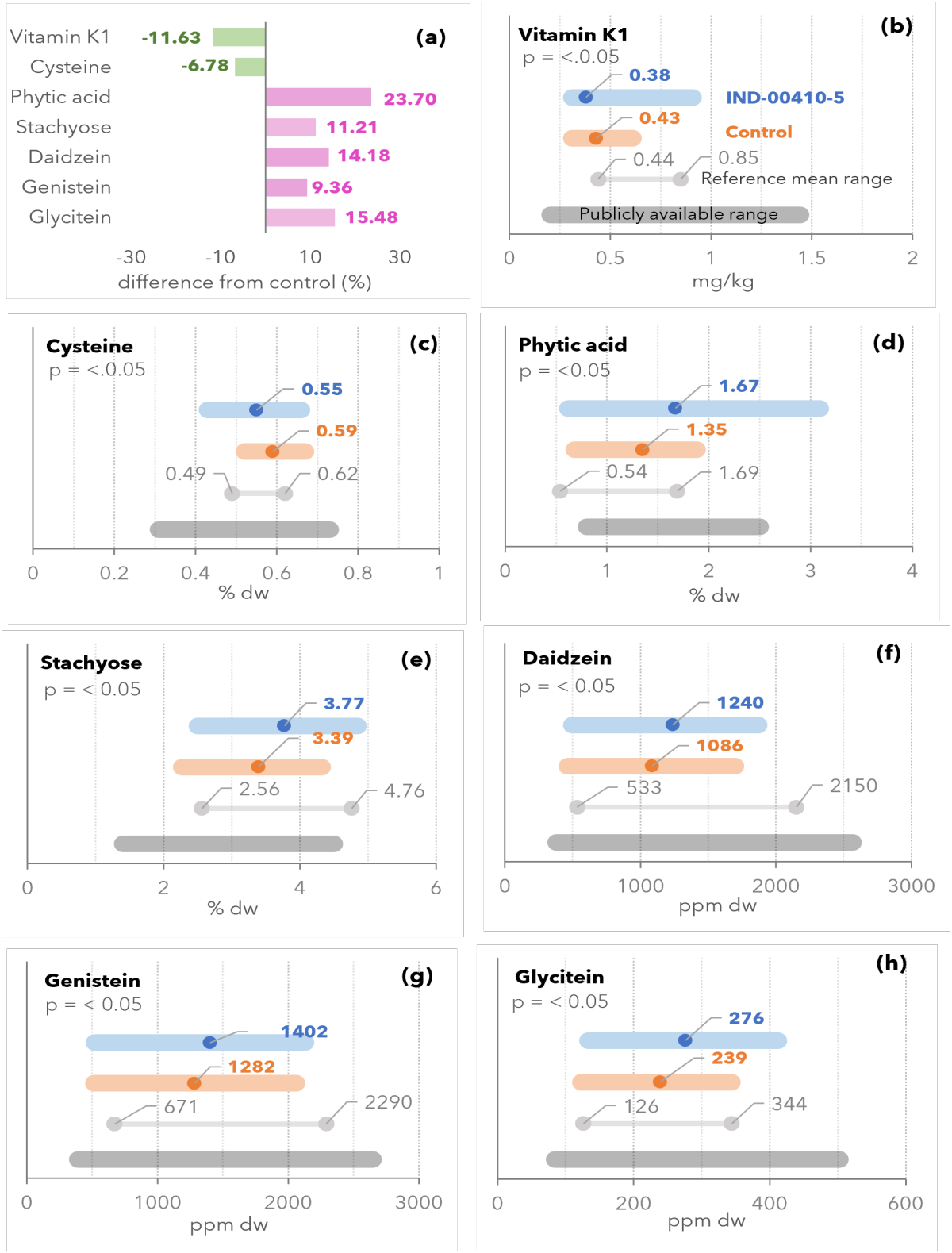


Figure 7: Visual summary of statistically significantly analyte differences in IND-00410-5 compared to the conventional control. ***(a)*** *Percentage deviation of the mean IND-00410-5 value from the mean control value for each of the 7 analytes for which significant differences were found.* ***(b) - (h)*** *Measured means (dots) and ranges for IND-00410-5 (blue bars) and the control (orange bars) for the 7 analytes as labelled. The grey dots joined by lines represent the mean range in the commercial non-GM soybean reference varieties grown in the same field trials. The dark grey bars represent the publicly-available ranges for each analyte. Note that the x-axes vary in scale and unit for each analyte.*

6 Nutritional impact

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

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

IND-00410-5 is the result of genetic modifications to confer tolerance to drought and the herbicide glufosinate, with no intention to significantly alter nutritional parameters in the food. The compositional analyses have demonstrated that the genetic modifications have not altered the nutrient composition of IND-00410-5 compared with that of conventional non-GM soybean varieties. The introduction of food derived from IND-00410-5 into the food supply is therefore expected to have negligible nutritional impact.

7 References

Adeleke BS, Babalola OO (2020) Oilseed crop sunflower (Helianthus annuus) as a source of food: Nutritional and health benefits. Food Science & Nutrition. 8(9):4666-4684.

AFSI (2023) Formerly the International Life Sciences Institute, ILSI). Crop composition database, version 9.1. Agriculture & Food Systems Institute. https://www.cropcomposition.org/ Accessed 17 March 2023

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol. 215(3):403-410.

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

Australian Oilseeds Federation (2023) Soy Australia. http://www.australianoilseeds.com/soy_australia/Soybean_Production. Acessed 12 February 2023.

Bartholomaeus A, Batista JC, Burachik M, Parrott W (2015) Recommendations from the workshop on Comparative Approaches to Safety Assessment of GM Plant Materials: A road toward harmonized criteria? GM Crops Food. 6(2):69-79.

Bartholomaeus A, Parrott W, Bondy G, Walker K, ILSI (2013) The use of whole food animal studies in the safety assessment of genetically modified crops: limitations and recommendations. Crit Rev Toxicol. 43 Suppl 2(Suppl 2):1-24.
Block MD, Botterman J, Vandewiele M, et al. (1987) Engineering herbicide resistance in plants by expression of a detoxifying enzyme. EMBO J. 6(9):2513-2518.

Carrington JC, Freed DD (1990) Cap-independent enhancement of translation by a plant potyvirus 5' nontranslated region. J Virol. 64(4):1590-1597.

CERA (2011) A review of the environmental safety of the PAT protein. Environ Biosafety Res. 10(4):73-101.

Chan RL, Gonzalez DH (1994) A cDNA encoding an HD-zip protein from sunflower. Plant Physiol. 106(4):1687-1688.

Chiozza MV, Burachik M, Miranda PV (2020) Compositional analysis of soybean event IND-OO41O-5. GM Crops Food. 11(3):154-163.

Delaney B, Astwood JD, Cunny H, et al. (2008) Evaluation of protein safety in the context of agricultural biotechnology. Food Chem Toxicol. 46 Suppl 2:S71-97.

Depicker A, Stachel S, Dhaese P, Zambryski P, Goodman HM (1982) Nopaline synthase: transcript mapping and DNA sequence. J Mol Appl Genet. 1(6):561-573.

Dezar CA, Fedrigo GV, Chan RL (2005a) The promoter of the sunflower HD-Zip protein gene Hahb4 directs tissue-specific expression and is inducible by water stress, high salt concentrations and ABA. Plant Science 169(2)

Dezar CA, Gago GM, Gonzalez DH, Chan RL (2005b) Hahb-4, a sunflower homeobox-leucine zipper gene, is a developmental regulator and confers drought tolerance to Arabidopsis thaliana plants. Transgenic Res. 14(4):429-440.

FAOSTAT (2023) FAOSTAT Statistics Division Food and Agriculture Organization of the United Nations. http://www.fao.org/faostat/en/#home. Accessed 12 February 2023

Fast BJ, Galan MP, Schafer AC (2016) Event DAS-444O6-6 soybean grown in Brazil is compositionally equivalent to non-transgenic soybean. GM Crops Food. 7(2):79-83.

Frame BR, Shou H, Chikwamba RK, et al. (2002) Agrobacterium tumefaciens-mediated transformation of maize embryos using a standard binary vector system. Plant Physiol. 129(1):13-22.

Gago GM, Almoguera C, Jordano J, Gonzalez DH, Chan RL (2002) Hahb-4, a homeobox-leucine zipper gene potentially involved in abscisic acid-dependent responses to water stress in sunflower. Plant, Cell & Environment. 25

Gallie DR, Tanguay RL, Leathers V (1995) The tobacco etch viral 5' leader and poly(A) tail are functionally synergistic regulators of translation. Gene. 165(2):233-238.

Gelvin SB (2021) Plant DNA Repair and Agrobacterium T-DNA Integration. Int J Mol Sci. 22(16)

Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A. 100(12):6940-6945.

Gheysen G, Villarroel R, Van Montagu M (1991) Illegitimate recombination in plants: a model for T-DNA integration. Genes Dev. 5(2):287-297.

Gonzalez FG, Capella M, Ribichich KF, et al. (2019) Field-grown transgenic wheat expressing the sunflower gene HaHB4 significantly outyields the wild type. J Exp Bot. 70(5):1669-1681.

Hajdukiewicz P, Svab Z, Maliga P (1994) The small, versatile pPZP family of Agrobacterium binary vectors for plant transformation. Plant Mol Biol. 25(6):989-994.

Hammond BG, Jez JM (2011) Impact of food processing on the safety assessment for proteins introduced into biotechnology-derived soybean and corn crops. Food Chem Toxicol. 49(4):711-721.

Haq TA, Mason HS, Clements JD, Arntzen CJ (1995) Oral immunization with a recombinant bacterial antigen produced in transgenic plants. Science. 268(5211):714-716.

Hara O, Murakami T, Imai S, et al. (1991) The bialaphos biosynthetic genes of Streptomyces viridochromogenes: cloning, heterospecific expression, and comparison with the genes of Streptomyces hygroscopicus. J Gen Microbiol. 137(2):351-359.

Harrison JM, Howard D, Malven M, et al. (2013) Principal variance component analysis of crop composition data: a case study on herbicide-tolerant cotton. J Agric Food Chem. 61(26):6412-6422.

Heeb S, Itoh Y, Nishijyo T, et al. (2000) Small, stable shuttle vectors based on the minimal pVS1 replicon for use in gram-negative, plant-associated bacteria. Mol Plant Microbe Interact. 13(2):232-237.

Herouet C, Esdaile DJ, Mallyon BA, et al. (2005) Safety evaluation of the phosphinothricin acetyltransferase proteins encoded by the pat and bar sequences that confer tolerance to glufosinate-ammonium herbicide in transgenic plants. Regul Toxicol Pharmacol. 41(2):134-

149.

Hymowitz T (1970) On the domestication of the soybean. Economic Botany. 24:408-421.

ILSI (2016) A Review of the Food and Feed Safety of the PAT Protein. Foundation, IR. Washington, D.C. USA.

Lentz DL, Pohl MD, Alvarado JL, Tarighat S, Bye R (2008) Sunflower (Helianthus annuus L.) as a pre-Columbian domesticate in Mexico. Proc Natl Acad Sci U S A. 105(17):6232-6237.

Lepping MD, Herman RA, Potts BL (2013) Compositional equivalence of DAS-444O6-6 (AAD-12 + 2mEPSPS + PAT) herbicide-tolerant soybean and nontransgenic soybean. J Agric Food Chem. 61(46):11180-11190.

Manavella PA, Dezar CA, Ariel FD, Chan RL (2008a) Two ABREs, two redundant root-specific and one W-box cis-acting elements are functional in the sunflower HAHB4 promoter. Plant Physiol Biochem. 46(10):860-867.

Manavella PA, Dezar CA, Bonaventure G, Baldwin IT, Chan RL (2008b) HAHB4, a sunflower HD-Zip protein, integrates signals from the jasmonic acid and ethylene pathways during wounding and biotic stress responses. Plant J. 56(3):376-388.

Mayerhofer R, Koncz-Kalman Z, Nawrath C, et al. (1991) T-DNA integration: a mode of illegitimate recombination in plants. EMBO J. 10(3):697-704.

Mir R, Aranda LZ, Biaocchi T, et al. (2017) A DII Domain-Based Auxin Reporter Uncovers Low Auxin Signaling during Telophase and Early G1. Plant Physiol. 173(1):863-871.

Murakami T, Anzai H, Imai S, et al. (1986) The bialaphos biosynthetic genes of Streptomyces hygroscopicus: Molecular cloning and characterization of the gene cluster. Molecular and General Genetics 205:42-50.

Murphy E (1985) Nucleotide sequence of a spectinomycin adenyltransferase AAD(9) determinant from Staphylococcus aureus and its relationship to AAD(3") (9). Mol Gen Genet. 200(1):33-39.

Oberdoerfer RB, Shillito RD, de Beuckeleer M, Mitten DH (2005) Rice (Oryza sativa L.) containing the bar gene is compositionally equivalent to the nontransgenic counterpart. J Agric Food Chem. 53(5):1457-1465.

Odell JT, Nagy F, Chua NH (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature. 313(6005):810-812.

OECD (2003) Considerations for the Safety Assessment of Animal Feedstuffs Derived from Genetically Modified Plants: Series on the Safety of Novel Foods and Feeds. Report No: 9. Development, OFEC-OA. Paris.

OECD (2006) Soybean (GLYCINE MAX (L.) MARR.). Organisation for Economic Co-operation and Development. Paris.

OECD (2012) Revised Consensus Document on Compositional Considerations for New Varieties of Soybean [Glycine Max (L.) Merr.]: Key Food and Feed Nutrients, Antinutrients, Toxicants and Allergens: Series on the Safety of Novel Foods and Feeds. Report No: 25. Organisation for Economic Cooperation and Development. Paris.

Rapp WD, Lilley GG, Nielsen NC (1990) Characterization of soybean vegetative storage

proteins and genes. Theor Appl Genet. 79(6):785-792.

Ribichich KF, Chiozza M, Avalos-Britez S, et al. (2020) Successful field performance in warm and dry environments of soybean expressing the sunflower transcription factor HB4. J Exp Bot. 71(10):3142-3156.

Skinner C, Patfield S, Stanker L, He X (2013) Development of monoclonal antibodies and immunoassays for sensitive and specific detection of Shiga toxin Stx2f. PLoS One. 8(9):e76563.

Thompson CJ, Movva NR, Tizard R, et al. (1987) Characterization of the herbicide-resistance gene bar from Streptomyces hygroscopicus. EMBO J. 6(9):2519-2523.

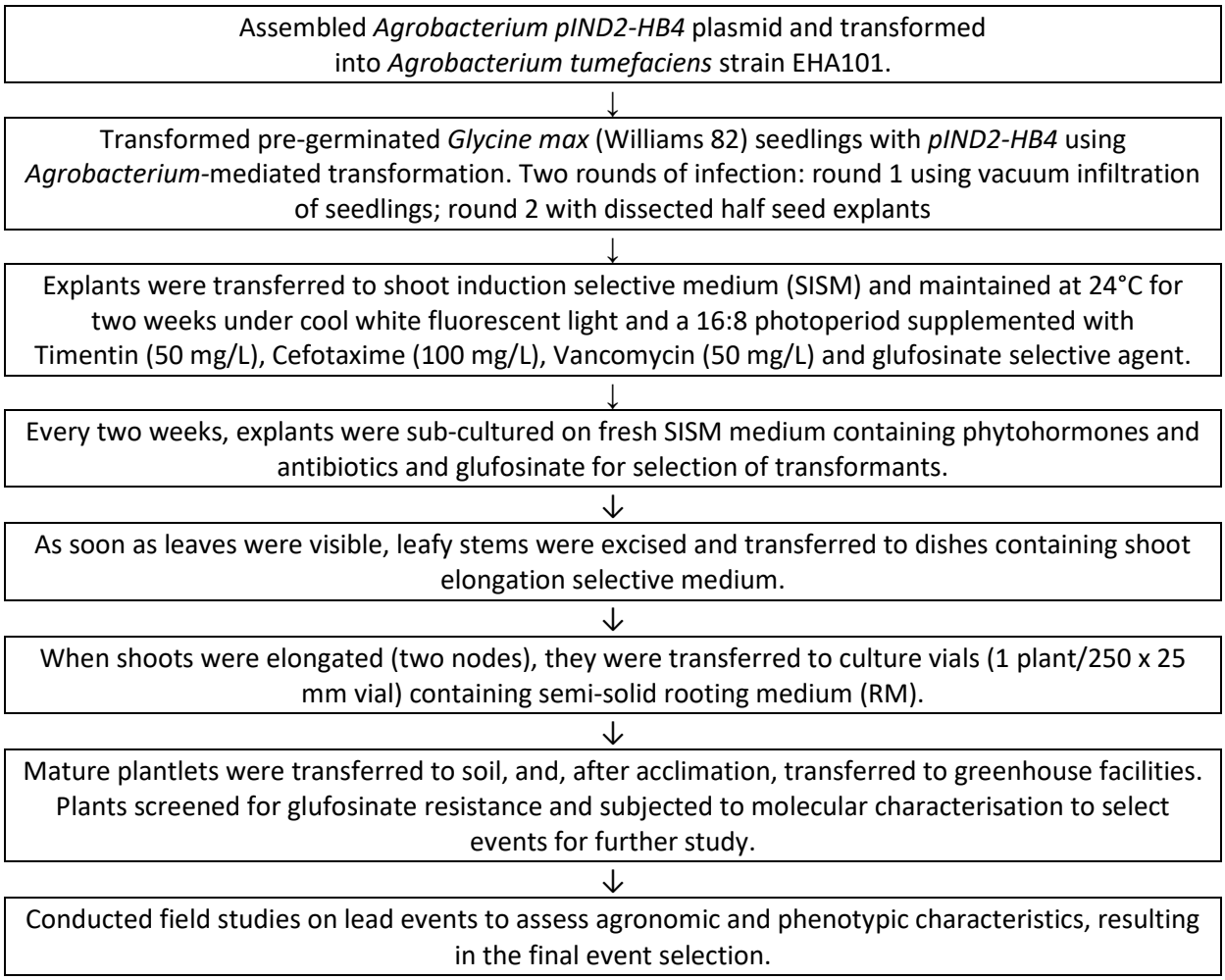
van Haaren MJ, Sedee NJ, de Boer HA, Schilperoort RA, Hooykaas PJ (1989) Mutational analysis of the conserved domains of a T-region border repeat of Agrobacterium tumefaciens. Plant Mol Biol. 13(5):523-531.

Wehrmann A, Van Vliet A, Opsomer C, Botterman J, Schulz A (1996) The similarities of bar and pat gene products make them equally applicable for plant engineers. Nat Biotechnol. 14(10):1274-1278.

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene. 33(1):103-119.

Yue H, Shu D, Wang M, et al. (2018) Genome-Wide Identification and Expression Analysis of the HD-Zip Gene Family in Wheat (Triticum aestivum L.). Genes (Basel). 9(2)

Appendix 1


Assembled *Agrobacterium pIND2-HB4* plasmid and transformed into *Agrobacterium tumefaciens* strain EHA101.
Transformed pre-germinated *Glycine max* (Williams 82) seedlings with *pIND2-HB4* using *Agrobacterium*-mediated transformation. Two rounds of infection: round 1 using vacuum infiltration of seedlings; round 2 with dissected half seed explants
Explants were transferred to shoot induction selective medium (SISM) and maintained at 24°C for two weeks under cool white fluorescent light and a 16:8 photoperiod supplemented with Timentin (50 mg/L), Cefotaxime (100 mg/L), Vancomycin (50 mg/L) and glufosinate selective agent.
Every two weeks, explants were sub-cultured on fresh SISM medium containing phytohormones and antibiotics and glufosinate for selection of transformants.
As soon as leaves were visible, leafy stems were excised and transferred to dishes containing shoot elongation selective medium.
When shoots were elongated (two nodes), they were transferred to culture vials (1 plant/250 x 25 mm vial) containing semi-solid rooting medium (RM).
Mature plantlets were transferred to soil, and, after acclimation, transferred to greenhouse facilities. Plants screened for glufosinate resistance and subjected to molecular characterisation to select events for further study.
Conducted field studies on lead events to assess agronomic and phenotypic characteristics, resulting in the final event selection.

Appendix 2

Genetic elements present in the pIND2 HB4 plasmid

Genetic elements	Relative position	Size (bp)	Source	Description & Function
IS	1-1186	1186	Binary vector pPZP202	Sequence used for cloning (Hajdukiewicz et al. 1994)
Stabilising protein from pVS1	1187-1816	630		Plasmid stability in culture (Heeb et al. 2000)
IS	1187-2244	1058		Sequence used for cloning (Hajdukiewicz et al. 1994)
Replication protein from pVS1	2245-3318	1074		Replication protein (Heeb et al. 2000)
IS	3319-3383	65		Sequence used for cloning (Hajdukiewicz et al. 1994)
Origin of replication from pVS1	3384-3578	195		Origin of replication from pVS1 (Heeb et al. 2000)
IS	3579-3921	343		Sequence used for cloning (Hajdukiewicz et al. 1994)
Basis of mobility region from plasmid pBR322	3922-4062	141		
IS	4063-4247	185		Sequence used for cloning (Hajdukiewicz et al. 1994)
Plasmid origin of replication	4248-4836	589		High copy number ColE1/pMB1/pBR322/pUC origin of replication. Plasmid origin of replication (Yanisch-Perron et al. 1985)
IS	4837-5081	245		Sequence used for cloning (Hajdukiewicz et al. 1994)
SmR Antibiotic resistance gene	5082-5873	792		Confers resistance to spectinomycin and streptomycin-aminoglycoside adenylyl transferase (Murphy 1985)
IS	5874-6396	522		Sequence used for cloning (Hajdukiewicz et al. 1994)
Left Border (LB) sequence	6397-6421	25		Secondary cleavage site releases ssDNA insert from pIND2-HB4 (van Haaren et al. 1989)
IS	6422-6660	239		Sequence used for cloning (Hajdukiewicz et al. 1994)
bar expression cassette				
Tvsp; poly(A)signal of a	6661-7206	546	*Glycine max*	Poly (A) signal for the termination of *bar* transcription (Rapp et al. 1990)
bar coding sequence	7207-7770	564	*Streptomyces hygroscopicus*	Generates mRNA that leads to phosphinothricin acetyltransferase (PAT) providing herbicide tolerance (Frame et al. 2002; Mir et al. 2017)
IS	7771-7781	11		
Tobacco Etch Virus (TEV) 5´ leader sequence	7782-7911	130	*Tobacco Etch Virus*	Directs efficient translation of the bar gene (Carrington and Freed 1990; Gallie et al. 1995)
IS	7912-7983	72		
2x35S Promoter	7984-8670	687	*Cauliflower Mosaic Virus*	*De novo* expression of the *bar* gene (Odell et al. 1985; Haq et al. 1995)
IS	8671-8682	12		
HaHB4 expression cassette				
LPF Promoter	8683-9891	1209	*Helianthus annus*	*De novo* expression of the *HaHB4* gene (Dezar et al. 2005a; Manavella et al. 2008a)

Genetic elements	Relative position	Size (bp)	Source	Description & Function
IS	9892-9902	11		
HaHB4 coding sequence	9903-10433	531	*Helianthus annus*	Generates mRNA that leads to HaHB4 providing environmental stress tolerance (Chan and Gonzalez 1994; Gago et al. 2002; Dezar et al. 2005b; Manavella et al. 2008b)
IS	10434-10542	109		
NOS-ter; poly(A)signal of nopaline synthase gene	10453-10707	253	*Agrobacterium tumefaciens*	Poly (A) signal for the termination of *HaHB4* transcription (Depicker et al. 1982)
IS	10704-10993	290		
Right Border (RB) sequence	10994-11018	25	Binary vector pPZP202	Primary cleavage site releases ssDNA insert from pIND2-HB4 (van Haaren et al. 1989)
IS	11019-11133	115	Binary vector pPZP202	Vector sequence used for DNA cloning (Hajdukiewicz et al. 1994)