Application to FSANZ for the Inclusion of soybean event IND-ØØ41Ø-5 with increased tolerance to environmental stresses in Standard 1.5.2 Food Produced Using Gene Technology

Submitting Company:
Bioceres Crop Solutions

Ocampo 210 bis, Rosario, S2000 Province of Santa Fe, Argentina

Part 1 General Requirements (3.1.1)

A. Executive Summary

Bioceres Crop Solutions has developed a genetically modified (GM) soybean line using the sunflower *HaHB4* gene to confer increased tolerance to environmental stresses avoiding reduction of crop yield. The HAHB4 protein belongs to the HD-Zip family of transcription factors, characterised by the presence of two functional domains: the homeodomain (HD), responsible for DNA binding, and a leucine zipper motif (LZ) involved in protein-protein interaction and dimerisation. The soybean event described in this application has the unique OECD code: IND-ØØ41Ø-5 and is referred to as 'HB4 soybean' in this submission.

HB4 soybean was developed using *Agrobacterium*-mediated transformation of the soybean (*Glycine max*) variety Williams 82 (Bernard and Cremeens, 1988) with the binary plasmid *pIND2-HB4*. The selected event (IND-ØØ41Ø-5) has been field evaluated over many growing seasons in Argentina and the United States with data supporting the conclusion that the *HaHB4* gene confers increased tolerance to environmental stresses that reduce crop yields, and that soybean event IND-ØØ41Ø-5 also exhibits tolerance to glufosinate-based herbicides.

Molecular characterisation of the event was performed to determine the number of copies, arrangement, and stability of the inserted DNA. Molecular analysis shows a single T-DNA locus comprised of a single copy of the selectable bar marker-gene, a single copy of the *HaHB4* gene, and their respective regulatory sequences. No unintended components from the binary vector DNA are present in IND-ØØ41Ø-5.

Field trials were undertaken with soybean event IND-ØØ41Ø-5 to compare agronomic performance and biosafety with the conventional variety and other cultivated varieties used as controls. Results from these trials confirmed no changes were observed in soybean event IND-ØØ41Ø-5 that could have an impact on the environment. Stability of the genetic modification was assessed and confirm that the HB4 trait is stably inherited and conforms to Mendelian segregation principles.

Compositional analysis was performed following the OECD Consensus Document recommendations for soybean (OECD, 2012). Comparison of nutritional and anti-nutritional compounds showed no biologically relevant differences exist that could result in increased harm to humans or other non-target organisms. Analysis of the HAHB4 and PAT proteins as well as putative polypeptides produced from the inserted DNA indicated there are no sequences with significant homology to known allergens or toxins in HB4 soybean.

Analysis of the HB4 soybean has not revealed any biologically relevant differences compared to the conventional variety, except for the intended tolerance to abiotic stress and herbicide tolerance. Collectively, results of the molecular characterisation, agronomic assessment, and composition analysis support this application for amendment to the *Australia New Zealand Food Standards Code* to allow inclusion of HB4 soybean in **Standard 1.5.2**-*Food Produced Using Gene Technology*.

B. Applicant Details

(a)	Applicant's name/s	Dr Patricia Miranda
(b)	Company/organisation name	Bioceres Crop Solutions
(c)	Address (street and postal)	c/o Ocampo 210 bis, 2000 Rosario, Santa Fe, Argentina
(d)	Telephone number	+54 341 486 1100 extension 1109
(e)	Email address	patricia.miranda@indear.com
(f)	Nature of the applicant's business	Bioceres Crop Solutions is a fully - integrated provider of crop productivity solutions, including high - impact, patented technologies for seeds and microbial ag-inputs, as well as next - generation crop nutrition and protection solutions, each of which offers substantial economic and environmental benefits and anchored by the HB4® technology, which is behind the world's only drought - tolerant soybeans and wheat.
(g)	Details of other individuals, companies or organisations associated with the application	PTM Solutions Australia Pty Ltd 11 Moras Court Gisborne, VICTORIA 3437 Carl@ptmsolutions.com.au Or Rock@ptmsolutions.com.au

C. Purpose of the Application

This application seeks to amend the *Australia New Zealand Food Standards Code* to allow for the inclusion of soybean event IND-ØØ41Ø-5 in **Standard 1.5.2**-*Food Produced Using Gene Technology*.

Bioceres Crop Solutions has developed and evaluated soybean events that have increased yield opportunity under conditions of environmental stress. The soybean event described in this application has the unique OECD code: IND-ØØ41Ø-5 and is referred to as 'HB4 soybean' in this submission.

Currently, Bioceres Crop Solutions does not intend to import HB4 soybean into Australia or New Zealand for food consumption. The primary aim of this application is to obtain a food safety approval to protect international trade. Previously, Trigall Genetics, a partnership between Bioceres Crop Solutions and Florimond Desprez, received approval for the inclusion of an HB4 wheat (IND-ØØ412-7) in **Standard 1.5.2**-*Food Produced Using Gene Technology* (A1232). Bioceres Crop Solutions are exploring opportunities to introgress the HB4 trait into Australian wheat germplasm and would seek to obtain import approval from the Department of Agriculture, Forestry and Fisheries and relevant cultivation approvals through other regulatory agencies such as the Office of the Gene Technology Regulator (OGTR) and the Australian Pesticides and Veterinary Medicines Authority (APVMA).

This submission is consistent with Bioceres Crop Solutions commitment to global stewardship, adhering to industry best practice by obtaining regulatory approvals in production and import markets.

D. Justification for the Application

Bioceres Crop Solutions has developed a new soybean event, IND-ØØ41Ø-5. The new soybean event was created using the sunflower *HaHB4* gene that confers increased yield opportunity under conditions of environmental (abiotic) stress. The event also contains the herbicide tolerance *bar* gene from *Streptomyces hygroscopicus*, expressing the glufosinate-inactivating enzyme phosphinothricin N-acetyl transferase (PAT). These genes have recently been assessed by FSANZ (A1232) in genetically modified wheat.

Soybean (*Glycine max* L) is a species of legume native to East Asia and an important global crop providing a low-cost source of vegetable oil and protein. Soybean oil is primarily consumed as table oil or further processed into a wide variety of products such as baked goods, dressings, and sauces. Soybean is an important industrial crop utilised in producing edible oils, wax, paints, dyes, and fibre (Rezaei et al., 2002; Raghuvanshi and Bisht, 2010). More recently, soybean is being used as a meat substitute extensively used by vegan and vegetarian consumers (Messina et al., 2022; Bryant 2022). The seed is used mainly to produce meal that accounts for approximately 80% of the seed and is predominantly for use in animal feed.

The global production area for soybean (130 million hectares in 2021/2022; FAOSTAT 2022) continues to increase to meet the demand for meal to support livestock production. Soybean production (353 million metric tonnes in 2021/2022; FAOSTAT 2022) is dominated by Brazil (35.7%), the USA (34.2%), Argentina (12.47%), China (4.65%), and India (3.37%). Australia remains a relatively small player in the market in both production (50,000 tonnes from 25,000ha) and consumption, yet both are increasing (Australian Oilseeds Federation 2022).

Drought is the most significant environmental stress which limits crop productivity around the world. Low water availability at critical stages of crop development leads to great yield losses (Duque et al., 2013). ABARES research has shown that changes in climate conditions over the last 20 years have had an adverse effect on the productivity of Australian cropping farms (Hughes et al. 2017). Similarly, New Zealand has experienced several major droughts during the last decades, leading to significant agricultural production losses (Pourzand and Noy 2019).

It is predicted that the shift in climate toward higher temperatures and altered rainfall patterns (predominantly drier) are expected to lead to more frequent and intense drought conditions. As such, tolerance to drought stress is a highly desired goal of soybean genetic improvement and significant efforts are being made to understand the effects of drought and develop varieties with drought tolerance through various breeding strategies (e.g., Arya et al 2021; Carter et al., 2016; Chen et al., 2021; Dayoub et al., 2021; Du et al., 2020)

Members of the HD-Zip family of transcription factors (TFs), unique to plants, have been shown to be involved in regulating the response of plants to environmental stress (Schena and Davis, 1992). Expression of genes of the HD-Zip subfamily I is regulated by external factors such as drought, extreme temperatures, osmotic stresses, and light conditions (Ariel et al., 2007; Chan, 2009). The *HaHB4* (*Helianthus annuus* homeobox 4) gene is a member of the HD-Zip sub-family I, coding for the sunflower transcription factor HAHB4 (González et al., 2020). The introduction of *HaHB4* gene in soybean event IND-ØØ41Ø-5 led to the drought stress tolerance phenotype. Phenotypic and field performance selection of several *HaHB4*-containing lines allowed the development of a transgenic soybean (termed IND-ØØ41Ø-5), which was shown to provide an increased yield opportunity under conditions of environmental stress.

E. Information to Support the Application

This application consists of 2 parts containing information in accordance with the following checklists:

- Part 1: General requirements (3.1.1)

- Part 2: Foods produced using gene technology (3.5.1) main document, Part 2: Specific Data Requirements for Safety Assessment.

F. Assessment Procedure

Bioceres Crop Solutions is anticipating that this application will be considered under the **General Procedure** for Administrative Assessment process by Food Standards Australia New Zealand.

G. Confidential Commercial Information (CCI)

There is no Confidential Commercial Information (CCI) included in this submission document.

Release of Information

Bioceres Crop Solutions is submitting the information in this application for review by Food Standards Australia New Zealand (FSANZ) for amendment to the Food **Standard 1.5.2** *Food Produced Using Gene Technology*. Bioceres Crop Solutions holds proprietary rights to the extent allowable by law to all such information and by submitting this information, Bioceres Crop Solutions does not authorise its release to any third party except to the extent it is duly requested under the *Freedom of Information Act 1982* (*FOI Act*) or in compliance with the responsibility of FSANZ to publish documents required under Sections 8, 8(A), 8(C) and 8(D) of the *FOI Act*; and this information is responsive to the specific aforementioned request. Accordingly, except as specifically stated above, Bioceres Crop Solutions does not authorise the release, publication, or other distribution of this information (including website posting or otherwise), nor does Bioceres Crop Solutions authorise any third party to use, obtain, or rely upon this information, directly or indirectly, as part of any other application or for any other use, without Bioceres Crop Solutions prior notice and written consent. Submission of this information does not in any way waive Bioceres Crop Solutions rights (including rights to exclusivity and compensation) to such information.

H. Other Confidential Information

No additional confidential material is included in this submission document.

I. Exclusive Capturable Commercial Benefit

Bioceres Crop Solutions acknowledges that the proposed amendment to the Standard will likely result in an exclusive capturable commercial benefit being accrued to Bioceres Crop Solutions as defined in Section 8 of the *FSANZ Act*.

J. International and Other National Standards

A list of current applications and approval status is provided in Table 1. Responsible environmental stewardship and deployment of biotechnology-derived products are important to Bioceres Crop Solutions. The joint venture partner Bioceres uses INDEAR as its Research and Development Company. INDEAR is a member of Excellence Through Stewardship® (ETS), an industry-coordinated initiative that promotes the global adoption of stewardship programs and quality management systems for the full life cycle of biotechnology-derived plant products. The ETS "Guide for Product Launch Stewardship of Biotechnology-Derived Products" (ETS, 2013) also references and is consistent with the product launch policies of the Biotechnology Industry Organisation and Crop Life International.

Table 1: Current Applications and Approval Status for IND-ØØ41Ø-5

Country/Region	Competent National Authority	Type of Authorisation	Approval Status
Argentina	Ministerio de Ganadería Agricultura y Pesca (MAGyP)	Food, Feed and Cultivation/Production	Deregulated (2015)
Uruguay	Ministerio de Ganadería, Agricultura y Pesca (GNBio)	Food, Feed and Cultivation/Production	Submitted (2015)
USA	Food and Drug Administration (FDA)	Food and feed	Deregulated (2019)
	United States Department of Agriculture (USDA)	Determination of non-regulated status	
China	Ministry of Agriculture and Rural Affairs (MARA)	Food and feed	Approved (2022)
Brazil	Comissão Técnica Nacional de Biossegurança (CTNBio)	Food, Feed and Cultivation/Production	Deregulated (2019)
Paraguay	The National Commission of Agricultural and Forestry Biosafety (CONBIO)	Food, Feed and Cultivation/Production	Deregulated (2019)
Bolivia	Ministerio de Medio Ambiente y Agua (MMAyA)	Food, Feed and Cultivation/Production	Submitted (2022)
Canada	Canadian Food Inspection Agency (CFIA) – Plant Biosafety Office (PBO)	Cultivation/Production	Deregulated (2021)
	Health Canada	Food	
	Canadian Food Inspection Agency (CFIA) – Animal Feed Division (AFD)	Feed	
India	Genetic Engineering Appraisal Committee (GEAC)	Food and feed	Submitted (2019)
European Union	European Food Safety Authority (EFSA)	Food and feed	Submitted (2020)
Malaysia	Department of Biosafety (DoB)	Food and feed	Submitted (2020)
Indonesia	Ministry of Agriculture (MoA)	Feed	Submitted (2021)
	National Agency of Drug and Food Control (BPOM)	Food	Submitted (2022)
South Africa	Department of Agriculture, Land Reform and Rural Development (DALRRD)	Food and feed	Approved (2022)

K. Statutory Declaration – Australia

See attached statutory declaration provided separately.

L. Checklists Provided With Application

General Requirements

General requirements (3.1.1)		
Check	**Page No.**	**Mandatory requirements**
☒	2	A Form of application ☒ Application in English ☒ Executive Summary (separated from main application electronically) ☒ Relevant sections of Part 3 clearly identified ☒ Pages sequentially numbered ☒ Electronic copy (searchable) ☒ All references provided
☒	3	B Applicant details
☒	3	C Purpose of the application
☒	4	D Justification for the application ☒ Regulatory impact information ☒ Impact on international trade
☒	4	E Information to support the application ☒ Data requirements
☒	5	F Assessment procedure ☒ General ☐ Major ☐ Minor ☐ High level health claim variation
☒	5	G Confidential commercial information ☒ CCI material separated from other application material ☒ Formal request including reasons ☒ Non-confidential summary provided
☒	5	H Other confidential information ☒ Confidential material separated from other application material ☒ Formal request including reasons
☒	5	I Exclusive Capturable Commercial Benefit ☒ Justification provided
☒	6	J International and other national standards ☒ International standards ☒ Other national standards
☒	7	K Statutory Declaration
☒	8	L Checklist/s provided with application ☒ 3.1.1 Checklist ☒ All page number references from application included ☒ Any other relevant checklists for Chapters 3.2–3.7

Foods Produced Using Gene Technology

Foods produced using gene technology (3.5.1)		
Check	**Page No.**	**Mandatory requirements**
☒	18	A.1 Nature and identity
☒	19	A.2 History of use of host and donor organisms
☒	22	A.3 Nature of genetic modification
☒	38	B.1 Characterisation and safety assessment
☒	50	B.2 New proteins
☒	54	B.3 Other (non-protein) new substances
☒	54	B.4 Novel herbicide metabolites in GM herbicide-tolerant plants
☒	55	B.5 Compositional analyses
☒	60	C Nutritional impact of GM food
☒	60	D Other information

Part 2: Specific Data Requirements for Safety Assessment

The following information is provided to support an application for a new genetically modified food. The details presented are in accordance with Section 3.5.1. of the FSANZ Application Handbook as at, 1 July 2019.

Table of Contents

List of Tables

List of Figures

List of Supplement Reports

1	HB4 Soybean#01010301-Ev2_Identity of Genes and Expressed Proteins
2	HB4 soybean Report#01010289-Ev1_plant transformation
3	HB4 Soybean_Report#01010290-Ev4_Molecular characterization
4	HB4 Soybean_Report #01010271-Ev2-HAHB4 Protein Detection
5	HB4 Soybean_Report#01010297-Ev3-PAT protein detection
6	HB4_Report#01010273-Ev2_Protein Safety
7	HB4_Report#01010296-Ev2_HAHB4 Protein Production
8	HB4 Soybean_Report#01010291-Ev6_BioinformaticAnalysis
9	HB4 Soybean_Report #01010298-Ev2-Endogenous Allergen Levels

Abbreviations, Acronyms and Definitions[1]

Abbreviation	Definition
ADF	Acid detergent fibre
ADP	Adenosine diphosphate
ANOVA	Analysis of Variance
ATP	Adenosine triphosphate
AUG / ATG	Start codon
Backbone DNA	DNA associated with construct backbone
bar	Gene from *Streptomyces hygroscopicus*
bp	Base pair
CBI	Confidential Business Information
CDS	Coding sequence
Chr	Chromosome
CONABIA	Argentina National Advisory Commission on Agricultural Biotechnology
C-t	Carboxy terminal region
DW / dwt	Dry weight
DIG	Digitonin
DNA insert	DNA sequence from *pIND2-HB4* integrated into the soybean genome
dNTP	Deoxy nucleotide triphosphate
dsRNA	Double-stranded RNA
ELISA	Enzyme Linked Immunosorbent Assay
EPA	Environmental Protection Agency
ETS	Excellence Through Stewardship
EU	European Union
FAO	Food and Agriculture Organization of the United Nations
FARRP	Food Allergy Resource Research Program and the University of Nebraska Lincoln
FAS	United States Department of Agriculture, Foreign Agriculture Service
FDA	Food and Drug Administration
FW / fwt	Fresh weight
GE, GM, GMO	Genetically-engineered/modified/modified organism
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
Kb	Kilobase
kDa	Kilodaltons
LC-MS	Liquid Chromatography Mass Spectrometry
LOD	Limit of detection
LOQ	Limit of quantification
LZ	Leucine zipper
MALDI-TOF	Matrix-Assisted Laser Desorption/Ionization - Time of flight
Mt	Metric tonnes
NDF	Neutral Detergent Fibre

[1] NOTE: Abbreviations of units of measure and of physical and chemical quantities are used according to the standard format described in Instructions to Authors in the Journal of Biological Chemistry (http://www.jbc.org/).

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
RT-qPCR	Reverse transcription-qualitative polymerase chain reaction
SDAP	Structural Database of Allergenic Proteins
TF	Transcription factor
US / USA	United States of America
WHO	World Health Organization
WT	Wild type

A. Technical Information on the Food Produced Using Gene Technology

A.1. Nature and Identity of the Genetically Modified Food

A.1(a) A description of the GM organism from which the new GM food is derived. The description must include the nature and purpose of the genetic modification.

The soybean event IND-ØØ41Ø-5 was developed by transforming the soybean variety Williams 82, with the plasmid vector *pIND2-HB4* using *Agrobacterium*-mediated transformation. The event was developed to confer increased tolerance to environmental stresses avoiding reduction of crop yield and exhibits tolerance to glufosinate-based herbicides (Table 2).

The plasmid *pIND2-HB4* contains a single cassette for the expression of the transcription factor *HaHB4* that confers tolerance to environmental stress and an additional cassette for the expression of the *bar* gene coding for the enzyme phosphinothricin N-acetyl transferase (PAT), providing herbicide tolerance.

Table 2: Summary of Genes, Intended Traits, and Benefits in HB4 Soybean

Construct	Gene Target	Mechanism	Intended Trait	Intended Benefit
pIND2-HB4	*HaHB4*: codes for a sunflower transcription factor belonging to the homeodomain-leucine zipper I subfamily[1]	*De novo* expression	Environmental stress tolerance	Yield protection under abiotic stress
	Bar: phosphinothricin N-acetyl transferase[2]	*De novo* expression	Tolerance to glufosinate-based herbicides	Provides post emergence herbicide tolerance for the on-farm management of weeds

[1] The HAHB4 protein has previously been evaluated by FSANZ in wheat (A12132), and by US FDA EFSE: early food safety assessment: NPC 00016 (FDA, 2015) and in soybean (FDA, 2017).
[2] The PAT protein has previously been evaluated by FSANZ in several crops. For example: Soybean (A481, A1046, A1073, A1081); Canola (A372, A1140); Maize (A375, A380, A385, A386, A446, A543, A1106, A1116, A1118, A1192); Cotton, A518, A533, A1028, A1040, A1080); Rice (A589) and wheat (A1232).

Details of the identity of genes and expression products are provided in the Supplement Report **HB4 Soybean#01010301-Ev2_Identity of Genes and Expressed Proteins**

A.1(b) The name, number or other identifier of each of the new lines or strains of GM organism from which the food is derived.

In accordance with OECD '*Guidance for the Designation of a Unique Identifier for Transgenic Plants*', the OECD Unique Identification Code for the soybean event is IND-ØØ41Ø-5.

A.1(c) The name the food will be marketed under (if known).

The soybean containing the environmental stress tolerance technology is marketed as:

- HB4 Soybean

This soybean will be marketed under a variety of labels depending on the background soybean variety and licenced user of the event.

A.2. History of use of the host and donor organisms

A.2(a) For the donor organism(s) from which the genetic elements are derived:

A.2(a)(i) Any known pathogenicity, toxicity or allergenicity of relevance to the food

The donor organisms of all the genetic elements included in the constructions used to obtain IND-ØØ41Ø-5 soybean have a history of use and/or exposure, as described in the next section.

Although some of the donor organisms may be related with pathogenicity (e.g., *A. tumefaciens* is a plant pathogen, some *E. coli* strains are pathogenic), none of the genetic elements used to obtain soybean event IND-ØØ41Ø-5 is associated to pathogenic properties.

No toxicity or pathogenicity has been reported for any of the donors and/or elements used to obtain HB4 soybean.

Concerning allergenicity, none of the donor organisms is recognised as a major allergen source and, even when allergenic components have been reported for some of them (i.e., sunflower), the genetic elements included in the constructions used to obtain HB4 soybean are not associated to them.

A.2(a)(ii) History of use of the organism in the food supply or history of human exposure to the organism through other than intended food use (e.g., as a normal contaminant)

Donor DNA of the insert for event IND-ØØ41Ø-5 consists of both coding and non-coding genetic elements from a single binary plasmid *pIND2-HB4* as described in Section A.3(b). The *HB4* gene coding sequence is from sunflower (*Helianthus annuus*) and for *the Bar* gene from *Streptomyces hygroscopicus.* The non-coding elements of the *pIND2-HB4* plasmid are from the cauliflower mosaic virus (CaMV), sunflower (*Helianthus annuus*), soybean (*Glycine max* L.), tobacco etch virus (TEV) and *Agrobacterium tumefaciens*.

***Helianthus annus* L. (*HaHB4*- donor)**

The biology and history of sunflower has been widely reviewed (see for example: CFIA 2015; Putnam et al., 2021). The development of the commercial sunflower has been a multi-national effort spanning continents and thousands of years. The sunflower is native to North America and was first grown as a crop by indigenous tribes over 4,500 years ago. Native Americans cultivated the sunflower from its original bushy, multi-headed type to produce a single-stemmed plant bearing a large flower.

The crop's multiple uses included milling for flour or meal production to make bread and cakes. Seeds were roasted, cracked, and eaten whole, either as a snack or mixed with other grains, nuts, and pulses into a type of granola.

The early Americans also discovered that sunflower oil could be extracted and used for cooking. Aside from the crop's value as a food, archaeologists have shown sunflower had a variety of non-food uses. The sunflower's oils and pigments were used as a sunscreen or the basis for a purple dye for skin, hair, or textile decoration, while the plant's sturdy, fibrous stem was exploited in construction. The HAHB4 protein has been previously assessed by FSANZ (A1232).

***Streptomyces hygroscopicus* (*Bar* donor)**

The *bar* gene in HB4 soybean is like that originally cloned from *Streptomyces hygroscopicus* (Murakami et al., 1986) and demonstrated to be useful as a selectable marker in other bacteria (Thompson et al., 1987) and in plants (de Block et al., 1987; Takano and Dayan 2020). It is identical to the bar gene described in Frame et al., (2002) and Mir et al., 2017).

Streptomyces hygroscopicus is a common saprophytic bacterial species that is found worldwide. Soil is the predominant habitat, but these organisms may also be isolated from water.

Streptomyces hygroscopicus produces a variety of useful antimicrobial and herbicidal compounds (Dunne et al., 1998), of which the PAT enzyme confers phosphinothricin tripeptide (phosphinothricin or bialaphos) tolerance. This tolerance is conferred through inactivation by transfer of an acetyl group. Acetyltransferase activity has been identified in six other bacterial species from five different genera of common soil bacteria. This is thought to have evolved as a protective mechanism to protect these microorganisms from antimicrobials produced by both themselves, and other competing microorganisms. Consequently, natural resistance to phosphinothricin and N-acetyltransferase has also been reported in various genera of soil bacteria (Bartsch and Tebbe, 1989).

Recently, numerous works report their important symbiotic relationships with plants and animals (Kaltenpoth et al., 2005; Behie et al., 2016). A recent work describes the first clearly documented case of their mutualism with vertebrates, sea turtles (Sarmiento-Ramirez et al., 2014). In almost all reported cases the streptomycetes protect the host or its food resources from pathogenic fungi.

Streptomyces species very rarely cause human disease but can be detected as common colonisers of human bodies, especially the skin, the respiratory tract, the guts, and the genital tract using molecular techniques (Herbrik et al., 2020). In general, streptomycetes cause suppurative granulomatous tissue changes (Dunne et al., 1998; Herbrik et al., 2020). However, their clinical manifestations and isolations are rare. It is expected that humans would be exposed to these microorganisms and anti-microbial compounds directly through the consumption of roots and other vegetables that are eaten fresh.

The PAT protein is expressed by several transgenic crops that have been in commercial production for many years. FSANZ have not identify any public health or safety concerns associated with the expression of PAT, as encoded by the *pat* or *bar* gene, in numerous assessments (for example, Soybean (A481, A1046, A1073, A1081); Canola (A372, A1140); Maize (A375, A380, A385, A386, A446, A543, A1106, A1116, A1118, A1192); Cotton, A518, A533, A1028, A1040, A1080); Rice (A589) and wheat (A1232). Therefore, this protein has been well characterised and demonstrated to be non-toxic to humans and animals.

Non-coding sequences

The promoter and terminator sequences used in HB4 soybean are derived from common plants or plant pathogens. These genetic elements constitute a minute component of their respective genomes, no genetic elements associated with human, or animal pathogenicity have been used in the construction of HB4 soybean.

Many of the organisms from which these elements are derived are model species in plant science with a history of safe use.

Expression of the *HaHB4* gene is driven by two different allelic promoter regions in sunflower (large and short promoter fragments, LPF and SPF respectively; Dezar et al., 2005b; Manavella et al., 2006). In soybean event IND-ØØ41Ø-5, the LPF sequence has been used to direct *HaHB4* gene expression. Expression of the *bar* gene is driven by the cauliflower mosaic virus (CaMV) 35S promoter and enhanced with a 5' leader sequence from tobacco etch virus.

A poly (A) signal for the termination of *HaHB4* gene transcription is derived from *Agrobacterium tumefaciens*, a soil born, gram-negative bacterium that has been extensively studied since it was identified as the causative agent of crown gall disease in plants (Depicker et al., 1982). A poly (A) signal for the termination of *bar* gene expression is derived from the 3′ untranslated region of a soybean vegetative storage protein (VSP) gene.

A.2(b) A description of the host organism into which the genes were transferred:

A.2(b)(i) Its history of safe use for food

The host organism is a conventional soybean (*Glycine max* L.), belonging to the family Leguminosae. Soybean is grown as a commercial food and feed crop in many countries worldwide, with some 100 countries listed as producers in 2020 (FAOSTAT 2022) and has a long history of safe use for both humans and livestock.

The biology of soybean is fully described in several OECD documents (OECD 2000; OECD 2006; OECD 2012) and other regulatory publications (CFIA, 2021).

Details of the pathogenicity, toxicity or allergenicity of soybean are described in the OECD Revised Consensus Document on Compositional Considerations for New Varieties of soybean [*Glycine max* (L.) Merr.]: Key Food and Feed Nutrients, Anti-nutrients and Toxicants and allergens (OECD 2012).

Soybean is one of eight foods that account for 90% of all IgE-mediated food allergies (Wang et al., 2022). The prevalence of soybean allergy in the general population is reported to be between 0.3% and 0.4% worldwide (Savage et al., 2010), with an increased prevalence reported in children with atopic eczema (Becker et al., 2004).

No sequences associated with known toxins or allergens were used in creating the soybean event proposed in this application.

ORDER:	*Fabales*
FAMILY:	*Fabaceae*
GENUS:	*Glycine*
SPECIES:	*G. max L*
COMMON NAME:	Soybean

A.2(b)(ii) The part of the organism typically used as food

The major soybean commodity products are seeds, oil, and meal. Soybean seeds are valued for their high levels of protein (38-45%) and high oil content (approximately 20%). Approximately 85% of the world's soybean crop is processed into soybean oil and meal with the remainder processed in other ways or eaten whole.

Soybean oils, both liquid and partially hydrogenated, are sold as vegetable oil, or end up in a wide variety of processed foods. Soybean meal, or soymeal, is the material remaining after solvent extraction of oil from soybean flakes, with a 50% soy protein content. The meal is heat treated and ground in a hammer mill. The majority of soybean meal production globally is used as livestock feed.

A.2(b)(iii) The types of products likely to include the food or food ingredient

In addition to their use in livestock feed, soybean products are widely used for human consumption. Common soybean products include soy sauce, soy milk, tofu, soy meal, soy flour, textured vegetable protein (TVP), soy curls, tempeh, soy lecithin and soybean oil. 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.

Crude oil is further refined to produce cooking oil, shortening and lecithin as well as being incorporated into a variety of edible and technical/industrial products. The flakes are dried and undergo further processing to form products such as meal (for use in livestock, pet, and poultry food), protein concentrate and isolate (for use in both edible and technical/industrial products), and textured flour (for edible uses). The hulls are used in mill feed.

More recently, soybean is being used as a meat substitute extensively used by vegan and vegetarian consumers (Messina et al., 2022; Bryant 2022).

A.2(b)(iv) Whether special processing is required to render food derived from the organism safe to eat

Unprocessed (raw) soybeans are not suitable for food use, and have only limited feed uses, as they contain toxicants and anti-nutritional factors, such as lectins and trypsin inhibitors (OECD 2012). Appropriate heat processing inactivates these compounds.

A.3. The nature of the genetic modification

A.3(a) A description of the method used to transform the host organism

The soybean event IND-ØØ41Ø-5 was developed by transforming the variety Williams 82 with the plasmid *pIND2-HB4* to produce the proteins HAHB4 and PAT. This plasmid is described in detail in Section A3(b) with the HAHB4 and PAT protein assessed previously by FSANZ (see Table 2). The transformation protocol is described in Figure 1 and detailed in the Supplement Report **HB4 soybean Report#01010289-Ev1_plant transformation.**

Conclusion of the Development of HB4 Soybean

Soybean event IND-ØØ41Ø-5 was developed by transforming the soybean variety Williams 82 with the binary vector *pIND2-HB4*. Transformation introduced DNA sequences (*HaHB4* and *bar*) intended to provide tolerance to environmental stresses and tolerance to herbicides containing glufosinate.

All genetic elements used to create HB4 soybean were derived from the genomes of organisms present in the natural environment. Soybean and wheat events containing the coding sequences of *HaHB4* and *bar* have been assessed and approved by other regulatory agencies and GM events from a range of food crops containing the coding sequence of the *bar* gene have been assessed and approved by FSANZ from numerous independent submissions.

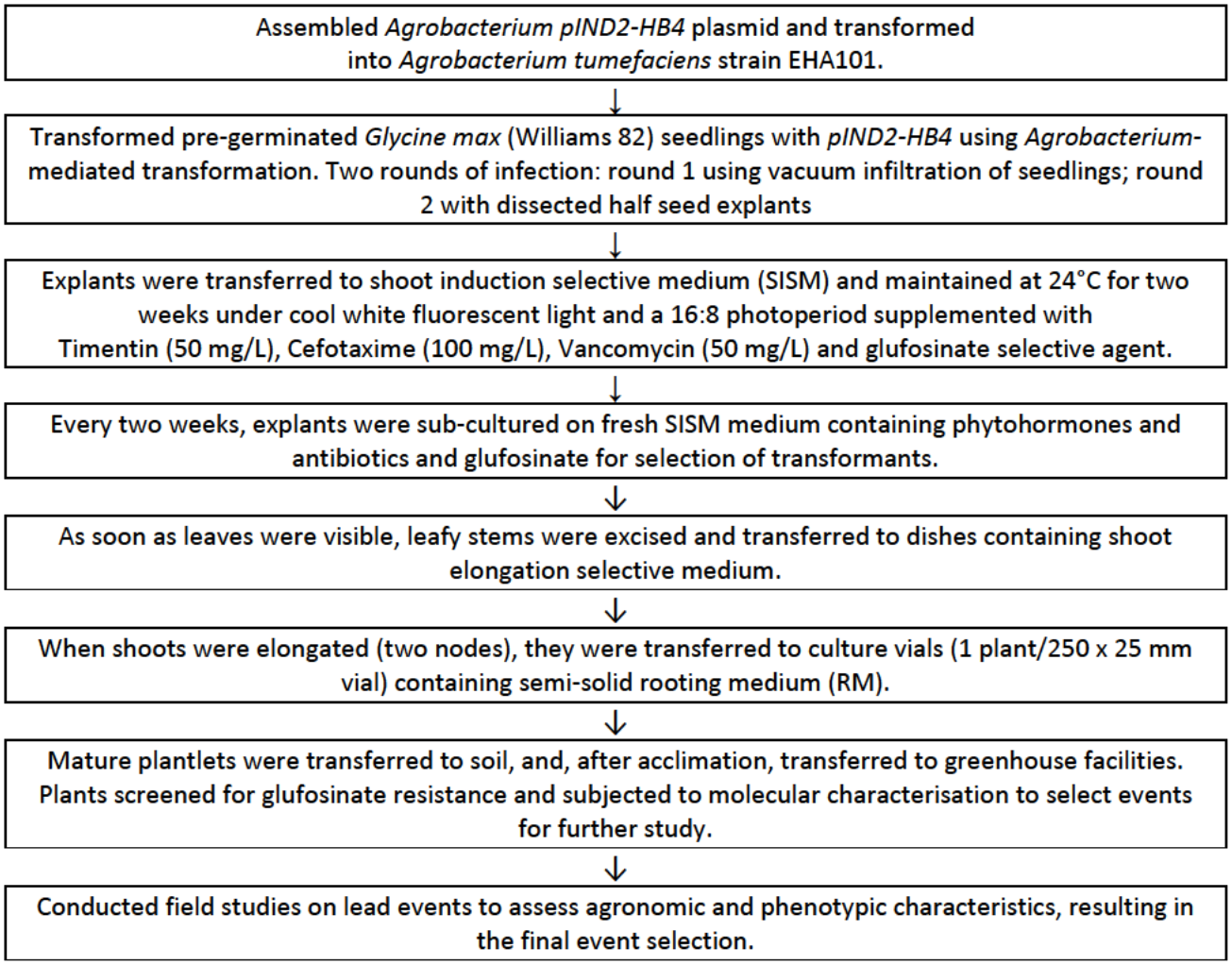


Figure 1. The Development and Selection of HB4 Soybean Transformed with *pIND2-HB4*

A.3(b) A description of the construct and the transformation vectors used

HB4 soybean was developed by transforming the variety Williams 82 with *pIND2-HB4*. All genetic elements were derived from the genomes of species commonly found in the environment and/or the food chain.

The *pIND2-HB4* binary plasmid was constructed using the small versatile pPZP family of *Agrobacterium* binary vectors for plant transformation (Hajdukiewicz et al., 1994). A map of *pIND2-HB4* is provided in Figure 2 with corresponding descriptions of the genetic elements provided in (Table 3).

pIND2-HB4 is an 11,133 bp plasmid vector carrying the genetic elements to deliver the expression of the HAHB4 transcription factor and *bar* gene in soybean (Figure 2 and Figure 3).

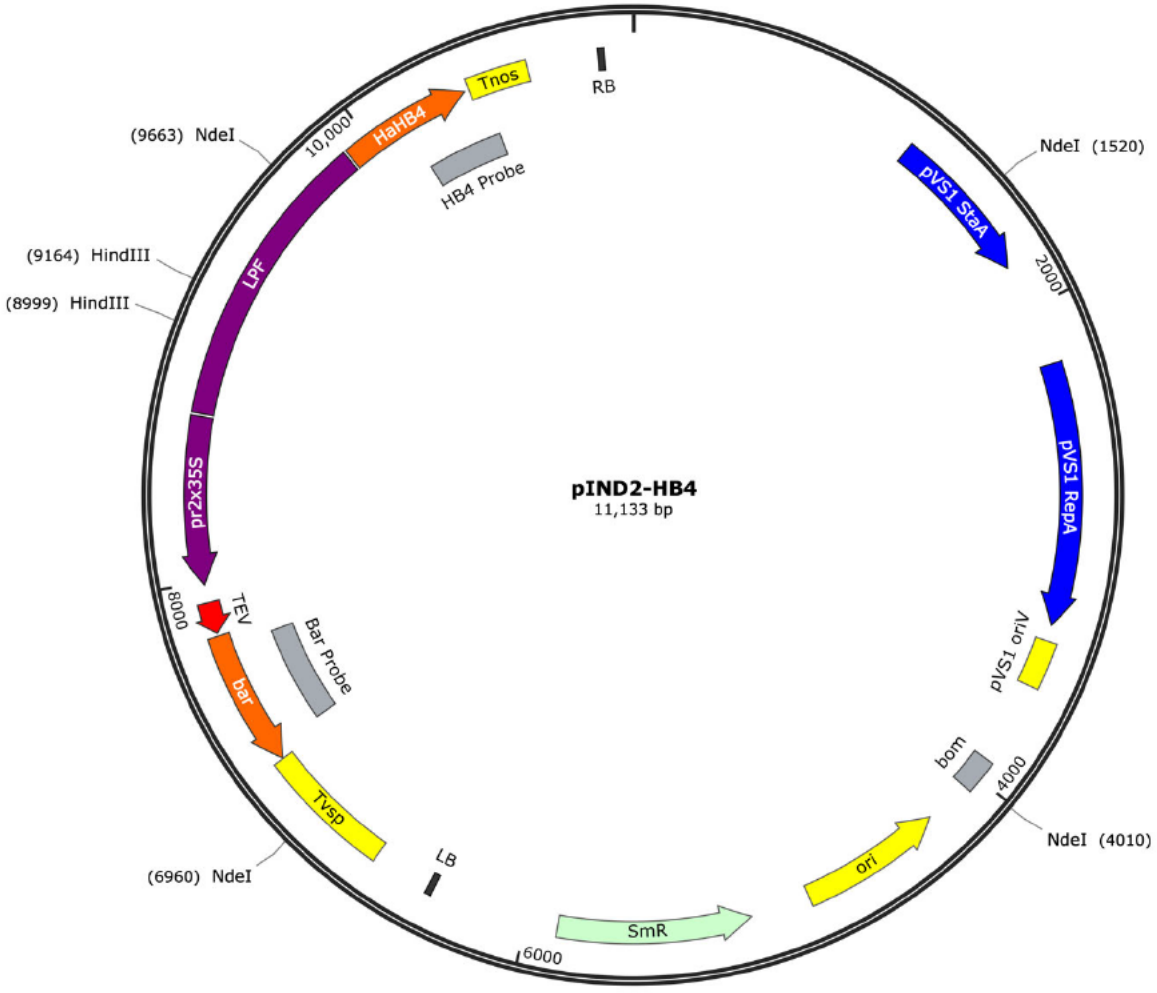


Figure 2. Plasmid Map of *pIND2-HB4*

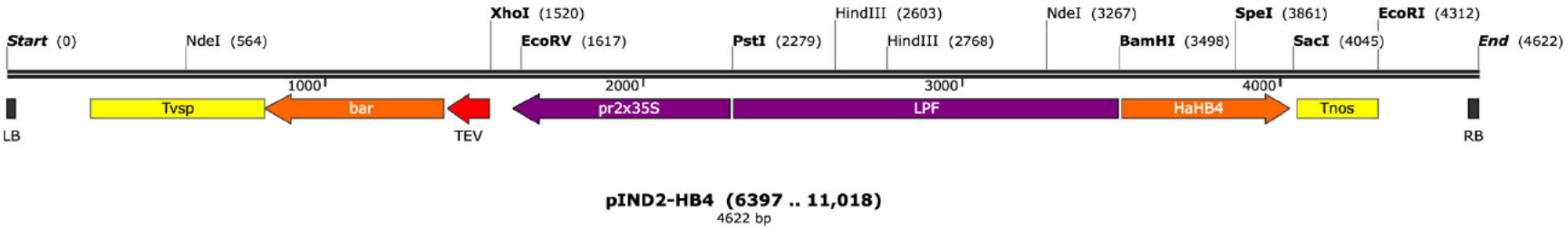


Figure 3. T-DNA region of *pIND2-HB4*

Table 3. Genetic Elements of *pIND2-HB4*

Genetic Element	*Origin*	Accession Number	Position (*pIND2-HB4*)	Size (bp)	Intended Function
1. Intervening Sequence	Binary vector pPZP202	U10461.1	1-1,186	1,186	Vector sequence used for DNA cloning (Hajdukiewicz et al., 1994)
2. Stabilising protein from the *Pseudomonas* plasmid pVS1			1,187-1,816	630	Plasmid stability in culture (Heeb et al., 2000)
3. Intervening Sequence			1,817-2,244		Vector sequence used for DNA cloning (Hajdukiewicz et al., 1994)
4. Replication protein from the *Pseudomonas* plasmid pVS1			2,245-3,318	1,074	Replication protein (Heeb et al., 2000)
5. Intervening Sequence			3,319-3,383		Vector sequence used for DNA cloning (Hajdukiewicz et al., 1994)
6. Origin of replication from the *Pseudomonas* plasmid pVS1			3,384-3,578	195	Origin of replication from pVS1 (Heeb et al., 2000)
7. Intervening Sequence			3,579-3,921		Vector sequence used for DNA cloning
8. Basis of mobility region from plasmid pBR322			3,922-4,062	141	
9. Intervening Sequence			4,063-4,247		Vector sequence used for DNA cloning (Hajdukiewicz et al., 1994)
10. Plasmid origin of replication			4,248-4,836	589	High copy number ColE1/pMB1/pBR322/pUC origin of replication. Plasmid origin of replication (Yanisch-Perron et al., 1985)
11. Intervening Sequence			4,837-5,081		Vector sequence used for DNA cloning (Hajdukiewicz et al., 1994)
12. SmR Antibiotic resistance gene			5,082-5,873	792	Confers resistance to spectinomycin and streptomycin-aminoglycoside adenylyl transferase (Murphy, 1985)
13. Intervening Sequence			5,874-6,396	522	Vector sequence used for DNA cloning (Hajdukiewicz et al., 1994)
14. Left Border sequence			6,397-6,421	25	Secondary cleavage site releases ssDNA insert from pIND2-HB4 (van Haaren et al., 1989)
15. Intervening Sequence			6,422-6,660		Vector sequence used for DNA cloning (Hajdukiewicz et al., 1994)

Genetic Element	*Origin*	Accession Number	Position (*pIND2-HB4*)	Size (bp)	Intended Function
16. Tvsp; poly(A)signal of a soybean vegetative storage protein (vspB) gene	*Glycine max*	M76980.1	6,661-7,206	546	Poly (A) signal for the termination of *bar* transcription (Rapp et al., 1990)
17. *bar* coding sequence	*Streptomyces hygroscopicus*	APM87886.1	7,207-7,770	564	Generates mRNA that leads to phosphinothricin acetyltransferase (PAT) providing herbicide tolerance (Frame et al., 2002; Mir et al., 2017)
18. Intervening Sequence			7,771-7,781		
19. Tobacco Etch Virus (TEV) 5′ leader sequence	*Tobacco Etch Virus*		7,782-7,911	130	Directs efficient translation of the *bar* gene (Carrington and Freed, 1990; Gallie et al., 1995)
20. Intervening Sequence			7,912-7,983		
21. Pr2x35S Promoter	*Cauliflower Mosaic Virus*		7,984-8,670		*De novo* expression of the *HaHB4* gene (Odell et al., 1985; Haq et al., 1995)
22. Intervening Sequence			8,971-8,682		
23. LPF Promoter	*Helianthus annus*	AF339749.1	8,683-9,891	1,209	*De novo* expression of the *HaHB4* gene (Dezar et al., 2005b; Manavella et al., 2006)
24. Intervening Sequence			9,892-9,902		
25. *HaHB4* coding sequence	*Helianthus annus*	AF339748.1	9,903-10,433	531	Generates mRNA that leads to HAHB4 providing environmental stress tolerance (Chan and Gonzalez 1994; Gago et al., 2002; Dezar et al., 2005a; Manavella et al., 2006)
26. Intervening Sequence			10,434-10,542		
27. NOS-ter; poly(A)signal of nopaline synthase gene	*Agrobacterium tumefaciens*	V00087.1	10,453-10,705	253	Poly (A) signal for the termination of *HaHB4* transcription (Depicker et al., 1982)
28. Intervening Sequence			10,704-10,993		
29. RB sequence	Binary vector pPZP202	U10461.1	10,994-11,018	25	Primary cleavage site releases ssDNA insert from pIND2-HB4 (van Haaren et al., 1989)
30. Intervening Sequence			11,019-11,133		Vector sequence used for DNA cloning (Hajdukiewicz et al., 1994)

A.3(c) A full molecular characterisation of the genetic modification in the new organism

This Section provides information that addresses the requirements for Part A.3(c) A full molecular characterisation of the genetic modification in the new organism, including:

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

(ii) A determination of the number of insertion sites, and the number of copies at each insertion site

(iii) Full DNA sequence of each insertion site, including junction regions with the host DNA

(iv) A map depicting the organisation of the inserted genetic material at each insertion site; and

(v) Details of an analysis of the insert and junction regions for the occurrence of any open reading frames (ORFs).

Further information is provided in the Supplement Report **HB4 Soybean_Report#01010290-Ev4_Molecular characterization.**

A.3(c)(i) to (iii) Structure of the Insertion in HB4 Soybean

Soybean event IND-ØØ41Ø-5 was generated by *Agrobacterium*-mediated transformation of soybean explants with the plasmid *pIND2-HB4*. Molecular characterisation was undertaken to determine the number of loci and the sequence of the T-DNA insertion. Analysis via Southern blot hybridisation and Next Generation Sequencing (NGS) demonstrates a single insertion of the T-DNA located on chromosome 9.

1. Southern blot hybridisation

The conventional approach to determine the copy number and integration patterns of transgenic events is to use Southern blot hybridisation. Genomic DNA of homozygous T5 IND-ØØ41Ø-5 plants was digested with the restriction endonucleases *Hind*III and *Nde*I (Figure 4).

There are four *Nde*I restriction sites in plasmid *pIND2-HB4* (Figure 2), two within the T-DNA and two in the binary vector backbone. Complete *Nde*I digestion in the T-DNA should release a DNA segment of 2,703 bp long that contains the binding target for the *bar* probe (Figure 4A). The *HaHB4* probe was expected to detect a DNA fragment with a minimum size of 1.35 kb, assuming a single, intact T-DNA (Figure 4A).

There were only two *Hind*III sites within the T-DNA region, located near each other, and no other *Hind*III sites were present in the *pIND2-HB4* plasmid backbone (Figure 2). Assuming the occurrence of a single, intact T-DNA in the genome of IND-ØØ41Ø-5, the minimum fragment size detected by hybridisation of the *HaHB4* probe would be 1.85 kb. The probe for the selectable marker *bar* gene, on the other hand, would detect digests extending over the left border into the soybean genome. These fragments should be longer than 2.6 kb (Figure 4B).

Assuming the occurrence of a single intact copy of each of the relevant sequences, the minimum predicted sizes of the bands detected in each digest with the *HaHB4* (226 bp) and *bar* (448 bp) probes are summarised in Table 4. The analysis of hybridisation obtained with the *HaHB4* and *bar* probes indicated the presence of a single gene copy T-DNA in soybean event IND-ØØ41Ø-5 (Figure 5).

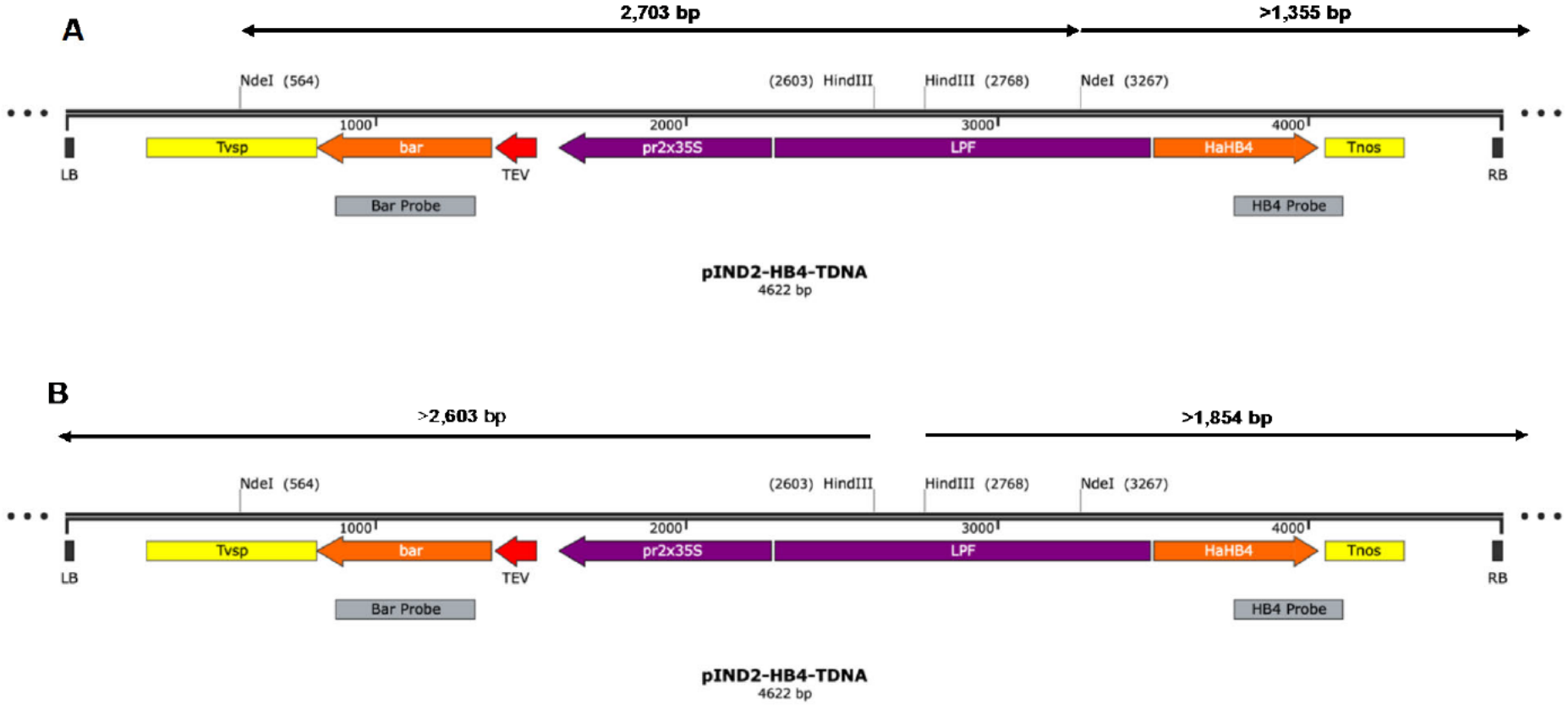


Figure 4. Restriction map of the T-DNA region of the plasmid *pIND2-HB4*

The plasmid T-DNA region containing the CDSs of *HaHB4* and *bar* with their regulatory elements. **A**. Fragments resulting from digestion with the restriction enzyme *Nde*I; **B**. Fragments resulting from digestion with restriction enzyme *Hind*III.

Table 4. Predicted band sizes from Southern blot hybridisation

Restriction enzyme	Minimum fragment size predicted (bp)	
	***HaHB4* probe**	***bar* probe**
*Hind*III	1,854	2,603
*Nde*I	1,355	2,703

Minimum band sizes based on hybridisation with the *HaHB4* (226bp) and *bar* (448bp) probes

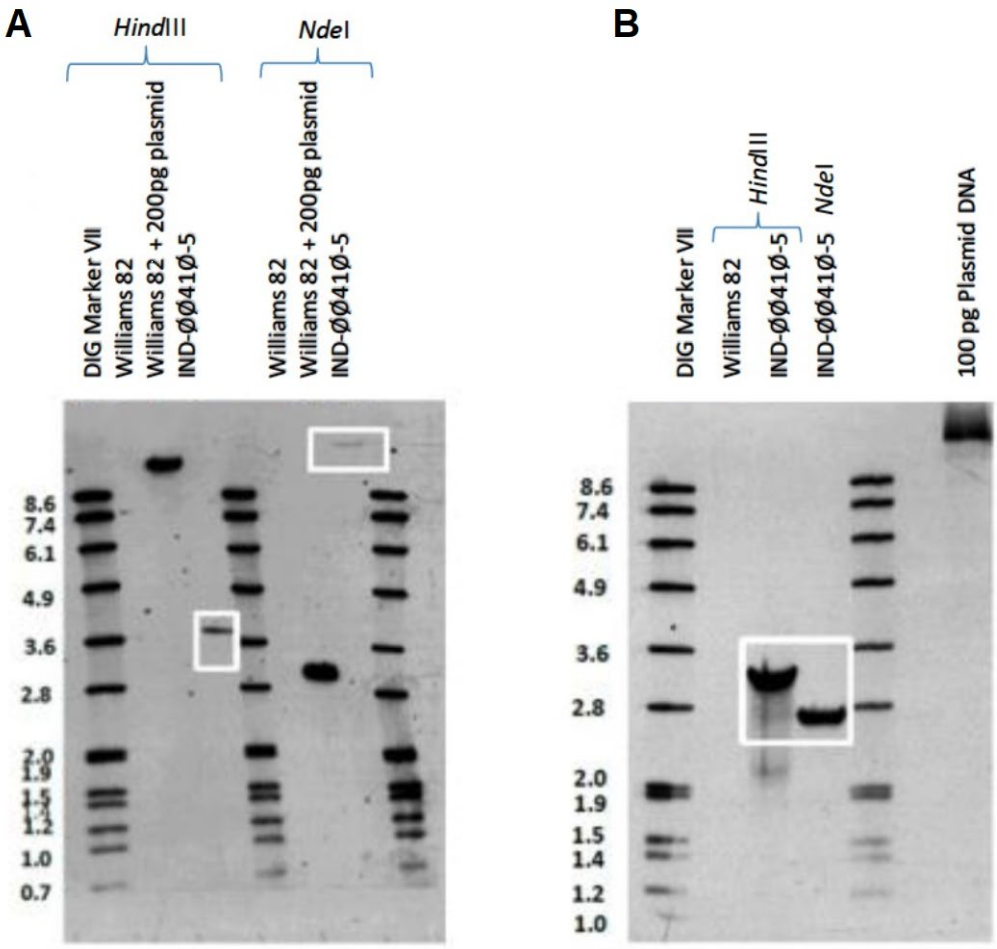


Figure 5. Southern Blots of IND-ØØ41Ø-5 Plant DNA Digested with *Hind*III and *Nde*I hybridised with DIG-labelled probes for *HaHB4* and *bar* detection.

A. DNA bands in IND-ØØ41Ø-5 digests hybridising to the *HaHB4* probe demonstrating a single band greater than 1.8Kb (*Hind*III digest) and greater than 2.6Kb (*Nde*I digest); **B.** DNA bands in IND-ØØ41Ø-5 digests hybridising to the *bar* probe demonstrating a single band greater than 2.6Kb (*Hind*III digest) and the predicted 2.7Kb (*Nde*I digest). DIG-labelled Marker VII ladder band sizes are indicated on the left of the blots in kb.

2. HB4 Soybean insertion sequence analysis

Recent advances in genome sequencing have led to the development of next generation sequencing (NGS) technologies (Morey et al., 2013; Reuter et al., 2015; Heather and Chain, 2016). NGS refers to a collection of technologies that utilise massively parallel sequencing approaches producing millions of short read sequences in a much shorter time, at a much cheaper cost and with higher throughput compared to Sanger sequencing. In combination with bioinformatics, NGS technology was used to characterise the insertion of *pIND2-HB4* T-DNA in event IND-ØØ41Ø-5.

NGS was used to determine the whole genome sequence of soybean event IND-ØØ41Ø-5. The DNA insertion was assembled *de novo* from the Illumina sequence reads**.** and two junction sequences (JS) between the inserted DNA and the soybean genome were identified.

The results from the Southern blot hybridisation, concluding a single T-DNA insert, were confirmed by mapping the Illumina-generated sequence data against the whole sequence of the plasmid vector used for transformation (Figure 6). Data generated using both NGS and Southern hybridisation technologies support the insertion of a single copy of the *pIND2-HB4* T-DNA in event IND-ØØ41Ø-5.

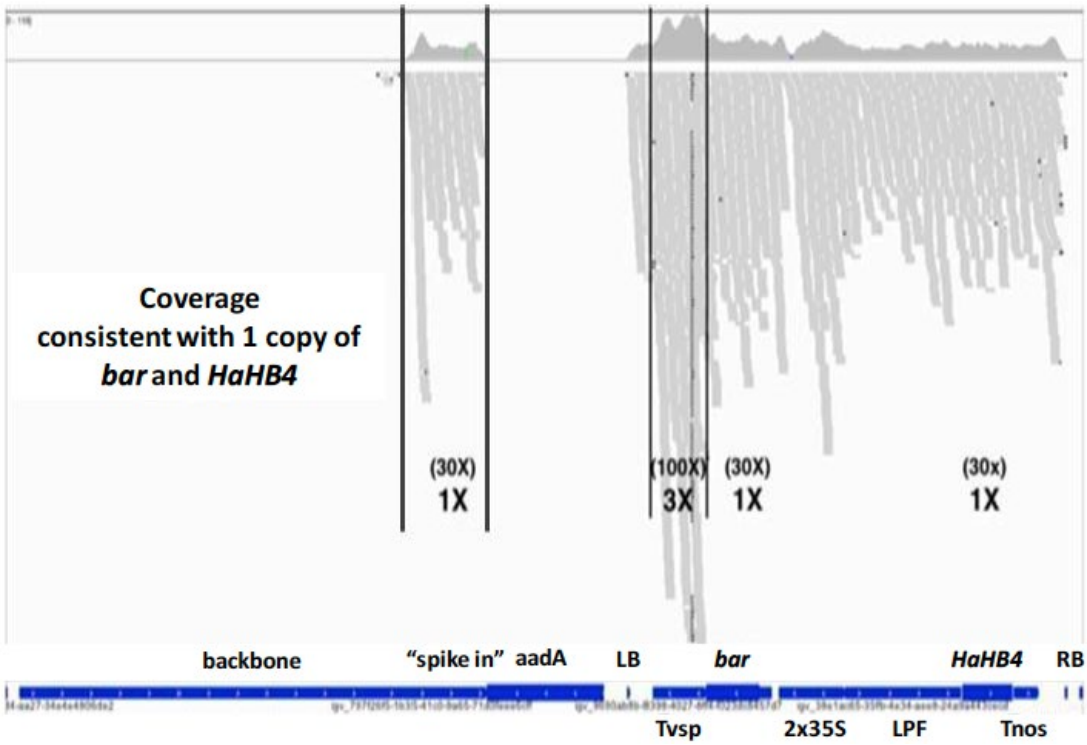


Figure 6. Sequence analysis of Event IND-ØØ41Ø-5
Event IND-ØØ41Ø-5. DNA sequence reads were mapped against the complete plasmid vector sequence of *pIND2-HB4* (shown in blue). The total read coverage is presented in parentheses immediately above the normalised read coverage for each element in the plasmid vector. The normalised read coverage provided an estimate of the copy number of each element of the transformation vector present in the IND-ØØ41Ø-5 genome. Mapped sequence reads between the positions labelled "backbone" and "aadA" correspond to a sequence identical with control DNA included in the sequencing run to check for the sequencing error rate—this is labelled "spike in". The resulting coverage of mapped reads indicated that no backbone elements from the vector were present in event IND-ØØ41Ø-5. The 3X normalised coverage of the Tvsp element (soybean vegetative storage protein terminator) is due to the additional reads generated from two endogenous copies of the element in the soybean genome.

The sequence of the insert in event IND-ØØ41Ø-5 corresponded to the sequence of T-DNA in the binary vector, with a single copy of each gene and each regulatory element (Figure 3). The left and right border sequencies were not transferred completely. Three bp of the left border and the complete right border, including three bp upstream are missing in IND-ØØ41Ø-5.

The complete sequence of the insertion and the flanking soybean sequences is provided in **Appendix 1**.

Conventional Sanger sequencing of multiple amplicons covering the whole insertion and its flanking sequences corroborated the JSA analysis of the Illumina-generated sequence.

3. Localisation of the HB4 soybean insert

Junction Sequence Analysis (JSA) of event IND-ØØ41Ø-5 using the Illumina-generated sequence data was consistent with the integration of a single T-DNA copy at a single locus (Figure 6) (Kovalic et al., 2012). This result was supported by the finding of only two junction sequences in the whole-genome sequencing of event IND-ØØ41Ø-5.

The flanking sequences were mapped to the soybean genome by homology search using BLASTN (Zhang et al., 2000) against the *Glycine max* v4.0 reference, Annotation Release 104, NCBI Accession NC_038245.2 (Figure 7 and Figure 8).

```
>Glycine max cultivar Williams 82 chromosome 9, Glycine_max_v4.0
Sequence ID: NC_038245.2 Length: 50572668
Range 1: 36258367 to 36258728

Score:664 bits(359), Expect:0.0,
Identities:361/362(99%),  Gaps:0/362(0%), Strand: Plus/Minus

Query  1         TTCATTTTTTAAGAAGTGAATATCAACGTCTTCCCTTATGTATCGTATCCTGTCATCATA  60
                 |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct  36258728  TTCATTTTTTAAGAAGTCAATATCAACGTCTTCCCTTATGTATCGTATCCTGTCATCATA  36258669

Query  61        GACTGGCTGCAAGTTTTGGTCAATGTAAAAAGATATTGACACTCTATTCTTGTATCTTAA  120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  36258668  GACTGGCTGCAAGTTTTGGTCAATGTAAAAAGATATTGACACTCTATTCTTGTATCTTAA  36258609

Query  121       GATTTTGTGGAACTTGGAAATCTTTTTCCTTTGTACGTGACTCCCCTCTCAGTTGGGTCA  180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  36258608  GATTTTGTGGAACTTGGAAATCTTTTTCCTTTGTACGTGACTCCCCTCTCAGTTGGGTCA  36258549

Query  181       GCCTGAgtgatttttttctcaaatcaagaaactttatttataaatctaacattataatat  240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  36258548  GCCTGAGTGATTTTTTTCTCAAATCAAGAAACTTTATTTATAAATCTAACATTATAATAT  36258489

Query  241       taaaaaaacaaatattaaaatattcatgatatttttaaatctaaataatattctaaaaat  300
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  36258488  TAAAAAAACAAATATTAAAATATTCATGATATTTTTAAATCTAAATAATATTCTAAAAAT  36258429

Query  301       ttgaaacaaataaattcttgaaaataaactaaattattcTTTTCCAAACTAACTAAAGAT  360
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  36258428  TTGAAACAAATAAATTCTTGAAAATAAACTAAATTATTCTTTTCCAAACTAACTAAAGAT  36258369

Query  361       AT  362
                 ||
Sbjct  36258368  AT  36258367
```

Figure 7. Left flanking sequence alignment to a soybean reference genome
BLASTN alignment of the left flanking sequence (Query 1) with the reference genome (Sbjct). A single nucleotide difference between the left flanking region sequence and the soybean reference genome currently available in NCBI is highlighted at position 17.

```
>Glycine max cultivar Williams 82 chromosome 9, Glycine_max_v4.0
```

```
Sequence ID: NC_038245.2 Length: 50572668
Range 1: 36257980 to 36258224

Score:453 bits(245), Expect:6e-126,
Identities:245/245(100%),  Gaps:0/245(0%), Strand: Plus/Minus

Query  1         ACCCTCAATCATCTCACTTCATTATctcctatatttttttattaacttctcttttatacta  60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  36258224  ACCCTCAATCATCTCACTTCATTATCTCCTATATTTTTTATTAACTTCTCTTTTATACTA  36258165

Query  61        ttttaaaaaaataaaaagtgagaatTTAAACAGAAAAAACCTCTCTCAAGTCTTTCTCTC  120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  36258164  TTTTAAAAAAATAAAAAGTGAGAATTTAAACAGAAAAAACCTCTCTCAAGTCTTTCTCTC  36258105

Query  121       TATTTCAGTGGTCTGAGTTCAGTTGCGTCTCTTAATCTTTTAGGTTGGGAAAACATCATC  180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  36258104  TATTTCAGTGGTCTGAGTTCAGTTGCGTCTCTTAATCTTTTAGGTTGGGAAAACATCATC  36258045

Query  181       TTCTTTTGGGAGATTGGCTCCACCCACAACAGTTGTTAACTTGTTTacataaataattga  240
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  36258044  TTCTTTTGGGAGATTGGCTCCACCCACAACAGTTGTTAACTTGTTTACATAAATAATTGA  36257985

Query  241       tattc  245
                 |||||
Sbjct  36257984  TATTC  36257980
```

Figure 8. Right flanking sequence alignment to a soybean reference genome
BLASTN alignment of the right flanking sequence (Query 1) with the reference genome (Sbjct).

The insertion within the soybean genome (Figure 9) occurred in a single location of chromosome 9 between genomic positions 36,258,367 and 36,258,224. The insertion is located 3' to a putative F-box/LRR-repeat protein like the one coded by At3g26922 gene (LOC100806405 in the NCBI *Glycine max* v4.0), and 4.4Kb 5' of an uncharacterised protein (LOC102666618 in the NCBI *Glycine max* v4.0). The junction sequence alignment indicated that 142 bp of genomic DNA was deleted because of the T-DNA insertion from *pIND2-HB4* (Figure 9). The insertion site does not indicate interruption of any gene.

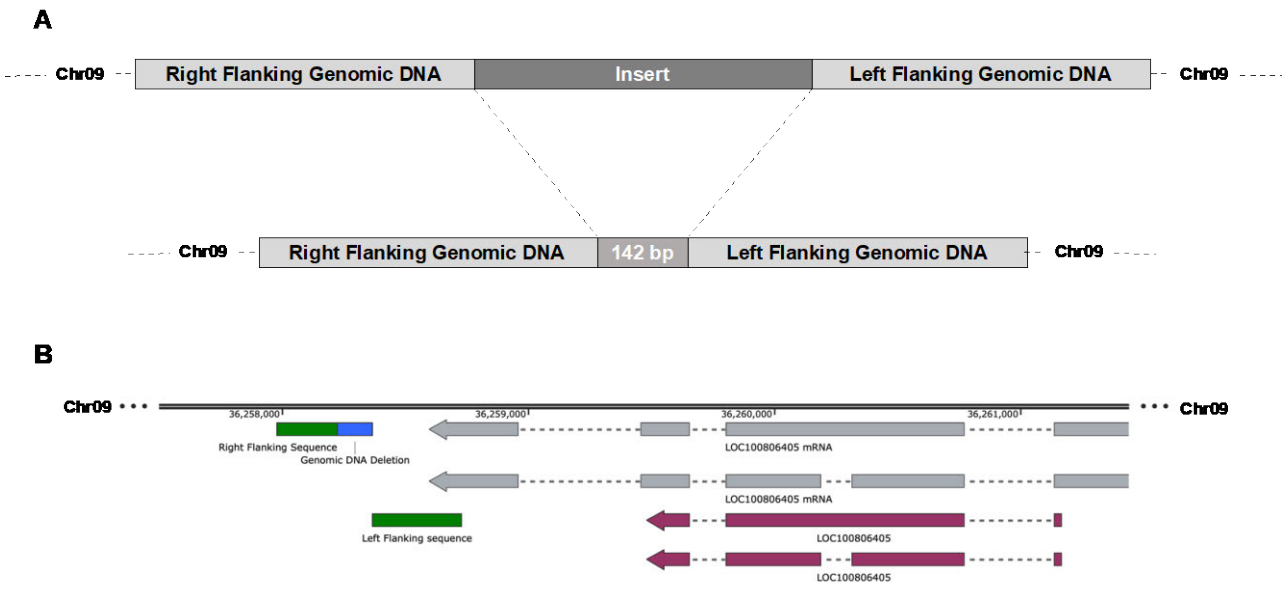


Figure 9. Insertion site on HB4 soybean
A. Sequence analysis indicated that 142 bp of genomic DNA was deleted upon insertion of the T-DNA. **B.** Position of the insertion within the soybean genome. The deletion is shown in blue. The nearest annotated gene (LOC100806405) is shown in grey.

4. Absence of vector backbone DNA

The absence of vector backbone sequences in event IND-ØØ41Ø-5 was determined through whole-genome sequence obtained with the Illumina NGS method (see Figure 6). In addition, Southern blot analyses were employed to provide a second assay for vector backbone sequences. Three probes (aadA, STA, and REP) were used to detect vector backbone sequences. None of these probes detected vector backbone sequences when hybridised to IND-ØØ41Ø-5 DNA.

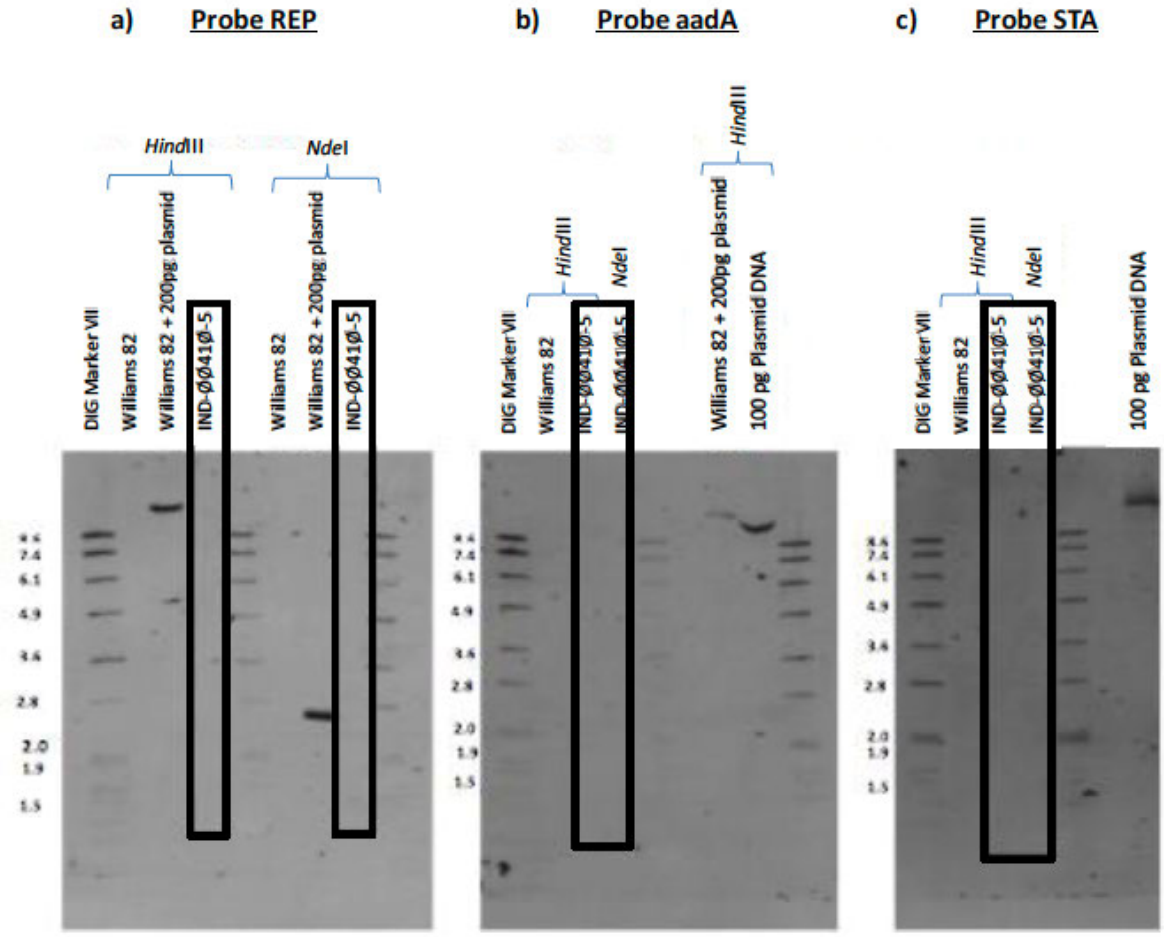


Figure 10. DNA gel blot analyses for detection of vector backbone sequence in event IND-ØØ41Ø-5
Genomic DNA from leaves of either T5 generation event IND-ØØ41Ø-5or non-transgenic control Williams 82 was digested with *Hind*III and *Nde*I and hybridised with DIG-labelled probes for a) REP, b) aadA, and c) STA sequence. Williams 82 + 200 pg plasmid DNA and 100 pg plasmid DNA were used as positive controls. DIG labelled Marker VII ladder band sizes are indicated on the left of the blots in kb. Rectangles highlight lanes showing no probe hybridization signal in event IND-ØØ41Ø-5.

Conclusions

A combination of Sanger and Illumina NGS sequencing corroborated studies using Southern blots and showed the presence of a single insert in soybean IND-ØØ41Ø-5. The structure and sequence of the insert in IND-ØØ41Ø-5 are provided, with flanking DNA sequences. No backbone DNA was integrated into the Soybean genome. No annotated genes were disrupted by the insertion of the T-DNA.

A.3(c)(iv) A map depicting the organisation of the inserted genetic material at each insertion site

Details of the organisation of the inserted genetic material at the integration site are described above. Specifically:

HB4 Soybean–Detailed organisation of the genetic elements of the T-DNA insert in accordance with the insert organisation of the *pIND4-HB4* vector (Figure 3) and integration into the soybean genome (Figure 9).

A.3(c)(v) Details of an analysis of the insert and junction regions for the occurrence of any open reading frames (ORFs)

The sequence of the insertion and the soybean flanking regions was subjected to an ORF analysis (see Section B1(d) and the Supplement Report **HB4 Soybean_Report#01010291-Ev6_BioinformaticAnalysis**). None of the peptides that might be hypothetically produced from these ORFs were identified as homologs of known toxins or allergens (see Section B1(d)).

A.3(d) A description of how the line or strain from which food is derived was obtained from the original transformant (i.e. provide a family tree or describe the breeding process) including which generations have been used.

The process of development and selection event IND-ØØ41Ø-5 is summarised in Figure 1 and and detailed in Supplemental Report **HB4 soybean Report#01010289-Ev1_plant transformation**.

A schematic representation of the development of HB4 soybean and the generations used for analysis is presented in Figure 11.

The original event IND-ØØ41Ø-5 and its derivatives will continue to be crossed into elite soybean varieties through conventional breeding programs. Commercial varieties of soybean containing the HB4 trait will be used for food.

A.3(e) Evidence of the stability of the genetic changes, including:

(i) The pattern of inheritance of the transferred gene(s) and the number of generations over which this has been monitored
(ii) The pattern of inheritance and expression of the phenotype over several generations and, where appropriate, across different environments

Several approaches were used to assess the stability of the insertion in HB4 soybean. Firstly, stability of the T-DNA insertion was assessed over six generations by sequencing the T-DNA and flanking sequences (**Appendix 2**). Except for the Illumina-derived sequence (T6), each sequence presented for each generation (T1, T3, and T5) is a consensus sequence obtained from Sanger sequencing of 3 different plant amplicons. No changes in DNA sequence were detected across the six tested generations.

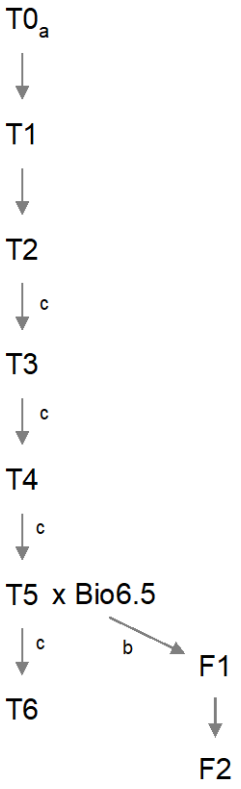
Secondly, segregation of the T-DNA was assessed using a PCR diagnostic in F_2 progeny plants from crosses between IND-ØØ41Ø-5 and a commercial soybean cultivar (Bio 6.5). The PCR diagnostic examined the T-DNA Left Border Junction compared to the native allele (Figure 12).

A homozygous IND-ØØ41Ø-5 plant was crossed with Bio 6.5 to produce F_1 progeny. Four F_1 plants were self-pollinated to produce F_2 seeds that were used for the segregation analysis. In total 73 F_2 plants were assessed for zygosity (Table 5). F_2 plants were scored as homozygous for the IND-ØØ41Ø-5 insertion (I) when the amplicon for the Left Border Junction was present and the amplicon for the native allele was absent. F_2 plants were scored as hemizygous (H) when both amplicons described above were present. F_2 plants were

scored as homozygous for the native Williams 82 cultivar allele (W) when the amplicon for the Left Border Junction was absent and the amplicon for the native allele was present.

Chi-square goodness of fit tests indicated that there was no significant departure from the predicted 1:2:1 (I:H:W) genotypic segregation ratio (X^2[2,N=73] = 0.05, P=0.8426; Table 5).

Protein expression of both HAHB4 and PAT at the T6 generation confirm the functionality of the gene cassette (see Section B.1).



Experiment	Generation	Report
Sequencing insert and flanking sequencies	T6	#01010290 Molecular Characterization
Southern analysis	T5	#01010290 Molecular Characterization
Segregation analysis	F2	#01010290 Molecular Characterization
Stability analysis	T1, T3, T5, T6	#01010290 Molecular Characterization
Protein expression analysis	T6	#01010271 HB4 detection and #01010297 PAT detection
Composition field trials	T6	Chiozza, Burachik and Miranda (2020) Compositional Assessment and #01010298 Endogenous Allergen Levels
Agronomic and phenotypic field trials	T6	HB4 Soybean Agronomic Characterization AR&USA

T: transformation generation. T0 is the transformed plant; T1 is the first generation of transformed plant.
F: Filial generation. F1 is the first filial generation
[a] Williams 82 was used for transformation
[b] Single plant used
[c] Homozygous generation

Figure 11. Schematic representation of the development of event IND-ØØ41Ø-5 and the generations used in the different studies.

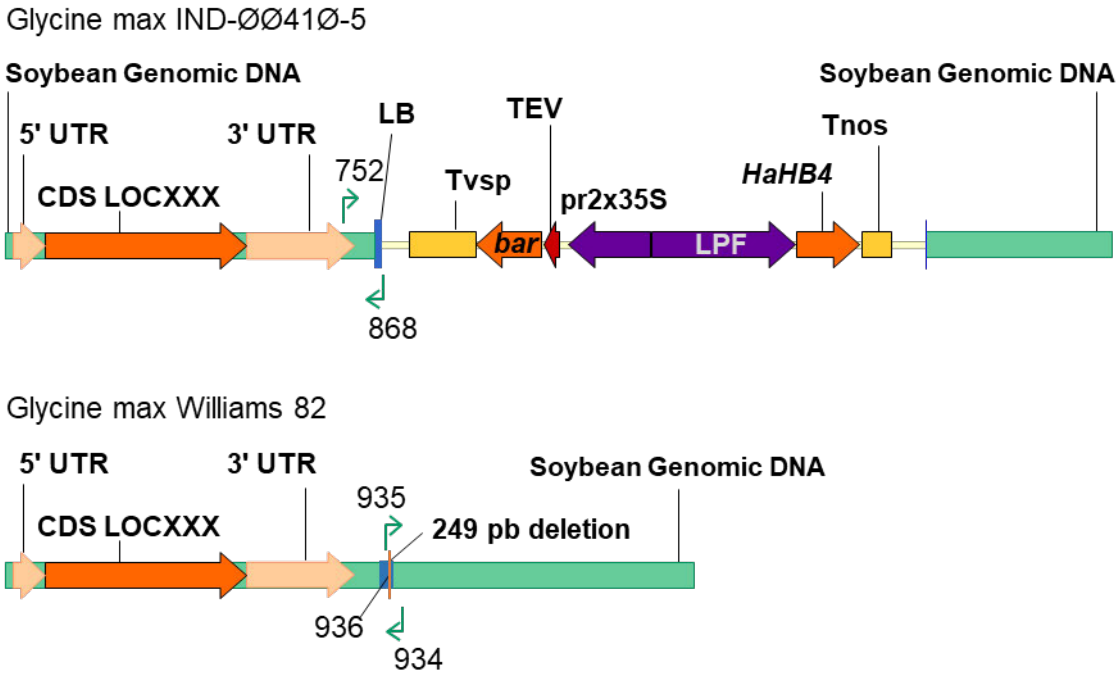


Figure 12. PCR diagnostic to test for segregation of the T-DNA insertion of event IND-ØØ41Ø-5
Upper panel: scheme of the insertion in soybean event IND-ØØ41Ø-5 showing the elements present within the T-DNA, and the primers (indicated as green broken arrows) used for segregation analysis. The primers 868 and 752 were used to test for the presence of the Left Border Junction. **Lower panel:** scheme of native allele showing the elements present in insertion region (without the T-DNA). Primers 934 and 935 and probe 936 (indicated as a light orange vertical line) were used to test for the presence of the native allele. UTR: untranslated region, CDS LOCXXX: coding sequence of the gene LOC100806405.

Table 5. Analysis of segregation of IND-ØØ41Ø-5 T-DNA in F_2 plants.

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

I: IND-ØØ41Ø-5 homozygous; H: hemizygous; W: Williams 82 homozygous.

Summary of genetic stability studies

The results of the segregation pattern in the F_2 generation, as well as the presence of all the genetic elements in the different generations analysed support the conclusion that the IND-ØØ41Ø-5 insertion resides at a single locus within the soybean genome, it is stable and is inherited according to Mendelian principles.

A.3(g) An analysis of the expressed RNA transcripts, where RNA interference has been used

Not applicable to this application.

B. Characterisation and Safety Assessment of New Substances

B.1. Characterisation and Safety Assessment of New Substances

B.1(a) a full description of the biochemical function and phenotypic effects of all new substances (e.g. a protein or an untranslated RNA) that are expressed in the new GM organism, including their levels and site of accumulation, particularly in edible portions

The soybean event IND-ØØ41Ø-5 was developed by transforming the soybean variety Williams 82, with the plasmid vector *pIND2-HB4* using Agrobacterium-mediated transformation. The event was developed to confer increased tolerance to environmental stresses avoiding reduction of crop yield and exhibits tolerance to glufosinate-based herbicides (Table 2).

Two new proteins are expressed in HB4 soybean; the transcription factor HAHB4 that confers tolerance to environmental stress, and the enzyme phosphinothricin N-acetyl transferase (PAT), providing herbicide tolerance.

Further information is provided in the following Supplement Reports:

- **HB4 Soybean#01010301-Ev2_Identity of Genes and Expressed Proteins**
- **HB4 Soybean_Report #01010271-Ev2-HAHB4 Protein Detection**
- **HB4 Soybean_Report#01010297-Ev3-PAT protein detection**
- **HB4_Report#01010273-Ev2_Protein Safety**
- **HB4 Soybean_Report#01010291-Ev6_BioinformaticAnalysis**
- **HB4 Soybean_Report #01010298-Ev2-Endogenous Allergen Levels.**

Identity and function of the HAHB4 protein

The HAHB4 protein has recently been assessed by FSANZ (A1232) in genetically modified wheat. The homeodomain-leucine zipper (HD-Zip) gene family is an important class of transcription factors only found in plants (Henriksson et al., 2005; Ariel et al., 2007). Members of this gene family play vital roles in plant growth and development and participate in responding to various biotic and abiotic stresses (Liu et al., 2013; Li et al., 2019).

The *HaHB4* (*Helianthus annuus* homeobox 4) gene is a member of the HD-Zip sub-family I coding for the sunflower transcription factor HAHB4 (Dezar et al., 2005a; Harris et al., 2011; González et al., 2020).

Transgenic *Arabidopsis thaliana* plants expressing *HaHB4* exhibit a characteristic phenotype that includes a strong tolerance to water stress, are less sensitive to external ethylene and enter the senescence pathway later (Manavella et al., 2006). Expression studies in sunflower indicate that HaHB4 transcript levels are elevated in mature/senescent leaves and again demonstrated the action of this TF in the regulation of ethylene-related genes. Stable transformation of *Arabidopsis* plants as well as transient transformation of sunflower leaves, further confirmed the involvement of HAHB4 in direct and indirect regulation of multiple stresses including water deficit, saline exposure, ABA and ethylene responses, photosynthesis, mechanical damage, and herbivory. This and subsequent research (Manavella et al., 2008a, 2008b, 2008c, Dezar et al., 2005a, 2005b) led to the proposal that HAHB4 is involved in a mechanism related to ethylene-mediated senescence and that this TF participates in the regulation of the expression of genes involved in developmental responses of plants to desiccation.

The sunflower HAHB4 protein was identified by using a degenerate oligonucleotide derived from the conserved HD amino acid sequence WFQNRRA to screen a cDNA library generated from sunflower stem

(Chan and Gonzalez, 1994). HAHB4 was later shown to preferentially bind as a dimer to the dyad-symmetrical sequence CAAT(A/T)ATTG (Palena et al., 1999).

The amino acid sequence of the HAHB4 protein expressed in soybean event IND-ØØ41Ø-5 differs slightly from the one deduced from the nucleotide sequence of the cDNA of the mRNA transcript of the native sunflower *HaHB4* gene that was annotated in the NCBI GenBank, Accession number AF339748.1 (Chan and Gonzalez, 1994; Gago et al., 2002; González et al., 2019; Figure 13). It is, however, the same protein as assessed by FSANZ in A1232.

The differences between the native sunflower gene and the HAHB4 protein in soybean event IND-ØØ41Ø-5 include:

1. A deletion of amino acids 7-10 (as numbered by the NCBI original sequence, accession AAA63768.2).
2. A Lys to Arg substitution at position 22 (K22⇒R18)
3. A Phe to Leu substitution at position 159 (F159⇒L155)
4. A Phe to Leu substitution at position 175 (F175⇒L171)

```
HAHB4Sunflower     MSLQQVPTTETTTRKNRNEGRKRFTDKQISFLEYMFETQSRPELRMKHQL  50
HAHB4Wheat(A1232)  MSLQQV----TTTRKNRNEGRRRFTDKQISFLEYMFETQSRPELRMKHQL  46
HAHB4Soybean       MSLQQV----TTTRKNRNEGRRRFTDKQISFLEYMFETQSRPELRMKHQL  46

HAHB4Sunflower     AHKLGLHPRQVAIWFQNKRARSKSRQIEQEYNALKHNYETLASKSESLKK 100
HAHB4Wheat(A1232)  AHKLGLHPRQVAIWFQNKRARSKSRQIEQEYNALKHNYETLASKSESLKK  96
HAHB4Soybean       AHKLGLHPRQVAIWFQNKRARSKSRQIEQEYNALKHNYETLASKSESLKK  96

HAHB4Sunflower     ENQALLNQLEVLRNVAEKHQEKTSSSGSGEESDDRFTNSPDVMFGQEMNV 150
HAHB4Wheat(A1232)  ENQALLNQLEVLRNVAEKHQEKTSSSGSGEESDDRFTNSPDVMFGQEMNV 146
HAHB4Soybean       ENQALLNQLEVLRNVAEKHQEKTSSSGSGEESDDRFTNSPDVMFGQEMNV 146

HAHB4Sunflower     PFCDGFAYFEEGNSLLEIEEQLPDPQKWWEF 181
HAHB4Wheat(A1232)  PFCDGFAYLEEGNSLLEIEEQLPDLQKWWEF 177
HAHB4Soybean       PFCDGFAYLEEGNSLLEIEEQLPDLQKWWEF 177
```

Figure 13. Alignment of HAHB4 protein sequences
Alignment of the amino acid sequence of sunflower HAHB4 (Accession AAA63768.2) (HAHB4Sunflower); HAHB4 wheat (FSANZ A1232) and the sequence translated in IND-ØØ41Ø-5 (HAHB4Soybean). Differences between the wheat and soybean to the native sunflower accession are highlighted. Numbers correspond to amino acid positions and are in frame with the GenBank HAHB4 accession.

The introduction of *HaHB4* gene in soybean event IND-ØØ41Ø-5 led to the environmental stress tolerance phenotype. Phenotypic and field performance evaluation of several *HaHB4*-containing lines allowed the selection of a transgenic soybean (termed IND-ØØ41Ø-5), which was shown to provide an increased yield opportunity under conditions of environmental stress (Ribichich et al., 2020).

HAHB4 is homologous to proteins with a history of safe use

Proteins with a history of safe use, or that are structurally and functionally related to proteins with a history of safe use, generally are considered safe to consume (Hammond and Cockburn, 2008). As a component of the safety assessment of HAHB4, bioinformatic analyses were conducted to identify sequence homology between the HAHB4 protein and proteins with a history of safe use.

HAHB4 is a member of the homeodomain-leucine zipper (HD-Zip) gene family and is found in sunflower and genetically modified HB4 wheat and soybean. In 2015, the US Food and Drug Administration (FDA) has completed the Early Food Safety Evaluation (EFSE) process for HAHB4. In the EFSE process, the FDA reviewed safety data provided and supported the conclusion that the inadvertent presence of low levels of the HAHB4 protein would not raise food safety concerns (FDA 2015).

A thoughtful analysis of the HD-Zip superfamily performed by Harris et al. (2011), provides an unrooted phylogenetic tree of the HD-Zip protein superfamily. The tree contains over 50 selected sequences, grouped into clades (Figure 14).

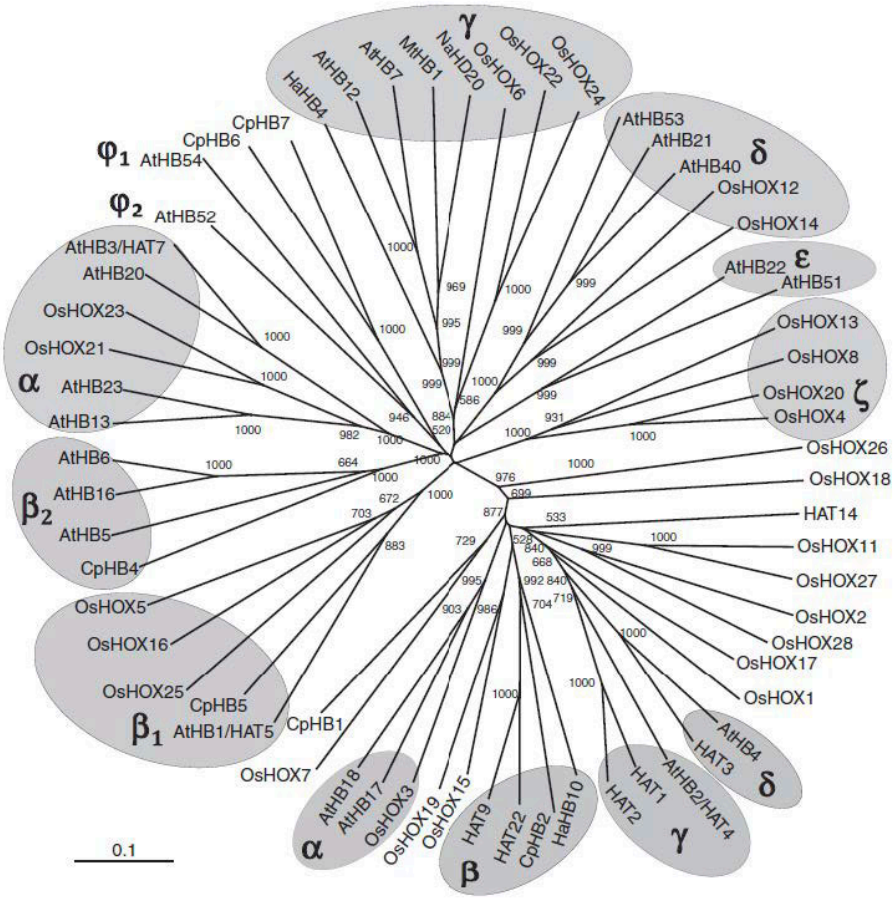


Figure 14. Unrooted phylogenetic tree of the HD-Zip protein superfamily
CLUSTALX alignment (Thompson et al., 1997) based on full-length amino acid sequences. The different clades within HD-Zip I and II family of proteins are circled and identified as α, β1, β2, γ, δ, ε, ζ, ϕ1 and ϕ2 (Agalou et al., 2008; Ciarbelli et al., 2008; Henriksson et al., 2005). Branch lengths are drawn to scale. Two-letter prefixes for sequence identifiers indicate species of origin. At, *Arabidopsis thaliana*; Cp, *Craterostigma plantagineum*; Mt, *Medicago truncatula*; Na, *Nicotiana attenuata*; Os, *Oryza sativa*; HB, homeobox; HOX, homeobox. Taken from Harris et al. (2011).

Identity and function of the PAT protein

The *bar* gene in HB4 soybean is like that originally cloned from *Streptomyces hygroscopicus* (Murakami et al., 1986) and demonstrated to be useful as a selectable marker in other bacteria (Thompson et al., 1987) and in plants (de Block et al., 1987). It is the same as that described in Frame et al., (2002) and Mir et al., (2017). Importantly, the *bar* gene produces the enzyme phosphinothricin acetyl transferase (PAT), which breaks down phosphinothricin (also known as glufosinate), a broad-spectrum herbicide that acts as a competitive inhibitor of glutamine synthetase. As such, plants modified to contain the *bar* gene can tolerate herbicides that contain glufosinate ammonium.

Details on the common soil bacterium *Streptomyces hygroscopicus* are provided in Section A2(a)(i).

The PAT protein is expressed by several transgenic crops that have been in commercial production for many years. FSANZ have not identify any public health or safety concerns associated with the expression of PAT, as encoded by the *pat* or *bar* gene, in numerous assessments (for example, Soybean (A481, A1046, A1073, A1081); Canola (A372, A1140); Maize (A375, A380, A385, A386, A446, A543, A1106, A1116, A1118, A1192); Cotton, A518, A533, A1028, A1040, A1080); Rice (A589) and wheat (A1232). The history of safe use of *S. hygroscopicus*, and safety data for the PAT protein are also provided in Herouet et al. (2005) and ILSI (2016). Therefore, this protein has been well characterised and demonstrated to be non-toxic to humans and animals.

HAHB4 Protein Expression in Soybean Event IND-ØØ41Ø-5

Members of the HD-Zip family of transcription factors (TFs), unique to plants, have been shown to be involved in regulating the response of plants to environmental stress (Schena and Davis, 1992). TFs control gene expression by binding to genomic DNA in a sequence-specific manner.

Expression of genes of the HD-Zip subfamily I is regulated by external factors such as drought, extreme temperatures, osmotic stresses, and light conditions (Ariel et al., 2007; Chan, 2009). As such their expression levels under optimal growing conditions can be either non existing or extremely low (Suárez-López et al., 2001).

The *HaHB4* (*Helianthus annuus* homeobox 4) gene is a member of the HD-Zip sub-family I, coding for the sunflower transcription factor HAHB4 (see Section A.3(b)). The HAHB4 expression level in soybean event IND-ØØ41Ø-5 proved to be too low to be measured using Western blot or ELISA methodologies. Therefore, a specific targeted LC-MS method based on HAHB4-specific proteotypic peptides was developed and validated using recombinant HAHB4 (Supplement Report **HB4 Soybean_Report #01010271-Ev2-HAHB4 Protein Detection**).

Seed and leaf samples from the IND-ØØ41Ø-5 soybean event and the Williams 82 conventional comparator were collected from field trials performed during the 2013 season in the United States and the 2012-2013 growing season in Argentina. Briefly, field trials were established at each site following a completely randomized block design that include four replicates at each location. Leaf samples were collected at the BBCH 71-75 stage (Meier, 2001) from at least three soybean plants located in the interior rows of each plot.

The levels of HAHB4 in tissues of soybean event IND-ØØ41Ø-5 were determined by the absolute quantification (AQUA) method of protein quantification by targeted LC-MS/MS (Gerber et al., 2003). The proteotypic peptides were detected and quantified using stable isotope labelled peptide standards. Stable isotope-labelled peptides were used as internal standards and spiked into the sample to accurately quantify the endogenous levels of transgenic protein. This workflow is like other targeted proteomic workflows for the identification of biomarkers and low-level endogenous proteins in complex matrices (Fortin et al., 2009; Yocum and Chinnaiyan, 2009).

The analytical method used to quantify expression of transgenic HAHB4 protein was validated using rHAHB4 as an analytical reference standard to fortify the Williams 82 soybean tissue and protein samples prior to analysis.

LC-MS/MS analyses of HAHB4 protein levels from leaf and seed tissue samples collected from each plot from six field sites in Argentina and five sites in the US were carried out on 70 µg protein equivalents. The limits of detection (LOD) for these LTQ-MS analyses of 70 µg protein samples were 0.026 and 0.027 µg/g DW for seed and leaf tissue, respectively. Despite the high sensitivity of this LTQ-MS method, no HAHB4 protein was detected in any of the analysed samples. Therefore, a more sensitive method was developed using a triple-quadrupole MS.

Due to the lack of detection of HAHB4 protein in the 70 µg protein samples, processing of leaf and seed tissue samples was scaled up 6-fold for all six of the field sites from Argentina and for all five of the field sites in the US. For these analyses, protein samples from each of the four plots for each field site were pooled before loading an SDS-PAGE gel with 420 µg of protein (six lanes of 70 µg each). To further increase sensitivity, these scaled-up samples were analysed on a Thermo Fisher Scientific TSQ Vantage triple-quadrupole mass spectrometer. The LOD for these scaled-up samples was enhanced to 0.007 µg/g DW and 0.003 µg/g DW for seed and leaf tissue, respectively. The LLOQ for these increased scale analyses on the TSQ MS was determined to be 3 fmol per 420 µg of both seed and leaf protein, equivalent to 0.027 µg/g DW seed and 0.041 µg/g DW leaf tissue.

Among all the samples analysed, only two leaf extracts showed a signal above the LOD but, even in these cases, the amount of HAHB4 was below the LLOQ. The samples were leaf tissue from sites in the US (Ladoga, Indiana and Pemberton, Ohio) that contained 5 and 4 ng/g DW, respectively.

Conclusion of HAHB4 protein detection

HAHB4 protein expression in the transgenic soybean event IND-ØØ41Ø-5 analysed using a specific and sensitive LC-MS/MS method was found to be extremely low, consistent with expected levels for TFs. Even with the use of a highly sensitive method, HAHB4 was only detected in two of the field samples.

PAT Expression in HB4 Soybean

The safety of PAT proteins has been well established. They are widely consumed since the very beginning of the development of genetically modified crops and shown not to raise concerns from a food/feed safety perspective (Hérouet *et al*, 2005; ILSI, 2016). The PAT and HAHB4 proteins have been previously assessed and deemed safe by FSANZ in other crops (e.g., Soybean: A481, A1046, A1073, A1081; Canola: A372, A1140; Maize: A375, A380, A385, A386, A446, A543, A1106, A1116, A1118, A1192; Cotton: A518, A533, A1028, A1040, A1080; Rice A589 and wheat A1232).

The determination of the levels of PAT protein in soybean event IND-ØØ41Ø-5 was determined using a commercially available ELISA kit, which was turned quantitative by the addition of a standard curve of recombinant PAT protein (See Supplement Report **HB4 Soybean_Report#01010297-Ev3-PAT protein detection)**.

Leaf/tissue samples (as described for HAHB4 quantification) were collected from multiple sites across the USA and Argentina. Measurements of PAT levels were performed using a commercially available ELISA kit (Cat. No. AP 013 NW V10) from Envirologix (Portland, Maine, USA). PAT protein was detected in IND-ØØ41Ø-5 soybean seeds and leaves (Table 6). There was no measurable PAT protein in any of the non-transgenic Williams 82 control samples, except for a few that showed faint background signal. The highest value measured in the IND-ØØ41Ø-5 seed samples was 69 µg/g FW at Cordoba in Argentina and 12.68 µg/g FW in leaves at Pemberton Ohio in the US. These values are like those previously reported for PAT protein in glufosinate-tolerant soybeans or other GM crops, such as 127 µg/g FW (cotton seed) and 935 µg/g FW (corn

leaf) (CERA, 2011). Values varied across the different locations due to slight differences in weather conditions as described for other PAT expressing crop plants (de Block et al., 1987; CERA, 2011). PAT protein was detected in IND-ØØ41Ø-5 seeds and leaves, but not in Williams 82 samples, as expected.

Table 6. PAT protein levels in leaf and seed from HB4 soybean field trials

Site[1]	IND-ØØ41Ø-5		Williams 82	
	Leaf	Seed	Leaf	Seed
US Sites				
IL3	10.19 ± 0.39 (9.30 – 11.01)	48.46 ± 1.82 (43.39 - 51.87)	0	0
IN	12.14 ± 0.14 (11.90 – 12.53)	46.47 ± 6.25 (27.76 - 53.76)	0.03 ± 0.02 (0 – 0.10)	0
OH2	12.68 ± 0.93 (11.21 ± 15.37)	58.68 ± 2.69 (53.66 - 63.79)	0	0
IA	8.87 ± 1.09 (6.18 – 11.48)	58.31 ± 0.74 (57.33 - 60.50)	0	0
KS	9.90 ± 1.69 6.14 – 13.62)	50.80 ± 6.03 (33.75 – 59.36)	0	0
Argentina Sites				
A	7.49 ± 1.39 (4.29 – 10.50)	69.05 ± 1.07 (68.17 – 71.17)	0	0.02 ± 0.02 (0.01 – 0.04)
D2	9.51 ± 0.56 (8.64 – 11.14)	34.49 ± 1.55 30.15 – 37.15)	0	0
G1	6.72 ± 1.00 (4.90 – 9.37)	30.33 ± 1.20 (27.96 – 32.68)	0	0
Q1	5.44 ± 0.74 (3.49 – 6.62)	65.57 ± 1.54 (61.58 – 68.85)	0	0.01 ± 0.00 (0 – 0.01)
Q2	7.46 ± 1.61 (3.77 – 11.11)	68.70 ± 1.35 (64.74 – 70.66)	0	0.03 ± 0.01 (0.01 – 0.07)
W1	7.74 ± 0.65 (6.00 – 9.18)	23.00 ± 2.83 (17.57 – 28.91)	0	0.01 ± 0.01 (0 – 0.06)

[1]The field locations in the United States were Effingham, IL (IL3); Ladoga, IN (IN); Pemberton, OH (OH2); Richland, IA (IA); and Troy KS (KS). 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).
Values are expressed in µg/g fresh weight and represent the average results of four plots ± standard error (Range)

B.1(b) Information about prior history of human consumption of the new substances, if any, or their similarity to substances previously consumed in food.

See the relevant parts of Section B.1(a) above on history of safe use and refer to the relevant supplemental reports. Further, details on composition analysis is presented below in Section B5.

Comparison of grain and forage composition between the transgenic event and the control demonstrated that the levels of most of the nutrients, micronutrients, anti-nutrients, and other bioactive compounds were similar. In the few cases in which there were statistically significant differences between the event and the control, levels measured in IND-ØØ41Ø-5 soybean were either within the range of the reference varieties and/or the values reported in the literature, revealing that these differences were within the natural compositional variability of soybean. When analysed within the context of the natural variability provided by the commercial varieties cultivated along in the test sites and the range of values reported in the literature, it can be concluded that the transgenic event IND-ØØ41Ø-5 is compositionally equivalent to conventional soybean. As such, the substantial equivalence of soybean event IND-ØØ41Ø-5 composition coupled with the low levels of protein expression support the claim that exposure of humans and livestock to HAHB4 and PAT from HB4 soybean is negligible.

B.1(c) information on whether any new protein has undergone any unexpected post-translational modification in the new host

Glycosylation of proteins has been suggested as a distinguishing structural feature of allergenic proteins (Altmann 2007). Post-translational modifications (PTMs) to HAHB4 cannot be directly evaluated as protein expression levels are below the limit of detection. Further, the structure of the HAHB4 protein from soybean event IND-ØØ41Ø-5 was searched for the signal sequence required for transport to the endoplasmic reticulum, a pre-requisite for glycosylation (Pattison and Amtmann et al., 2009) and other glycosylation sites. No such signal peptides were found in HAHB4 using the public algorithms SignalP-5.0 (Almagro Armenteros et al., 2019a), TargetP-2.0 (Almegro Armenteros et al., 2019b) and Predotar v1.3 (Small et al., 2004).

Additionally, glycosylation-acceptor sites were assessed using EnsembleGly software (Caragea et al., 2007; Gomord et al., 2010) and SPRINT-Gly (Taherzadeh et al., 2019). No consensus sequences for glycosylation were found.

The absence of both signal sequences for transport to ER and glycosylation acceptor sites suggests that glycosylation in HAHB4 from soybean event IND-ØØ41Ø-5 is unlikely.

B.1(d) where any ORFs have been identified (in subparagraph A.3(c)(v) of this Guideline (3.5.1)), bioinformatics analyses to indicate the potential for allergenicity and toxicity of the ORFs

The sequence of the insertion and 200 bp of the flanking sequences in soybean event IND-ØØ41Ø-5 was analysed in search of any putative open reading frames (ORF). An ORF was defined as a contiguous, ≥8 amino acid sequence between start- and subsequent in-frame stop-codons. Nucleotide sequences were translated in three reading frames from two directions. All six reading frames within the IND-ØØ41Ø-5 insert and flanking regions were analysed for ORFs. Further details, including the sequence and location of the hypothetical peptides along the inserts and the procedures followed to carry out the different bioinformatic analysis are provided in the Supplement Report **HB4 Soybean_Report#01010291-Ev6_BioinformaticAnalysis**.

This analysis identified 74 putative peptides including the two new expression products, HAHB4 and PAT (Figure 15). A total of 26 ORFs were ≥30 amino acids with only seven of 100 or more amino acids and the largest 187 amino acids.

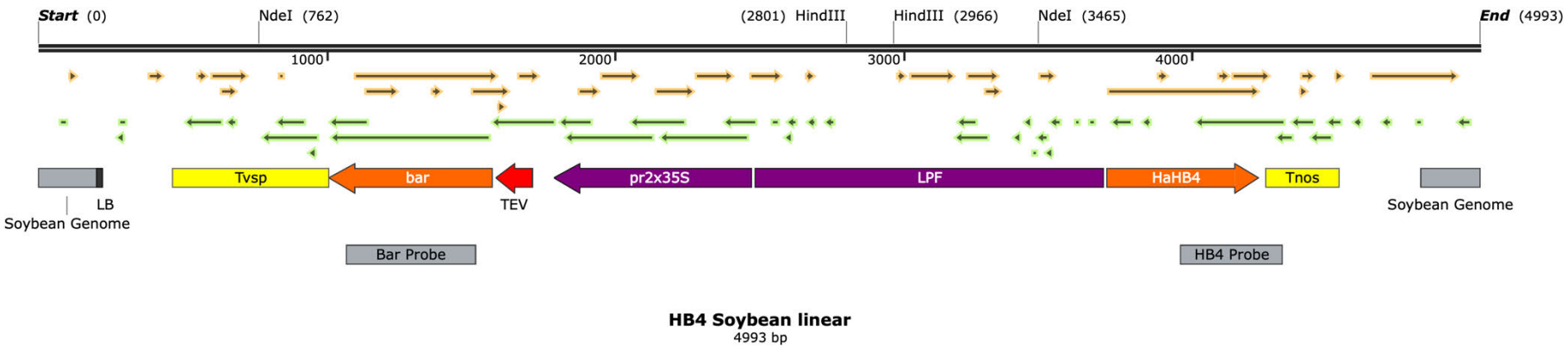


Figure 15. Open reading frames identified for the event IND-ØØ41Ø-5 insertion

ORFs were analysed to determine homology between known toxins or allergens and the hypothetical peptides that might be generated from these ORFs. A summary of the methods used to identify ORF sequences and evaluate the sequences against known allergens or toxins is provided in Table 7.

Table 7. Overview of analyses using bioinformatics

Analysis	Purpose	Approach
Start-to-stop ORF Analysis	Identify all open reading frames associated with the IND-ØØ41Ø-5 soybean insert, including flanking sequences (200 bp).	Python script: systematically identify all ORFs (≥8 amino acids) located between a start codon and a stop codon where all six reading frames are considered (Cock et al 2009).
Allergenicity Analysis	Confirm that known allergenic sequences have not been generated by the genetic modification.	AllergenOnline (FASTA Search): identify any small regions of identity or larger regions of homology between ORFs and known allergens. Structural homology with allegens was also tested with SDAP
Toxicity Analysis	Confirm that sequences similar to known toxins have not been generated by the genetic modification.	BLAST (blastp) search: identify any ORFs with homology to proteins with "toxin" in its NCBI annotation. TADB2.0 (http://bioinfo-mml.sjtu.edu.cn/TADB2/) and T3DB (http://www.t3db.ca) were also searched.

The search for homologies between the sequence of interest and those associated to known proteins was done using BLASTp (version 2.10.0+; Altschul et al. 1990) and default parameters, against the National Center for Biotechnology Information (NCBI) non-redundant protein database (updated in April, 2021) (http://blast.ncbi.nlm.nih.gov). This was repeated during drafting of this submission with version 2.13.0+ (updated March 2022) with no differences identified.

Similarity to antinutrient proteins was examined looking for related terms in the BLASTp search results (antiamylase, amylase inhibitor, Kunitz, enzyme inhibitor, lectin, lipase inhibitor, trypsin inhibitor, pepsin inhibitor).

No relevant homology was found between these putative peptides to any allergenic or toxic sequence indicating that none of the hypothetical translation products derived from IND-ØØ41Ø-5 soybean pose any safety concern.

Allergenicity Searches

Allergenicity potential was evaluated using the public, allergen-specific search engine (http://www.allergenonline.org/databasefasta.shtml) available through the Food Allergy Research and Resource Program (FARRP) at the University of Nebraska. All searches were performed using the most current database (version 21; February 14, 2021). Version 21 contains 2233 protein (amino acid) sequence entries that are categorised into 913 taxonomic-protein groups of unique proven or putative allergens (food, airway, venom/salivary and contact.

In accordance with the globally recognised regulatory recommendations (FAO/WHO, 2001, Codex Alimentarius, 2003), the homology of every ORF was analysed in fragments of 80 amino acids (Sliding 80mer window option) or identity of allergenic epitopes (8mer Exact Match).

To proceed with structural similarity analysis, the option "FASTA Search in SDAP" was selected on the Structural Database of Allergenic Proteins page (https://fermi.utmb.edu/SDAP/sdap_fas.html).

The analysis of potential allergenicity using the AllergenOnline database and tools confirmed no relevant homologies between the primary structure of the different amino acid sequences analysed with known allergens. This includes the absence of homology greater than or equal to 35% in 80 amino acids successive segments, as well as the absence of shared identity with allergenic epitopes when analysing successive peptides of 8 contiguous amino acids.

To detect a putative structural similarity with allergens, the SDAP database (Ivanciuc et al., 2002 and 2003) was used. No significant homology was found for any of the analysed sequences.

Toxicity Searches

To analyse potential toxicity of putative peptides, their homologous proteins, obtained from the alignment with the NCBI entries, was examined to search for the presence of any known toxin. Also, similarity between putative peptides and the toxins grouped in TADB2 and T3DB databases was evaluated. No homology was found with known toxins.

There was only one exception related to PAT-associated peptides, which presented some new homologies with toxins (not existing in previous bioinformatic studies) (Figure 16). The homology found is related to the presence of a common N-acetyltransferase (NAT) domain, present in PAT and in a novel family of proteins belonging to the type II toxin-antitoxin systems having a GNAT (GCN5-related NAT)-fold (Jurenas et al., 2017). This toxin-antitoxin system was initially discovered in plasmids and its function is associated to plasmid maintenance in the growing bacteria population (Jurenas et al., 2017).

Bioperl Reformatted HTML of BLASTP Search Report for PP_42

Gish, W. (1996-2006) http://blast.wustl.edu **Query=** PP_42 (Length: 229)

Database: TADB2_aa.fas

12,714 sequences; 1,560,970 total letters

Sequences producing significant alignments:	**Score (bits)**	**E value**
TADB|T5298 gi|194291114|ref|YP_002007021.1| phosphinothricin N-acety...	97.6	3.4e-24
TADB|T2094 gi|78065730|ref|YP_368499.1| N-acetyltransferase GCN5 [Bu...	76.2	1e-17
TADB|T4799 gi|161525329|ref|YP_001580341.1| N-acetyltransferase GCN5...	75.1	2.1e-17
TADB|T5352 gi|172060083|ref|YP_001807735.1| N-acetyltransferase GCN5...	74.4	3.4e-17
TADB|T2970 gi|115351078|ref|YP_772917.1| N-acetyltransferase GCN5 [B...	73.4	7e-17
TADB|T508 gi|17547852|ref|NP_521254.1| antibiotic resistance (acety...	70.9	3.9e-16
TADB|T838 gi|27376227|ref|NP_767756.1| phosphinothricin acetyltrans...	70.2	6.3e-16
TADB|T5113 gi|169633694|ref|YP_001707430.1| phosphinothricin N-acety...	67.7	3.5e-15
TADB|T5122 gi|169796191|ref|YP_001713984.1| phosphinothricin N-acety...	67.7	3.5e-15
TADB|T1353 gi|50084795|ref|YP_046305.1| phosphinothricin N-acetyltra...	66.7	7.3e-15
TADB|T2731 gi|107022223|ref|YP_620550.1| N-acetyltransferase GCN5 [B...	66.7	7.3e-15
TADB|T3104 gi|116689168|ref|YP_834791.1| GCN5-like N-acetyltransfera...	66.7	7.3e-15
TADB|T5233 gi|170732472|ref|YP_001764419.1| N-acetyltransferase GCN5...	64.5	3.1e-14
TADB|T1306 gi|53714970|ref|YP_100962.1| putative acetyltransferase [...	54.0	4.7e-11
TADB|T1765 gi|60682936|ref|YP_213080.1| putative acetyltransferase [...	54.0	4.7e-11

Figure 16. Toxicity Analysis with Toxin Antitoxin Database.
The complete amino acid sequence of each putative peptide was introduced into the search tool available in the "Toxin Antitoxin Database" (WU-BLAST 2.0). The result obtained for the PAT protein. Upper panel: Alignment of the PAT protein with sequences in TADB. Lower panel: List of sequences with significant (E score < 10^{-5}) homology.

Conclusion

The performed analysis allowed the identification of putative expression products that could be generated by the genetic modification introduced in soybean event IND-ØØ41Ø-5. The results of the ORF and the bioinformatic analysis included the new expressed proteins in HB4 soybean (i.e., HAHB4 and PAT).

The bioinformatic analysis demonstrated no relevant similarity between the expression products or the putative peptides and known allergens or toxins.

Some homologies were found between PAT and a novel family of proteins belonging to the type II toxin-antitoxin systems since they possess a common NAT catalytic domain. This type of domain is known to be present in proteins from many species. For example, N-acetyltransferases catalyze the transfer of an acyl moiety from acyl coenzyme A (acyl-CoA) to a diverse group of substrates and are widely distributed in all domains of life (Salah Ud-Din et al., 2016). Furthermore, PAT protein safety has been established by scientific (Herouet et al., 2005) as well as regulatory precedents (CERA, 2011; ILSI, 2016). Also, it is expressed in commercial GM crops approved in many countries (ISAAA, 2021), incorporated into glufosinate-tolerant crops since the very beginning of the GMO development (Stringam et al., 2003; CFIA, 1995). Based on the above, there is no evidence of a risk with the use of PAT protein in soybean event IND-ØØ41Ø-5.

Summary Safety assessment of HAHB4 protein

A weight-of-evidence approach using risk assessment principles was used to evaluate the safety of the HAHB4 protein.

The weight-of-evidence strongly supports HAHB4 safety:

- The prevalence of HD-Zip family of transcription factors in edible crops, including Sunflower (*Helianthus annuus*), is widespread in nature, and the HB4 protein is like proteins already present in the food supply with a history of safe consumption
- Bioinformatic analysis confirms that HAHB4 lack sequence similarity to known toxins and allergens (see above)
- Homology of HAHB4 to other proteins in plants with a history of safe use provides additional evidence that HAHB4 in soybean event IND-ØØ41Ø-5 is as safe for human consumption as HD-Zip proteins like HAHB4 in other foods; and
- The potential exposure for humans and livestock to HAHB4 is negligible.

Based on the weight-of-evidence and considering the close-to-zero risk associated to the HAHB4 protein, soybean event IND-ØØ41Ø-5 is as safe as conventional varieties for humans, livestock, and the environment.

Summary Safety Assessment of the PAT Protein

A weight-of-evidence approach using risk assessment principles was used to evaluate the safety of the PAT protein. This approach has been presented and assessed by FSANZ in numerous applications and considered all data in a comprehensive manner to evaluate the safety of PAT, including risk assessment results (potential hazard X potential exposure = potential risk).

The biosafety of PAT and HAHB4 proteins have been previously assessed and deemed safe by FSANZ in other crops (e.g., Soybean: A481, A1046, A1073, A1081; Canola: A372, A1140; Maize: A375, A380, A385, A386, A446, A543, A1106, A1116, A1118, A1192; Cotton: A518, A533, A1028, A1040, A1080; Rice A589 and wheat A1232).

B.2. New Proteins

The PAT proteins have been previously assessed by FSANZ. An updated bioinformatics comparison including the amino acid sequence of PAT expressed in soybean IND-ØØ41Ø-5 to known protein toxins, anti-nutrients and allergens is presented above in Section B1(d).

As HAHB4 has also been assessed by FSANZ previously (A1232), the following information is provided in accordance with the FSANZ Handbook.

B.2 (a) Information on potential toxicity

Details of the potential toxicity of the protein HAHB4 protein as well as other putative ORFs are presented in Supplement Report **HB4 Soybean_Report#01010291-Ev6_BioinformaticAnalysis** and the following Sections:

- Section A.2(a)(i) and
- Section B.1(d)

The bioinformatic analysis demonstrated no relevant similarity between the putative peptides and known toxins. The HAHB4 protein is from sunflower and shares homology with numerous proteins found in food plants and therefore has a history of safe use.

B.2(a)(ii) information on the stability of the protein to proteolysis in appropriate gastrointestinal model systems

Details on the stability of the HAHB4 protein are provided in Supplement Report **HB4_Report#01010273-Ev2_Protein Safety.**

The HAHB4 protein is a transcription factor and is present at extremely low concentrations in sunflower as well as in the soybean event IND-ØØ41Ø-5. Therefore, isolation of sufficient quantities of protein from HB4 soybean was not feasible. Consequently, protein stability analysis was carried out using *E. coli* produced HAHB4 (rHAHB4), which proved to be equivalent to the native protein expressed in IND-ØØ41Ø-5 soybean (see Supplemental Report **HB4_Report#01010296-Ev2_HAHB4 Protein Production**).

Recombinant HAHB4 was subjected to simulated gastric fluid (SGF) assays performed following pre-established protocols (Thomas et al., 2004). Under these conditions, rHAHB4 protein was rapidly degraded as observed by the absence of the respective protein band 0.5 min after initiation of the assay (Figure 17). These results show that the HAHB4 protein is rapidly digested by pepsin *in vitro*.

B.2(a)(iii) an animal toxicity study if the bioinformatic comparison and biochemical studies indicate either a relationship with known protein toxins/anti-nutrients or resistance to proteolysis.

The bioinformatic analyses did not indicate any relationships with known protein toxins/anti-nutrients and the protein did not show any resistance to proteolysis.

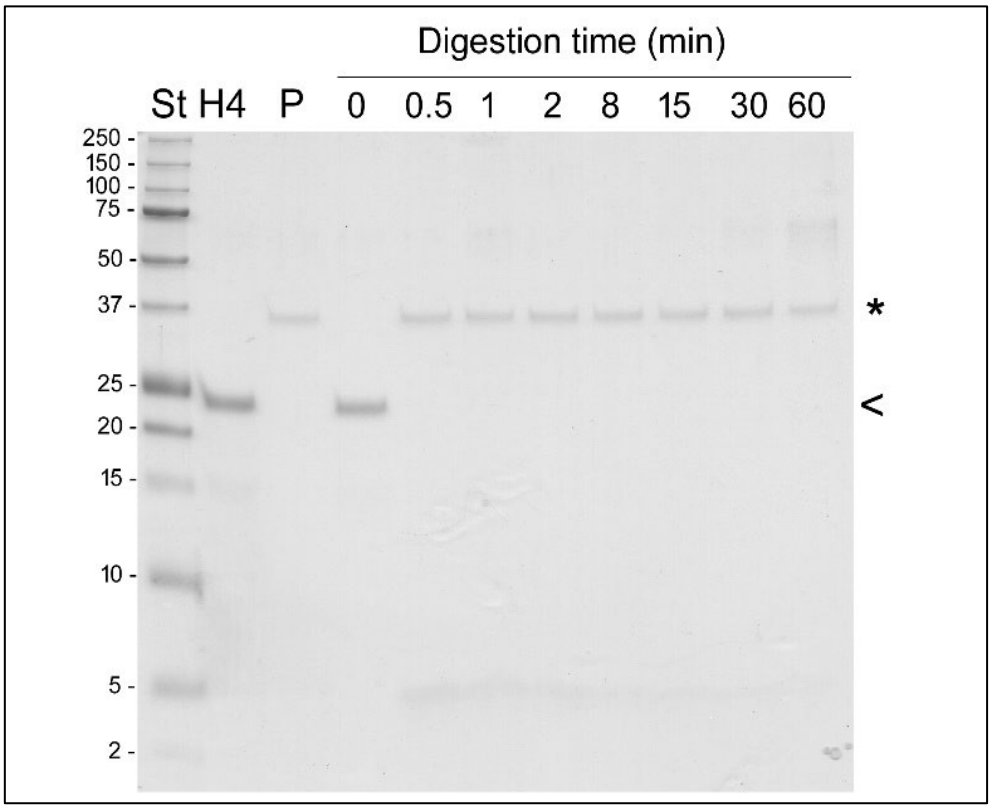


Figure 17. Digestibility of HAHB4
Recombinant HAHB4 (H4) incubated with pepsin (P) and analysed by SDS-PAGE and protein staining. <: indicates the location of the rHAHB4 protein band. *: indicates the location of the pepsin band. St indicates the molecular weight standard lane.

B.2(b) information on the potential allergenicity of any new proteins, including:

Details of the potential allergenicity of the protein HAHB4 protein as well as other putative ORFs are presented in Supplement Report **HB4 Soybean_Report#01010291-Ev6_BioinformaticAnalysis**and the following Sections:

- Section A.2(a)(i) and
- Section B.1(d)

The bioinformatic analysis demonstrated no relevant similarity between the putative peptides and known allergens. The HAHB4 protein is from sunflower and shares homology with numerous proteins found in food plants and therefore has a history of safe use.

Additional information is provided below.

B.2(b)(iii) source of the new protein the new protein's structural properties, including, but not limited to, its susceptibility to enzymatic degradation (e.g. proteolysis), heat and/or acid stability

Details on the thermal stability of the HAHB4 protein are provided in Supplement Report **HB4_Report#01010273-Ev2_Protein Safety.**

A sample of rHAHB4 protein was incubated at different temperatures (60, 75 or 90 °C) for up to 60 min. Aliquots were taken after 10, 30 and 60 min of incubation and analysed by SDS-PAGE followed by protein staining (1.2 µg/lane) and ELISA. Results indicate that rHAHB4 integrity is not affected by heating. Incubation

at 90 °C produced a slightly lower signal than the other tested temperatures even at short incubation times, but final absorbance values at 60 min did not show a significant difference from the control incubated at room temperature (Figure 18). These results suggest that the HAHB4 protein is not significantly degraded by high temperatures.

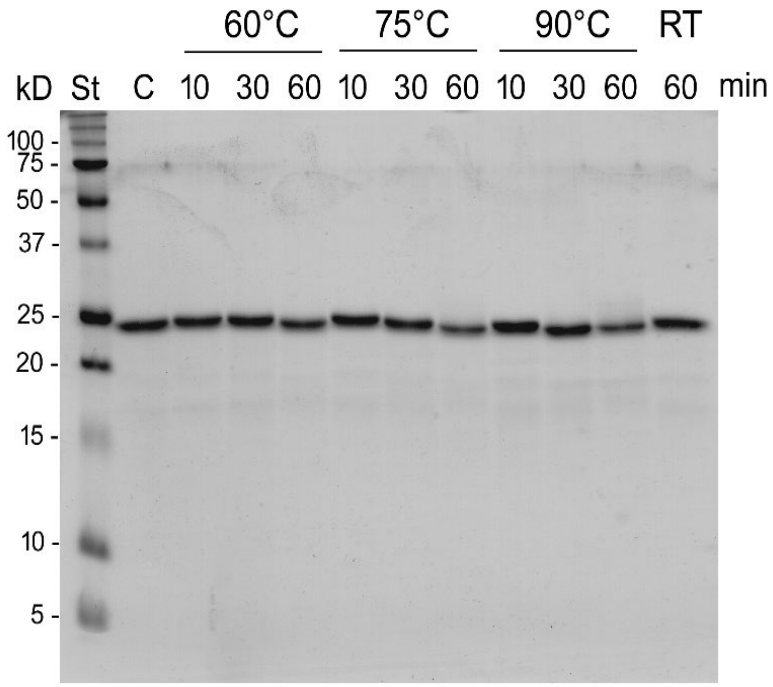


Figure 18. Effect of thermal treatment on rHAHB4 electrophoretic mobility.
rHAHB4 protein was incubated at different temperatures for up to 60 min and analysed by SDS-PAGE and protein staining. Original samples kept at 4°C (C) or room temperature (RT) were included as controls. St indicates the molecular weight standard lane.

B.2(b)(iv) specific serum screening where a new protein is derived from a source known to be allergenic or has sequence homology with a known allergen

Not applicable. The HAHB4 protein is not from a source known to be allergenic nor does it display sequence homology with known allergens.

B.2(b)(v) information on whether the new protein(s) have a role in the elicitation of gluten-sensitive enteropathy, in cases where the introduced genetic material is obtained from wheat, rye, barley, oats, or related cereal grains.

Not applicable. The HAHB4 protein is not from wheat, rye, barley, oats, or related cereal grains.

Where the new protein has been produced from an alternative source (e.g., microbial expression system) to obtain sufficient quantities for analysis, information **must** be provided to demonstrate that the protein tested is biochemically, structurally and functionally equivalent to that expressed in the food produced using gene technology.

Details of rHAHB4 are provided in Supplemental Report **HB4_Report#01010296-Ev2_HAHB4 Protein Production.** To ensure the recombinant protein was produced in *E. coli* as expected, HAHB4 was characterised for N-terminal sequence and protein mass analysis.

Protein purified from *E. coli* had the same sequence as the protein present in the IND-ØØ41Ø-5 event (Figure 19). N-terminal sequencing of HAHB4 produced from the soluble and insoluble fractions confirmed no N terminal modifications, correct N-terminal amino acid sequence for the first seven amino acids (MSLQQVT) and confirmed the polyhistidine tag had been removed. Detection of the peptides from both the soluble and insoluble fractions demonstrated 47% coverage of the HAHB4 protein with each peptide scoring a probability greater than 90% that the sequence had been correctly identified. Taken together, the algorithms in Protein Prophet assigned a 99% probability the protein was correctly identified in the samples. MALDI-TOF analysis further showed the HAHB4 produced in *E. coli* were of the expected molecular mass. Based on the collective data from LC-MS analysis, MALDI-TOF mass detection, and N-terminal sequencing, HAHB4 protein produced in *E. coli* was shown to be equivalent to the protein present in IND-ØØ41Ø-5 soybean. These data support the conclusion that HAHB4 protein is suitable for use in safety evaluations and to serve as a reliable standard for further studies.

```
HB4E.coli     MSLQQVTTTRKNRNEGRRRFTDKQISFLEYMFETQSRPELRMKHQLAHKL  50
HAHB4Soybean MSLQQVTTTRKNRNEGRRRFTDKQISFLEYMFETQSRPELRMKHQLAHKL  50

HB4E.coli     GLHPRQVAIWFQNKRARSKSRQIEQEYNALKHNYETLASKSESLKKENQA  100
HAHB4Soybean GLHPRQVAIWFQNKRARSKSRQIEQEYNALKHNYETLASKSESLKKENQA  100

HB4E.coli     LLNQLEVLRNVAEKHQEKTSSSGSGEESDDRFTNSPDVMFGQEMNVPFCD  150
HAHB4Soybean LLNQLEVLRNVAEKHQEKTSSSGSGEESDDRFTNSPDVMFGQEMNVPFCD  150

HB4E.coli     GFAYLEEGNSLLEIEEQLPDLQKWWEF 177
HAHB4Soybean GFAYLEEGNSLLEIEEQLPDLQKWWEF 177
```

Figure 19. HAHB4 and rHAHB4 protein comparison

Sequence alignment showing the translated *E. coli*-produced HAHB4 sequence (HB4E.coli) is identical to the protein sequence translated from the plasmid transformed into IND-ØØ41Ø-5 (HAHB4Soybean) and used to clone the gene sequence present in the *E. coli* expression vector pARC666.1 B8. As reported in the [Southern/sequencing data] section, the T-DNA sequence found in IND-ØØ41Ø-5.

B.3. Other (non-protein) new substances

If other (non-protein) substances are produced as a result of the introduced DNA, information must be provided on the following:

B.3(a) the identity and biological function of the substance

B.3(b) whether the substance has previously been safely consumed in food

B.3(c) potential dietary exposure to the substance

Only two proteins are added to the HB4 soybean (HAHB4 and PAT). The HAHB4 protein belongs to a large class of TFs unique to plants, which are associated to plant stress-response pathways. Therefore, being a component of the plant natural physiological response, no new proteins, or metabolites other than the natural ones would be expected to arise from its activity. The PAT protein has been used extensively to provide herbicide tolerance.

B.3(d)(i) where RNA interference has been used: the role of any endogenous target gene and any changes to the food as a result of silencing that gene

Not applicable to this submission.

B.3(d)(ii) where RNA interference has been used: the expression levels of the RNA transcript

Not applicable to this submission.

B.3(d)(iii) where RNA interference has been used: the specificity of the RNA interference

Not applicable to this submission.

B.4. Novel herbicide metabolites in GM herbicide tolerant plants

The identity and levels of herbicide and any novel metabolites that may be present in the food produced using gene technology.

If novel metabolites are present then the application should address the following, where appropriate:

(a) toxicokinetics and metabolism

(b) acute toxicity

(c) short-term toxicity

(d) long-term toxicity and carcinogenicity

(e) reproductive and developmental toxicity

(f) genotoxicity.

The PAT enzyme is not anticipated to function within HB4 soybean any differently to the way that it functions within a range of other crops containing the PAT enzymes and previously assessed by FSANZ (e.g., Soybean (A481, A1046, A1073, A1081); Canola (A372, A1140); Maize (A375, A380, A385, A386, A446, A543, A1106, A1116, A1118, A1192); Cotton, A518, A533, A1028, A1040, A1080); Rice (A589); and wheat (A1232). Specifically, no novel metabolites would be expected to be formed and therefore, glufosinate-ammonium metabolism studies submitted to FSANZ previously in association with other crops are expected to sufficiently describe the metabolism of glufosinate-ammonium in HB4 soybean.

B.5 Compositional analyses of the food produced using gene technology

This must include all of the following:

B.5(a) the levels of relevant key nutrients, toxicants and anti-nutrients in the food produced using gene technology compared with the levels in an appropriate comparator (usually the non-GM counterpart). A statistical analysis of the data must be provided.

B.5(b) information on the range of natural variation for each constituent measured to allow for assessment of biological significance should any statistically significant differences be identified

B.5(c) the levels of any other constituents that may potentially be influenced by the genetic modification, as a result, for example, of downstream metabolic effects, compared with the levels in an appropriate comparator as well as the range of natural variation.

> ***In the case of herbicide-tolerant plants, the levels of each constituent in the food produced using gene technology must be determined using plants sprayed with the herbicide.***

Verifying the compositional equivalence between genetically modified crops and their non-transgenic counterparts has been a main component in the safety evaluation of GM crops (Kuiper et al., 2001; Privalle et al., 2013). Following the outline of the OECD Revised Consensus Document on Compositional Considerations for New Varieties of Soybean (OECD, 2012). Details from compositional analysis of HB4 Soybean have been published by Chiozza, Burachik and Miranda (2020). A summary of compositional analysis is presented below.

Soybean field trials were conducted in Argentina and the United States during the 2012 and 2013 growing seasons. Six locations in Argentina (in the Provinces of Buenos Aires, Córdoba, Entre Ríos, and Santa Fe) and five locations in the US (in the States of Illinois, Indiana, Iowa, Kansas, and Ohio) were chosen representing major soybean production areas and covering a diversity of environmental conditions. A randomised complete block design with four replicate blocks was used in each trial. Entries were soybean event IND-ØØ41Ø-5, the near isogenic control variety Williams 82, and a set of commercial reference varieties used by farmers and adapted for each location. These local varieties were used to estimate the natural compositional variability for the crop, giving the appropriate context for the interpretation of the experimental results in terms of their biological significance.

Compositional analyses were conducted following the OECD Revised Consensus Document on Compositional Considerations for New Varieties of Soybean (OECD, 2012). Nutrients and micronutrients measured in grain (a total of 36 analytes) included proximates (moisture, protein, fat, ash, and carbohydrates), fibre (crude fibre, acid detergent fibre, ADF, and neutral detergent fibre, NDF), minerals (calcium and phosphorous), main fatty acids profile, vitamins (E and K1), and amino acid composition. Nutrients measured in forage (a total of 9 analytes) included proximates, fibre (ADF and NDF), and minerals (calcium and phosphorous). Anti-nutrients and other bioactive compounds measured in grain (8 in total) included isoflavones (daidzein, genistein, and glycitein), stachyose, raffinose, phytic acid, lectin, and trypsin inhibitors.

Comparison of grain contents of proximates, ADF, NDF, crude fibre, minerals and vitamins has shown only one (Vitamin K1) statistically significant difference between the soybean event IND-ØØ41Ø-5 and the near-isogenic control variety Williams 82 (Table 8). However, the value for the event was within the range reported in the literature (OECD 2012; ILSI 2019). Significant differences were not found between the event IND-ØØ41Ø-5 and the control soybean for any of the six fatty acids measured (Table 9).

Table 8. Proximates, fibre, minerals and vitamins of soybean grain

Component[a]	IND-ØØ41Ø-5 (Range)	Williams 82 (Range)	Commercial Reference Range[b]	Literature Range[c]
Ash	5.69 ± 0.05 (5.20–6.36)	5.68 ± 0.05 (5.03–6.42)	4.83–6.35	3.9–7.0
Carbohydrates	35.84 ± 0.35 (32.27–41.52)	35.19 ± 0.38 (32.16–40.97)	31.46–38.11	29.6–50.2
Moisture	9.46 ± 0.18 (7.05–11.6)	9.28 ± 0.18 (7.41–11.6)	7.78–11.82	4.7–34.4
Protein	39.03 ± 0.30 (34.58–42.74)	39.78 ± 0.24 (36.49–43.93)	36.60–43.10	33.2–45.5
Total Fat	19.98 ± 0.19 (17.55–21.80)	19.56 ± 0.28 (15.90–22.48)	16.60–21.64	8.1–23.6
Acid Detergent Fibre	12.51 ± 0.40 (6.69–16.0)	12.99 ± 0.34 (9.18–18.3)	10.50–17.77	7.8–18.6
Neutral Detergent Fibre	16.88 ± 0.27 (14.30–21.23)	16.83 ± 0.23 (13.80–21.16)	14.10–18.07	8.5–21.3
Crude Fibre	7.35 ± 0.44 (3.21–12.50)	7.74 ± 0.40 (4.66–13.20)	4.61–13.60	4.12–18.5[d]
Phosphorus	0.56 ± 0.01 (0.35–0.69)	0.57 ± 0.01 (0.38–0.68)	0.36–0.61	0.50–0.94
Calcium	0.26 ± 0.01 (0.20–0.37)	0.25 ± 0.01 (0.18–0.35)	0.20–0.31	0.12–0.31
Vitamin E	1.87 ± 0.06 (0.11–2.78)	1.81 ± 0.07 (0.95–2.93)	1.37–3.13	0.19–6.17
Vitamin K1	0.38 ± 0.02 * (0.31–0.91)	0.43 ± 0.02 (0.31–0.61)	0.44–0.85	0.06–1.76[d]

Numbers represent mean ± standard error of 44 values measured in samples from field trials developed during 2012–2013 in 11 different locations, except for vitamin K1, which was only measured in the 20 samples from the 5 US trials.
a: Results are expressed as % dry weight, except for moisture (% fresh weight), vitamins E (mg/100 gr dwt) and K1 (mg/kg); b: Values measured in commercial varieties grown in the same trials; c: ILSI values within OECD 2012, unless otherwise indicated; d: ILSI Crop Composition database V7.0.34; *Significant difference ($p < 0.05$).

Table 9. Fatty acid profile of soybean grain

Component[a]	IND-ØØ41Ø-5 (Range)	Williams 82 (Range)	Commercial Reference Range[b]	Literature Range[c]
Palmitic acid	2.17 ± 0.03 (1.82–2.88)	2.12 ± 0.03 (1.74–2.61)	1.76–2.52	0.67–2.78
Stearic acid	0.82 ± 0.01 (0.68–1.02)	0.84 ± 0.01 (0.62–1.06)	0.61–1.15	0.28–1.13
Oleic acid	4.31 ± 0.07 (3.35–5.25)	4.46 ± 0.10 (3.03–5.37)	2.86–5.52	1.36–6.56
Linoleic acid	10.85 ± 0.10 (9.51–12.40)	10.43 ± 0.14 (8.62–12.31)	8.33–11.72	3.46–13.36
Linolenic acid	1.42 ± 0.02 (1.14–1.87)	1.37 ± 0.02 (1.04–1.69)	1.20–1.66	0.30–2.19
Arachidic acid	0.06 ± 0.00 (0.04–0.09)	0.06 ± 0.00 (0.03–0.11)	0.03–0.07	0.02–0.11

Numbers represent mean ± standard error of 44 values measured in samples from field trials developed during 2012–2013 in 11 different locations.
a: Results are expressed as % dry weight; b: Values measured in commercial varieties grown in the same trials; c: ILSI values within OECD 2012

Analysis of the amino acids profile has shown only one (cysteine) statistically significant difference between the event IND-ØØ41Ø5 and the non-transgenic control line Williams 82 (Table 10). However, the value for the event fell within the range provided by both the reference varieties and the literature (OECD 2012; ILSI 2019).

Table 10. Amino acid composition of soybean grain

Component[a]	IND-ØØ41Ø-5 (Range)	Williams 82 (Range)	Commercial Reference Range[b]	Literature Range[c]
Alanine	1.85 ± 0.02 (1.57 – 2.14)	1.86 ± 0.02 (1.65 – 2.17)	1.63 – 2.10	1.51 – 2.10
Arginine	2.83 ± 0.04 (2.43 – 3.22)	2.96 ± 0.02 (2.57 – 3.19)	2.67 – 3.27	2.28 – 3.4
Aspartic Acid	4.51 ± 0.04 (4.04 – 5.07)	4.56 ± 0.04 (4.04 – 5.08)	4.17 – 4.91	3.81 – 5.12
Cysteine	0.55 ± 0.01* (0.43 – 0.66)	0.59 ± 0.01 (0.52 – 0.67)	0.49 – 0.62	0.37 – 0.81
Glycine	1.73 ± 0.01 (1.59 – 2.00)	1.69 ± 0.01 (1.53 – 1.83)	6.28 – 7.41	1.46 – 1.99
Glutamic Acid	7.00 ± 0.05 (6.33 – 7.72)	6.97 ± 0.05 (6.20 – 7.56)	1.58 – 1.79	5.84 – 8.20
Histidine	1.00 ± 0.01 (0.85 – 1.19)	1.01 ± 0.01 (0.84 – 1.13)	0.90 – 1.10	0.87 – 1.17
Isoleucine	1.77 ± 0.02 (1.50 – 1.95)	1.83 ± 0.01 (1.63 – 1.95)	1.60 – 1.87	1.53 – 2.07
Leucine	3.02 ± 0.02 (2.62 – 3.30)	3.02 ± 0.02 (2.70 – 3.26)	2.80 – 3.15	2.59 – 3.62
Lysine	2.20 ± 0.05 (1.68 – 2.60)	2.33 ± 0.02 (1.99 – 2.61)	2.14 – 2.59	2.28 – 2.83
Methionine	0.52 ± 0.01 (0.44 – 0.60)	0.52 ± 0.00 (0.44 – 0.57)	0.45 – 0.55	0.43 – 0.68
Phenylalanine	1.96 ± 0.02 (1.56 – 2.18)	1.99 ± 0.02 (1.67 – 2.19)	1.77 – 2.20	1.63 – 2.34
Proline	2.00 ± 0.02 (1.69 – 2.37)	2.01 ± 0.02 (1.70 – 2.30)	1.85 – 2.29	1.68 – 2.28
Serine	1.94 ± 0.02 (1.64 – 2.33)	2.03 ± 0.01 (1.78 – 2.24)	1.80 – 2.19	1.10 – 2.48
Threonine	1.54 ± 0.02 (1.30 – 1.69)	1.50 ± 0.02 (1.34 – 1.71)	1.35 – 1.64	1.14 – 1.86
Tryptophan	0.50 ± 0.01 (0.34 – 0.61)	0.52 ± 0.01 (0.40 – 0.62)	0.41 – 0.60	0.36 – 0.50
Tyrosine	1.38 ± 0.03 (1.04 – 1.63)	1.39 ± 0.03 (1.03 – 1.62)	1.03 – 1.61	1.01 – 1.61
Valine	1.83 ± 0.02 (1.60 – 2.07)	1.88 ± 0.01 (1.69 – 2.17)	1.71 – 2.10	1.59 – 2.20

Numbers represent mean ± standard error of 44 values measured in samples from field trials developed during 2012–2013 in 11 different locations.

a: Results are expressed as % dry weight; b: Values measured in commercial varieties grown in the same trials; c: ILSI values within OECD 2012; *Significant difference ($p < 0.05$).

Data of the levels of anti-nutrients and other bioactive components showed five significant differences between soybean IND-ØØ41Ø-5 and Williams 82. These include phytic acid, stachyose, and the three isoflavones (Table 11). However, the values of all these analytes in IND-ØØ41Ø-5 soybean were within the range of the commercial reference varieties (Table 11).

No significant differences between IND-ØØ41Ø-5 and Williams 82 were found for the levels of any of the 9 analytes measured in forage (Table 12).

Table 11. Anti-nutrients and isoflavones composition of soybean grain

Component[a]	IND-ØØ41Ø-5 (Range)	Williams 82 (Range)	Commercial Reference Range[b]	Literature Range[c]
Phytic acid	1.67 ± 0.10 * (0.62 – 3.09)	1.35 ± 0.04 (0.68 – 1.88)	0.54 – 1.69	0.63 – 1.96
Lectins (mg/g)	4.78 ± 0.13 (2.43 – 6.34)	4.73 ± 0.15 (3.02 – 7.03)	1.29 – 6.09	0.11 – 9.04
Raffinose	0.88 ± 0.03 (0.55 – 1.39)	0.85 ± 0.02 (0.70 – 1.09)	0.64 – 1.2	0.21 – 0.66
Stachyose	3.77 ± 0.07 * (2.50 – 4.85)	3.39 ± 0.08 (2.27 – 4.32)	2.56 – 4.76	1.21 – 3.50
Trypsin Inhibitor (TIU/mg dwt)	35.04 ± 1.78 (18.60 – 60.30)	33.46 ± 1.60 (19.30 – 56.10)	18.6 – 56.1	19.59 – 118.68
Daidzein	1240 ± 53.0 * (497 – 1870)	1086 ± 48.0 (462 – 1700)	533 – 2150	60.0 – 2453.5
Genistein	1402 ± 64.0 * (518– 2130)	1282 ± 61.0 (515 – 2060)	671 – 2290	144.3 – 2837.2
Glycitein	276 ± 11.0 * (133 – 412)	239 ± 8.0 (123 – 344)	126 – 344	15.3 – 310.4

Numbers represent mean ± standard error of 44 values measured in samples from field trials developed during 2012–2013 in 11 different locations.

a: Results are expressed as % dry weight, except for lectins (mg/g), Trypsin Inhibitor units (TIU/mg dwt) and isoflavones (ppm dwt); b: Values measured in commercial varieties grown in the same trials; c: ILSI values within OECD 2012; *Significant difference ($p < 0.05$).

Table 12. Proximates, fibre and minerals of soybean forage

Component[a]	IND-ØØ41Ø-5 (Range)	Williams 82 (Range)	Commercial Reference Range[b]	Literature Range[c]
Ash	9.12 (0.28) (6.43 – 15.90)	9.04 (0.37) (6.50 – 20.50)	6.96 – 19.10	6.71 - 1078
Carbohydrates	49.76 (2.69) (30.43 – 75.40)	50.54 (2.67) (30.72 – 77.30)	32.40 – 73.90	27.8 - 80.6
Moisture (% fwt)	76.72 (0.72) (65.31 – 85.50)	76.94 (0.67) (65.32 – 85.60)	64.30 – 84.20	73.5 – 81.6[d]
Protein	20.85 (0.44) (14.80 – 26.90)	20.68 (0.48) (13.70 – 29.20)	15.60 – 24.70	14.37 – 24.71
Total Fat	2.46 (0.13) (1.15 – 4.70)	2.47 (0.12) (1.32 – 4.33)	1.38– 3.48	1.30 – 5.13
Acid Detergent Fiber	33.08 (0.65) (24.50 – 42.50)	33.17 (0.55) (27.30 – 41.20)	20.30 – 36.78	12.85 – 64.10[d]
Neutral Detergent Fiber	41.64 (1.01) (29.50 – 52.40)	41.87 (0.98) (26.30 – 53.70)	25.60 – 52.30	19.26 – 82.00[d]
Phosphorus	0.25 (0.01) (0.20 – 0.35)	0.26 (0.01) (0.18 – 0.35)	0.21 – 0.37	N/A
Calcium	1.21 (0.02) (1.03 – 1.56)	1.26 (0.03) (0.96 – 1.58)	0.97 – 1.51	N/A

Numbers represent mean ± standard error of 44 values measured in samples from field trials developed during 2012–2013 in 11 different locations.
a: Results are expressed as % dry weight, b: Values measured in commercial varieties grown in the same trials; c: ILSI values within OECD 2012 unless otherwise stated; d: ILSI Crop Composition database V7.0.34; N/A data not available.

Quantitative analysis of soy allergens

Quantitative comparison of known soy allergen levels in soybean event IND-ØØ41Ø-5 and its parental comparator variety, Williams 82, were performed to assess the potential for effects on endogenous allergen expression (see Supplement Report **HB4 Soybean_Report #01010298-Ev2-Endogenous Allergen Levels**). The soy allergens tested, and the analysis method reflected the guidance listed in the most current EFSA consensus document (EFSA, 2017). The quantitative analysis was conducted following the Good Laboratory Practice (GLP) Standards, Code of Federal Regulations, Title 40 Part 160 and the United States Environmental Protection Agency Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA). The levels of eleven soy allergens in seeds collected from eight field trial sites were determined.

The only allergen (Gly m Bd 28K) that was significantly different between HB4 and Williams 82 soy at multiple field sites was lower in HB4 than Williams 82. Levels of Gly m Bd 28K were also within the range found in co-harvested commercial comparators.

These analyses support the conclusion that HB4 soybean does not pose any increased allergenic potential to humans or animals. The novel trait introduced by HB4 soybean does not alter its potential for allergenicity compared to other commercial soybeans varieties.

Conclusions from compositional analysis

In summary, the nutrient and anti-nutrient contents in grain and forage from the soybean event IND-ØØ41Ø-5 were found to be equivalent to those measured in the non-transgenic parental line and like the levels displayed by commercial soybean reference varieties planted in the same locations, and comparable to the values reported in the literature. These results confirm that the transgenic event IND-ØØ41Ø-5 is compositionally equivalent to conventional soybean.

C. Information related to the nutritional impact of the genetically modified food

Soybean has a long history of safe use. Global production in 2020 was more than 350 million tonnes over 127 million ha (FAOSTAT, 2022). Most was consumed directly by humans and the remaining fed to animals.

The soybean event IND-ØØ41Ø-5 in this submission has been transformed with gene cassettes designed to express the stress tolerance gene *HaHB4* and the *bar* gene to produce the PAT protein for herbicide tolerance. The introduction of the genetic modification had no nutritional impact on the soybean. This is supported by the fact that:

- Molecular characterisation demonstrated stability of the inserts during numerous generations
- The HAHB4 protein is part of an HD-Zip 1 family found across all plants, with a history of safe consumption and no significant homology to known allergens and toxins; and
- Compositional analysis did not indicate biologically significant changes to the levels of nutrients or anti-nutrients in the event compared to its conventional counterpart. Event composition is within the normal variation of soybean varieties and is substantially equivalent to conventional soybean.

The difference between the HB4 soybean and the untransformed control, relates to low levels of the newly expressed HAHB4 protein and the PAT protein. However, the expression of these two new proteins did not alter the compositional profile. Thus, food products derived from HB4 soybean are anticipated to be nutritionally equivalent to food products derived from other commercially available soybean, except that HB4 soybean is tolerant to environmental stress and has herbicide tolerance.

D. Other Information

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

The new polypeptide produced by the insert in soybean event IND-ØØ41Ø-5 have been well characterised and are prevalent in the food chain. The proteins are non-toxic and occurs at very low levels in the transformed plant. Its safety is supported by a weight-of-evidence that indicates safety for human consumption. Considering the compositional equivalence between the soybean event and its conventional variety, and the lack of any observed phenotypic characteristics indicative of unintended effects arising from the genetic modification process, there was no plausible risk hypothesis that would indicate the need for animal feeding studies.

References Cited

1. Agalou A, Purwantomo S, Övernäs E, Johannesson H, Zhu X, Estiati A, De Kam RJ, Engström P, Slamet-Loedin IH, Zhu Z, Wang M, Xiong L, Meijer AH and Ouwerkerk PBF (2008). A genome-wide survey of HD-Zip genes in rice and analysis of drought-responsive family members. Plant Molecular Biology Springer 66, 87–103. https://doi.org/10.1007/s11103-007-9255-7
2. Almagro Armenteros JJ, Salvatore M, Winther O, Emanuelsson O, von Heijne G, Elofsson A, Nielsen H. (2019a) Detecting Sequence Signals in Targeting Peptides Using Deep Learning Life Science Alliance 2 (5), e201900429. doi:10.26508/lsa.201900429
3. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H (2019b) SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37: 420–423. doi:10.1038/s41587-019-0036-z
4. Altmann F. (2007). The role of protein glycosylation in allergy. Int Arch Allergy Immunol. 2007;142(2):99-115. doi: 10.1159/000096114. Epub 2006 Oct 9. PMID: 17033195.
5. Altschul S, Gish W, Miller W, Myers E. and Lipman D. (1990). Basic local alignment search tool. J. Mole. Biol. 215:403-410.
6. Ariel FD, Manavella PA, Dezar C and Chan RL (2007). The true story of the HD-Zip family. Trends Plant Sci.,12(9): 419-426
7. Arya H, Singh MB, Bhalla PL. Towards Developing Drought-smart Soybeans. Front Plant Sci. 2021 Oct 6;12:750664. doi: 10.3389/fpls.2021.750664.
8. Australian Oilseeds Federation (2022). Crop Report July 2022. Accessed 12 September 2022 from http://www.australianoilseeds.com/__data/assets/pdf_file/0018/34317/AOF_Crop_Report_July_2022.pdf
9. Bartholomaeus, A., Parrott, W., Bondy, G., and 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.
10. Bartsch, K., Tebbe, C. (1989) Initial steps in the degradation of the phosphinothricin (glufosinate) by soil bacteria. Applied and Environmental Microbiology 63, 711-716.
11. Becker, W., D. Brasseur, J.L. Bresson, A. Flynn, A.A. Jackson, P. Lagiou, G. Mingrone, B. Moseley, A. Palou, H. Przyrembel, S. Salminen, S. Strobel and H. van Loveren (2004). "Opinion of the Scientific Panel on Dietetic Products, Nutrition and Allergies on a request from the Commission relating to the Evaluation of Allergenic Foods for Labelling Purposes", EFSA Journal Vol. 32, pp. 1-197
12. Behie, S. W., Bonet, B., Zacharia, V. M., McClung, D. J., and Traxler, M. F. (2016). Molecules to ecosystems: actinomycete natural products in situ. Front. Microbiol. 7:2149. doi: 10.3389/fmicb.2016.02149
13. Bernard R and Cremeens C (1988). Registration of 'Williams 82' soybean. Crop Science 28, 1027-1028.
14. de Block MD, Botterman J, Vandewiele M, Dockx J, Thoen C, Gossele V, Movva NR, Thompson C, Montagu MV and Leemans J (1987). Engineering herbicide resistance in plants by expression of a detoxifying enzyme. EMBO J., 6: 2513-2518.
15. Bryant C.J (2022). Plant-based animal product alternatives are healthier and more environmentally sustainable than animal products. Future Foods 6 December 2022 100174 https://doi.org/10.1016/j.fufo.2022.100174

16. Caragea C, Sinapov J, Silvescu A, Dobbs D and Honavar V (2007). Glycosylation site prediction using ensembles of Support Vector Machine classifiers. BMC Bioinformatics BioMed Central 8(1), 438.
17. Carrington J and Freed, (1990). Cap-independent enhancement of translation by a plant potyvirus 5' nontranslated region. Journal of Virology 64(4), 1590-1597
18. Carter, T. E., Todd, S. M., and Gillen, A. M. (2016). Registration of 'USDA-N8002' soybean cultivar with high yield and abiotic stress resistance traits. J. Plant. Regist. 10, 238–245. doi: 10.3198/jpr2015.09.0057crc
19. CERA (2011). A review of the environmental safety of the PAT protein. ILSI Research Foundation, Washington, DC. USA.
20. CFIA (1995). Decision Document DD95-01: Determination of Environmental Safety of Agrevo Canada Inc.'s Glufosinate Ammonium-Tolerant Canola. Canadian Food Inspection Agency (CFIA) Ottawa, Canada.
21. CFIA (2015) The Biology of Helianthus annuus L. Biology Document BIO2005-01: A companion document to the Directive 94-08 (Dir94-08), Assessment Criteria for Determining Environmental Safety of Plant with Novel Traits. https://inspection.canada.ca/plant-varieties/plants-with-novel-traits/applicants/directive-94-08/biology-documents/helianthus-annuus-l-/eng/1330977236841/1330977318934 Accessed 25 April 2021
22. CFIA (2021), The Biology of *Glycine max* L. Merr. (Soybean), Biology Document BIO2021-01, Plant Biosafety Office. Canadian Food Inspection Agency. Ottawa, Ont. Canada.
23. Chan RL (2009). The use of sunflower transcription factors as biotechnological tools to improve yield and stress tolerance in crops. Int. J. Exp. Bot., 78: 5-10.
24. Chan RL and Gonzalez DH (1994). A cDNA encoding an HD-zip protein from sunflower. Plant Physiol., 106(4): 1687-1688.
25. Chen, L., Yang, H., Fang, Y., Guo, W., Chen, H., Zhang, X., et al. (2021). Overexpression of GmMYB14 improves high-density yield and drought tolerance of soybean through regulating plant architecture mediated by the brassinosteroid pathway. *Plant Biotechnol. J.* 19:33098207, 702–716. doi: 10.1111/pbi.13496
26. Chew W, Hrmova M and Lopato S, 2013. Role of homeodomain leucine zipper (HD-Zip) IV transcription factors in plant development and plant protection from deleterious environmental factors. International Journal of Molecular Sciences 14(4), 8122-8147. doi:10.3390/ijms14048122
27. Chiozza MV., Burachik M., Miranda PV (2020). Compositional analysis of soybean event IND-ØØ41Ø-5, GM Crops & Food, 11:3, 154-163
28. Ciarbelli AR, Ciolfi A, Salvucci S, Ruzza V, Possenti, M, Carabelli M, Fruscalzo A, Sessa G, Morelli G and Ruberti I (2008). The Arabidopsis Homeodomain-leucine Zipper II gene family: Diversity and redundancy. Plant Molecular Biology, 68(4-5), 465–478. https://doi.org/10.1007/s11103-008-9383-8
29. Cock PA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B and de Hoon MJL (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25, 1422-1423
30. Codex Alimentarius (2003). Guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants (CAC/GL 45-2003). Annex: assessment of possible allergenicity.
31. Dayoub, E., Lamichhane, J. R., Schoving, C., Debaeke, P., and Maury, P. (2021). Early-stage phenotyping of root traits provides insights into the drought tolerance level of soybean cultivars. Agronomy 11:188. doi: 10.3390/agronomy11010188

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

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

34. Dezar CA, Fedrigo GV and Chan RL (2005b). 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 Sci., 169: 447-459.

35. Du, Y., Zhao, Q., Chen, L., Yao, X., and Xie, F. (2020). Effect of drought stress at reproductive stages on growth and nitrogen metabolism in soybean. Agronomy 10:302. doi: 10.3390/agronomy10020302

36. Dunne, E.F., Burman, W.J., Wilson, M.J. (1998) Streptomyces pneumonia in a patient with human immunodeficiency virus infection: case report and review of the literature on invasive streptomyces. Clinical Infectious Disease 27, 93-96

37. Duque AS, de Almeida AM, da Silva AB, da Silva JM, Farinha AP, Santos D, Fevereiro P and de Sousa Araújo S (2013). Abiotic Stress Responses in Plants: Unraveling the Complexity of Genes and Networks to Survive. Chapter 3, In "Agricultural and Biological Sciences. Abiotic Stress - Plant Responses and Applications in Agriculture". Kourosh Vahdati and Charles Leslie (Eds.). ISBN 978-953-51-1024-8.

38. ETS (2013). Excellence Through Stewardship. Guide for Stewardship of Biotechnology-Derived Plant Products. Excellence Through Stewardship.

39. FAO/WHO (2001). Joint Report of the FAO/WHO Expert Consultation on Allergenicity of Foods Derived from Biotechnology (22-25 January 2001, Rome, Italy).

40. FAOSTAT (2022) FAOSTAT Statistics Division Food and Agriculture Organization of the United Nations. http://www.fao.org/faostat/en/#home Accessed 12 September 2022

41. FARRP (2022). Food Allergy Research and Resource Program. UNL Allergenic Foods and Their Allergens. Http://Farrp.Unl.Edu/Resources/Gi-Fas/Informall. Accessed September 24 2022

42. FDA (2015). NPC 000016: Agency Response Letter CFSAN/Office of Food Additive Safety. FDA New Protein Consultation (NPC) 000016, U.S. Food and Drug Administration. Division of Biotechnology and GRAS Notice Review. http://www.fda.gov/Food/FoodScienceResearch/Biotechnology/Submissions/ucm222595.htm

43. FDA (2017). Biotechnology Consultation - Note to File. Biotechnology Notification File No. 000155. Subject: HB4 Soybean with altered tolerance to environmental stresses. July 28, 2017. https://www.fda.gov/downloads/Food/IngredientsPackagingLabeling/GEPlants/Submissions/ucm572729.pdf

44. Fortin T, Salvador A, Charrier J, Lenz C, Lacoux X, Morla A, Choquet-Kastylevsky G and Lemoine J 2009. Clinical quantitation of prostate-specific antigen biomarker in the low nanogram/milliliter range by conventional bore liquid chromatography-tandem mass spectrometry (multiple reaction monitoring) coupling and correlation with ELISA tests Molecular & Cellular Proteomics MCP 8, 1006-1015. doi: 10.1074/mcp.m800238-mcp200

45. Frame BR, Shou H, Chikwamba RK, Zhang Z, Xiang C, Fonger TM, Pegg SE, Li B, Nettleton DS, Pei D, Wang K. (2002). *Agrobacterium tumefaciens*-mediated transformation of maize embryos using a standard binary vector system. Plant Physiol. 2002 May;129(1):13-22.

46. Gago GM, Almoguera C, Jordano J, González DH and Chan RL (2002). Hahb-4, a homeobox-leucine zipper gene potentially involved in ABA-dependent responses to water stress in sunflower. Plant Cell Environ., 25: 633-640.

47. Gallie D, Tanguay R and Leathers V, 1995. The Tobacco etch virus 5' leader and poly (A) tail are functionally synergistic regulators of translation. Gene 165(2), 233-238

48. Gerber SA, Rush J, Stemman O, Kirschner MW and Gygi SP (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proceedings of the National Academy of Sciences 100,12 , 6940-6945. https://doi.org/10.1073/pnas.0832254100

49. Gomord V, Fitchette AC, Menu-Bouaouiche L, Saint-Jore-Dupas C, Plasson C, Michaud D and Faye L (2010). Plant-specific glycosylation patterns in the context of therapeutic protein production. Plant biotechnology journal 8(5), 564-587. https://doi.org/10.1111/j.1467-7652.2009.00497.x

50. González FG, Capella M, Ribichich KF, Curín F, Giacomelli JI, Ayala F, Watson G, Otegui ME and Chan RL, 2019. Field-grown transgenic wheat expressing the sunflower gene HaHB4 significantly outyields the wild type. Journal of Experimental Botany, 70(5):1669–1681. doi:10.1093/jxb/erz037. Advance Access Publication 6 February 2019

51. González FG, Rigalli N, Miranda PV, Romagnoli M, Ribichich KF, Trucco F, Portapila M, Otegui ME and Chan RL, 2020. An Interdisciplinary Approach to Study the Performance of Second-generation Genetically Modified Crops in Field Trials: A Case Study with Soybean and Wheat Carrying the Sunflower HaHB4 Transcription Factor. Frontiers in Plant Science. www.frontiersin.org. March 2020, Volume 11, Article 178.

52. Gupta PK, Balyan HS, Gahlaut V and Kulwal P (2012). Phenotyping, genetic dissection, and breeding for drought and heat tolerance in common wheat: status and prospects. Plant Breed. Rev., 36: 85-168.

53. Hajdukiewicz P, Svab Z, Maliga P. (1994). The small, versatile pPZP family of *Agrobacterium* binary vectors for plant transformation. Plant Mol. Biol. 25:989-94.

54. Hammond B and Cockburn A. (2008). The safety assessment of proteins introduced into crops developed through agricultural biotechnology: a consolidated approach to meet current and future needs B.G. Hammond (Ed.), Food Safety of Proteins in Agricultural Biotechnology, CRC Press, Boca Raton, FL (2008), pp. 259-288

55. Harris JC, Hrmova M, Lopato S and Langridge P (2011). Modulation of plant growth by HD-Zip class I and II transcription factors in response to environmental stimuli. New Phytologist 190(4), 823-837. https://doi.org/10.1111/j.1469-8137.2011.03733.x

56. Haq T, Mason H, Clements J and Arntzen C, (1995). Oral immunization with a recombinant bacterial antigen produced in transgenic plants. Science 268, 714-716

57. Heather, J. M., and Chain, B. (2016). The sequence of sequencers: the history of sequencing DNA. Genomics 107, 1-8. doi: 10.1016/j.ygeno.2015.11.003

58. Heberle-Bors E, Charvat B, Thompson D, Schernthaner JP, Barta A, Matzke AJM and Matzke MA (1988). Genetic analysis of T-DNA insertions into the tobacco genome. Plant Cell Rep., 7: 571-574.

59. Heeb S, Itoh Y, Nishijyo T, Schnider U, Keel C, Wade J, Walsh U, O'Gara F, Haas D. (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):232-7.

60. Henriksson E, Olsson A S, Johannesson H, Johansson H, Hanson J, Engström P and Söderman E (2005). Homeodomain leucine zipper class I genes in Arabidopsis. Expression patterns and phylogenetic relationships. Plant physiology 139, 1, 509-518. https://doi.org/10.1104/pp.105.063461

61. Herbrík A, Corretto E, Chron˘ áková A, Langhansová H, Petrásková P, Hrdý J, C˘ ihák M, Krištu° fek V, Bobek J, Petr˘íc˘ ek M and Petr˘íc˘ ková K (2020) A Human Lung-Associated Streptomyces sp. TR1341 Produces Various Secondary Metabolites Responsible for Virulence, Cytotoxicity and Modulation of Immune Response. Front. Microbiol. 10:3028. doi: 10.3389/fmicb.2019.03028

62. Herman, R.A., and 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.

63. Hérouet C, Esdaile DJ, Mallyon BA, Debruyne E, Schulz A, Currier T, Hendrickx K, van der Klis R-J and Rouan D (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: 134–149.

64. Hughes, N, Lawson, K & Valle, H (2017). Farm performance and climate: Climate-adjusted productivity for broadacre cropping farms, Canberra, April. CC BY 3.0.

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

66. ILSI (2019). Crop composition database. 2019 [accessed May 8]. https://www.cropcomposition.org/

67. ISAAA (2021). GM Events with Glufosinate Herbicide Tolerance. International Service for The Acquisition of Agri-Biotech Applications. https://www.isaaa.org/gmapprovaldatabase/gmtrait/default.asp?TraitID=1&GMTrait=Glufosinate%20herbicide%20tolerance Accessed May 12 2021

68. Ivanciuc O, Schein C H and Braun W (2002). Data Mining of Sequences and 3D Structures of Allergenic Proteins. Bioinformatics, 18(10): 1358-1364.

69. Ivanciuc O, Schein CH and Braun W (2003). SDAP: Database and Computational Tools for Allergenic Proteins. Nucleic Acids Res., 31(1): 359-362.

70. Jurėnas D, Garcia-Pino A, van Melderen L (2017). Novel toxins from type II toxin-antitoxin systems with acetyltransferase activity. Plasmid 93, 30-35.

71. Kaltenpoth, M., Gottler, W., Herzner, G., and Strohm, E. (2005). Symbiotic bacteria protect wasp larvae from fungal infestation. Curr. Biol. 15, 475–479. doi:10.1016/j.cub.2004.12.084

72. Kovalic D, Garnaat C, Yan Y, Groat J, Silvanovich A, Ralston L, Huang M, Tian Q, Christian A, Cheikh N, Hjelle J, Padgette S and Bannon G (2012). The Use of Next Generation Sequencing and Junction Sequence Analysis Bioinformatics to Achieve Molecular Characterization of Crops Improved Through Modern Biotechnology. Plant Genome, 5: 149-163.

73. Kuiper HA, Kleter GA, Noteborn HPJM and Kok EJ (2001). Assessment of the food safety issues related to genetically modified foods. Plant J., 27: 503-528.

74. Li Y, Bai B, Wen F, Zhao M, Xia Q, Yang DH, Wang G. (2019) Genome-Wide Identification and Expression Analysis of HD-ZIP I Gene Subfamily in Nicotiana tabacum. Genes (Basel). 2019 Jul 30;10(8):575. doi: 10.3390/genes10080575. PMID: 31366162; PMCID: PMC6723700.

75. Liu, W.; Fu, R.; Li, Q.; Li, J.; Wang, L.; Ren, Z. (2013) Genome-wide identification and expression profile of homeodomain-leucine zipper Class I gene family in Cucumis sativus. Gene, 531, 279–287.

76. Manavella PA, Arce AL, Dezar CA, Bitton F, Renou JP, Crespi M and Chan RL (2006). Cross-Talk Between Ethylene and Drought Signaling Pathways is Mediated by the Sunflower Hahb-4 Transcription Factor. Plant J., 48: 125-137.

77. Manavella PA, Dezar CA and 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: 860-867.

78. Manavella PA, Dezar CA, Ariel FD, Drincovich MF and Chan RL (2008b). The sunflower HD-Zip transcription factor HAHB4 is up regulated in darkness acting as a repressor of photosynthesis related genes transcription. J. Exp. Bot., 59: 3143-3155.

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

80. Meier U, 2001. Growth stages of mono-and dicotyledonous plants, 2nd edition. BBCH Monograph. Federal Biological Research Centre for Agriculture and Forestry, Bonn, Germany

81. Messina, M., Sievenpiper, J.L., Williamson, P., Kiel, J., Erdman, J.W., (2022). Perspective: soy-based meat and dairy alternatives, despite classification as ultra-processed foods, deliver high-quality nutrition on par with unprocessed or minimally processed animal-based counterparts. Adv. Nutr.

82. Mir R, Aranda LZ, Biaocchi T, Luo A, Sylvester AW, Rasmussen CG. (2017). A DII Domain-Based Auxin Reporter Uncovers Low Auxin Signaling during Telophase and Early G1. Plant Physiol. 2017 Jan;173(1):863-871.

83. Morey, M., Fernández-Marmiesse, A., Castiñeiras, D., Fraga, J. M., Couce, M. L., and Cocho, J. A. (2013). A glimpse into past, present, and future DNA sequencing. Mol. Genet. Metab. 110, 3-24. doi: 10.1016/j.ymgme.2013.04.024

84. Murakami, T., Anzai, H., Imai, S., Satoh, A., Nagaoka, K., Thompson, C. J. (1986) The bialaphos biosynthetic genes of Streptomyces hygroscopicus: Molecular cloning and characterization of the gene cluster. Molecular and General Genetics 205, 42-50.

85. 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. 200:33-39.

86. Odell J, Nagy F and Chua N 1985. Identification of DNA sequences required for activity of the Cauliflower mosaic virus 35S promoter. Nature 313, 810-812

87. OECD (2000). Consensus Document on the Biology of *Glycine max* (L.) Merr. (Soybean). ENV/JM/MONO(2000)9. Series on harmonization of regulatory oversight in biotechnology No. 15. Organization for Economic Co-operation and Development, Paris, France.

88. OECD (2003). Organization for Economic Co-Operation and Development. The Working Party on Chemicals, Pesticides and Biotechnology ENV/JM/MONO. Organization for Economic Co-Operation and Development 9, 1–46.

89. OECD (2006) "Section 2 - Soybean (GLYCINE MAX (L.) MARR.)", in *Safety Assessment of Transgenic Organisms, Volume 1: OECD Consensus Documents*, OECD Publishing, Paris, https://doi.org/10.1787/9789264095380-5-en.

90. 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. ENV/JM/MONO(2012)24. Series on the Safety of Novel Foods and Feeds No. 25 (Paris, France: Organisation for Economic Cooperation and Development).

91. OECD (1999). Consensus document on general information concerning the genes and their enzymes that confer tolerance to phosphinothricin herbicide. ENV/JM/MONO(99)13. Series on harmonization of regulatory oversight in biotechnology No. 11. Organisation for Economic Co-operation and Development, Paris, France.

92. Oltmanns H, Frame B, Lee L-Y, Johnson S, Li B, Wang K and Gelvin SB (2010). Generation of Backbone-Free, Low Transgene Copy Plants by Launching T-DNA from the Agrobacterium Chromosome. Plant Physiol., 152: 1158–1166.

93. Palena CM, González DH and Chan RL (1999). A monomer-dimer equilibrium modulates the interaction of the sunflower homeodomain leucine-zipper protein Hahb-4 with DNA. Biochem. J., 341: 81-87.

94. Pattison RJ and Amtmann A (2009). N-glycan production in the endoplasmic reticulum of plants. Trends in plant science 14, 2, 92-99. https://doi.org/10.1016/j.tplants.2008.11.008

95. Pourzand F and Noy I (2019). Impacts of droughts on agricultural productivity and profitability in New Zealand: A micro-level study. Contributed paper prepared for presentation at the 63rd AARES Annual Conference, Melbourne, Vic 12-15 February 2019. Australasian Agricultural and Resource Economics Society

96. Privalle LS, Gillikin N and Wandelt CH (2013). Bringing a Transgenic Crop to Market: Where Compositional Analysis Fits. J. Agric. Food Chem., 61: 8260 - 8266.

97. Putnam DH, Oplinger ES, Hicks DR, Durgan BR, Noetzel DM, Meronuck RA, Doll JD, Schulte EE (2021). Sunflower. In Alternative Field Crops Manual. https://hort.purdue.edu/newcrop/afcm/sunflower.html Accessed April 23 2021

98. Raghuvanshi, R. S., and Bisht, K. (2010). "Chapter 18- Uses of Soybean: Products and Preparation," in The soybean: botany, production and uses. ed. G. Singh (CAB International), 404–426

99. Rapp W, Lilley G and Nielsen N, (1990). Characterization of soybean vegetative storage proteins and genes. Theoretical and Applied Genetics 79(6), 785-792. https://doi.org/10.1007/BF00224246

100. Reuter, J. A., Spacek, D. V., and Snyder, M. P. (2015). High-throughput sequencing technologies. *Mol. Cell.* 58, 586-597. doi: 10.1016/j.molcel.2015.05.004

101. Rezaei, K., Wang, T., and Johnson, L. A. (2002). Hydrogenated vegetable oils as candle wax. J. Am. Oil Chem. Soc. 79, 1241–1247. doi: 10.1007/s11746-002-0634-z

102. Ribichich KF., Chiozza M., Avalos-Britex S., Cabello JV., Arce AL., Watson G., Arias C., Portapila M., Trucco F, Otegui ME., Chan RL (2020). Successful field performance in warm and dry environments of soybean expressing the sunflower transcription factor HB4. J. Exp Bot 71: No. 10 3142-3156

103. Salah Ud-Din, A. I. M., Tikhomirova, A., & Roujeinikova, A. (2016). Structure and functional diversity of GCN5-related N-acetyltransferases (GNAT). Int. J Molec. Sci., 17(7): 1018.

104. Sarmiento-Ramirez, J. M., van der Voort, M., Raaijmakers, J. M., and Dieguez-Uribeondo, J. (2014). Unravelling the microbiome of eggs of the endangered sea turtle Eretmochelys imbricata identifies bacteria with activity against the emerging pathogen Fusarium falciforme. PLoS One 9:e95206. doi: 10.1371/journal.pone.0095206

105. Savage, J. H., A. J. Kaeding, E. C. Matsui, and R. A. Wood. (2010). The natural history of soy allergy. Journal of Allergy and Clinical Immunology 125 (3):683–6. doi: 10.1016/j.jaci.2009.12.994

106. Schena M and Davis RW (1992). HD-Zip proteins: members of an Arabidopsis homeodomain protein superfamily. Proc. Nat. Acad. Sci., 89: 3894-3898.

107. Shou H, Frame BR, SA and Wang K (2004). Assessment of transgenic maize events produced by particle bombardment or Agrobacterium-mediated transformation. Molecular Breed., 13: 201-208.

108. Small I, Peeters N, Legeai F and Lurin C (2004). Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics 4(6), 1581-1590. https://doi.org/10.1002/pmic.200300776

109. Stringam, G. R., Ripley, V. L., Love, H. K. and Mitchell, A. (2003). Transgenic herbicide tolerant canola – the Canadian experience. Crop Sci. 43: 1590–1593.

110. Suárez-López, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., Coupland, G (2001). CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. Nature, 410: 1116–1120

111. Taherzadeh G, Dehzangi A, Golchin M, Zhou Y, Campbell MP. (2019). SPRINT-Gly: predicting N- and O-linked glycosylation sites of human and mouse proteins by using sequence and predicted structural properties. Bioinformatics. 2019 Oct 15;35(20):4140-4146. doi: 10.1093/bioinformatics/btz215. PMID: 30903686.

112. Takano HK and Dayan FE (2020). Glufosinate-ammonium: a review of the current state of knowledge. Pest Management Science, 76(12):3911-3925.

113. Thomas K, Aalbers M, Bannon GA, Bartels M, Dearman RJ, Esdaile DJ, Fu TJ, Glatt CM, Hadfield N, Hatzos C, Hefle SL, Heylings JR, Goodman RE, Henry B, Herouet C, Holsapple M, Ladics GS, Landry TD, MacIntoshj SC, Ricec EA, Privallek LS, Steinerk HY, Teshimal R, van Reeb R, Woolhiserd M and Zawodnyk 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(2), 87-98. https://doi.org/10.1016/j.yrtph.2003.11.003

114. Thompson CJ, Rao Movva N, Tizard R, Crameri R, Davies JE, Lauwereys M and Botterman J (1987). Characterization of the herbicide-resistance gene bar from Streptomyces hygroscopicus. EMBO J., 6(9): 2519-2523.

115. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F and Higgins DG (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic acids research 25, 24, 4876-4882. https://doi.org/10.1093/nar/25.24.4876

116. USDA (2019). Petition for Determination of Non-Regulated Status for The New Plant Variety Soybean (IND-00410-5) Intended for Environmental Release and Food and Feed Use.

117. USDA (2011). National Environmental Policy Act Decision and Finding of No Significant Impact. Bayer CropScience Insect Resistant and Glufosinate Ammonium-Tolerant (TwinLinkTM) Cotton, Events T304-40 x GHBl19 United States Department of Agriculture Animal and Plant Health Inspection Service Biotechnology Regulatory Services. References are made to: 40 CFR § 174.522; US-EPA, 2010c.

118. VanHaaren, M.J.J., Sedee, N.J.A.J.A., de Boer, H.A.A., Schilperoort, R.A.A., Hooykaas, P.J.J.J.J., (1989). Mutational Analysis of the Conserved Domains of a T-Region Border Repeat of Agrobacterium Tumefaciens. Plant Molecular Biology *13*, 523–531.

119. Wang J., He Z., Raghavan V (2022). Soybean allergy: characteristics, mechanisms, detection and its reduction through novel food processing techniques, Critical Reviews in Food Science and Nutrition, DOI: 10.1080/10408398.2022.2029345

120. Yanisch-Perron C, Vieira J and Messing J (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequencing of the M13mp18 and pUC9 vectors. Gene, 33: 103-119.

121. Yocum A and Chinnaiyan A, (2009). Current affairs in quantitative targeted proteomics: multiple reaction monitoring-mass spectrometry. Briefings in Functional Genomics & Proteomics 8, 145-157.

122. Zhang Z, Schwartz S, Wagner L, Miller W (2000). A greedy algorithm for aligning DNA sequences, J Comput Biol 2000; 7(1-2):203-14.

Appendix 1.

Figure 20. Sequence of the insert and flanking soybean sequence in event IND-ØØ41Ø-5

Nucleotides corresponding to specific elements of the insert are highlighted in different colours. Soybean genome is indicated in green, *HaHB4* in red and *bar* in light blue. Sequence of the T-DNA locus assembled *de novo* from Illumina sequence reads**,** including: the soybean genome flanking sequences, in upper case highlighted in green; the *bar* sequence, in lower case highlighted in light blue; and the *HaHB4* sequence, in lower case highlighted in red. The regulatory sequences: 2X35S promoter and nos terminator for *bar*, LPF promoter and vsp terminator for *HaHB4* are in upper case. The left border (minus 3 bp) is highlighted in yellow.

TCCCCTCTCAGTTGGGTCAGCCTGAGTGATTTTTTTCTCAAATCAAGAAACTTTATTTATAAATCTAACATT
ATAATATTAAAAAAACAAATATTAAAATATTCATGATATTTTTAAATCTAAATAATATTCTAAAAATTTGAA
ACAAATAAATTCTTGAAAATAAACTAAATTATTCTTTTCCAAACTAACTAAAGATATCAGGATATATTGTGG
TGTAAACAAATTGACGCTTAGACAACTTAATAACACATTGCGGACGTTTTTAATGTACTGAATTAACGCCGA
ATTGCTCTAGCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATT
ACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACG
ACGTTGTAAAACGACGGCCAGTGCCAAGCTAATTCGCTTCAAGACGTGCTCAAATCACTATTTCCACACCCC
TATATTTCTATTGCACTCCCTTTTAACTGTTTTTTATTACAAAAATGCCCTGGAAAATGCACTCCCTTTTTG
TGTTTGTTTTTTTGTGAAACGATGTTGTCAGGTAATTTATTTGTCAGTCTACTATGGTGGCCCATTATATTA
ATAGCAACTGTCGGTCCAATAGACGACGTCGATTTTCTGCATTTGTTTAACCACGTGGATTTTATGACATTT
TATATTAGTTAATTTGTAAAACCTACCCAATTAAAGACCTCATATGTTCTAAAGACTAATACTTAATGATAA
CAATTTTCTTTTAGTGAAGAAAGGGATAATTAGTAAATATGGAACAAGGGCAGAAGATTTATTAAAGCCGCG
TAAGAGACAACAAGTAGGTACGTGGAGTGTCTTAGGTGACTTACCCACATAACATAAAGTGACATTAACAAA
CATAGCTAATGCTCCTATTTGAATAGTGCATATCAGCATACCTTATTACATATAGATAGGAGCAAACTCTAG
ctagattgttgagcagatctcggtgacgggcaggaccggacggggcggtaccggcaggctgaagtccagctg
ccagaaacccacgtcatgccagttcccgtgcttgaagccggccgcccgcagcatgccgcgggggggcatatcc
gagcgcctcgtgcatgcgcacgctcgggtcgttgggcagcccgatgacagcgaccacgctcttgaagccctg
tgcctccagggacttcagcaggtgggtgtagagcgtggagcccagtcccgtccgctggtggcgggggggagac
gtacacggtcgactcggccgtccagtcgtaggcgttgcgtgccttccaggggcccgcgtaggcgatgccggc
gacctcgccgtccacctcggcgacgagccagggatagcgctcccgcagacggacgaggtcgtccgtccactc
ctgcggttcctgcggctcggtacggaagttgaccgtgcttgtctcgatgtagtggttgacgatggtgcagac
cgccggcatgtccgcctcggtggcacggcggatgtcggccgggcgtcgttctgggctcatGGTAGATCCCCC
GTTCGTAAATGGTGAAAATTTTCAGAAAATTGCTTTTGCTTTAAAAGAAATGATTTAAATTGCTGCAATAGA
AGTAGAATGCTTGATTGCTTGAGATTCGTTTGTTTTGTATATGTTGTGTTGAGAATTAATTCTCGAGGTCCT
CTCCAAATGAAATGAACTTCCTTATATAGAGGAAGGGTCTTGCGAAGGATAGTGGGATTGTGCGTCATCCCT
TACGTCAGTGGAGATATCACATCAATCCACTTGCTTTGAAGACGTGGTTGGAACGTCTTCTTTTTCCACGAT
GCTCCTCGTGGGTGGGGGTCCATCTTTGGGACCACTGTCGGCAGAGGCATCTTCAACGATGGCCTTTCCTTT
ATCGCAATGATGGCATTTGTAGGAGCCACCTTCCTTTTCCACTATCTTCACAATAAAGTGACAGATAGCTGG
GCAATGGAATCCGAGGAGGTTTCCGGATATTACCCTTTGTTGAAAAGTCTCAATTGCCCTTTGGTCTTCTGA
GACTGTATCTTTGATATTTTTGGAGTAGACAAGTGTGTCGTGCTCCACCATGTTATCACATCAATCCACTTG
CTTTGAAGACGTGGTTGGAACGTCTTCTTTTTCCACGATGCTCCTCGTGGGTGGGGGTCCATCTTTGGGACC
ACTGTCGGCAGAGGCATCTTCAACGATGGCCTTTCCTTTATCGCAATGATGGCATTTGTAGGAGCCACCTTC
CTTTTCCACTATCTTCACAATAAAGTGACAGATAGCTGGGCAATGGAATCCGAGGAGGTTTCCGGATATTAC
CCTTTGTTGAAAAGTCTCAATTGCCCTTTGGTCTTCTGAGACTGTATCTTTGATATTTTTGGAGTAGACAAG
TGTGTCGTGCTCCACCATGTTGACCTGCAGGTCGACACCTGGCACATCGTATCTTATCTCTTTTGTCGTTTC
CAACACACCACAACACACCTACAAACGTGTCAATTCACACTTCACCAATTTCATTTCCTTTTAGTCAATCAT
ATTAAAAGTAGTAGCCCCCACCCCCATTTGTTACCTACCATTTCCCACTTTAATAATCACCCACGCTATGTC
CACTTGTACTTTTGTTTGCACACAACTCTTCCCATAAAATATCAAACCAAATTTTTTTTAGTGGAAAACAAA
TTCCCCAAATAGAATACTAACGAAATTCATCGCATCAGAATACACTCATCTCTGAACAGTGGCGAAGCTTGA
CGTTTTCGACGGGGGGTCGGAAAACGTATGTACCCGAAATTTCTATAGAATCGGGGGGTCGAAAACGTATAT
ACCCAAAATTTCTATACGAAAACTACATATATAACACTACTGAGCAAAAAGTTCGGGGGTTCGGGCGCCCCT
CCCGGCCCCTTCAAAGCTTCGCCAATGTCTCTGAACCGAAGAAAACCCTCACTCGTCTACTAGCCAATGAAT
CCTCACCAGGGAAAACCCTCACTCGTCTTACTGGACTATTGGCGCTTCCAAATGGACTACTTGCGAAATTCA
CCACATTGGGATACACTCGTCTACTGCGGTGAGGTAAAACCCGCTTGGTTCAAGGATCGAACTAGCGATTGC
TGCCTACTCGCCTAATCTCCCATCATCAACAGGTGCCGCCGAAACAAAATGCTGGGGGCGGGAGTTGAACCT
AGGTCCAGTGACGCACCCATGAATTTTTTTTCTAGGGATGCGAACGAGTGGTTTAACCATACTTTTAAGAGG
TGCGATCGGAAATTTTACCTATAAAATACACTAAAAAAGTTCCAAGGGTCCACCCACCCCTTAACCTAAGTC

CGCCTTTGTCTGGATCACGTGAAACATCAGGTCTCTCCCTTACCAGTCCAGCTACGACTCATTGACAAAATA
TCAAAACCATATGATTTTGAGTTTTATCTCAACCGAAAGTGACATCATGACAGAGAATCGACATAACCAAAA
CGTGTAAACGTACAACTCACCATTGCGTTGAAAAGGACAAAACAGGTAGGATTCTTGTCAAATTCAACGCGT
ACACCTGTGCTTCATCTAAACCCCATACTTTTAAGAACCTTTATAAAGACCACTCACTATATATACACATAT
ATAATATCACTTATCAAACCCTCGGATCCACCatgtctcttcaacaagtaacaaccaccaggaagaaccgaa
acgaggggcggagacgatttaccgacaaacaaataagtttcctagagtacatgtttgagacacagtcgagac
ccgagttaaggatgaaacaccagttggcacataaactcgggcttcatcctcgtcaagtggcgatatggttcc
agaacaaacgcgcgcgatcaaagtcgaggcagattgagcaagagtataacgcgctaaagcataactacgaga
cgcttgcgtctaaatccgagtctctaaagaaagagaatcaggccctactcaatcaattggaggtgctgagaa
atgtagccgaaaagcatcaagagaaaactagtagtagtggcagcggtgaagaatcggatgatcggtttacga
actctccggacgttatgtttggtcaagaaatgaatgttccgttttgcgacggttttgcgtaccttgaagaag
gaaacagtttgttggagattgaagaacaactgccagaccttcaaaagtggtgggagttcTAAGAGCTCGAAT
TTCCCCGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTTGCGATGATT
ATCATATAATTTCTGTTGAATTACGTTAAGCATGTAATAATTAACATGTAATGCATGACGTTATTTATGAGA
TGGGTTTTTATGATTAGAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGCAAAC
TAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGGGAATTCGTAATCATGTCATAGCTGTTT
CCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGG
GGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTG
TCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGAGCTTGAGCTTGG
ATCAGATTGTCGTTTCCCGCCTTCAGTTTAAACTATCAGTACCCTCAATCATCTCACTTCATTATCTCCTAT
ATTTTTTATTAACTTCTCTTTTATACTATTTTAAAAAAATAAAAAGTGAGAATTTAAACAGAAAAAACCTCT
CTCAAGTCTTTCTCTCTATTTCAGTGGTCTGAGTTCAGTTGCGTCTCTTAATCTTTTAGGTTGGGAAAACAT
CATCTTCTTTTGGGAGATTGGCTCC

Appendix 2.

Figure 21. IND-ØØ41Ø-5 T-DNA with soybean chromosome flanking sequences from across six generations
Except for the Illumina-derived sequence (T6), each sequence presented for each generation (T1 to T5) was a consensus sequence obtained from sequencing of 3 different plant amplicons. The T1 to T5 generation sequences were contigs obtained through conventional Sanger capillary sequencing. Soybean flanking sequences (green); *HaHB4* sequence (red); *bar* gene (light blue); left border (pink).

```
                      1                                                  50
IND-ØØ41Ø-5 T1      (1) ACGCAACTGAACTCAGACCACTGAAATAGAGAGAAAGACTTGAGAGAGGT
IND-ØØ41Ø-5 T3      (1) ACGCAACTGAACTCAGACCACTGAAATAGAGAGAAAGACTTGAGAGAGGT
IND-ØØ41Ø-5 T5      (1) ACGCAACTGAACTCAGACCACTGAAATAGAGAGAAAGACTTGAGAGAGGT
IND-ØØ41Ø-5 T6      (1) ACGCAACTGAACTCAGACCACTGAAATAGAGAGAAAGACTTGAGAGAGGT
     Consensus      (1) ACGCAACTGAACTCAGACCACTGAAATAGAGAGAAAGACTTGAGAGAGGT
                     51                                                 100
IND-ØØ41Ø-5 T1     (51) TTTTTCTGTTTAAATTCTCACTTTTTATTTTTTTAAAATAGTATAAAAGA
IND-ØØ41Ø-5 T3     (51) TTTTTCTGTTTAAATTCTCACTTTTTATTTTTTTAAAATAGTATAAAAGA
IND-ØØ41Ø-5 T5     (51) TTTTTCTGTTTAAATTCTCACTTTTTATTTTTTTAAAATAGTATAAAAGA
IND-ØØ41Ø-5 T6     (51) TTTTTCTGTTTAAATTCTCACTTTTTATTTTTTTAAAATAGTATAAAAGA
     Consensus     (51) TTTTTCTGTTTAAATTCTCACTTTTTATTTTTTTAAAATAGTATAAAAGA
                    101                                                 150
IND-ØØ41Ø-5 T1    (101) GAAGTTAATAAAAAATATAGGAGATAATGAAGTGAGATGATTGAGGGTAC
IND-ØØ41Ø-5 T3    (101) GAAGTTAATAAAAAATATAGGAGATAATGAAGTGAGATGATTGAGGGTAC
IND-ØØ41Ø-5 T5    (101) GAAGTTAATAAAAAATATAGGAGATAATGAAGTGAGATGATTGAGGGTAC
IND-ØØ41Ø-5 T6    (101) GAAGTTAATAAAAAATATAGGAGATAATGAAGTGAGATGATTGAGGGTAC
     Consensus    (101) GAAGTTAATAAAAAATATAGGAGATAATGAAGTGAGATGATTGAGGGTAC
                    151                                                 200
IND-ØØ41Ø-5 T1    (151) TGATAGTTTAAACTGAAGGCGGGAAACGACAATCTGATCCAAGCTCAAGC
IND-ØØ41Ø-5 T3    (151) TGATAGTTTAAACTGAAGGCGGGAAACGACAATCTGATCCAAGCTCAAGC
IND-ØØ41Ø-5 T5    (151) TGATAGTTTAAACTGAAGGCGGGAAACGACAATCTGATCCAAGCTCAAGC
IND-ØØ41Ø-5 T6    (151) TGATAGTTTAAACTGAAGGCGGGAAACGACAATCTGATCCAAGCTCAAGC
     Consensus    (151) TGATAGTTTAAACTGAAGGCGGGAAACGACAATCTGATCCAAGCTCAAGC
                    201                                                 250
IND-ØØ41Ø-5 T1    (201) TCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCA
IND-ØØ41Ø-5 T3    (201) TCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCA
IND-ØØ41Ø-5 T5    (201) TCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCA
IND-ØØ41Ø-5 T6    (201) TCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCA
     Consensus    (201) TCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCA
                    251                                                 300
IND-ØØ41Ø-5 T1    (251) GCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCA
IND-ØØ41Ø-5 T3    (251) GCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCA
IND-ØØ41Ø-5 T5    (251) GCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCA
IND-ØØ41Ø-5 T6    (251) GCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCA
     Consensus    (251) GCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCA
                    301                                                 350
IND-ØØ41Ø-5 T1    (301) ATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTAT
IND-ØØ41Ø-5 T3    (301) ATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTAT
IND-ØØ41Ø-5 T5    (301) ATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTAT
IND-ØØ41Ø-5 T6    (301) ATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTAT
     Consensus    (301) ATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTAT
                    351                                                 400
IND-ØØ41Ø-5 T1    (351) GCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACA
IND-ØØ41Ø-5 T3    (351) GCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACA
IND-ØØ41Ø-5 T5    (351) GCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACA
IND-ØØ41Ø-5 T6    (351) GCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACA
     Consensus    (351) GCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACA
                    401                                                 450
IND-ØØ41Ø-5 T1    (401) CAGGAAACAGCTATGACATGATTACGAATTCCCGATCTAGTAACATAGAT
IND-ØØ41Ø-5 T3    (401) CAGGAAACAGCTATGACATGATTACGAATTCCCGATCTAGTAACATAGAT
IND-ØØ41Ø-5 T5    (401) CAGGAAACAGCTATGACATGATTACGAATTCCCGATCTAGTAACATAGAT
IND-ØØ41Ø-5 T6    (401) CAGGAAACAGCTATGACATGATTACGAATTCCCGATCTAGTAACATAGAT
     Consensus    (401) CAGGAAACAGCTATGACATGATTACGAATTCCCGATCTAGTAACATAGAT

                    451                                                 500
```

```
IND-ØØ41Ø-5 T1    (451) GACACCGCGCGCGATAATTTATCCTAGTTTGCGCGCTATATTTTGTTTTC
IND-ØØ41Ø-5 T3    (451) GACACCGCGCGCGATAATTTATCCTAGTTTGCGCGCTATATTTTGTTTTC
IND-ØØ41Ø-5 T5    (451) GACACCGCGCGCGATAATTTATCCTAGTTTGCGCGCTATATTTTGTTTTC
IND-ØØ41Ø-5 T6    (451) GACACCGCGCGCGATAATTTATCCTAGTTTGCGCGCTATATTTTGTTTTC
     Consensus    (451) GACACCGCGCGCGATAATTTATCCTAGTTTGCGCGCTATATTTTGTTTTC
                      501                                               550
IND-ØØ41Ø-5 T1    (501) TATCGCGTATTAAATGTATAATTGCGGGACTCTAATCATAAAAACCCATC
IND-ØØ41Ø-5 T3    (501) TATCGCGTATTAAATGTATAATTGCGGGACTCTAATCATAAAAACCCATC
IND-ØØ41Ø-5 T5    (501) TATCGCGTATTAAATGTATAATTGCGGGACTCTAATCATAAAAACCCATC
IND-ØØ41Ø-5 T6    (501) TATCGCGTATTAAATGTATAATTGCGGGACTCTAATCATAAAAACCCATC
     Consensus    (501) TATCGCGTATTAAATGTATAATTGCGGGACTCTAATCATAAAAACCCATC
                      551                                               600
IND-ØØ41Ø-5 T1    (551) TCATAAATAACGTCATGCATTACATGTTAATTATTACATGCTTAACGTAA
IND-ØØ41Ø-5 T3    (551) TCATAAATAACGTCATGCATTACATGTTAATTATTACATGCTTAACGTAA
IND-ØØ41Ø-5 T5    (551) TCATAAATAACGTCATGCATTACATGTTAATTATTACATGCTTAACGTAA
IND-ØØ41Ø-5 T6    (551) TCATAAATAACGTCATGCATTACATGTTAATTATTACATGCTTAACGTAA
     Consensus    (551) TCATAAATAACGTCATGCATTACATGTTAATTATTACATGCTTAACGTAA
                      601                                               650
IND-ØØ41Ø-5 T1    (601) TTCAACAGAAATTATATGATAATCATCGCAAGACCGGCAACAGGATTCAA
IND-ØØ41Ø-5 T3    (601) TTCAACAGAAATTATATGATAATCATCGCAAGACCGGCAACAGGATTCAA
IND-ØØ41Ø-5 T5    (601) TTCAACAGAAATTATATGATAATCATCGCAAGACCGGCAACAGGATTCAA
IND-ØØ41Ø-5 T6    (601) TTCAACAGAAATTATATGATAATCATCGCAAGACCGGCAACAGGATTCAA
     Consensus    (601) TTCAACAGAAATTATATGATAATCATCGCAAGACCGGCAACAGGATTCAA
                      651                                               700
IND-ØØ41Ø-5 T1    (651) TCTTAAGAAACTTTATTGCCAAATGTTTGAACGATCGGGGAAATTCGAGC
IND-ØØ41Ø-5 T3    (651) TCTTAAGAAACTTTATTGCCAAATGTTTGAACGATCGGGGAAATTCGAGC
IND-ØØ41Ø-5 T5    (651) TCTTAAGAAACTTTATTGCCAAATGTTTGAACGATCGGGGAAATTCGAGC
IND-ØØ41Ø-5 T6    (651) TCTTAAGAAACTTTATTGCCAAATGTTTGAACGATCGGGGAAATTCGAGC
     Consensus    (651) TCTTAAGAAACTTTATTGCCAAATGTTTGAACGATCGGGGAAATTCGAGC
                      701                                               750
IND-ØØ41Ø-5 T1    (701) TCTTAGAACTCCCACCACTTTTGAAGGTCTGGCAGTTGTTCTTCAATCTC
IND-ØØ41Ø-5 T3    (701) TCTTAGAACTCCCACCACTTTTGAAGGTCTGGCAGTTGTTCTTCAATCTC
IND-ØØ41Ø-5 T5    (701) TCTTAGAACTCCCACCACTTTTGAAGGTCTGGCAGTTGTTCTTCAATCTC
IND-ØØ41Ø-5 T6    (701) TCTTAGAACTCCCACCACTTTTGAAGGTCTGGCAGTTGTTCTTCAATCTC
     Consensus    (701) TCTTAGAACTCCCACCACTTTTGAAGGTCTGGCAGTTGTTCTTCAATCTC
                      751                                               800
IND-ØØ41Ø-5 T1    (751) CAACAAACTGTTTCCTTCTTCAAGGTACGCAAAACCGTCGCAAAACGGAA
IND-ØØ41Ø-5 T3    (751) CAACAAACTGTTTCCTTCTTCAAGGTACGCAAAACCGTCGCAAAACGGAA
IND-ØØ41Ø-5 T5    (751) CAACAAACTGTTTCCTTCTTCAAGGTACGCAAAACCGTCGCAAAACGGAA
IND-ØØ41Ø-5 T6    (751) CAACAAACTGTTTCCTTCTTCAAGGTACGCAAAACCGTCGCAAAACGGAA
     Consensus    (751) CAACAAACTGTTTCCTTCTTCAAGGTACGCAAAACCGTCGCAAAACGGAA
                      801                                               850
IND-ØØ41Ø-5 T1    (801) CATTCATTTCTTGACCAAACATAACGTCCGGAGAGTTCGTAAACCGATCA
IND-ØØ41Ø-5 T3    (801) CATTCATTTCTTGACCAAACATAACGTCCGGAGAGTTCGTAAACCGATCA
IND-ØØ41Ø-5 T5    (801) CATTCATTTCTTGACCAAACATAACGTCCGGAGAGTTCGTAAACCGATCA
IND-ØØ41Ø-5 T6    (801) CATTCATTTCTTGACCAAACATAACGTCCGGAGAGTTCGTAAACCGATCA
     Consensus    (801) CATTCATTTCTTGACCAAACATAACGTCCGGAGAGTTCGTAAACCGATCA
                      851                                               900
IND-ØØ41Ø-5 T1    (851) TCCGATTCTTCACCGCTGCCACTACTACTAGTTTTCTCTTGATGCTTTTC
IND-ØØ41Ø-5 T3    (851) TCCGATTCTTCACCGCTGCCACTACTACTAGTTTTCTCTTGATGCTTTTC
IND-ØØ41Ø-5 T5    (851) TCCGATTCTTCACCGCTGCCACTACTACTAGTTTTCTCTTGATGCTTTTC
IND-ØØ41Ø-5 T6    (851) TCCGATTCTTCACCGCTGCCACTACTACTAGTTTTCTCTTGATGCTTTTC
     Consensus    (851) TCCGATTCTTCACCGCTGCCACTACTACTAGTTTTCTCTTGATGCTTTTC
                      901                                               950
IND-ØØ41Ø-5 T1    (901) GGCTACATTTCTCAGCACCTCCAATTGATTGAGTAGGGCCTGATTCTCTT
IND-ØØ41Ø-5 T3    (901) GGCTACATTTCTCAGCACCTCCAATTGATTGAGTAGGGCCTGATTCTCTT
IND-ØØ41Ø-5 T5    (901) GGCTACATTTCTCAGCACCTCCAATTGATTGAGTAGGGCCTGATTCTCTT
IND-ØØ41Ø-5 T6    (901) GGCTACATTTCTCAGCACCTCCAATTGATTGAGTAGGGCCTGATTCTCTT
     Consensus    (901) GGCTACATTTCTCAGCACCTCCAATTGATTGAGTAGGGCCTGATTCTCTT
                      951                                              1000
IND-ØØ41Ø-5 T1    (951) TCTTTAGAGACTCGGATTTAGACGCAAGCGTCTCGTAGTTATGCTTTAGC
IND-ØØ41Ø-5 T3    (951) TCTTTAGAGACTCGGATTTAGACGCAAGCGTCTCGTAGTTATGCTTTAGC
IND-ØØ41Ø-5 T5    (951) TCTTTAGAGACTCGGATTTAGACGCAAGCGTCTCGTAGTTATGCTTTAGC
IND-ØØ41Ø-5 T6    (951) TCTTTAGAGACTCGGATTTAGACGCAAGCGTCTCGTAGTTATGCTTTAGC
     Consensus    (951) TCTTTAGAGACTCGGATTTAGACGCAAGCGTCTCGTAGTTATGCTTTAGC

                     1001                                              1050
```

```
IND-ØØ41Ø-5 T1   (1001) GCGTTATACTCTTGCTCAATCTGCCTCGACTTTGATCGCGCGCGTTTGTT
IND-ØØ41Ø-5 T3   (1001) GCGTTATACTCTTGCTCAATCTGCCTCGACTTTGATCGCGCGCGTTTGTT
IND-ØØ41Ø-5 T5   (1001) GCGTTATACTCTTGCTCAATCTGCCTCGACTTTGATCGCGCGCGTTTGTT
IND-ØØ41Ø-5 T6   (1001) GCGTTATACTCTTGCTCAATCTGCCTCGACTTTGATCGCGCGCGTTTGTT
     Consensus   (1001) GCGTTATACTCTTGCTCAATCTGCCTCGACTTTGATCGCGCGCGTTTGTT
                      1051                                              1100
IND-ØØ41Ø-5 T1   (1051) CTGGAACCATATCGCCACTTGACGAGGATGAAGCCCGAGTTTATGTGCCA
IND-ØØ41Ø-5 T3   (1051) CTGGAACCATATCGCCACTTGACGAGGATGAAGCCCGAGTTTATGTGCCA
IND-ØØ41Ø-5 T5   (1051) CTGGAACCATATCGCCACTTGACGAGGATGAAGCCCGAGTTTATGTGCCA
IND-ØØ41Ø-5 T6   (1051) CTGGAACCATATCGCCACTTGACGAGGATGAAGCCCGAGTTTATGTGCCA
     Consensus   (1051) CTGGAACCATATCGCCACTTGACGAGGATGAAGCCCGAGTTTATGTGCCA
                      1101                                              1150
IND-ØØ41Ø-5 T1   (1101) ACTGGTGTTTCATCCTTAACTCGGGTCTCGACTGTGTCTCAAACATGTAC
IND-ØØ41Ø-5 T3   (1101) ACTGGTGTTTCATCCTTAACTCGGGTCTCGACTGTGTCTCAAACATGTAC
IND-ØØ41Ø-5 T5   (1101) ACTGGTGTTTCATCCTTAACTCGGGTCTCGACTGTGTCTCAAACATGTAC
IND-ØØ41Ø-5 T6   (1101) ACTGGTGTTTCATCCTTAACTCGGGTCTCGACTGTGTCTCAAACATGTAC
     Consensus   (1101) ACTGGTGTTTCATCCTTAACTCGGGTCTCGACTGTGTCTCAAACATGTAC
                      1151                                              1200
IND-ØØ41Ø-5 T1   (1151) TCTAGGAAACTTATTTGTTTGTCGGTAAATCGTCTCCGCCCCTCGTTTCG
IND-ØØ41Ø-5 T3   (1151) TCTAGGAAACTTATTTGTTTGTCGGTAAATCGTCTCCGCCCCTCGTTTCG
IND-ØØ41Ø-5 T5   (1151) TCTAGGAAACTTATTTGTTTGTCGGTAAATCGTCTCCGCCCCTCGTTTCG
IND-ØØ41Ø-5 T6   (1151) TCTAGGAAACTTATTTGTTTGTCGGTAAATCGTCTCCGCCCCTCGTTTCG
     Consensus   (1151) TCTAGGAAACTTATTTGTTTGTCGGTAAATCGTCTCCGCCCCTCGTTTCG
                      1201                                              1250
IND-ØØ41Ø-5 T1   (1201) GTTCTTCCTGGTGGTTGTTACTTGTTGAAGAGACATGGTGGATCCGAGGG
IND-ØØ41Ø-5 T3   (1201) GTTCTTCCTGGTGGTTGTTACTTGTTGAAGAGACATGGTGGATCCGAGGG
IND-ØØ41Ø-5 T5   (1201) GTTCTTCCTGGTGGTTGTTACTTGTTGAAGAGACATGGTGGATCCGAGGG
IND-ØØ41Ø-5 T6   (1201) GTTCTTCCTGGTGGTTGTTACTTGTTGAAGAGACATGGTGGATCCGAGGG
     Consensus   (1201) GTTCTTCCTGGTGGTTGTTACTTGTTGAAGAGACATGGTGGATCCGAGGG
                      1251                                              1300
IND-ØØ41Ø-5 T1   (1251) TTTGATAAGTGATATTATATATGTGTATATATAGTGAGTGGTCTTTATAA
IND-ØØ41Ø-5 T3   (1251) TTTGATAAGTGATATTATATATGTGTATATATAGTGAGTGGTCTTTATAA
IND-ØØ41Ø-5 T5   (1251) TTTGATAAGTGATATTATATATGTGTATATATAGTGAGTGGTCTTTATAA
IND-ØØ41Ø-5 T6   (1251) TTTGATAAGTGATATTATATATGTGTATATATAGTGAGTGGTCTTTATAA
     Consensus   (1251) TTTGATAAGTGATATTATATATGTGTATATATAGTGAGTGGTCTTTATAA
                      1301                                              1350
IND-ØØ41Ø-5 T1   (1301) AGGTTCTTAAAAGTATGGGGTTTAGATGAAGCACAGGTGTACGCGTTGAA
IND-ØØ41Ø-5 T3   (1301) AGGTTCTTAAAAGTATGGGGTTTAGATGAAGCACAGGTGTACGCGTTGAA
IND-ØØ41Ø-5 T5   (1301) AGGTTCTTAAAAGTATGGGGTTTAGATGAAGCACAGGTGTACGCGTTGAA
IND-ØØ41Ø-5 T6   (1301) AGGTTCTTAAAAGTATGGGGTTTAGATGAAGCACAGGTGTACGCGTTGAA
     Consensus   (1301) AGGTTCTTAAAAGTATGGGGTTTAGATGAAGCACAGGTGTACGCGTTGAA
                      1351                                              1400
IND-ØØ41Ø-5 T1   (1351) TTTGACAAGAATCCTACCTGTTTTGTCCTTTTCAACGCAATGGTGAGTTG
IND-ØØ41Ø-5 T3   (1351) TTTGACAAGAATCCTACCTGTTTTGTCCTTTTCAACGCAATGGTGAGTTG
IND-ØØ41Ø-5 T5   (1351) TTTGACAAGAATCCTACCTGTTTTGTCCTTTTCAACGCAATGGTGAGTTG
IND-ØØ41Ø-5 T6   (1351) TTTGACAAGAATCCTACCTGTTTTGTCCTTTTCAACGCAATGGTGAGTTG
     Consensus   (1351) TTTGACAAGAATCCTACCTGTTTTGTCCTTTTCAACGCAATGGTGAGTTG
                      1401                                              1450
IND-ØØ41Ø-5 T1   (1401) TACGTTTACACGTTTTGGTTATGTCGATTCTCTGTCATGATGTCACTTTC
IND-ØØ41Ø-5 T3   (1401) TACGTTTACACGTTTTGGTTATGTCGATTCTCTGTCATGATGTCACTTTC
IND-ØØ41Ø-5 T5   (1401) TACGTTTACACGTTTTGGTTATGTCGATTCTCTGTCATGATGTCACTTTC
IND-ØØ41Ø-5 T6   (1401) TACGTTTACACGTTTTGGTTATGTCGATTCTCTGTCATGATGTCACTTTC
     Consensus   (1401) TACGTTTACACGTTTTGGTTATGTCGATTCTCTGTCATGATGTCACTTTC
                      1451                                              1500
IND-ØØ41Ø-5 T1   (1451) GGTTGAGATAAAACTCAAAATCATATGGTTTTGATATTTTGTCAATGAGT
IND-ØØ41Ø-5 T3   (1451) GGTTGAGATAAAACTCAAAATCATATGGTTTTGATATTTTGTCAATGAGT
IND-ØØ41Ø-5 T5   (1451) GGTTGAGATAAAACTCAAAATCATATGGTTTTGATATTTTGTCAATGAGT
IND-ØØ41Ø-5 T6   (1451) GGTTGAGATAAAACTCAAAATCATATGGTTTTGATATTTTGTCAATGAGT
     Consensus   (1451) GGTTGAGATAAAACTCAAAATCATATGGTTTTGATATTTTGTCAATGAGT
                      1501                                              1550
IND-ØØ41Ø-5 T1   (1501) CGTAGCTGGACTGGTAAGGGAGAGACCTGATGTTTCACGTGATCCAGACA
IND-ØØ41Ø-5 T3   (1501) CGTAGCTGGACTGGTAAGGGAGAGACCTGATGTTTCACGTGATCCAGACA
IND-ØØ41Ø-5 T5   (1501) CGTAGCTGGACTGGTAAGGGAGAGACCTGATGTTTCACGTGATCCAGACA
IND-ØØ41Ø-5 T6   (1501) CGTAGCTGGACTGGTAAGGGAGAGACCTGATGTTTCACGTGATCCAGACA
     Consensus   (1501) CGTAGCTGGACTGGTAAGGGAGAGACCTGATGTTTCACGTGATCCAGACA

                      1551                                              1600
```

```
IND-ØØ41Ø-5 T1   (1551) AAGGCGGACTTAGGTTAAGGGGTGGGTGGACCCTTGGAACTTTTTTAGTG
IND-ØØ41Ø-5 T3   (1551) AAGGCGGACTTAGGTTAAGGGGTGGGTGGACCCTTGGAACTTTTTTAGTG
IND-ØØ41Ø-5 T5   (1551) AAGGCGGACTTAGGTTAAGGGGTGGGTGGACCCTTGGAACTTTTTTAGTG
IND-ØØ41Ø-5 T6   (1551) AAGGCGGACTTAGGTTAAGGGGTGGGTGGACCCTTGGAACTTTTTTAGTG
     Consensus   (1551) AAGGCGGACTTAGGTTAAGGGGTGGGTGGACCCTTGGAACTTTTTTAGTG
                      1601                                              1650
IND-ØØ41Ø-5 T1   (1601) TATTTTATAGGTAAAATTTCCGATCGCACCTCTTAAAAGTATGGTTAAAC
IND-ØØ41Ø-5 T3   (1601) TATTTTATAGGTAAAATTTCCGATCGCACCTCTTAAAAGTATGGTTAAAC
IND-ØØ41Ø-5 T5   (1601) TATTTTATAGGTAAAATTTCCGATCGCACCTCTTAAAAGTATGGTTAAAC
IND-ØØ41Ø-5 T6   (1601) TATTTTATAGGTAAAATTTCCGATCGCACCTCTTAAAAGTATGGTTAAAC
     Consensus   (1601) TATTTTATAGGTAAAATTTCCGATCGCACCTCTTAAAAGTATGGTTAAAC
                      1651                                              1700
IND-ØØ41Ø-5 T1   (1651) CACTCGTTCGCATCCCTAGAAAAAAAATTCATGGGTGCGTCACTGGACCT
IND-ØØ41Ø-5 T3   (1651) CACTCGTTCGCATCCCTAGAAAAAAAATTCATGGGTGCGTCACTGGACCT
IND-ØØ41Ø-5 T5   (1651) CACTCGTTCGCATCCCTAGAAAAAAAATTCATGGGTGCGTCACTGGACCT
IND-ØØ41Ø-5 T6   (1651) CACTCGTTCGCATCCCTAGAAAAAAAATTCATGGGTGCGTCACTGGACCT
     Consensus   (1651) CACTCGTTCGCATCCCTAGAAAAAAAATTCATGGGTGCGTCACTGGACCT
                      1701                                              1750
IND-ØØ41Ø-5 T1   (1701) AGGTTCAACTCCCGCCCCCAGCATTTTGTTTCGGCGGCACCTGTTGATGA
IND-ØØ41Ø-5 T3   (1701) AGGTTCAACTCCCGCCCCCAGCATTTTGTTTCGGCGGCACCTGTTGATGA
IND-ØØ41Ø-5 T5   (1701) AGGTTCAACTCCCGCCCCCAGCATTTTGTTTCGGCGGCACCTGTTGATGA
IND-ØØ41Ø-5 T6   (1701) AGGTTCAACTCCCGCCCCCAGCATTTTGTTTCGGCGGCACCTGTTGATGA
     Consensus   (1701) AGGTTCAACTCCCGCCCCCAGCATTTTGTTTCGGCGGCACCTGTTGATGA
                      1751                                              1800
IND-ØØ41Ø-5 T1   (1751) TGGGAGATTAGGCGAGTAGGCAGCAATCGCTAGTTCGATCCTTGAACCAA
IND-ØØ41Ø-5 T3   (1751) TGGGAGATTAGGCGAGTAGGCAGCAATCGCTAGTTCGATCCTTGAACCAA
IND-ØØ41Ø-5 T5   (1751) TGGGAGATTAGGCGAGTAGGCAGCAATCGCTAGTTCGATCCTTGAACCAA
IND-ØØ41Ø-5 T6   (1751) TGGGAGATTAGGCGAGTAGGCAGCAATCGCTAGTTCGATCCTTGAACCAA
     Consensus   (1751) TGGGAGATTAGGCGAGTAGGCAGCAATCGCTAGTTCGATCCTTGAACCAA
                      1801                                              1850
IND-ØØ41Ø-5 T1   (1801) GCGGGTTTTACCTCACCGCAGTAGACGAGTGTATCCCAATGTGGTGAATT
IND-ØØ41Ø-5 T3   (1801) GCGGGTTTTACCTCACCGCAGTAGACGAGTGTATCCCAATGTGGTGAATT
IND-ØØ41Ø-5 T5   (1801) GCGGGTTTTACCTCACCGCAGTAGACGAGTGTATCCCAATGTGGTGAATT
IND-ØØ41Ø-5 T6   (1801) GCGGGTTTTACCTCACCGCAGTAGACGAGTGTATCCCAATGTGGTGAATT
     Consensus   (1801) GCGGGTTTTACCTCACCGCAGTAGACGAGTGTATCCCAATGTGGTGAATT
                      1851                                              1900
IND-ØØ41Ø-5 T1   (1851) TCGCAAGTAGTCCATTTGGAAGCGCCAATAGTCCAGTAAGACGAGTGAGG
IND-ØØ41Ø-5 T3   (1851) TCGCAAGTAGTCCATTTGGAAGCGCCAATAGTCCAGTAAGACGAGTGAGG
IND-ØØ41Ø-5 T5   (1851) TCGCAAGTAGTCCATTTGGAAGCGCCAATAGTCCAGTAAGACGAGTGAGG
IND-ØØ41Ø-5 T6   (1851) TCGCAAGTAGTCCATTTGGAAGCGCCAATAGTCCAGTAAGACGAGTGAGG
     Consensus   (1851) TCGCAAGTAGTCCATTTGGAAGCGCCAATAGTCCAGTAAGACGAGTGAGG
                      1901                                              1950
IND-ØØ41Ø-5 T1   (1901) GTTTTCCCTGGTGAGGATTCATTGGCTAGTAGACGAGTGAGGGTTTTCTT
IND-ØØ41Ø-5 T3   (1901) GTTTTCCCTGGTGAGGATTCATTGGCTAGTAGACGAGTGAGGGTTTTCTT
IND-ØØ41Ø-5 T5   (1901) GTTTTCCCTGGTGAGGATTCATTGGCTAGTAGACGAGTGAGGGTTTTCTT
IND-ØØ41Ø-5 T6   (1901) GTTTTCCCTGGTGAGGATTCATTGGCTAGTAGACGAGTGAGGGTTTTCTT
     Consensus   (1901) GTTTTCCCTGGTGAGGATTCATTGGCTAGTAGACGAGTGAGGGTTTTCTT
                      1951                                              2000
IND-ØØ41Ø-5 T1   (1951) CGGTTCAGAGACATTGGCGAAGCTTTGAAGGGGCCGGGAGGGGCGCCCGA
IND-ØØ41Ø-5 T3   (1951) CGGTTCAGAGACATTGGCGAAGCTTTGAAGGGGCCGGGAGGGGCGCCCGA
IND-ØØ41Ø-5 T5   (1951) CGGTTCAGAGACATTGGCGAAGCTTTGAAGGGGCCGGGAGGGGCGCCCGA
IND-ØØ41Ø-5 T6   (1951) CGGTTCAGAGACATTGGCGAAGCTTTGAAGGGGCCGGGAGGGGCGCCCGA
     Consensus   (1951) CGGTTCAGAGACATTGGCGAAGCTTTGAAGGGGCCGGGAGGGGCGCCCGA
                      2001                                              2050
IND-ØØ41Ø-5 T1   (2001) ACCCCCGAACTTTTTGCTCAGTAGTGTTATATATGTAGTTTTCGTATAGA
IND-ØØ41Ø-5 T3   (2001) ACCCCCGAACTTTTTGCTCAGTAGTGTTATATATGTAGTTTTCGTATAGA
IND-ØØ41Ø-5 T5   (2001) ACCCCCGAACTTTTTGCTCAGTAGTGTTATATATGTAGTTTTCGTATAGA
IND-ØØ41Ø-5 T6   (2001) ACCCCCGAACTTTTTGCTCAGTAGTGTTATATATGTAGTTTTCGTATAGA
     Consensus   (2001) ACCCCCGAACTTTTTGCTCAGTAGTGTTATATATGTAGTTTTCGTATAGA
                      2051                                              2100
IND-ØØ41Ø-5 T1   (2051) AATTTTGGGTATATACGTTTTCGACCCCCCGATTCTATAGAAATTTCGGG
IND-ØØ41Ø-5 T3   (2051) AATTTTGGGTATATACGTTTTCGACCCCCCGATTCTATAGAAATTTCGGG
IND-ØØ41Ø-5 T5   (2051) AATTTTGGGTATATACGTTTTCGACCCCCCGATTCTATAGAAATTTCGGG
IND-ØØ41Ø-5 T6   (2051) AATTTTGGGTATATACGTTTTCGACCCCCCGATTCTATAGAAATTTCGGG
     Consensus   (2051) AATTTTGGGTATATACGTTTTCGACCCCCCGATTCTATAGAAATTTCGGG

                      2101                                              2150
```

```
IND-ØØ41Ø-5 T1   (2101) TACATACGTTTTCCGACCCCCCGTCGAAAACGTCAAGCTTCGCCACTGTT
IND-ØØ41Ø-5 T3   (2101) TACATACGTTTTCCGACCCCCCGTCGAAAACGTCAAGCTTCGCCACTGTT
IND-ØØ41Ø-5 T5   (2101) TACATACGTTTTCCGACCCCCCGTCGAAAACGTCAAGCTTCGCCACTGTT
IND-ØØ41Ø-5 T6   (2101) TACATACGTTTTCCGACCCCCCGTCGAAAACGTCAAGCTTCGCCACTGTT
     Consensus   (2101) TACATACGTTTTCCGACCCCCCGTCGAAAACGTCAAGCTTCGCCACTGTT
                      2151                                            2200
IND-ØØ41Ø-5 T1   (2151) CAGAGATGAGTGTATTCTGATGCGATGAATTTCGTTAGTATTCTATTTGG
IND-ØØ41Ø-5 T3   (2151) CAGAGATGAGTGTATTCTGATGCGATGAATTTCGTTAGTATTCTATTTGG
IND-ØØ41Ø-5 T5   (2151) CAGAGATGAGTGTATTCTGATGCGATGAATTTCGTTAGTATTCTATTTGG
IND-ØØ41Ø-5 T6   (2151) CAGAGATGAGTGTATTCTGATGCGATGAATTTCGTTAGTATTCTATTTGG
     Consensus   (2151) CAGAGATGAGTGTATTCTGATGCGATGAATTTCGTTAGTATTCTATTTGG
                      2201                                            2250
IND-ØØ41Ø-5 T1   (2201) GGAATTTGTTTTCCACTAAAAAAAATTTGGTTTGATATTTTATGGGAAGA
IND-ØØ41Ø-5 T3   (2201) GGAATTTGTTTTCCACTAAAAAAAATTTGGTTTGATATTTTATGGGAAGA
IND-ØØ41Ø-5 T5   (2201) GGAATTTGTTTTCCACTAAAAAAAATTTGGTTTGATATTTTATGGGAAGA
IND-ØØ41Ø-5 T6   (2201) GGAATTTGTTTTCCACTAAAAAAAATTTGGTTTGATATTTTATGGGAAGA
     Consensus   (2201) GGAATTTGTTTTCCACTAAAAAAAATTTGGTTTGATATTTTATGGGAAGA
                      2251                                            2300
IND-ØØ41Ø-5 T1   (2251) GTTGTGTGCAAACAAAAGTACAAGTGGACATAGCGTGGGTGATTATTAAA
IND-ØØ41Ø-5 T3   (2251) GTTGTGTGCAAACAAAAGTACAAGTGGACATAGCGTGGGTGATTATTAAA
IND-ØØ41Ø-5 T5   (2251) GTTGTGTGCAAACAAAAGTACAAGTGGACATAGCGTGGGTGATTATTAAA
IND-ØØ41Ø-5 T6   (2251) GTTGTGTGCAAACAAAAGTACAAGTGGACATAGCGTGGGTGATTATTAAA
     Consensus   (2251) GTTGTGTGCAAACAAAAGTACAAGTGGACATAGCGTGGGTGATTATTAAA
                      2301                                            2350
IND-ØØ41Ø-5 T1   (2301) GTGGGAAATGGTAGGTAACAAATGGGGGTGGGGGCTACTACTTTTAATAT
IND-ØØ41Ø-5 T3   (2301) GTGGGAAATGGTAGGTAACAAATGGGGGTGGGGGCTACTACTTTTAATAT
IND-ØØ41Ø-5 T5   (2301) GTGGGAAATGGTAGGTAACAAATGGGGGTGGGGGCTACTACTTTTAATAT
IND-ØØ41Ø-5 T6   (2301) GTGGGAAATGGTAGGTAACAAATGGGGGTGGGGGCTACTACTTTTAATAT
     Consensus   (2301) GTGGGAAATGGTAGGTAACAAATGGGGGTGGGGGCTACTACTTTTAATAT
                      2351                                            2400
IND-ØØ41Ø-5 T1   (2351) GATTGACTAAAAGGAAATGAAATTGGTGAAGTGTGAATTGACACGTTTGT
IND-ØØ41Ø-5 T3   (2351) GATTGACTAAAAGGAAATGAAATTGGTGAAGTGTGAATTGACACGTTTGT
IND-ØØ41Ø-5 T5   (2351) GATTGACTAAAAGGAAATGAAATTGGTGAAGTGTGAATTGACACGTTTGT
IND-ØØ41Ø-5 T6   (2351) GATTGACTAAAAGGAAATGAAATTGGTGAAGTGTGAATTGACACGTTTGT
     Consensus   (2351) GATTGACTAAAAGGAAATGAAATTGGTGAAGTGTGAATTGACACGTTTGT
                      2401                                            2450
IND-ØØ41Ø-5 T1   (2401) AGGTGTGTTGTGGTGTGTTGGAAACGACAAAAGAGATAAGATACGATGTG
IND-ØØ41Ø-5 T3   (2401) AGGTGTGTTGTGGTGTGTTGGAAACGACAAAAGAGATAAGATACGATGTG
IND-ØØ41Ø-5 T5   (2401) AGGTGTGTTGTGGTGTGTTGGAAACGACAAAAGAGATAAGATACGATGTG
IND-ØØ41Ø-5 T6   (2401) AGGTGTGTTGTGGTGTGTTGGAAACGACAAAAGAGATAAGATACGATGTG
     Consensus   (2401) AGGTGTGTTGTGGTGTGTTGGAAACGACAAAAGAGATAAGATACGATGTG
                      2451                                            2500
IND-ØØ41Ø-5 T1   (2451) CCAGGTGTCGACCTGCAGGTCAACATGGTGGAGCACGACACACTTGTCTA
IND-ØØ41Ø-5 T3   (2451) CCAGGTGTCGACCTGCAGGTCAACATGGTGGAGCACGACACACTTGTCTA
IND-ØØ41Ø-5 T5   (2451) CCAGGTGTCGACCTGCAGGTCAACATGGTGGAGCACGACACACTTGTCTA
IND-ØØ41Ø-5 T6   (2451) CCAGGTGTCGACCTGCAGGTCAACATGGTGGAGCACGACACACTTGTCTA
     Consensus   (2451) CCAGGTGTCGACCTGCAGGTCAACATGGTGGAGCACGACACACTTGTCTA
                      2501                                            2550
IND-ØØ41Ø-5 T1   (2501) CTCCAAAAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCAATTGAGA
IND-ØØ41Ø-5 T3   (2501) CTCCAAAAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCAATTGAGA
IND-ØØ41Ø-5 T5   (2501) CTCCAAAAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCAATTGAGA
IND-ØØ41Ø-5 T6   (2501) CTCCAAAAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCAATTGAGA
     Consensus   (2501) CTCCAAAAATATCAAAGATACAGTCTCAGAAGACCAAAGGGCAATTGAGA
                      2551                                            2600
IND-ØØ41Ø-5 T1   (2551) CTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCCCA
IND-ØØ41Ø-5 T3   (2551) CTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCCCA
IND-ØØ41Ø-5 T5   (2551) CTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCCCA
IND-ØØ41Ø-5 T6   (2551) CTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCCCA
     Consensus   (2551) CTTTTCAACAAAGGGTAATATCCGGAAACCTCCTCGGATTCCATTGCCCA
                      2601                                            2650
IND-ØØ41Ø-5 T1   (2601) GCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTA
IND-ØØ41Ø-5 T3   (2601) GCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTA
IND-ØØ41Ø-5 T5   (2601) GCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTA
IND-ØØ41Ø-5 T6   (2601) GCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTA
     Consensus   (2601) GCTATCTGTCACTTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTA

                      2651                                            2700
```

```
IND-ØØ41Ø-5 T1   (2651) CAAATGCCATCATTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTG
IND-ØØ41Ø-5 T3   (2651) CAAATGCCATCATTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTG
IND-ØØ41Ø-5 T5   (2651) CAAATGCCATCATTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTG
IND-ØØ41Ø-5 T6   (2651) CAAATGCCATCATTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTG
     Consensus   (2651) CAAATGCCATCATTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTG
                      2701                                            2750
IND-ØØ41Ø-5 T1   (2701) CCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAA
IND-ØØ41Ø-5 T3   (2701) CCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAA
IND-ØØ41Ø-5 T5   (2701) CCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAA
IND-ØØ41Ø-5 T6   (2701) CCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAA
     Consensus   (2701) CCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAA
                      2751                                            2800
IND-ØØ41Ø-5 T1   (2751) AAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATAA
IND-ØØ41Ø-5 T3   (2751) AAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATAA
IND-ØØ41Ø-5 T5   (2751) AAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATAA
IND-ØØ41Ø-5 T6   (2751) AAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATAA
     Consensus   (2751) AAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATAA
                      2801                                            2850
IND-ØØ41Ø-5 T1   (2801) CATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAG
IND-ØØ41Ø-5 T3   (2801) CATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAG
IND-ØØ41Ø-5 T5   (2801) CATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAG
IND-ØØ41Ø-5 T6   (2801) CATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAG
     Consensus   (2801) CATGGTGGAGCACGACACACTTGTCTACTCCAAAAATATCAAAGATACAG
                      2851                                            2900
IND-ØØ41Ø-5 T1   (2851) TCTCAGAAGACCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCC
IND-ØØ41Ø-5 T3   (2851) TCTCAGAAGACCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCC
IND-ØØ41Ø-5 T5   (2851) TCTCAGAAGACCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCC
IND-ØØ41Ø-5 T6   (2851) TCTCAGAAGACCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCC
     Consensus   (2851) TCTCAGAAGACCAAAGGGCAATTGAGACTTTTCAACAAAGGGTAATATCC
                      2901                                            2950
IND-ØØ41Ø-5 T1   (2901) GGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTTATTGTGAA
IND-ØØ41Ø-5 T3   (2901) GGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTTATTGTGAA
IND-ØØ41Ø-5 T5   (2901) GGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTTATTGTGAA
IND-ØØ41Ø-5 T6   (2901) GGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTTATTGTGAA
     Consensus   (2901) GGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTTATTGTGAA
                      2951                                            3000
IND-ØØ41Ø-5 T1   (2951) GATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAG
IND-ØØ41Ø-5 T3   (2951) GATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAG
IND-ØØ41Ø-5 T5   (2951) GATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAG
IND-ØØ41Ø-5 T6   (2951) GATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAG
     Consensus   (2951) GATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCATTGCGATAAAG
                      3001                                            3050
IND-ØØ41Ø-5 T1   (3001) GAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGA
IND-ØØ41Ø-5 T3   (3001) GAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGA
IND-ØØ41Ø-5 T5   (3001) GAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGA
IND-ØØ41Ø-5 T6   (3001) GAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGA
     Consensus   (3001) GAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGA
                      3051                                            3100
IND-ØØ41Ø-5 T1   (3051) CCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTC
IND-ØØ41Ø-5 T3   (3051) CCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTC
IND-ØØ41Ø-5 T5   (3051) CCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTC
IND-ØØ41Ø-5 T6   (3051) CCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTC
     Consensus   (3051) CCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTC
                      3101                                            3150
IND-ØØ41Ø-5 T1   (3101) TTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACG
IND-ØØ41Ø-5 T3   (3101) TTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACG
IND-ØØ41Ø-5 T5   (3101) TTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACG
IND-ØØ41Ø-5 T6   (3101) TTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACG
     Consensus   (3101) TTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACG
                      3151                                            3200
IND-ØØ41Ø-5 T1   (3151) CACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCA
IND-ØØ41Ø-5 T3   (3151) CACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCA
IND-ØØ41Ø-5 T5   (3151) CACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCA
IND-ØØ41Ø-5 T6   (3151) CACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCA
     Consensus   (3151) CACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCA

                      3201                                            3250
```

```
IND-ØØ41Ø-5 T1   (3201) TTTCATTTGGAGAGGACCTCGAGAATTAATTCTCAACACAACATATACAA
IND-ØØ41Ø-5 T3   (3201) TTTCATTTGGAGAGGACCTCGAGAATTAATTCTCAACACAACATATACAA
IND-ØØ41Ø-5 T5   (3201) TTTCATTTGGAGAGGACCTCGAGAATTAATTCTCAACACAACATATACAA
IND-ØØ41Ø-5 T6   (3201) TTTCATTTGGAGAGGACCTCGAGAATTAATTCTCAACACAACATATACAA
     Consensus   (3201) TTTCATTTGGAGAGGACCTCGAGAATTAATTCTCAACACAACATATACAA
                      3251                                            3300
IND-ØØ41Ø-5 T1   (3251) AACAAACGAATCTCAAGCAATCAAGCATTCTACTTCTATTGCAGCAATTT
IND-ØØ41Ø-5 T3   (3251) AACAAACGAATCTCAAGCAATCAAGCATTCTACTTCTATTGCAGCAATTT
IND-ØØ41Ø-5 T5   (3251) AACAAACGAATCTCAAGCAATCAAGCATTCTACTTCTATTGCAGCAATTT
IND-ØØ41Ø-5 T6   (3251) AACAAACGAATCTCAAGCAATCAAGCATTCTACTTCTATTGCAGCAATTT
     Consensus   (3251) AACAAACGAATCTCAAGCAATCAAGCATTCTACTTCTATTGCAGCAATTT
                      3301                                            3350
IND-ØØ41Ø-5 T1   (3301) AAATCATTTCTTTTAAAGCAAAAGCAATTTTCTGAAAATTTTCACCATTT
IND-ØØ41Ø-5 T3   (3301) AAATCATTTCTTTTAAAGCAAAAGCAATTTTCTGAAAATTTTCACCATTT
IND-ØØ41Ø-5 T5   (3301) AAATCATTTCTTTTAAAGCAAAAGCAATTTTCTGAAAATTTTCACCATTT
IND-ØØ41Ø-5 T6   (3301) AAATCATTTCTTTTAAAGCAAAAGCAATTTTCTGAAAATTTTCACCATTT
     Consensus   (3301) AAATCATTTCTTTTAAAGCAAAAGCAATTTTCTGAAAATTTTCACCATTT
                      3351                                            3400
IND-ØØ41Ø-5 T1   (3351) ACGAACGGGGGATCTACCATGAGCCCAGAACGACGCCCGGCCGACATCCG
IND-ØØ41Ø-5 T3   (3351) ACGAACGGGGGATCTACCATGAGCCCAGAACGACGCCCGGCCGACATCCG
IND-ØØ41Ø-5 T5   (3351) ACGAACGGGGGATCTACCATGAGCCCAGAACGACGCCCGGCCGACATCCG
IND-ØØ41Ø-5 T6   (3351) ACGAACGGGGGATCTACCATGAGCCCAGAACGACGCCCGGCCGACATCCG
     Consensus   (3351) ACGAACGGGGGATCTACCATGAGCCCAGAACGACGCCCGGCCGACATCCG
                      3401                                            3450
IND-ØØ41Ø-5 T1   (3401) CCGTGCCACCGAGGCGGACATGCCGGCGGTCTGCACCATCGTCAACCACT
IND-ØØ41Ø-5 T3   (3401) CCGTGCCACCGAGGCGGACATGCCGGCGGTCTGCACCATCGTCAACCACT
IND-ØØ41Ø-5 T5   (3401) CCGTGCCACCGAGGCGGACATGCCGGCGGTCTGCACCATCGTCAACCACT
IND-ØØ41Ø-5 T6   (3401) CCGTGCCACCGAGGCGGACATGCCGGCGGTCTGCACCATCGTCAACCACT
     Consensus   (3401) CCGTGCCACCGAGGCGGACATGCCGGCGGTCTGCACCATCGTCAACCACT
                      3451                                            3500
IND-ØØ41Ø-5 T1   (3451) ACATCGAGACAAGCACGGTCAACTTCCGTACCGAGCCGCAGGAACCGCAG
IND-ØØ41Ø-5 T3   (3451) ACATCGAGACAAGCACGGTCAACTTCCGTACCGAGCCGCAGGAACCGCAG
IND-ØØ41Ø-5 T5   (3451) ACATCGAGACAAGCACGGTCAACTTCCGTACCGAGCCGCAGGAACCGCAG
IND-ØØ41Ø-5 T6   (3451) ACATCGAGACAAGCACGGTCAACTTCCGTACCGAGCCGCAGGAACCGCAG
     Consensus   (3451) ACATCGAGACAAGCACGGTCAACTTCCGTACCGAGCCGCAGGAACCGCAG
                      3501                                            3550
IND-ØØ41Ø-5 T1   (3501) GAGTGGACGGACGACCTCGTCCGTCTGCGGGAGCGCTATCCCTGGCTCGT
IND-ØØ41Ø-5 T3   (3501) GAGTGGACGGACGACCTCGTCCGTCTGCGGGAGCGCTATCCCTGGCTCGT
IND-ØØ41Ø-5 T5   (3501) GAGTGGACGGACGACCTCGTCCGTCTGCGGGAGCGCTATCCCTGGCTCGT
IND-ØØ41Ø-5 T6   (3501) GAGTGGACGGACGACCTCGTCCGTCTGCGGGAGCGCTATCCCTGGCTCGT
     Consensus   (3501) GAGTGGACGGACGACCTCGTCCGTCTGCGGGAGCGCTATCCCTGGCTCGT
                      3551                                            3600
IND-ØØ41Ø-5 T1   (3551) CGCCGAGGTGGACGGCGAGGTCGCCGGCATCGCCTACGCGGGCCCCTGGA
IND-ØØ41Ø-5 T3   (3551) CGCCGAGGTGGACGGCGAGGTCGCCGGCATCGCCTACGCGGGCCCCTGGA
IND-ØØ41Ø-5 T5   (3551) CGCCGAGGTGGACGGCGAGGTCGCCGGCATCGCCTACGCGGGCCCCTGGA
IND-ØØ41Ø-5 T6   (3551) CGCCGAGGTGGACGGCGAGGTCGCCGGCATCGCCTACGCGGGCCCCTGGA
     Consensus   (3551) CGCCGAGGTGGACGGCGAGGTCGCCGGCATCGCCTACGCGGGCCCCTGGA
                      3601                                            3650
IND-ØØ41Ø-5 T1   (3601) AGGCACGCAACGCCTACGACTGGACGGCCGAGTCGACCGTGTACGTCTCC
IND-ØØ41Ø-5 T3   (3601) AGGCACGCAACGCCTACGACTGGACGGCCGAGTCGACCGTGTACGTCTCC
IND-ØØ41Ø-5 T5   (3601) AGGCACGCAACGCCTACGACTGGACGGCCGAGTCGACCGTGTACGTCTCC
IND-ØØ41Ø-5 T6   (3601) AGGCACGCAACGCCTACGACTGGACGGCCGAGTCGACCGTGTACGTCTCC
     Consensus   (3601) AGGCACGCAACGCCTACGACTGGACGGCCGAGTCGACCGTGTACGTCTCC
                      3651                                            3700
IND-ØØ41Ø-5 T1   (3651) CCCCGCCACCAGCGGACGGGACTGGGCTCCACGCTCTACACCCACCTGCT
IND-ØØ41Ø-5 T3   (3651) CCCCGCCACCAGCGGACGGGACTGGGCTCCACGCTCTACACCCACCTGCT
IND-ØØ41Ø-5 T5   (3651) CCCCGCCACCAGCGGACGGGACTGGGCTCCACGCTCTACACCCACCTGCT
IND-ØØ41Ø-5 T6   (3651) CCCCGCCACCAGCGGACGGGACTGGGCTCCACGCTCTACACCCACCTGCT
     Consensus   (3651) CCCCGCCACCAGCGGACGGGACTGGGCTCCACGCTCTACACCCACCTGCT
                      3701                                            3750
IND-ØØ41Ø-5 T1   (3701) GAAGTCCCTGGAGGCACAGGGCTTCAAGAGCGTGGTCGCTGTCATCGGGC
IND-ØØ41Ø-5 T3   (3701) GAAGTCCCTGGAGGCACAGGGCTTCAAGAGCGTGGTCGCTGTCATCGGGC
IND-ØØ41Ø-5 T5   (3701) GAAGTCCCTGGAGGCACAGGGCTTCAAGAGCGTGGTCGCTGTCATCGGGC
IND-ØØ41Ø-5 T6   (3701) GAAGTCCCTGGAGGCACAGGGCTTCAAGAGCGTGGTCGCTGTCATCGGGC
     Consensus   (3701) GAAGTCCCTGGAGGCACAGGGCTTCAAGAGCGTGGTCGCTGTCATCGGGC

                      3751                                            3800
```

```
IND-ØØ41Ø-5 T1   (3751) TGCCCAACGACCCGAGCGTGCGCATGCACGAGGCGCTCGGATATGCCCCC
IND-ØØ41Ø-5 T3   (3751) TGCCCAACGACCCGAGCGTGCGCATGCACGAGGCGCTCGGATATGCCCCC
IND-ØØ41Ø-5 T5   (3751) TGCCCAACGACCCGAGCGTGCGCATGCACGAGGCGCTCGGATATGCCCCC
IND-ØØ41Ø-5 T6   (3751) TGCCCAACGACCCGAGCGTGCGCATGCACGAGGCGCTCGGATATGCCCCC
     Consensus   (3751) TGCCCAACGACCCGAGCGTGCGCATGCACGAGGCGCTCGGATATGCCCCC
                      3801                                            3850
IND-ØØ41Ø-5 T1   (3801) CGCGGCATGCTGCGGGCGGCCGGCTTCAAGCACGGGAACTGGCATGACGT
IND-ØØ41Ø-5 T3   (3801) CGCGGCATGCTGCGGGCGGCCGGCTTCAAGCACGGGAACTGGCATGACGT
IND-ØØ41Ø-5 T5   (3801) CGCGGCATGCTGCGGGCGGCCGGCTTCAAGCACGGGAACTGGCATGACGT
IND-ØØ41Ø-5 T6   (3801) CGCGGCATGCTGCGGGCGGCCGGCTTCAAGCACGGGAACTGGCATGACGT
     Consensus   (3801) CGCGGCATGCTGCGGGCGGCCGGCTTCAAGCACGGGAACTGGCATGACGT
                      3851                                            3900
IND-ØØ41Ø-5 T1   (3851) GGGTTTCTGGCAGCTGGACTTCAGCCTGCCGGTACCGCCCCGTCCGGTCC
IND-ØØ41Ø-5 T3   (3851) GGGTTTCTGGCAGCTGGACTTCAGCCTGCCGGTACCGCCCCGTCCGGTCC
IND-ØØ41Ø-5 T5   (3851) GGGTTTCTGGCAGCTGGACTTCAGCCTGCCGGTACCGCCCCGTCCGGTCC
IND-ØØ41Ø-5 T6   (3851) GGGTTTCTGGCAGCTGGACTTCAGCCTGCCGGTACCGCCCCGTCCGGTCC
     Consensus   (3851) GGGTTTCTGGCAGCTGGACTTCAGCCTGCCGGTACCGCCCCGTCCGGTCC
                      3901                                            3950
IND-ØØ41Ø-5 T1   (3901) TGCCCGTCACCGAGATCTGCTCAACAATCTAGCTAGAGTTTGCTCCTATC
IND-ØØ41Ø-5 T3   (3901) TGCCCGTCACCGAGATCTGCTCAACAATCTAGCTAGAGTTTGCTCCTATC
IND-ØØ41Ø-5 T5   (3901) TGCCCGTCACCGAGATCTGCTCAACAATCTAGCTAGAGTTTGCTCCTATC
IND-ØØ41Ø-5 T6   (3901) TGCCCGTCACCGAGATCTGCTCAACAATCTAGCTAGAGTTTGCTCCTATC
     Consensus   (3901) TGCCCGTCACCGAGATCTGCTCAACAATCTAGCTAGAGTTTGCTCCTATC
                      3951                                            4000
IND-ØØ41Ø-5 T1   (3951) TATATGTAATAAGGTATGCTGATATGCACTATTCAAATAGGAGCATTAGC
IND-ØØ41Ø-5 T3   (3951) TATATGTAATAAGGTATGCTGATATGCACTATTCAAATAGGAGCATTAGC
IND-ØØ41Ø-5 T5   (3951) TATATGTAATAAGGTATGCTGATATGCACTATTCAAATAGGAGCATTAGC
IND-ØØ41Ø-5 T6   (3951) TATATGTAATAAGGTATGCTGATATGCACTATTCAAATAGGAGCATTAGC
     Consensus   (3951) TATATGTAATAAGGTATGCTGATATGCACTATTCAAATAGGAGCATTAGC
                      4001                                            4050
IND-ØØ41Ø-5 T1   (4001) TATGTTTGTTAATGTCACTTTATGTTATGTGGGTAAGTCACCTAAGACAC
IND-ØØ41Ø-5 T3   (4001) TATGTTTGTTAATGTCACTTTATGTTATGTGGGTAAGTCACCTAAGACAC
IND-ØØ41Ø-5 T5   (4001) TATGTTTGTTAATGTCACTTTATGTTATGTGGGTAAGTCACCTAAGACAC
IND-ØØ41Ø-5 T6   (4001) TATGTTTGTTAATGTCACTTTATGTTATGTGGGTAAGTCACCTAAGACAC
     Consensus   (4001) TATGTTTGTTAATGTCACTTTATGTTATGTGGGTAAGTCACCTAAGACAC
                      4051                                            4100
IND-ØØ41Ø-5 T1   (4051) TCCACGTACCTACTTGTTGTCTCTTACGCGGCTTTAATAAATCTTCTGCC
IND-ØØ41Ø-5 T3   (4051) TCCACGTACCTACTTGTTGTCTCTTACGCGGCTTTAATAAATCTTCTGCC
IND-ØØ41Ø-5 T5   (4051) TCCACGTACCTACTTGTTGTCTCTTACGCGGCTTTAATAAATCTTCTGCC
IND-ØØ41Ø-5 T6   (4051) TCCACGTACCTACTTGTTGTCTCTTACGCGGCTTTAATAAATCTTCTGCC
     Consensus   (4051) TCCACGTACCTACTTGTTGTCTCTTACGCGGCTTTAATAAATCTTCTGCC
                      4101                                            4150
IND-ØØ41Ø-5 T1   (4101) CTTGTTCCATATTTACTAATTATCCCTTTCTTCACTAAAAGAAAATTGTT
IND-ØØ41Ø-5 T3   (4101) CTTGTTCCATATTTACTAATTATCCCTTTCTTCACTAAAAGAAAATTGTT
IND-ØØ41Ø-5 T5   (4101) CTTGTTCCATATTTACTAATTATCCCTTTCTTCACTAAAAGAAAATTGTT
IND-ØØ41Ø-5 T6   (4101) CTTGTTCCATATTTACTAATTATCCCTTTCTTCACTAAAAGAAAATTGTT
     Consensus   (4101) CTTGTTCCATATTTACTAATTATCCCTTTCTTCACTAAAAGAAAATTGTT
                      4151                                            4200
IND-ØØ41Ø-5 T1   (4151) ATCATTAAGTATTAGTCTTTAGAACATATGAGGTCTTTAATTGGGTAGGT
IND-ØØ41Ø-5 T3   (4151) ATCATTAAGTATTAGTCTTTAGAACATATGAGGTCTTTAATTGGGTAGGT
IND-ØØ41Ø-5 T5   (4151) ATCATTAAGTATTAGTCTTTAGAACATATGAGGTCTTTAATTGGGTAGGT
IND-ØØ41Ø-5 T6   (4151) ATCATTAAGTATTAGTCTTTAGAACATATGAGGTCTTTAATTGGGTAGGT
     Consensus   (4151) ATCATTAAGTATTAGTCTTTAGAACATATGAGGTCTTTAATTGGGTAGGT
                      4201                                            4250
IND-ØØ41Ø-5 T1   (4201) TTTACAAATTAACTAATATAAAATGTCATAAAATCCACGTGGTTAAACAA
IND-ØØ41Ø-5 T3   (4201) TTTACAAATTAACTAATATAAAATGTCATAAAATCCACGTGGTTAAACAA
IND-ØØ41Ø-5 T5   (4201) TTTACAAATTAACTAATATAAAATGTCATAAAATCCACGTGGTTAAACAA
IND-ØØ41Ø-5 T6   (4201) TTTACAAATTAACTAATATAAAATGTCATAAAATCCACGTGGTTAAACAA
     Consensus   (4201) TTTACAAATTAACTAATATAAAATGTCATAAAATCCACGTGGTTAAACAA
                      4251                                            4300
IND-ØØ41Ø-5 T1   (4251) ATGCAGAAAATCGACGTCGTCTATTGGACCGACAGTTGCTATTAATATAA
IND-ØØ41Ø-5 T3   (4251) ATGCAGAAAATCGACGTCGTCTATTGGACCGACAGTTGCTATTAATATAA
IND-ØØ41Ø-5 T5   (4251) ATGCAGAAAATCGACGTCGTCTATTGGACCGACAGTTGCTATTAATATAA
IND-ØØ41Ø-5 T6   (4251) ATGCAGAAAATCGACGTCGTCTATTGGACCGACAGTTGCTATTAATATAA
     Consensus   (4251) ATGCAGAAAATCGACGTCGTCTATTGGACCGACAGTTGCTATTAATATAA

                      4301                                            4350
```

```
IND-ØØ41Ø-5 T1   (4301) TGGGCCACCATAGTAGACTGACAAATAAATTACCTGACAACATCGTTTCA
IND-ØØ41Ø-5 T3   (4301) TGGGCCACCATAGTAGACTGACAAATAAATTACCTGACAACATCGTTTCA
IND-ØØ41Ø-5 T5   (4301) TGGGCCACCATAGTAGACTGACAAATAAATTACCTGACAACATCGTTTCA
IND-ØØ41Ø-5 T6   (4301) TGGGCCACCATAGTAGACTGACAAATAAATTACCTGACAACATCGTTTCA
     Consensus   (4301) TGGGCCACCATAGTAGACTGACAAATAAATTACCTGACAACATCGTTTCA
                      4351                                            4400
IND-ØØ41Ø-5 T1   (4351) CAAAAAAACAAACACAAAAAGGGAGTGCATTTTCCAGGGCATTTTTGTAA
IND-ØØ41Ø-5 T3   (4351) CAAAAAAACAAACACAAAAAGGGAGTGCATTTTCCAGGGCATTTTTGTAA
IND-ØØ41Ø-5 T5   (4351) CAAAAAAACAAACACAAAAAGGGAGTGCATTTTCCAGGGCATTTTTGTAA
IND-ØØ41Ø-5 T6   (4351) CAAAAAAACAAACACAAAAAGGGAGTGCATTTTCCAGGGCATTTTTGTAA
     Consensus   (4351) CAAAAAAACAAACACAAAAAGGGAGTGCATTTTCCAGGGCATTTTTGTAA
                      4401                                            4450
IND-ØØ41Ø-5 T1   (4401) TAAAAAACAGTTAAAAGGGAGTGCAATAGAAATATAGGGGTGTGGAAATA
IND-ØØ41Ø-5 T3   (4401) TAAAAAACAGTTAAAAGGGAGTGCAATAGAAATATAGGGGTGTGGAAATA
IND-ØØ41Ø-5 T5   (4401) TAAAAAACAGTTAAAAGGGAGTGCAATAGAAATATAGGGGTGTGGAAATA
IND-ØØ41Ø-5 T6   (4401) TAAAAAACAGTTAAAAGGGAGTGCAATAGAAATATAGGGGTGTGGAAATA
     Consensus   (4401) TAAAAAACAGTTAAAAGGGAGTGCAATAGAAATATAGGGGTGTGGAAATA
                      4451                                            4500
IND-ØØ41Ø-5 T1   (4451) GTGATTTGAGCACGTCTTGAAGCGAATTAGCTTGGCACTGGCCGTCGTTT
IND-ØØ41Ø-5 T3   (4451) GTGATTTGAGCACGTCTTGAAGCGAATTAGCTTGGCACTGGCCGTCGTTT
IND-ØØ41Ø-5 T5   (4451) GTGATTTGAGCACGTCTTGAAGCGAATTAGCTTGGCACTGGCCGTCGTTT
IND-ØØ41Ø-5 T6   (4451) GTGATTTGAGCACGTCTTGAAGCGAATTAGCTTGGCACTGGCCGTCGTTT
     Consensus   (4451) GTGATTTGAGCACGTCTTGAAGCGAATTAGCTTGGCACTGGCCGTCGTTT
                      4501                                            4550
IND-ØØ41Ø-5 T1   (4501) TACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTT
IND-ØØ41Ø-5 T3   (4501) TACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTT
IND-ØØ41Ø-5 T5   (4501) TACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTT
IND-ØØ41Ø-5 T6   (4501) TACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTT
     Consensus   (4501) TACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTT
                      4551                                            4600
IND-ØØ41Ø-5 T1   (4551) GCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCAC
IND-ØØ41Ø-5 T3   (4551) GCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCAC
IND-ØØ41Ø-5 T5   (4551) GCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCAC
IND-ØØ41Ø-5 T6   (4551) GCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCAC
     Consensus   (4551) GCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCAC
                      4601                                            4650
IND-ØØ41Ø-5 T1   (4601) CGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGCTAGAGCA
IND-ØØ41Ø-5 T3   (4601) CGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGCTAGAGCA
IND-ØØ41Ø-5 T5   (4601) CGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGCTAGAGCA
IND-ØØ41Ø-5 T6   (4601) CGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGCTAGAGCA
     Consensus   (4601) CGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGCTAGAGCA
                      4651                                            4700
IND-ØØ41Ø-5 T1   (4651) ATTCGGCGTTAATTCAGTACATTAAAAACGTCCGCAATGTGTTATTAAGT
IND-ØØ41Ø-5 T3   (4651) ATTCGGCGTTAATTCAGTACATTAAAAACGTCCGCAATGTGTTATTAAGT
IND-ØØ41Ø-5 T5   (4651) ATTCGGCGTTAATTCAGTACATTAAAAACGTCCGCAATGTGTTATTAAGT
IND-ØØ41Ø-5 T6   (4651) ATTCGGCGTTAATTCAGTACATTAAAAACGTCCGCAATGTGTTATTAAGT
     Consensus   (4651) ATTCGGCGTTAATTCAGTACATTAAAAACGTCCGCAATGTGTTATTAAGT
                      4701                                            4750
IND-ØØ41Ø-5 T1   (4701) TGTCTAAGCGTCAATTTGTTTACACCACAATATATCCTGATATCTTTAGT
IND-ØØ41Ø-5 T3   (4701) TGTCTAAGCGTCAATTTGTTTACACCACAATATATCCTGATATCTTTAGT
IND-ØØ41Ø-5 T5   (4701) TGTCTAAGCGTCAATTTGTTTACACCACAATATATCCTGATATCTTTAGT
IND-ØØ41Ø-5 T6   (4701) TGTCTAAGCGTCAATTTGTTTACACCACAATATATCCTGATATCTTTAGT
     Consensus   (4701) TGTCTAAGCGTCAATTTGTTTACACCACAATATATCCTGATATCTTTAGT
                      4751                                            4800
IND-ØØ41Ø-5 T1   (4751) TAGTTTGGAAAAGAATAATTTAGTTTATTTTCAAGAATTTATTTGTTTCA
IND-ØØ41Ø-5 T3   (4751) TAGTTTGGAAAAGAATAATTTAGTTTATTTTCAAGAATTTATTTGTTTCA
IND-ØØ41Ø-5 T5   (4751) TAGTTTGGAAAAGAATAATTTAGTTTATTTTCAAGAATTTATTTGTTTCA
IND-ØØ41Ø-5 T6   (4751) TAGTTTGGAAAAGAATAATTTAGTTTATTTTCAAGAATTTATTTGTTTCA
     Consensus   (4751) TAGTTTGGAAAAGAATAATTTAGTTTATTTTCAAGAATTTATTTGTTTCA
                      4801                                            4850
IND-ØØ41Ø-5 T1   (4801) AATTTTTAGAATATTATTTAGATTTAAAAATATCATGAATATTTTAATAT
IND-ØØ41Ø-5 T3   (4801) AATTTTTAGAATATTATTTAGATTTAAAAATATCATGAATATTTTAATAT
IND-ØØ41Ø-5 T5   (4801) AATTTTTAGAATATTATTTAGATTTAAAAATATCATGAATATTTTAATAT
IND-ØØ41Ø-5 T6   (4801) AATTTTTAGAATATTATTTAGATTTAAAAATATCATGAATATTTTAATAT
     Consensus   (4801) AATTTTTAGAATATTATTTAGATTTAAAAATATCATGAATATTTTAATAT

                      4851                                            4900
```

```
IND-ØØ41Ø-5 T1   (4851) TTGTTTTTTTAATATTATAATGTTAGATTTATAAATAAAGTTTCTTGATT
IND-ØØ41Ø-5 T3   (4851) TTGTTTTTTTAATATTATAATGTTAGATTTATAAATAAAGTTTCTTGATT
IND-ØØ41Ø-5 T5   (4851) TTGTTTTTTTAATATTATAATGTTAGATTTATAAATAAAGTTTCTTGATT
IND-ØØ41Ø-5 T6   (4851) TTGTTTTTTTAATATTATAATGTTAGATTTATAAATAAAGTTTCTTGATT
     Consensus   (4851) TTGTTTTTTTAATATTATAATGTTAGATTTATAAATAAAGTTTCTTGATT
                      4901                        4930
IND-ØØ41Ø-5 T1   (4901) TGAGAAAAAAATCACTCAGGCTGACCCAAC
IND-ØØ41Ø-5 T3   (4901) TGAGAAAAAAATCACTCAGGCTGACCCAAC
IND-ØØ41Ø-5 T5   (4901) TGAGAAAAAAATCACTCAGGCTGACCCAAC
IND-ØØ41Ø-5 T6   (4901) TGAGAAAAAAATCACTCAGGCTGACCCAAC
     Consensus   (4901) TGAGAAAAAAATCACTCAGGCTGACCCAAC
```