Open Access

OBM Genetics

Original Research

Genetic Stability, Inheritance Patterns and Expression Stability in Biotech Crops

Laura Privalle[1, *], Patricia Back [2], Apurva Bhargava [1], Zach Bishop [1], Krystal Cisneros [1], Isabelle Coats [1], Ine Criel [2], Lien Dhondt [2], Travis Draughn [1], Barb Fowler [3], Brad Franklin [1], Durba Ghoshal [1], Jim Lor [1], Jennifer Massengil [1], Sofie Moens [2], Tyson Mooney [1], Dannyel Nelson [1], Karolien Peeters [2], Sashi Sathischandra [1], Caroline Staut [2], Yoonhui Sung [1], Ann Tuttle [1], Annelies Van Hoecke [2], Annelies Van Raemdonck [2], Marie-Laure Verdegem [2], Steven Verhaeghe [2], Shane Walsh [1], Ann Wierckx [2], Qiang Zhao [1], Rozemarijn Dreesen [2]

1. BASF, 407 Davis Drive, Morrisville, NC 27560, USA; E-Mails: laura.privalle@basf.com; Apurva.bhargava@basf.com; zachary.bishop@basf.com; Krystal.cisneros@basf.com; isabelle.coats@basf.com; travis.draughn@basf.com; roy.franklin@basf.com; durba.ghoshal@basf.com; jim.lor@basf.com; jennifer.massengil@basf.com; Tyson.mooney@basf.com; dannyel.nelson@basf.com; Sashiprabha.sathischyandra@basf.com; Yoonhui.sung@basf.com; Annmarie.tuttle@basf.com; Shane.walsh@basf.com; qiang.zhao@basf.com
2. BASF, Belgium Coordination Center CommV, Technologiepark 101, 9052 Gent Zwijnaarde, Belgium; E-Mails: patricia.back@basf.com; Ine.criel@basf.com; lien.dhondt@basf.com; sofie.moens@basf.com; Karolien.peeters@basf.com; caroline.staut@basf.com; Annelies.van-hoecke@basf.com; annelies.van-raemdonck@basf.com; marie-laure.verdegem@basf.com; steven.verhaeghe@basf.com; ann.wierckx@basf.com; Rozemarijn.dreesen@basf.com
3. BASF Canada Inc., 120-343 70th Street East, S7P 0E1 Saskatoon, Sk., Canada; E-Mail: barbara.fowler@basf.com

* **Correspondence:** Laura Privalle; E-Mail: laura.privalle@basf.com

Academic Editor: Yuri Shavrukov

Special Issue: Plant Genetics and Gene Analysis

OBM Genetics
2020, volume 4, issue 4
doi:10.21926/obm.genet.2004120

Received: September 11, 2020
Accepted: November 20, 2020
Published: December 02, 2020

Abstract

Demonstration of the stability of traits newly introduced into a plant genome via genetic engineering approaches comprise a significant portion of the safety assessment that these products undergo prior to receiving the requisite regulatory approvals enabling commercial authorization. Different regions of the world have different regulatory requirements and many ask similar questions from multiple and overlapping perspectives. The entire central dogma, that is stability at the DNA level, mRNA level and protein level, is assessed for each product, although only a few regulatory authorities request data at the mRNA level. In this article, we present inheritance data obtained during the safety assessment of biotech products representing specific transgenic events in several crop species including *Brassica napus* (canola); canola quality *Brassica juncea* (yellow seeded canola); *Glycine max* (soybean), and *Gossypium hirsutum* (cotton) in which different traits have been introduced. The data presented confirm that all events examined were nuclear insertions that resulted in typical Mendelian Inheritance patterns and that the proteins are expressed similarly across multiple generations regardless of whether they were from backcrossed or outcrossed generations. These results demonstrate that newly inserted genes are transmitted to their progeny in a stable manner similar to that of endogenous genes. Further, the findings demonstrate that assessments of multigenerational stability have very limited value to a safety assessment.

Keywords

Plant biotechnology; stability; expression; insert characterization; inheritance patterns

1. Introduction

Genetically modified crops such as maize, cotton, soybean and canola, containing biotechnology derived agronomic traits, have been rapidly adopted by growers around the world over the past 25 years [1]. The majority of these crops express novel proteins and have undergone pre-market regulatory assessments prior to product authorization and commercialization. To properly conduct a regulatory assessment, the safety of the newly expressed protein is integral [2], along with in depth characterization of the event at the molecular level and aspects of its phenotypic/agronomic performance.

In the context of this paper, an “event” is defined as a unique insertion occurrence, which includes the inserted DNA comprising at least one gene cassette, as well as the plant genomic flanking region. As part of the characterization of an event, expression of the protein(s) is determined. Information on the expression levels of the proteins in plants produced using biotechnology approaches is necessary so that safety margins can be defined for feeding and ecotoxicological studies that form a part of the safety assessment of such products; to generate information for product labels necessary for pesticidal products; and to develop product management practices, such as insect resistance management, to ensure product performance. In addition, a molecular characterization of the event is undertaken, which provides information on the structure and expression of the inserted DNA and on the stability of the intended trait(s) encoded by this DNA region. The following points are routinely addressed: 1) genetic stability of the

(trans)gene(s) and the integration locus; 2) inheritance pattern of the event; and 3) stability of expression at the transcript (required only in a few geographies), and protein level across multiple generations. These assessment points are addressed by multiple analytical approaches and comparable guidance is provided for such studies by different regulatory bodies across the world.

While stability studies form part of the product molecular characterization in the context of product risk assessment, the issue of genetic stability, inheritance patterns and expression stability is clearly related also to seed product quality and performance. If breeders and growers could not rely on the consistency of the product performance, the product would not be purchased.

To date, little has been published on the stability analyses of commercial biotech products. However, the research of Qin *et al,* [3] demonstrated stability of a rice event over three successive generations with respect to agronomic traits, Mendelian inheritance patterns, transgene integrity, flanking sequence, copy number and transgene expression. More recently, Betts *et al.* [4] showed the stability of NPTII protein concentrations in maize leaves across successive generations. In this article, we present inheritance data generated in the context of the molecular characterization for the regulatory assessment of specific commercial events in several crop species including *Brassica napus* (canola), canola quality *Brassica juncea* (yellow seeded canola), *Glycine max* (soybean), and *Gossypium hirsutum* (cotton) in which different traits have been introduced. All data presented have been included in regulatory submissions for some regions of the world.

The events and the newly introduced genes for which results are presented are summarized in Table 1 and include:

1) MS11 *B. napus*, containing 3 gene cassettes (*barnase, barstar, bar*) https://www.aphis.usda.gov/brs/aphisdocs/16_23501p_a1.pdf

2) RF3 *B. napus* containing 2 gene cassettes (*barstar, bar*) https://www.aphis.usda.gov/brs/aphisdocs/98_27801p.pdf

3) RF3 *B. juncea*, containing 2 gene cassettes (*barstar, bar*)

4) GHB811 *G. hirsutum* containing 2 gene cassettes (*hppdPfW336-1Pa, 2mepsps*) https://www.aphis.usda.gov/brs/aphisdocs/17_13801p.pdf

5) 5547-127 *G. max* containing 1 gene cassette (*pat*). https://www.aphis.usda.gov/brs/aphisdocs/98_01401p.pdf

All events, except for RF3 *B. juncea,* were obtained by *Agrobacterium*-mediated transformation. RF3 *B. juncea* was obtained by conventional breeding with RF3 *B. napus*. In the events described in this paper, different types of promoters were used to modulate the expression of the newly introduced genes (Table 1). The promoters are either tissue-specific (tapetum), weakly constitutive or strongly constitutive. The introduced traits allow for sterility (i.e., Barnase expression in the tapetum of *Brassica* sp.), enhanced transformation frequency (i.e., Barstar in MS11 *B. napus*), or herbicide tolerance to either glyphosate (expression of 5-enolypyruvylshikimate 3-phosphate synthase-, (2mEPSPS)), glufosinate (expression of phosphinothricin acetyltransferase (PAT)) or HPPD inhibitor herbicides such as isoxaflutole (expression of 4-hydroxyphenylpyruvate dioxygenase HPPD W336).

The stability of these events was assessed across different breeding generations, by generating data for 1) the sequence of the inserted DNA over generations; 2) size and copy number of all detectable inserts; 3) genotypic and phenotypic stability and 4) protein and mRNA expression.

Table 1 Gene Cassettes Contained in Events for Which Data is Included.

Event	USDA Petition No.	Crop	Gene of Interest	Origin/Reference	Promoter	Promoter Description
MS11	16-235-01p	*B. napus* Canola	*barnase*	*B. amyloliquefaciens* [5]	Pta29	Tapetum-specific [6]
			barstar	*B. amyloliquefaciens* [5]	Pnos	Weakly constitutive [7]
			bar	*S. hygroscopicus* [8]	PssuAt	Strongly constitutive, green tissues [9]
RF3	98-278-01p	*B. juncea* Canola *B. napus* Canola	*bar*	*S. hygroscopicus* [8]	PssuAt	Strongly constitutive, green tissues [9]
			barstar	*B. amyloliquefaciens* [5]	Pnos	Weakly constitutive [7]
GHB811	17-138-01p	*G. hirsutum* Cotton	*hppdPfW336-1Pa*	*P. fluorescens* [10]	Pcsmvmv	Strongly constitutive [11]
			2mepsps	*Z. mays* [12]	Ph4a748	Strongly constitutive [13]
A5547-127	96-068-01p	*G. max* Soybean	*pat*	*S. viridochromegenes* [14]	P35S	Strongly constitutive [15, 16]

2. Materials and Methods

2.1 Greenhouse Production of Plant Samples

To limit variation due to environmental factors, plant materials used in expression characterization studies were produced within a single greenhouse production for each event. Various tissues of young and flowering plants as well as mature seeds from multiple breeding generations were sampled at standardized maturity stages for each crop [17]. The combination of a given plant tissue and maturity stage was defined as a matrix. For protein analysis, corresponding matrices, such as leaf, root, etc., from at least 4 individual plants were sampled separately, while for RNA studies, corresponding matrices from 5 individual plants were composited prior to sampling, resulting in a single biological replicate from 5 individual plants.

2.2 Processing of Plant Samples

Plant samples were ground to a fine powder. Grinding was performed in the presence of dry ice and/or liquid nitrogen. Processed samples were lyophilized prior to protein extraction and analysis. The percent dry weight (% DW) of each sample was determined from the fresh weight (FW) of the sample prior to lyophilization and the dry weight (DW) of the sample after lyophilization. For protein expression analysis, pollen samples were not processed or lyophilized.

Leaf discs from greenhouse grown plants were used to extract gDNA for Mendelian inheritance analysis.

2.3 Over-generation Insert Stability Analysis of MS11 B. napus Using Southern Blot Analysis

DNA from the transforming plasmid pTC0113 (https://www.aphis.usda.gov/brs/aphisdocs/16_23501p_a1.pdf) was digested using the restriction enzyme *Eco*RI (New England BioLabs) and used as a positive control. Genomic DNA (gDNA) was isolated from leaf material from individual plants, according to Dellaporta *et al.* [18]. Individual gDNA samples were digested with *Eco*RV (New England BioLabs). A 1 % TAE agarose gel was prepared and loaded with three individual DNA samples for each of the five breeding generations investigated, a negative control (digested gDNA of non-GM counterpart), a positive control (equimolar amount of digested pTCO113 DNA), and DIG-labeled molecular mass marker VII (Roche Applied Science). The positive control and the molecular mass marker were spiked in digested non-GM counterpart gDNA.

Subsequent to electrophoresis, the DNA was transferred to a positively charged nylon membrane (Roche Applied Science) by neutral blotting and hybridized with a DIG-labeled probe (PCR DIG Probe Synthesis Kit; Roche Applied Science) covering the entire T-DNA region of the pTC0113 plasmid (comprising the *barstar*, *barnase* and *bar* gene cassettes). Hybridization and detection of the probe followed the instructions of the DIG labeling system manual (Roche Applied Science). Hybridizing fragments were visualized digitally. For stable integration of the T-DNA region, two fragments of 4400 bp and 4900 bp were expected.

2.4 Assessment of Segregation Patterns

gDNA was isolated from leaf discs of each individual plant using a Beadex™ maxi plant kit with a KingFisher Flex instrument (LGC Genomics).

Either event-specific PCR (PCR that crosses the junction between the insert and endogenous genome) or gene-specific PCR analyses were performed to track, respectively, the event or trait genes inserted in the plant to assess the Mendelian segregation pattern. Positive and negative analytical controls together with a no template control were included to demonstrate performance of each method. As an additional, endogenous positive control, the PCR analysis included the amplification of gene sequence specific for each crop to validate the quality of the DNA as compatible with the PCR conditions and avoid false negative scoring. Samples with signal corresponding to the endogenous sequence only were recorded as negative.

2.5 *mRNA Transcript Analysis by real-time Reverse Transcriptase PCR Analysis*

Total RNA was extracted from at least 100 mg of ground plant tissue using the Spectrum™ plant total RNA kit (Sigma-Aldrich) which included treatment with DNase I to eliminate traces of gDNA. The RNA was quantified using a DeNovix™ DS-11-FX spectrophotometer, and the integrity verified using agarose gel electrophoresis.

cDNA was synthesized using total RNA as a template using the Thermo Fisher Scientific™ Maxima™ H Minus cDNA Synthesis Master Mix. For reverse transcription, an oligo-dT primer and random hexamer primers were applied. An additional DNase I treatment was included. In parallel, a no reverse transcriptase control (no-RT control) counterpart sample was prepared for each sample as a negative control to verify the absence of gDNA contamination within the subsequent real-time RT-PCR analysis. For these counterpart samples, no reverse transcriptase enzyme mix was included in the cDNA synthesis reaction mixture.

Real-time reverse transcriptase PCR (RT-PCR) was performed using either a fluorescent dye (Fast SYBR™ Green Master Mix; Thermo Fisher Scientific) or a hydrolysis probe, either TaqMan™ Universal PCR Master Mix (ROX™; Thermo Fisher Scientific) or PerfeCTa™ qPCR FastMix™ II (ROX™; Quantabio). Information on the detection method applied for each of the target gene cassettes is specified in Table 2. Real-time PCR amplification and related C_t scoring were carried out in a LightCycler® 480 II (Roche Applied Science).

Transcriptional expression of the target gene cassettes was semi-quantified by comparing the expression levels of each target gene cassette with the expression levels of three endogenous reference genes. *GhUBQ14*, *GhPP24a* and *GhFBX6* were used as endogenous reference genes for cotton [19, 20]. *APT1*, *TIP41* and *GDI1* were used as endogenous reference genes for canola [21, 22]. Primer sequences used to amplify target gene cassettes are summarized in Table 2.

The relative expression levels of the target genes were calculated using a relative quantification method ($\Delta\Delta C_t$ method) [23].

Table 2 Primer Sequences for Real-Time RT-PCR Assessment of Newly Expressed Protein Genes.

Product	Target gene coding sequence	Primer and Probe sequence (5'-3')		qRT-PCR detection method used
GHB811 cotton	*2mepsps*	forward reverse	TGCTGAACAGTGAGGATGTC AGGACCGCATTGCGATTCCA	SYBR™ Green
	hppdw336-1Pa	forward reverse FAM-labeled probe	TGCTGGCTTGAAGGTTATTGATC TTGAAGAGTTTCTCATAGAAGTTAGCCC TGACACACAACGTCTATCGTGGACGAATG	Hydrolysis probe (TaqMan™)
MS11 *B. napus*	*bar*	forward reverse FAM-labeled probe	CGTCAACCACTACATCGAGACAA GTCCACTCCTGCGGTTCCT ACGGTCAACTTCCGTACCGAGCCG	
	barstar	forward reverse	CAGAAGTATCAGCGACCTCCAC AAACTGCCTCCATTCCAAAAC	SYBR™ Green
	barnase	forward reverse	CATGGCGTGAAGCGGATATTA GCCAGTCGCTTGAGTAAAGAA	

2.6 *Protein Expression Analysis by Means of Enzyme-Linked Immunosorbent Assay*

Proteins were extracted from sub-samples of lyophilized plant tissues and non-lyophilized pollen samples using buffers indicated in Table 2 and an Omni-Prep homogenizer (Omni International Inc.).

Enzyme-Linked Immunosorbent Assay (ELISA) analysis was conducted using the kits described in Table 3 following the manufacturer's instructions (Envirologix). Four independent samples were analyzed for each tissue matrix.

Table 3 Protein Extraction and Quantification Specifics.

Protein	Extraction Buffer	ELISA Detection Method	Envirologix Catalog No.
2mEPSPS	0.01 M phosphate, 0.138 M NaCl, 0.0027 M KCl, 2% PVP40 (w/v), 1% Tween®20 (v/v), pH 7.4	QualiPlate™ kit for 2mEPSPS	AP 084
HPPD W336	0.01 M phosphate, 0.138 M NaCl, 0.0027 M KCl, 1% Tween®20 (v/v), pH 7.4	QualiPlate™ kit for HPPD	AP 128 NW
PAT/*bar*	0.01 M phosphate, 0.138 M NaCl, 0.0027 M KCl, 1% Tween®20 (v/v), pH 7.4	QualiPlate™ kit for PAT/*bar*	AP 013
Barstar	0.01 M phosphate, 0.138 M NaCl, 0.0027 M KCl, 1% Tween®20 (v/v), pH 7.4	QualiPlate™ kit for Barstar	AP 125

2.7 Statistical analysis

Chi-square analysis was performed to compare expected Mendelian segregation patterns to observed segregation ratios. The inheritance stability of the T-DNA insertion, containing the traits, was based on testing the observed trait segregation ratios relative to the trait segregation ratios expected from Mendelian inheritance principles based on the generation of the seed lot. Tables 4-7 include the expected trait segregation ratios. The critical value used to reject the hypothesis of a 1:1 or 3:1 ratio at the 5 % confidence level with one degree of freedom is 3.84 and for 1:2:1 with 2 degrees of freedom is 5.99 [24]. A hypothetical breeding tree is included (Figure 1) to indicate the typical process followed for the preparation of seed lots.

For the transcriptional expression analysis by RT-PCR, descriptive statistics were applied to calculate average relative expression results together with the standard deviations.

For the protein expression analysis, means and standard deviations are presented.

3. Results

Many regulatory authorities throughout the world require insert stability data at the molecular (DNA) and protein expression levels over at least three generations of the event breeding tree (see Figure 1), representing different branches including selfing, as well as back cross introgression in genetic backgrounds different from the plant transformation background.

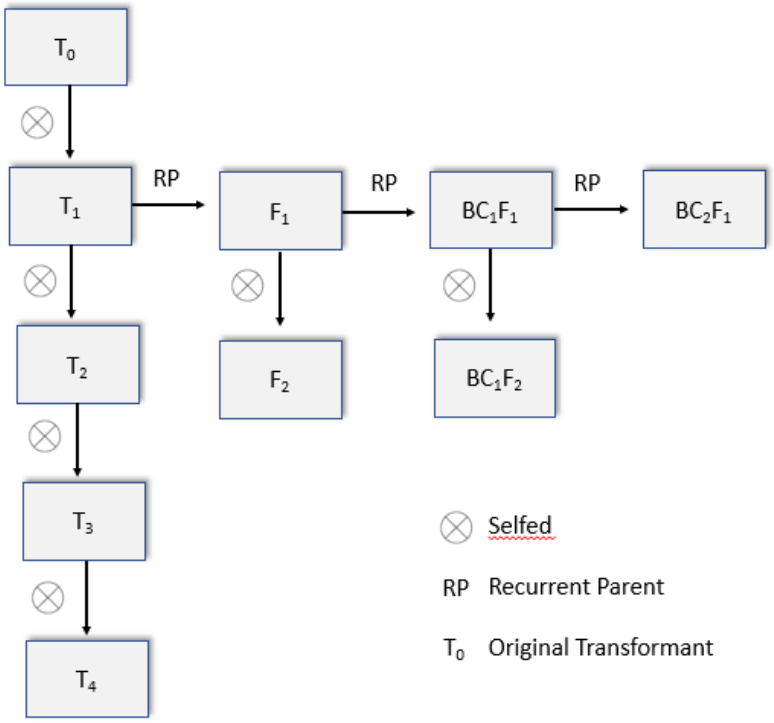

Figure 1 Pedigree Example. The original plant that has been regenerated from the transformed cell and that defines the event is referred to as the T_0 generation. When selfed (⊗) the seed produced is designated the T_1 generation. Backcrosses with a recurrent parent (RP) can be performed at any T generation, in the example here T_1 plants were used. The resultant hemizygous seed comprise the F_1 generation and if backcrossed again become the BC_1F_1 and so forth.

3.1 Genetic stability at the DNA level

All regulatory authorities request molecular characterization data, including information on the inserted sequences, the insertion site (and the surrounding host genome region), and demonstrating stability thereof in successive breeding generations. These regulatory requirements were traditionally and typically addressed by Sanger sequencing and Southern blot analysis. Newer technologies such as next generation sequencing have been accepted by regulatory agencies in many countries and have led to the gradual replacement of Southern blot analysis.

For all events discussed in this paper, the DNA stability over generations is demonstrated by Southern blot data. An example of over-generation stability of the insert as shown by Southern blot analysis is given in Figure 2. In this example for canola event MS11 *B. napus* gDNA from plants from 5 generations of seed lots (T_2, T_3, F_1, BC_1 and BC_2) were analyzed after digestion with a restriction enzyme and probed with the complete T-DNA region of the transformation plasmid. Consistency of the pattern was seen for all generations.

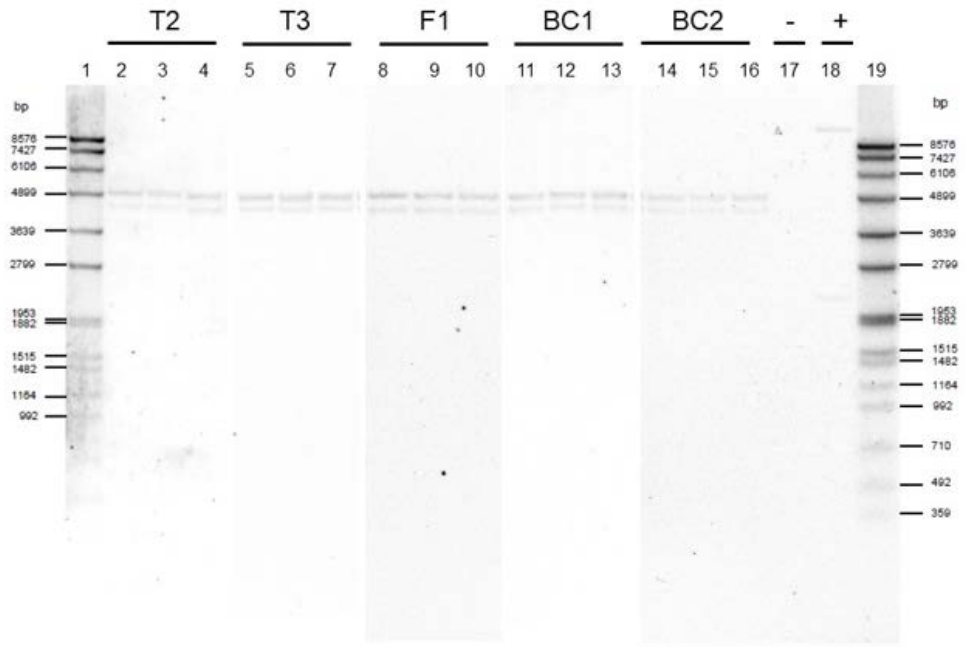

Figure 2 Southern blot analysis demonstrating stability of the inserted sequences and flanking genomic region in the *Brassica napus* event MS11 over different breeding generations. Genomic DNA from MS11 *B. napus* plants was digested with *Eco*RV and 4 µg of the resulting samples was subjected to Southern blot analysis. A specific banding pattern was observed following hybridization with a probe covering the entire T-DNA region (comprising the three gene cassettes, *barnase*, *barstar* and *bar*) of the plasmid used for transformation and was identical for all samples across the different breeding generations investigated. Each lane represents a single MS11 *B. napus* plant and results are presented for 3 plants of each generation (T_2, T_3, F_1, BC_1, BC_2). Molecular size markers are included in lanes 1 and 19 (7.5 ng DIG-labeled molecular mass marker VII spiked in digested non-GM counterpart gDNA). Lane 17 is the negative control (digested gDNA isolated from the non-GM counterpart), while lane 18 is the positive control (digested transforming plasmid pTCO113 spiked in digested non-GM counterpart gDNA).

Although out of the scope of this manuscript, resequencing of these events also occurs when they are incorporated into stacked trait products and in those cases no sequence differences were observed over different assessments in conventional breeding stacked trait products, nor did Southern blot data indicate any instability (GHB811 cotton and MS11 *B. napus*; data not shown). The stability of the RF3 *B. napus* locus was demonstrated in RF3 *B. juncea* by both sequencing and Southern blot data (data also not shown).

3.2 Inheritance patterns

While the data described in the previous section demonstrates stability of the insert over generations, some regulatory bodies also require information on the pattern of genetic and phenotypic stability of the event and resulting traits requiring a more quantitative approach requiring statistical analysis of segregation patterns. Data for such analyses can be recorded by plant breeders as they introgress the events into commercial (elite) germplasm as part of commercial product development. Nevertheless, specific regulatory studies are conducted to examine the inheritance patterns at both the genotypic and phenotypic level. Plants from seed from different generations, for which certain segregation ratios are expected, are characterized for the presence/absence of the transgenes at the molecular level using PCR and then confirmed qualitatively to be expressing the protein.

Results for MS11 *B. napus*, RF3 *B. juncea*, GHB811 cotton and A5542-127 soybean are shown in Tables 4-7. *B. napus* and *B. juncea* are largely self-pollinating (70 %), with the remaining 30 % attributed to wind and insect pollination, soybean is self-pollinated, and cotton is insect-pollinated. All crops/events examined here showed the expected segregation ratios and confirmed that the insertions are inherited in a predictable and stable manner following Mendelian principles associated with a single chromosomal locus within the nuclear genome.

Qualitative demonstration of the presence of the protein encoded by the transgenes using lateral flow strips confirmed the phenotypic inheritance as well (data not shown).

Table 4 MS11 *B. napus* Inheritance Pattern.

	T_3		T_4		T_5		BC_4		BC_5	
	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp
Pos	42	42	48	46	39	47.5	43	44.5	51	49
Neg	42	42	44	46	56	47.5	46	44.5	47	49
Expected Ratio*	1:1		1:1		1:1		1:1		1:1	
χ^2	0		0.174		3.042		0.101		0.163	

*positive: negative

Table 5 RF3 *B. juncea* Inheritance Pattern.

	BC_1		BC_2		BC_3		BC_2S_1	
	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp
Pos	109	98.5	108	99.5	95	92	104	116.25
Neg	88	98.5	91	99.5	86	92	51	37.75
Expected Ratio*	1:1		1:1		1:1		3:1	
χ^2	2.24		1.45		0.49		5.16	

*positive: negative

Table 6 GHB811 Cotton Inheritance Pattern.

	BC_1		BC_2		BC_3		BC_2S_1		BC_2F_5	
	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp
Pos	124	112.5	156	155.25	187	183	165	167.25	181	177
Neg	26	37.5	51	51.75	57	61	58	55.75	55	59
Expected Ratio*	3:1		3:1		3:1		3:1		3:1	
χ^2	4.702		0.014		0.350		0.121		0.362	

*positive: negative

Table 7 A5547-127 Soybean Inheritance Pattern.

	F_2		F_3	
	Obs	Exp	Obs	Exp
Homozygous	65	61.5	87	86.5
Hemizygous	112	123	181	173
Wildtype	69	61.5	78	86.5
Total	246		346	
Expected Ratio*	1:2:1		1:2:1	
χ^2	2.088		1.208	

*positive: negative

3.3 Genetic stability of expression

Many regulatory bodies require information on protein expression levels and evidence of the imparted trait should be sufficient indication that the insertion is performing as desired and dietary exposure assessments rely on the protein expression level, not the transcript. Therefore, it is not clear what additional information in support of risk assessment can be derived from measurements of mRNA expression. However, some regulatory authorities also request mRNA expression stability

studies. To address this requirement, the relative mRNA expression levels of the transgenes were assessed in various tissues of young and flowering plants as well as in mature seed.

In the GHB811 cotton event, the *2mepsps* gene cassette is driven by the Ph4a748 promoter from *Arabidopsis thaliana* and proved to be strongly and constitutively expressed in all cotton matrices, as expected based on the literature [13]. The *hppdPfW336 -1Pa* gene cassette under transcriptional control of the constitutive Pcsvmv promoter from the Cassava Vein Mosaic Virus [11] was also expressed in all cotton matrices. Since for all assessed matrices, similar expression patterns were observed over the three generations, the stability of transcriptional expression over generations was demonstrated (Figure 3). The difference in *2mepsps* transcript level in the T4 vs. the T3 and T5 generations is attributed to experimental noise and was not reflected by a difference in 2mEPSPS protein level (Figure 4).

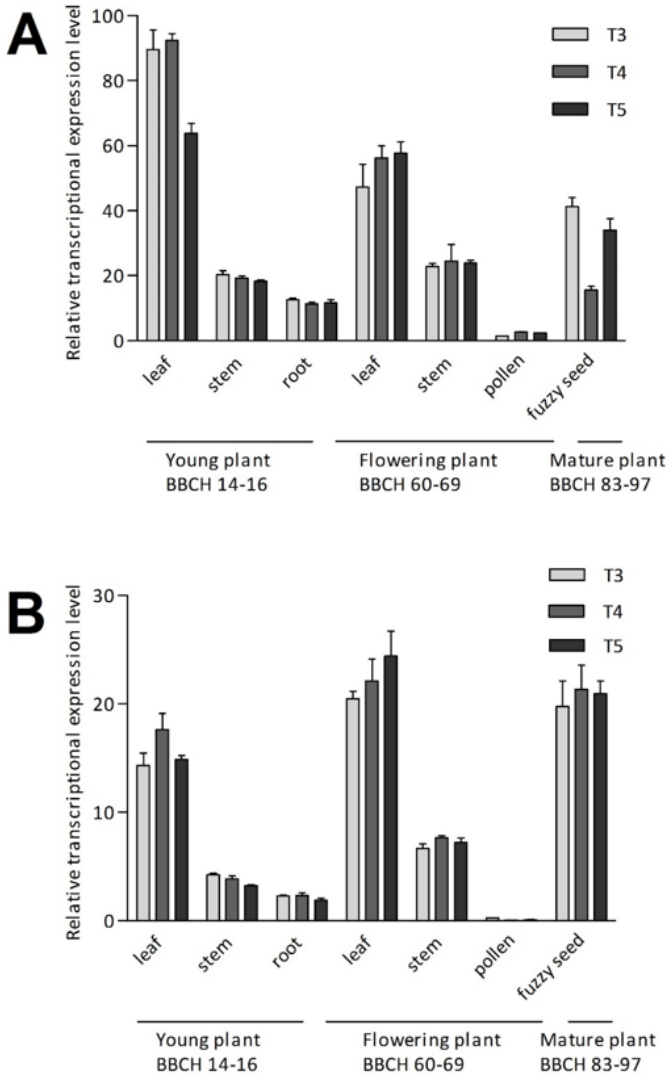

Figure 3 Graphical representation of determined *2mepsps* (panel A) and *hppdW336Pf-1Pa* (panel B) relative transcriptional expression levels in GHB811 cotton for assessed matrices. Error bars represent technical variation over six replicates (STD). Observed expression levels for the non-GM counterpart were below the quantitative range of the assay. All plants were homozygous with respect to the introduced traits.

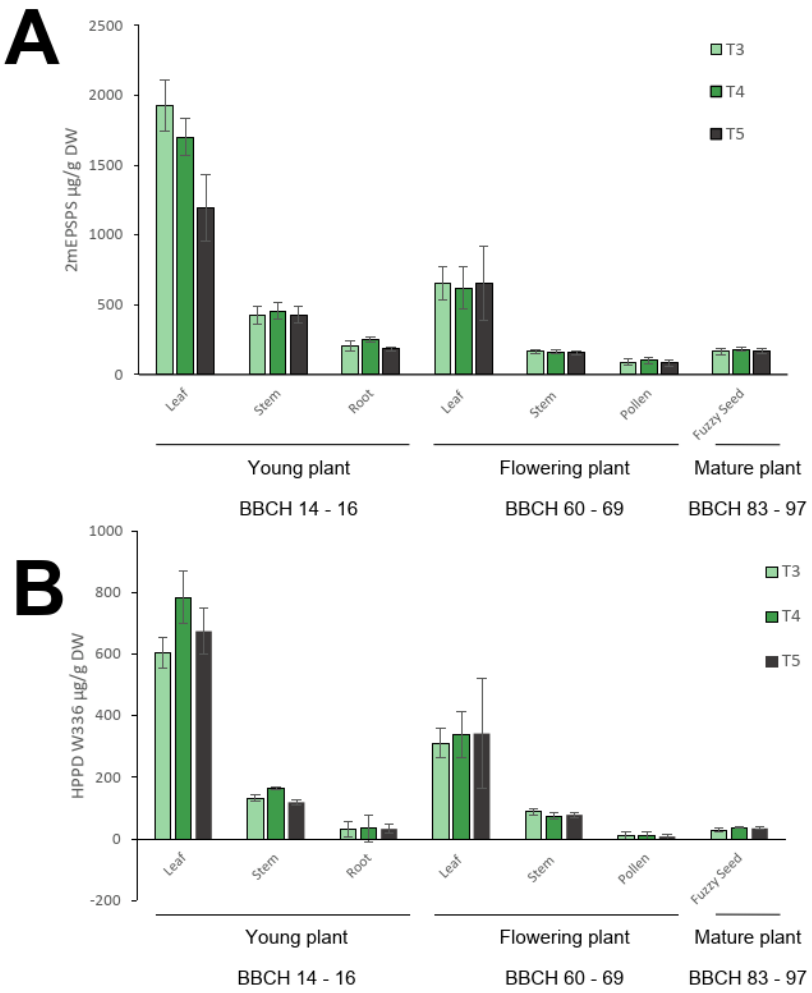

Figure 4 Protein expression levels in various matrices and developmental stages across three generations of GHB811 cotton plants. Protein levels were measured using ELISA and are presented on a dry weight basis for 2mEPSPS (panel A) and HPPD W336 (panel B). Standard deviations are indicated. All plants were homozygous with respect to the introduced traits.

Levels of the proteins 2mEPSPS and HPPDW336 expressed by the GHB811 cotton event were found to be consistent across generations and the relative (with respect to which tissues had the highest, lowest, etc) amounts correlate with the levels of transcripts (Figures 4 A&B). Absolute amounts of protein cannot be anticipated from the transcript level. Furthermore, the variation in the RT-PCR data reflects assay to assay variation as plant samples were pooled prior to analysis, the ELISA data represents both assay to assay and plant to plant variability. The lower value seen for the young leaf T5 sample is within the normal experimental variation seen for ELISA.

Similarly, the relative mRNA expression levels of the expressed transgenes of MS11 *B. napus* were assessed (Figures 5 A&B). mRNA expression of the *bar* and *barstar* genes, is driven by the constitutive promoters PssuAt [9] and Pnos [7], respectively. For both *bar* and *barstar*, mRNA expression was observed in all matrices assessed. While relative expression of *bar* is most pronounced in green tissues, *barstar* is mainly expressed in root tissue (of the matrices examined) from MS11 *B. napus* plants as expected. The variability of transcript levels in the stem tissue is considered to be due to expected noise in the data.

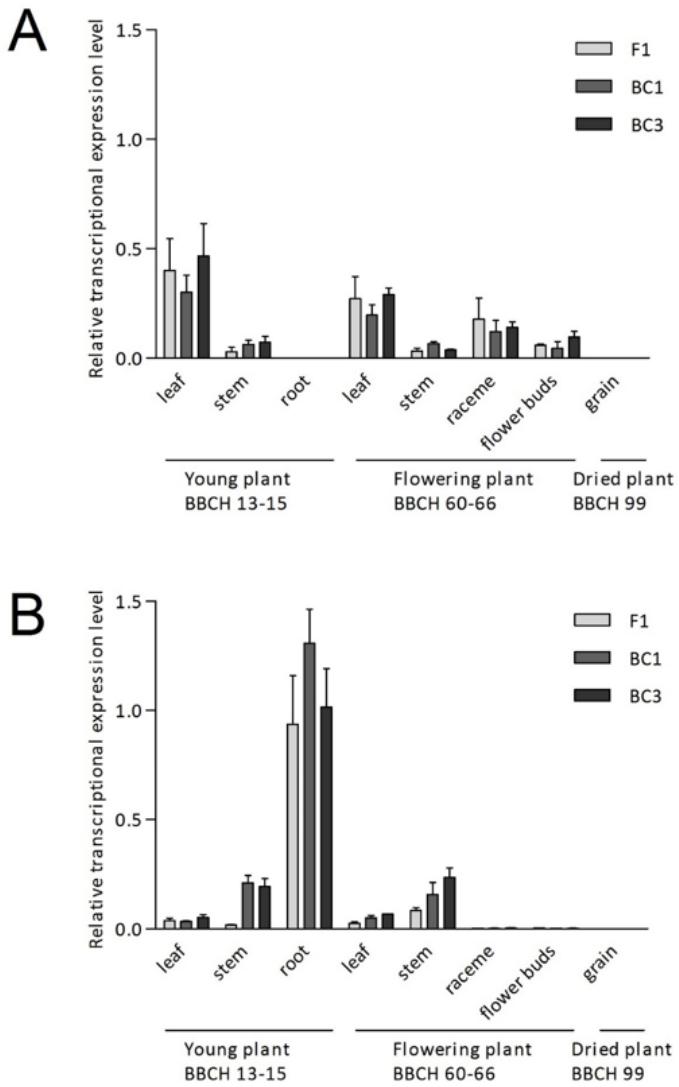

Figure 5 Graphical representation of determined *bar* (panel A) and *barstar* (panel B) relative transcriptional expression levels in MS11 *B. napus* for assessed matrices. Transcript levels of *bar* in root and grain tissue and of *barstar* in grain tissue were below LOQ and therefore not visualized due to the Y-axis scaling of the graph. Error bars represent technical variation over six replicates (STD). Observed expression levels for the non-GM counterpart were below the quantitative range of the assay. All plants were hemizygous with respect to the introduced traits.

Relative mRNA expression levels of the *barnase* gene cassette (tapetum-specific Pta29 promoter) [6] were only consistently detected at very low levels in flower buds (Table 8). For all other matrices, the data were below or at LOQ. These observations are as expected since the tissue specificity of the Pta29 promoter is restricted to flower buds, both temporally and spatially. Since the tapetum, where the Pta29 promoter is expressed, is a specialized layer within the flower bud, *barnase* expression is underestimated within a heterogenous flower bud matrix [25, 26]. Additionally, the ribonuclease activity of Barnase has been demonstrated to result in tapetal cell RNA hydrolysis and cell death, through which the RNA levels of *barnase* remained low [25]. The transcriptional expression patterns observed in the different MS11 *B. napus* plant matrices were similar over the three generations, demonstrating stability of transcriptional expression over generations.

Table 8 *Barnase* Relative Transcriptional Expression in MS11 *B. napus* and Non-GM Counterpart Samples.

Sample	Generation	Matrix	Qualitative scoring
MS11 *B. napus*	F1	BBCH 13-15 leaf	< LOD
		BBCH 13-15 stem	< LOQ
		BBCH 13-15 root	< LOQ
		BBCH 60-66 leaf	< LOQ
		BBCH 60-66 stem	< LOQ
		BBCH 60-66 raceme	< LOQ*
		BBCH 60-66 flower buds	Approx. LOQ**
		BBCH 99 grain	< LOQ*
	BC1	BBCH 13-15 leaf	< LOQ
		BBCH 13-15 stem	< LOQ
		BBCH 13-15 root	< LOQ
		BBCH 60-66 leaf	< LOQ
		BBCH 60-66 stem	< LOQ
		BBCH 60-66 raceme	Approx. LOQ**
		BBCH 60-66 flower buds	Very low
		BBCH 99 grain	Very low
	BC3	BBCH 13-15 leaf	< LOQ*
		BBCH 13-15 stem	< LOQ
		BBCH 13-15 root	< LOQ
		BBCH 60-66 leaf	< LOQ
		BBCH 60-66 stem	< LOQ
		BBCH 60-66 raceme	Approx. LOQ**
		BBCH 60-66 flower buds	Approx. LOQ**
		BBCH 99 grain	< LOQ
Non-GM counterpart	N.A.	BBCH 13-15 leaf	< LOD
		BBCH 13-15 stem	< LOD
		BBCH 13-15 root	< LOD
		BBCH 60-66 leaf	< LOD
		BBCH 60-66 stem	< LOD
		BBCH 60-66 raceme	< LOD
		BBCH 60-66 flower buds	< LOD
		BBCH 60-66 whole-plant	< LOD
		BBCH 99 grain	< LOD

N.A.: not applicable; STD: standard deviation; <LOD: below limit of detection; <LOQ: below limit of quantitation; * minority of replicates were in the quantitative range of the assay, therefore mean expression is below LOQ; ** at least half of the replicates were in the quantitative range of the assay, therefore the expression level was set to "approx. LOQ; 'Very low' indicates a fold change of expression compared to the set of endogenous reference genes lower than 0.001.

In MS11 *B. napus*, PAT protein (encoded by the *bar* gene) expression was found at consistent levels across three generations in the matrices examined (Figure 6). The Barstar protein was only detectable in the occasional root sample and Barnase was not detected (data not shown). This data corresponds to the transcriptional data and respective promoters as discussed above. Transcript detection is more sensitive than the protein detection method, since it was not possible to quantify Barstar protein levels in tissues other than root. Expression of Barnase leads to the death of the cells in which it was expressed. Hence, no protein was detected.

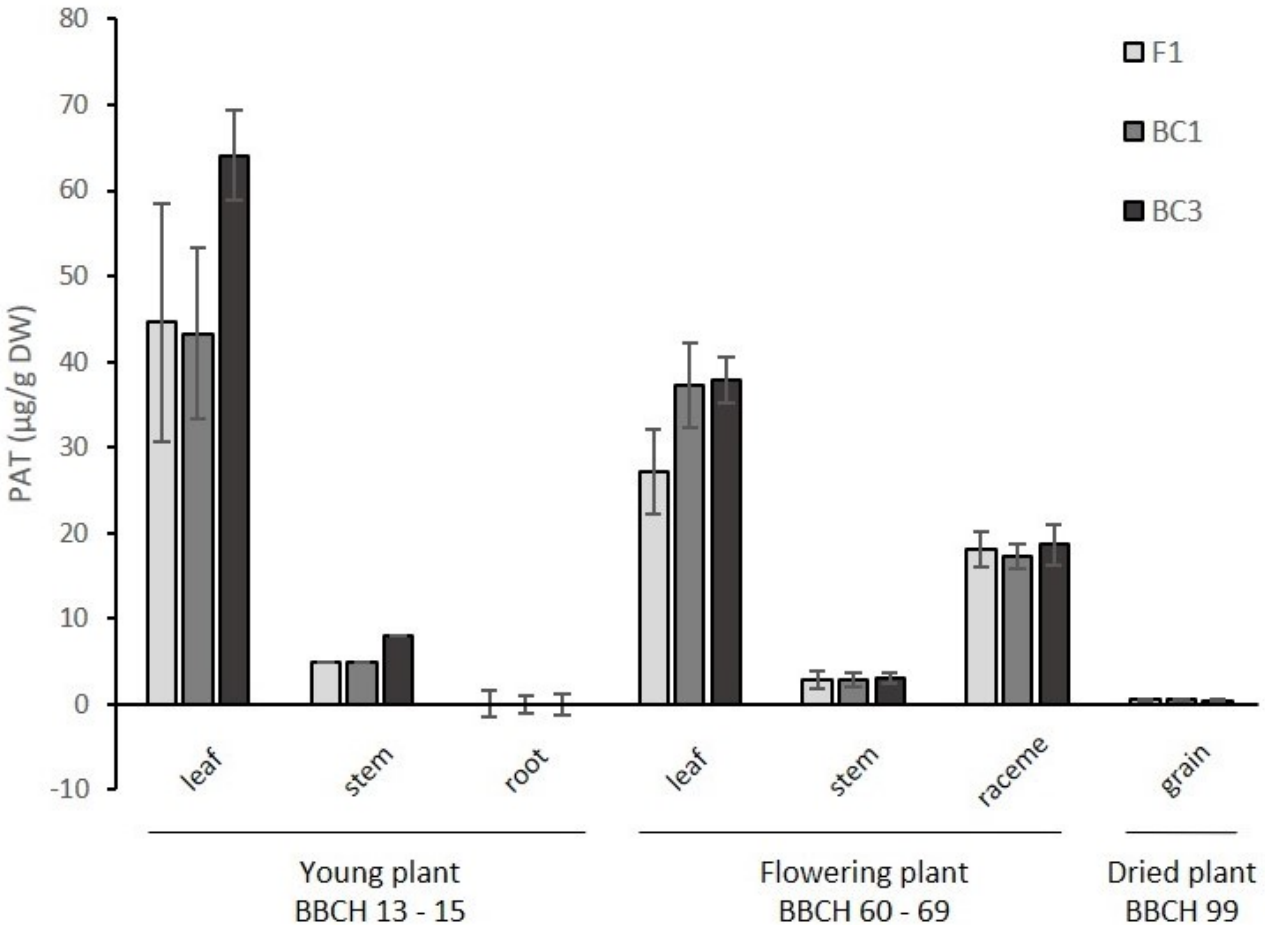

Figure 6 PAT/*bar* protein expression levels in various matrices and developmental stages across three generations of event MS11 *B. napus* plants. Protein levels were measured using ELISA and are presented on a dry weight basis. Standard deviations are indicated. All plants were hemizygous with respect to the introduced traits.

PAT protein expression in the RF3 *B. napus* expressed consistently across the three generations (Figure 7A). When the insert of RF3 *R. napus* was introgressed in RF3 *B. juncea*, PAT protein expression levels had the pattern seen in Figure 7B. Within the variability associated with ELISAs, PAT levels were found to be similar even when crossed into a different but related species (Figure 7B). The difference in the level of PAT protein observed for the hemizygous F1 generation vs. the other two homozygous generations may reflect the difference in the number of *bar* genes present and has been reported previously [27].

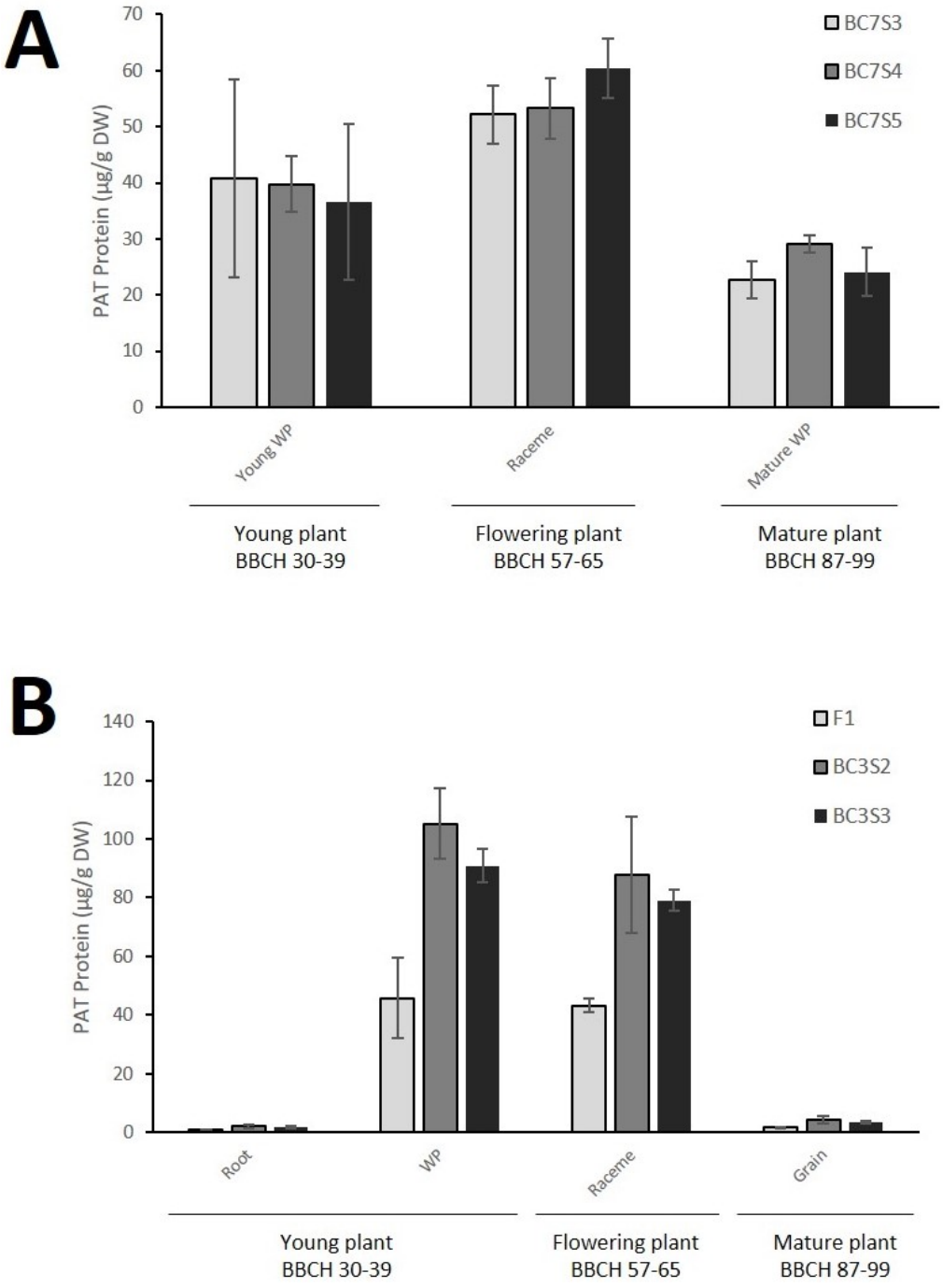

Figure 7 PAT/bar Protein expression levels in various matrices and developmental stages across three generations of RF3 *B. napus* (panel A) and B. RF3 *B. juncea* (panel B). Protein levels were measured using ELISA and are presented on a dry weight basis. WP = whole plant. Standard deviations are indicated. F1 plants were hemizygous whereas the BC3S2 and BC3S3 generations were homozygous with respect to the introduced traits.

4. Discussion

A requirement of many regulatory agencies as part of the risk assessment of biotech products is that the insertion in each event for which approval is being sought and, to some degree, its flanking plant genomic regions, be sequenced at the nucleotide level. For some regulatory authorities that

require renewal product applications, (re)sequencing of the events is required to demonstrate the absence of unfavorable mutations that may have occurred during the breeding process. Furthermore, if the event is sold as part of a conventionally bred "stacked trait product" with another single event, sequencing of the inserts of every single parental event is again required for those jurisdictions that require separate assessments for stacked trait products produced by conventional breeding of multiple single events. To date, the only differences in sequence after the first 10 years of a product on the market have been found to be due to improvements in the sequencing technologies and bioinformatics assembly of sequences which has allowed better analysis of hard to sequence regions, for example, regions that may have mononucleotide runs (i.e., AAAAAAAA vs. AAAAAAA) [28]. Strand slippage may also occur for poly A, T, C or G stretches which can lead to imperfect sequence outcomes [29].

There is no *a priori* reason that once a gene is introduced into the genome it should not be inherited in the same fashion as any endogenous gene and this would be dictated by which genome within the plant cells the insertion occurred. The individual nucleotides are indistinguishable from those within the endogenous genes. Transgene insertions are subject to the same tendency for mutations as all other genetic material within the plant [30, 31]. From the three genomes within a plant cell – the nuclear, plastidic and mitochondrial genomes, only the nuclear genome is inherited in a Mendelian fashion, while both the plastidic and mitochondrial genomes are maternally inherited (i.e., they are not inherited via pollen) [32, 33]. The Mendelian inheritance patterns observed for the events discussed in this article confirm that the events were integrated in the nuclear DNA.

Expression of the gene can be considered as the production of the mRNA or the production of the protein which implies that the mRNA was produced. Expression patterns are determined by the promoter which drives the gene. Generally, the trait (phenotype) is reflected by the presence of the protein (or absence of the endogenous protein, if the product was designed to abolish gene expression). Therefore, regulatory agencies generally focus data requirements on the levels of the protein produced. Consistency of the phenotype is certainly what the grower desires and therefore what the developer aims for. However, once the product (i.e., the protein and the crop) has been assessed as safe by a regulatory agency, the performance of the product is not a safety concern. Whereas it is possible that transgene inheritance could follow non-Mendelian inheritance patterns [34], any event that indicates a non-Mendelian inheritance is discarded early in the product development process and does not enter the product pipeline.

Unstable inheritance patterns could be attributed to gene silencing which is known to occur in plants. There is some knowledge of the mechanisms of silencing at either the transcript or post transcriptional level [35, 36]. During the product development and breeding processes [37], plant lines that do not perform consistently, or that are showing genetic instability, are discarded from further development. The results summarized in this paper confirm previous results and demonstrate that protein expression levels in commercial biotech products are consistent across generations. These results also reveal similar outcomes associated with measuring transcript levels, thus demonstrating that measuring protein levels are sufficient. As there is an additional level of translation control, on top of the variable half-life of gene transcripts and stability of proteins, there might not always be good correlation between transcript level and protein level. In addition, the applied RT-PCR approach to study transcriptional expression levels is much more sensitive

compared to ELISA as it involves repeated rounds of template amplification. With regard to hazard characterisation, the protein levels provide appropriate and sufficient information.

Generally speaking, there is very little published information on the expression levels of proteins in transgenic plants [27, 38-42]. Fearing *et al.* [38] published data on the expression levels of events expressing insecticidal Cry1Ab protein across successive generations during introgression into maize and reported consistent levels. Kramer *et al.* [39] confirmed similar protein expression levels between single and stacked trait products of maize, with differences in expression associated with gene copy number. Further, environmental and germplasm background variability was shown to result in more variation in expression than stacking of events. Gampala *et al.* [40] reported similar findings. Chinnadurai *et al.* [41] examined CP4 EPSPS levels, conferring herbicide tolerance, in diverse soybean germplasm and different environments in both single and stacked trait products, but they did not track expression levels by generation. Results from Fast *et al.* [42] demonstrated that herbicide treatment had no impact on expression levels for the maize, soybean and cotton events they examined. Wu *et al.* [27] showed for multiple cotton stacked trait products that expression levels were similar to the parental lines and may have been impacted by gene copy number.

In summary, the data presented in this paper show that the examined events were nuclear insertions presenting Mendelian inheritance patterns and that the proteins are expressed similarly across multiple generations regardless of whether they were from backcrossed or outcrossed generations. These results demonstrate that newly inserted genes as present in commercial biotech crops are transmitted to their progeny in a stable manner similar to that of endogenous genes. Furthermore, these data show that it is time to reconsider the relevance of the stability analyses for the overall risk assessment of the product. While stability of transgenes and inheritance patterns may be relevant research questions, in the context of commercial product development it is a matter of product quality and performance to ensure that the events are predictable and consistent over generations. This ensures that the product can be marketed and can provide value to farmers.

Acknowledgments

The authors would like to thank Jordan Sottosanto, Ana Atanassova and Felicity Keiper for critical review of the manuscript.

Author Contributions

Manuscript Development: Laura Privalle, Rozemarijn Dreesen

Technical contributions/data generation: Patricia Back, Apurva Bhargava, Zach Bishop, Krystal Cisneros, Ine Criel, Lien Dhondt, Travis Draughn, Brad Franklin, Durba Ghoshal, Shane Walsh, Jim Lor, Jennifer Massengil, Sofie Moens, Tyson Mooney, Dannyel Nelson, Karolien Peeters, Sashi Sathischandra, Caroline Staut, Annelies Van Hoecke, Annelies Van Raemdonck, Marie-Laure Verdegem, Steven Verhaeghe, Ann Wierckx, Qiang Zhao

Regulatory Affairs Management: Isabelle Coats, Barb Fowler, Yoonhui Sung, Ann Tuttle

Funding

BASF supplied total support of this work.

Competing Interests

The authors have declared that no competing interests exist.

References

1. Brookes G, Barfoot P. GM crop technology use 1996-2018: Farm income and production impacts. GM Crops Food. 2020; 11: 242-261.
2. Delaney B, Astwood JD, Cunny H, Conn RE, Herouet-Guicheney C, Macintosh S, et al. Evaluation of protein safety in the context of agricultural biotechnology. Food Chem Toxicol. 2008; 46: S71-S97.
3. Qin Y, Ahn HI, Park SY, Lim MH, Woo HJ, Shin KS, et al. T-DNA inheritance stability of resveratrol rice Iksan526 over multi-generations. Plant Breed Biotechnol. 2014; 2: 268-275.
4. Betts SD, Basu S, Bolar J, Booth R, Chang S, Cigan AM, et al. Uniform expression and relatively small position effects characterize sister transformants in maize and soybean. Front Plant Sci. 2019; 10: 1209.
5. Hartley RW. Barnase and barstar: Expression of its cloned inhibitor permits expression of a cloned ribonuclease. J Mol Biol. 1988; 202: 913-915.
6. Seurinck J, Truettner J, Goldberg RB. The nucleotide sequence of an anther-specific gene. Nucleic Acids Res. 1990; 11: 3403.
7. Depicker, A, Stachel, S, Dhaese, P, Zambryski P, Goodman HM. Nopaline synthase: Transcript mapping and DNA sequence. J Mol Appl Genet. 1982; 6: 561-573.
8. Thompson CJ, Movva NR, Tizard R, Crameri R, Davies JE, Lauwereys M, et al. Characterization of the herbicide-resistance gene *bar* from *Streptomyces hygroscopicus*. EMBO J. 1987; 9: 2519-2523.
9. Krebbers E, Seurinck J, Herdies A, Cashmore AR, Timko MP. Four genes in two diverged subfamilies encode the ribulose-1,5-biphosphate carboxylase small subunit polypeptides of *Arabidopsis thaliana*. Plant Mol Biol. 1988; 11: 745-759.
10. Boudec P, Rodgers M, Dumas F, Sailland A, Bourdon H. Mutated hydroxyphenylpyruvate dioxygenase, DNA sequence and isolation of plants which contain such a gene and which are tolerant to herbicides. Washington, DC: Patent and Trademark Office; 2001; United States Patent US62459698.
11. Verdaguer B, De Kochko A, Beachy RN, Fauquet C. Isolation and expression in transgenic tobacco and rice plants of the cassava vein mosaic virus (CVMV) promotor. Plant Mol Biol. 1996; 31: 1129-1139.
12. Lebrun M, Sailland A, Fressinet G, DeGryse E. Mutated 5-enol pyruvylshikimate-3 phosphate synthase, gene coding for said protein and transformed plants containing said gene. Washington, DC: Patent and Trademark Office; 1997.
13. Chabouté M, Chaubet N, Philipps G, Ehling M, Gigot C. Genomic organization and nucleotide sequences of two histone H3 and two histone H4 genes of *Arabidopsis thaliana*. Plant Mol Biol, 198; 8: 179-191.
14. Vinnemeier J, Dröge-Laser W, Pistorius EK, Broer I. Purification and partial characterization of the *Streptomyces viridochromogenes* Tii494 phosphinothricin-N-acetyltransferase mediating resistance to the herbicide phosphinothricin in transgenic plants. Z Naturforsch. 1995; 50: 796-805.

15. Harpster MH, Townsend JA, Jones JD, Bedbrook J, Dunsmuir P. Relative strengths of the 35S Cauliflower Mosaic Virus, 1', 2' and nopaline synthase promoters in transformed tobacco, sugarbeet and oilseed rape callus tissue. Mol Gen Genet. 1988; 212: 182-190.
16. Odell JT, Nagy F, Chua NH. Identification of DNA sequences required for activity of the Cauliflower Mosaic Virus 35S promoter. Nature. 1985; 313: 810-812.
17. Meier U. Growth stages of mono- and dicotyledonous plants- BBCH monograph. Berlin: German Federal Biological Research Centre for Agriculture and Forestry; 2001.
18. Dellaporta SL, Wood J, Hicks JB. A plant DNA minipreparation: Version II. Plant Mol Biol Rep. 1983; 1: 19-21.
19. Li XB, Cai L, Cheng ND, Liu JW. Molecular characterization of the cotton GhTUB1 Gene that is preferentially expressed in fiber. Plant Physiol. 2002; 130: 666-674.
20. Artico S, Nardeli SM, Brilhante O, Grossi-de-Sa MF, Alves-Ferreira M. Identification and evaluation of new reference genes in *Gossypium hirsutum* for accurate normalization of real-time quantitative RT-PCR data. BMC Plant Biol. 2010; 10: 49.
21. Rumlow A, Keunen E, Klein J, Palmann P, Riemenschneider A, Cuypers A, et al. Quantitative expression analysis in *Brassica napus* by northern blot analysis and reverse-transcription quantitative PCR in a complex experimental setting. PLOS One. 2016; 11: e0163679.
22. Yang HL, Liu J, Huang SM, Guo TT, Deng LB, Hua W. Selection and evaluation of novel reference genes for quantitative reverse transcription PCR (real-time RT-PCR) based on genome and transcriptome data in *Brassica napus* L. Gene. 2014; 538: 113-122.
23. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods. 2001; 25: 402-408.
24. Strickberger MW. In probability and statistical testing. 2nd ed. New York: MacMillan Publishing Company; 1976. 140p.-163p.
25. Mariani C, Gosselé V, De Beuckeleer M, De Block M, Goldberg RB, De Greef W, Leemans J. A chimaeric ribonuclease-inhibitor gene restores fertility to male sterile plants. Nature. 1992; 357: 384-387.
26. De Block M, Debrouwer D. Engineered fertility control in transgenic *Brassica napus* L.: Histochemical analysis of anther development. Planta. 1993; 189: 218-225.
27. Wu AJ, Chapman K, Satischandra S, Massengill J, Araujo R, Soria M, et al. GHB614 x T304-40 x HB119 x COT102 Cotton: Protein expression analyses of field-grown samples. J Agric Food Chem. 2019; 67: 275-281.
28. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. Genomics. 2016; 107: 1-8.
29. Clarke LA, Rebelo CS, Gonçalves J, Boavida MG, Jordan P. PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences. Mol Pathol. 2001; 54: 351-353.
30. De Maagd RA, Van de Wiel C, Schouten HJ. The plasticity of plant genomes causes and consequences: A survey of data on structural genome variation in plants. São Caetano do Sul: COGEM; 2020. Available from: https://cogem.net/app/uploads/2020/07/CGM-2020-04-COGEM-report_StructuralVariation.pdf.
31. Schnell J, Steele M, Bean J, Neuspiel M, Girard C, Dormann N, et al. A comparative analysis of insertional effects in genetically engineered plants: Considerations for pre-market assessments. Transgenic Res. 2015; 24: 1-17.

32. Greiner S, Sobanski J, Bock R. Why are most organelle genomes transmitted maternally? Bioessays. 2015; 37: 80-94.
33. Connett MB. Mechanisms of maternal inheritance of plastids and mitochondria: Developmental and ultrastructural evidence. Plant Mol Bio Rep. 1986; 4: 193-205.
34. Yin Z, Plader W, Malepszy S. Transgenic inheritance in plants. J Appl Genet. 2004; 45: 127-144.
35. De Alba AE, Elvira-Matelot E, Vaucheret H. Gene silencing in plants: A diversity of pathways. Biochim Biophys Acta. 2013; 1829: 1300-1308.
36. Guo QG, Liu Q, Smith NA, Liang GL, Wang MB. RNA silencing in plants: Mechanisms, technologies and applications in horticultural crops. Curr Genomics. 2016; 17: 476-489.
37. Privalle LS, Chen JW, Clapper G, Hunst P, Spiegelhalter F, Zhong X. Development of an agricultural biotechnology crop product: Testing from Discovery to commercialization. J Agri Food Chem. 2012; 60: 10179-10187.
38. Fearing PL, Brown D, Valchos D, Meghji M, Privalle L. Quantitative analysis of Cry1A(b) expression in Bt maize plants, tissues, and silage and stability of expression over successive generations. Mol Breeding. 1997; 3: 169-176.
39. Kramer C, Brune P, McDonald J, Nesbitt M, Sauve A, Storck-Weyhermueller S. Evolution of risk assessment strategies for food and feed uses of stacked GM events. Plant Biotechnol J. 2016; 14: 1899-1913.
40. Gampala SS, Fast BJ, Richey KA, Gao ZF, Hill R, Wulfkuhle B, et al. Single-event transgene product levels predict levels in genetically modified breeding stacks. J Agric Food Chem. 2017; 65: 7885-7892.
41. Chinnadurai P, Stojšin D, Liu K, Frierdich GE, Glenn KC, Geng T, et al. Variability of CP4 EPSPS expression in genetically engineered soybean (*Glycine max* L. Merrill). Transgenic Res. 2018; 27: 511-524.
42. Fast BJ, Shan G, Gampala S, Herman RA. Transgene expression in sprayed and non-sprayed herbicide-tolerant genetically engineered crops is equivalent. Regul Toxicol Pharmacol. 2020; 111: 104572.