Preprints.org

The Recombination of Insertion Sequence Results in Deletion of the 3′ End of the cry1Ac Gene with Concomitant Enhancement of Its Insecticidal Activity

Chunmeng Huang , Yongjun Zhang , Huilin Yu , Chenhui Hou , Haoyuan Guan , Xiuping Chen [*] , Jiajian Xie [*]

Article

The Recombination of Insertion Sequence Results in Deletion of the 3′ End of the *cry1Ac* Gene with Concomitant Enhancement of Its Insecticidal Activity

Chunmeng Huang [1,2,3]**, Yongjun Zhang** [1,4]**, Huilin Yu** [1,2,3]**, Chenhui Hou** [1,2,3]**, Haoyuan Guan** [1,2,3]**, Xiuping Chen** [1,2,3]***** and Jiajian Xie** [1,2,3]*****

[1] Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing 100193, People's Republic of China;

[2] Key Laboratory of Safety assessment of Genetically Modified Organisms (Environment), Ministry of Agriculture and Rural Affairs, Beijing 100193, People's Republic of China;

[3] Plant Ecological Environment Safety Inspection and Testing Center of Ministry of Agriculture and Rural Affairs (Beijing), Beijing 100193, People's Republic of China.

[4] State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing 100193, China.

*   Correspondence: xpchen@ippcaas.cn (X.C.); jjxie@ippcaas.cn (J.X.); Tel.: +86-010-62810797; +86-010-62810797

Abstract: The structure and expression of exogenous genes in transgenic crops are critical for the target traits. R7569 was a mutant event identified in this study with deletion at the 3' end of cry1Ac gene compared to the transgenic insect-resistant cotton MON531 event with commercial application. R7569 has the same exogenous insertion structure as MON531, but with a deletion in the 3' end of the cry1Ac gene and the terminator region. R7569 has a truncated cry1Ac gene with the length of 2,554 bp, encoding 881 amino acids. The transcription termination site was mainly concentrated downstream of the truncated position and extended 160-270 bp from the truncated position using rapid-amplification of cDNA ends (RACE). The transcript levels of cry1Ac genes of R7569 and MON531 decreased gradually at seedling, bud and bell stages, and the transcript levels of cry1Ac genes of R7569 were significantly higher than those of MON531 in seedling and bud stages, but there was no significant difference in the boll stage. The content of cry1Ac protein in R7569 gradually decreased with the seedling, bud and boll stage, and the content of cry1Ac protein in all three periods was higher than that of MON531. The insect resistance assay showed that the resistance levels of R7569 and MON531 were both at the high level, and the corrected mortality rate against bollworms was 99.5% and 95.2%, respectively, and there was no significant difference between them. The LC50 value of R7569 was 0.732ng/g dw, with a slope of 1.654 indicating a high level of resistance to bollworm. In this study, for the first time, we found a partial deletion of the target gene in commercially applied transgenic crops, and the partial deletion of the 3' end of the cry1Ac gene retained a better transcription, expression level and insecticidal activity, which can provide a specific case for the safety evaluation of transgenic crops.

Keywords: transgenic cotton; insect-resistant; MON531; R7569; RACE; RT-PCR; ELISA; LC_{50}

1. Introduction

Cotton (*Gossypium hirsutum* L.) is one of the most widely used cash crops worldwide [1]. The application of transgenic insect-resistant crop is beneficial for the control of target pests in cotton fields, helping to reduce environmental pollution and maintain human ecological health [2]. Exogenous insertion elements, target genes and exogenous protein expression, and traits are closely and intricately linked, so clarifying the molecular characterization of transgenic insertion structures is essential for conducting rigorous biosafety evaluation studies. One of the commonly used methods to introduce exogenous genes onto the plant genome is Agrobacterium transformation. Agrobacterium has a natural ability to transfer DNA, and the method is a complex process. The T-

DNA region of the Agrobacterium Ti plasmid is the region between the left and right borders, which has a 25 bp repetitive sequence that is the recognition sequence of the endonuclease complex encoded by Agrobacterium, which cuts between the third and fourth bases and produces a notch. The single-stranded T-strand enters the host plant cell and integrates into the host genome in single-copy full-length, broken incomplete copies, and multiple copy forms. Transgenic plant populations in which the exogenous gene was stably integrated into the genome were selected by trait and molecular characterization during the breeding selection process, and therefore, the progeny of transgenic plant populations do not generally show large variations, e.g. large fragment recombination or deletions.

However, studies have been reported on incomplete reading frames retaining the promoter in the exogenous insertion structure of GMOs, such as transgenic maize MON810 [3]. LaPaz and Rosati [4] found the truncated gene cry1Ab retained the promoter, while the 3' end sequence and the terminator region were deleted. The transcripts were analyzed by Northern Blot and Nucleic Acid Mismatch Endonuclease assay and found that there were differences in the length of the cDNAs in the transcripts, and that there was a read-through at the 3' end of the cry1Ab gene, and the termination site was close to the truncation position and extended downstream up to 1kb. Meanwhile, the fusion proteins thus produced did not have homologous similarity with the known proteins, and the fusion proteins were also analyzed for expression and insecticidal activity and showed normal insecticidal activity [4,5]. However, naturally occurring mutations in the truncated cry1Ac gene have rarely been reported in transgenic crops. Koul et al examined the expression of the full-length and truncated cry1Ac genes in transgenic tomatoes using western blot, confirming that both full-length and truncated cry1Ac genes were expressed in the transgenic tomato confirming that full-length and truncated cry1Ac genes were expressed as expected in transgenic tomato [6].

One of the main components of safety evaluation is molecular characteristics. The resolution of the integrated structure of exogenous genes is the basis and prerequisite for the safety evaluation of transgenic crops. Specific qualitative and quantitative assays were established according to the needs of testing. Qualitative tests can be used to distinguish between approved and unauthorized material or for certifying the purity of the material; quantitative tests can be used for labelling. Testing can play a role in GMO safety assessment and risk management as a tracing tool [7] or for traceability of GM and derivatives after consumption [8]. Currently, transgene detection at the nucleic acid level is based on four different detection strategies, namely, element screening, gene-specific PCR of the target, construct-specific PCR, and event-specific PCR. Among them, the event-specific method, which is recognized as the most specific method, targets the sequence of the linkage region between the exogenous integration vector and the plant genome (the exogenous gene flanking sequence). And since Agrobacterium transformation is commonly used to introduce exogenous genes so that the recombinant genes were randomly integrated into the plant genome, the flanking sequence was considered to be the more unique identity feature at present [9]. Palenzuela established a quantitative event assay for transgenic soybean GTS40-3-2 based on the fact that their primer amplification targets were located at the transgene insertion site and the linkage region of the soybean genome [10]. Marta established a quantitative assay for the event by thermal asymmetric interaction PCR (TAIL-PCR) to obtain the exogenous gene integration flanking sequence at the 3' end of the transgenic commercial maize MON810, and the structural analysis of the flanking sequence at the 5' end was found to have similarity with the known sequences in the database, but the 3' end of the flanking sequence was segregated into abnormal recombination that did not have any homology with the known sequences, thus providing the quantitative event detection method specific for the targeting of the 3' flanking sequence as the basis [11]. However, if the exogenous insertion structure is altered without changes in the flanking sequences at both ends, the event-specific method is not representative of its specificity, however, no cases have been found where deletions have been detected in commercially applied transgenic crops resulting in failure of specificity of the event method. However, this study has identified recombination deletions at the 3' end of exogenous insecticidal genes and terminators in commercially applied transgenic cotton. MON531 is a trans- insect-resistant cotton containing *cry1Ac* gene developed by Monsanto Company in the United States and has been approved for commercial cultivation in China. We found the that R7569 and MON531 have the same exogenous

insertion frame and flanking sequence, but the 3' end and the terminator sequence region of the *cry1Ac* gene sequence in R7569 had a recombination deletion. The deletion occurred at position 7569 at the 5' end of the whole insertion structure, so it was named R7569. In the present study, transcription, expression and insect resistance assays were carried out on a homozygous population to clarify the effects of the deletion at the 3' end of the target gene, so as to provide a case for the molecular characterization of transgenic crops. The study provided a case study for the risk assessment and monitoring of the upcoming large-scale release of transgenic crops in China.

2. Materials and Methods

2.1. Materials

In this study, an molecular structure different from the MON531 event was detected from domestic market, which was again verified from 240 market transgenic cotton seeds, planted in an incubator (temperature 24±2℃; photoperiod 16:8; humidity 65%-70%), and the plant was harvested called the T1-generation, which was transplanted to the greenhouse, and given normal conditions. The T2 generation seeds were harvested for exogenous insertion structure analysis. The T2 generation seeds were planted in the greenhouse also, and the first fully expanded leaves were collected from the top of the cotton plants at the seedling stage (4-6 leaves), the bud stage (the whole bud stage), and the boll stage (bolling stage), and were quickly frozen in liquid nitrogen for transcription, expression, and insect resistance analyses. Seeds of the T3 generation were harvested when the cotton was mature.

2.2. Validation of the Exogenous Insertion Structure of R7569

In order to clarify the difference between the insertion structure of R7569 and MON531, the sequence of insertion structure of R7569 was verified by qualitative PCR. Referring to the MON531 event structure, 5 pairs of primers were designed to cover its insertion structure using the software Vector NTI 11.5, and qualitative amplification was performed to verify its event-specific sequence, cry1Ac gene sequence and full-length insertion sequence, respectively. The reaction system was 20μL, of which 10μL of 2×Easy Taq PCR Super Mix (Beijing All Style Gold Biotechnology Co., Ltd., Beijing, China), 6μL of ddH2O, 1μL of each of the primers (10μmol/L), and 2μL of DNA. The PCR reaction procedure: 95℃ for 5 min; 35 cycles of 94℃ for 30 s, 58℃ for 30 s, 72℃ 2 min; 72℃ for 10 min, and 1.5% gel electrophoresis to detect the amplification products.

2.3. Establishment of Quantitative Testing Methods

The DNA of cotton seed of T2 generation in R7569 was used as the template, and five concentration gradients were set up, 100 ng/μl, 10 ng/μl, 1 ng/μl, 0.5 ng/μl, 0.2 ng/μl, and each gradient was set up three times in parallel. According to the calculation of 44,000 copies of 100 ng cotton genome, the corresponding copy numbers of five gradients were 44,000, 4,400, 440, 220, and 88. The logarithmic value of the copy number of the DNA template was used as the horizontal coordinate, and the Ct value was used as the vertical coordinate to establish the standard curve of the *cry1Ac* gene of R7569 and the standard gene of *ACP1* in cotton. The total reaction system was 20μL, and the RT-PCR reaction system was 20μL in total: 10μL of buffer SYBR® Premix Ex TaqTM (Tli RNaseH Plus), 1μL each of primers (10μmol/L), 0.4μL of fluorescent dye ROX, 2μL of cDNA, and 5.6μL of ddH2O. The real-time fluorescence quantification was carried out in a two-step method: 95℃ for 2 min; 94℃ for 5 s, 60℃ for 34 s, 40 cycles; collect the fluorescence signal at 60℃. The established real-time fluorescence quantitative PCR assay was used to quantitatively detect and analyze the genomic DNA of 476 transgenic cotton seeds nationwide in 2017, and three parallels were set for each concentration, and the experiment was repeated three times, and the inter-experimental standard deviation (SD) and relative standard deviation (SD) and relative standard deviation (RSD) between experiments. The content of R7569 was calculated with reference to the method reported by Yang et al [12]. Content of R7569 = number of insertion copies of R7569×2/number of copies of *ACP1* gene×100.

4

2.4. RT-PCR Analysis of cry1Ac Gene Transcript Analysis in R7569

The leaves of T2 generation of R7569 at seedling, bud and bell stages were selected and quickly ground into powder in liquid nitrogen, and about 0.1 g was dispensed into 2.0 mL RNAase-free centrifuge tubes to extract total RNA for analyzing the transcripts of cry1Ac genes and the transcript levels at different growth periods. Total RNA and reverse transcription were extracted according to the procedure of Plant Total RNA Extraction Kit (DP431) from TianGen Biochemical Technology Co. Smart® 5'and 3'RACE kit was purchased from Takara, qPCR SYBR Green Master Mix (Hieff UNICON® Universal Blue, YEASEN, Shanghai, China), Oligo (dT) Primer (Promega). Primary and secondary forward primers R7569-2387F1 and R7569-2426F2 were designed upstream of the recombination position of the cry1Ac gene to amplify with the universal long primer LP, and the universal short primer SP, respectively. For the transcripts of cry1Ac gene in R7569 needed to undergo two rounds of PCR amplification to obtain the high abundance of target sequences. Nested PCR amplification was used in the RACE experiments, and 41.5 µL of the Mix system for PCR was configured: 15.5 µL of ddH2O, 25.0 µL of 2×SeqAmp Buffer, and 1.0 µL of SeqAmp DNA polymerase; and 50 µL of the PCR system consisted of: 41.5 µL of Mix from step 1, 2.5 µL of RACE Ready cDNA, 5 µL of 10× UPM primer, 1 µL of gene-specific primer R7569-2387F1 or R7569-2426F2, and ddH2O supplemented to 50 µL. PCR procedures: 95℃ for 5 min; 25 cycles of 94℃ for 30 s; 68℃ for 30 s; 72℃ for 3 min; 72℃ for 10 min, and 10℃ for 2 min. 1.5% agarose gel electrophoresis was used to detect the amplification of the products, and the secondary PCR system and procedure were the same as that of the primary PCR. The recovered products of the nested PCR were ligated into the vector (pEASY-Blunt simple cloning vector, All-type Gold, Beijing, China) and transformed into the receptor cells, and the plates were coated to culture the clones. The positive clones with expected fragments were selected for sequencing, and the original sequences were removed from the vector sequences, imported into the analysis software Vector NTI, compared with the original sequences and counted the positions of the termination sites, and then plotted and analyzed.

The full-length *cry1Ac* gene in MON531 was used as a control to analyze the transcript levels of the *cry1Ac* gene of R7569 at three periods: seedling, bud and boll stage of cotton, and the *ACP1* gene was used as an internal reference gene, and the primer sequences were referred to the quantitative detection of the *ACP1* gene in Yang et al [12]. The transcription primers of *cry1Ac* gene in MON531were referred to the primer sequences of previous study [12]. The recombination region of cry1Ac gene of R7569 was used as the specific target region, and primers were designed for analyzing the transcript level of cry1Ac gene, and the expected amplified fragment was 281bp (not shown). The genomic DNA of R7569 was used as the template for the standard curve, six gradient dilutions (100, 10, 1, 0.5, 0.2, 0.1ng/µL) were performed, three parallels were set up, and a standard curve was made to calculate the amplification efficiency; the samples to be tested were diluted 10-fold to serve as the template, and three parallels were set up with the control samples and ddH2O as the quantitative control.

2.5. Determination of cry1Ac Insecticidal Protein by ELISA

T2 generation seeds, T2 generation seedling, bud and bell stage leaves, and T3 generation seeds were selected to determine the dynamic expression level of cry1Ac insecticidal protein using MON531 as a control. Field materials were sampled with reference to the national standard for insecticide resistance assay. The first fully expanded leaf on top was collected at seedling, bud and bell stages, and numbered and frozen in tin foil. They were ground and pulverised by liquid nitrogen grinding and divided into 2.0 mL RNAase-free centrifuge tubes (about 0.1 g of sample per tube) for total protein extraction. Preparation of protein extract (PBST): total protein was extracted and protein content was determined by the EnviroLogix cry1Ac/Ab Protein Quantification Kit (EnviroLogix, USA) procedure. The assay was set up three times in parallel. Exogenous insecticidal protein content was analyzed for significance of difference using Grafphad Prism 7.0 and plotted.

2.6. Determination of Resistance to Bollworm for Cry1Ac Proteins

Homozygous T3 seeds of transgenic cotton R7569, MON531 and non-transgenic cotton seeds (Zhongmian 49) were germinated and planted in seedling pots with nutrient soil, 10 pots of each type of cotton (20 cm in diameter), 5 seeds per pot, in an incubator at a temperature of 26±2℃ and a humidity of 60%-70%; photoperiod was 16L:8D. Leaves were collected from the first fully expanded leaf for insect resistance determination. The bollworm larvae of sensitive lines were provided by the professor Liang (Institute of Plant Protection, Chinese Academy of Agricultural Sciences). The bollworm larvae were treated with artificial feed for 24 h and used for indoor insect resistance assay.

Four groups of 10 leaves per treatment and five test insects per leaf were used in the insect resistance bioassay. Take a sterilised tube and fill it with 0.8% agar up to one-third of the volume of the tube, after the agar has condensed, take the first fully expanded leaf blade of a single fresh and clean cotton plant, insert the petiole into the agar to keep the leaf blade moist, gently access five heads of the first hatched larvae of the cotton bollworm with a moistened brush and plug the cotton plug tightly. The cotton plugs were sprayed with sterilised water and covered with a black cotton cloth (to prevent the escape of the test insects), and the bottom of the tubes were kept illuminated, and the bollworm larvae were observed and recorded feeding on the leaves in the incubator for six days, and no response to touching the insects with a brush was regarded as death, and the corrected mortality rate of the bollworm larvae was calculated.

The surface coating method was used for indoor bioassay as follows: to make artificial feed [2], use a syringe to suck up the unsolidified artificial feed to squeeze into 24-well culture plates, squeeze about 2 mL of feed into each well, so that there is no gap between the feed and the culture plate, and when the feed is solidified, the surface of the wall of the wells will be cleaned up to be flat and clean. Cotton leaf powder containing full-length cry1Ac, truncated cry1Ac protein, and no insecticidal protein was lyophilised and pulverised and added to the feed to make a poisoned feed, which was added to the corresponding 24-well plates for indoor bioassay. The insecticidal protein content was 0.03125%, 0.125%, 0.500%, 2.000% and 8%, respectively. In each well of the 24-well plate, 1 first-hatched larva of cotton bollworm (hatching time within 24 h) was accessed, and a total of 72 larvae were accessed in each concentration, with 24 larvae as one repetition and a total of three replications, and the plates with larvae were placed in a light incubator at a temperature of (26±1℃), a photoperiod of (16 L:8 D), and a relative humidity of (60%-70%) to cultivate the larvae, and the number of dead larvae was investigated and weighed after 7 d. The weights of the surviving larvae were recorded. The weight of the surviving larvae was recorded. Polo-plus software was used to analyze the raw data and calculate the slope and LC50 values, which indicated the concentration of Bt protein that caused 50% of the deaths of the grasshopper moths.

3. Results

R7569 Was Deleted at the 3′ Terminator Region of the cry1Ac Gene

The primers which can distinguish homozygosis and heterozygous MON531 was used to verify the flanking sequence of the insertion structure of R7569 (Figure 1a), the homozygosis sample amplified a single band of 529 bp, and the heterozygous sample amplified two bands of 529 bp and 615 bp, respectively. As can be seen in figure 1b, both R7569 and MON531 amplified a single band of 529 bp, which indicated that both of them were the homozygous samples of the transgene and both of them had the same flanking sequence (Figure 1b). The full-length sequence amplification of *cry1Ac* gene showed that MON531 contained the 3' end of the full-length *cry1Ac* gene sequence, whereas R7569 had a deletion of the 3' end of the *cry1Ac* gene sequence (Figure 1c). The exogenous insertion structure of R7569 was divided into five fragments, named F1, F2, F3, F4 and F5, and each fragment was used as a target sequence, and its insertion structure and sequence were analyzed by PCR amplification and clone sequencing and the full-length sequence of the insertion structure of R7569 was obtained (Figure 1d). From the integrated structure information of R7569, it can be seen that R7569 and MON531 had the same flanking sequence, but the 3' end of *cry1Ac* gene sequence in R7569 was deleted (Figure 1a). The amplification and sequencing results showed that the sequence of *cry1Ac* gene in R7569 was 2554 bp long, while that of MON531 was 3534 bp long, and the ORF of *cry1Ac*

gene in R7569 was predicted by bioinformatics, and it was found that the *cry1Ac* gene in R7569 coded for 881 amino acids, of which 851 amino acids were derived from the Bt gene, and another 30 amino acids were derived from the transcription termination site, which were derived from the amino acids produced by read-through after the transcription termination site.
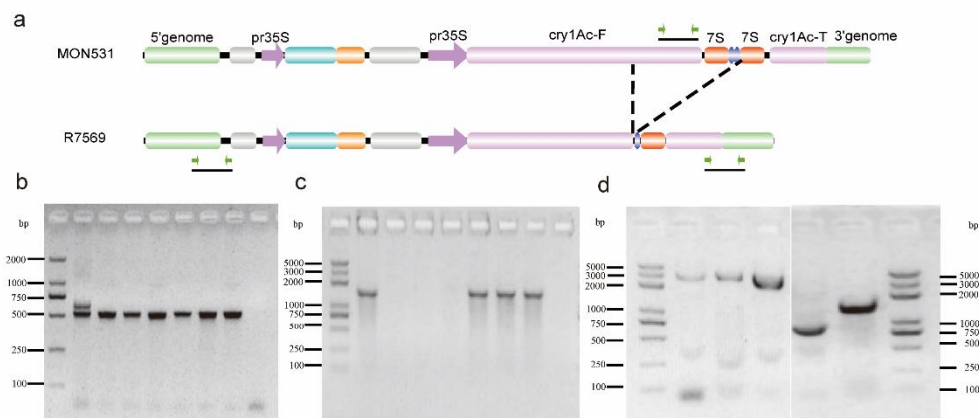
Figure 1. Exogenous insertion structure and verification results of R7569. (**a**). The whole exogenous insertion structure of R7569 and MON531. The black dashed line indicates the missing part of R7569. "F" indicated the full length of *cry1Ac* gene, and "T" indicated the segmental length of *cry1Ac* gene. The green arrow and the black line segment represent the PCR amplified fragment that validates the structure. (**b**). Verification of exogenous insertion structures in 3' terminal flanking sequence. (**c**). Verification of full length sequences for *cry1Ac* gene. (**d**). PCR amplified gel electrophoresis of full-length fragment of R7569.

Transcription Levels of the Truncated cry1Ac Gene in R7569 Were Higher than Those of the Full-Length cry1Ac Gene

Open reading frame prediction of the exogenous insert structure of R7569 and MON531 events were performed separately. The full-length *cry1Ac* gene in MON531 was 3534 bp long, encoding 1178 amino acids, and the *cry1Ac* gene of R7569 was 2554 bp long, encoding 881 amino acids. Amplification by primary and secondary PCR and agarose gel electrophoresis showed that the *cry1Ac* gene of R7569 obtained an expected amplified fragment of about 380 bp in secondary PCR, with diffuse bands in the expected amplified product due to transcript differences. All the single clones obtained were positively identified with an expected amplified fragment of around 500 bp, which would be amplified differently due to transcript differences. Sequencing (Figure 2a) analysis of all positive single clones showed that the transcriptional termination sites of truncated *cry1Ac* gene were mainly concentrated downstream of the truncation position and extended 160-270 bp from the truncation position, mainly at bases 161-173 and 233-265 downstream of the truncation position (Figure 2a,b).
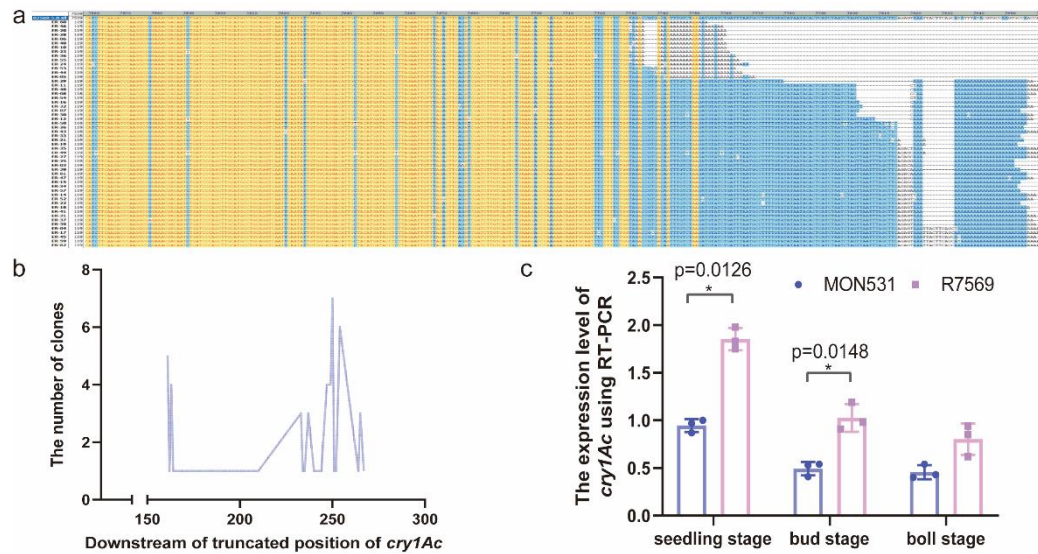
Figure 2. Transcription analysis of truncated *cry1Ac* gene in R7569. (**a**,**b**). The transcription termination site of truncated *cry1Ac* gene was analyzed by RACE. Compared with the full-length *cry1Ac* gene, the transcription termination site of truncated *cry1Ac* gene was located downstream of the truncated *cry1Ac* gene, and there were hot spots for transcription termination (54 sequences). (**c**). The transcription level of truncated *cry1Ac* gene compared to the full *cry1Ac* gene was analyzed by RT-qPCR. The stars on the horizontal line represented significant differences (t-test). The blue dot represented MON531 and the rose square represented R7569.

The genomic DNA content was measured and then diluted with ddH$_2$O into six concentration gradients as the template for establishing standard curve, and five of the valid points were selected to establish the standard curves of *cry1Ac* gene and the internal reference gene ubiquitin. Among them, the linear regression equation of *cry1Ac* gene in R7569 was Copies=10^(-0.2961*Ct+10.8149), with linear correlation coefficient R^2=0.9988, and the amplification efficiency was 97.7% (Fig S1a); and that of the full-length *cry1Ac* gene in MON531 was Copies=10^(- 0.2852*Ct+10.4909), linear correlation coefficient R^2=0.9949, and amplification efficiency was 92.9% (Figure S1b); the linear regression equation for the internal reference gene *ACP1* was Copies=10^(-0.2851*Ct+9.9019), linear correlation coefficient R^2=0.9949, and amplification efficiency was 92.8% (Figure S1c). All of the above data conformed to the international standard for quantitative transgene detection (-3.1 ≥ Slope ≥ -3.6; R^2 ≥0.98), and this fluorescence quantitative PCR was suitable for fluorescence quantitative analysis of *cry1Ac* gene and the internal reference gene *ACP1* in R7569 and MON531.

The trends of transcript levels of *cry1Ac* gene in R7569 and MON531 were compared at seedling, bud and bell stages, and t-tests were done to test the significance (Figure 2c). It can be seen that the transcript levels of *cry1Ac* genes of R7569 and MON531 gradually decreased with the seedling, bud and bell stages, among which the transcript levels of *cry1Ac* genes of R7569 were significantly higher than that of MON531 at seedling and bud stages (t-test, P<0.05), and there was no significant difference in the transcript levels of *cry1Ac* genes at the boll (t-test, P>0.05) (Figure 2c).

The Protein Content of the Truncated cry1Ac Gene Was Higher than That of the Full-Length cry1Ac Gene during the Cotton Growing Season

Measurement of insecticidal protein expression levels at seedling, bud and boll stage showed that both R7569 and MON531 showed a gradual decrease in insecticidal protein content in cotton leaves at seedling, bud and boll stage. The expression levels of exogenous insecticidal proteins of R7569 were all extremely significantly higher than those of MON531 at the same reproductive stage (t-test, P<0.01) (Figure 3a).
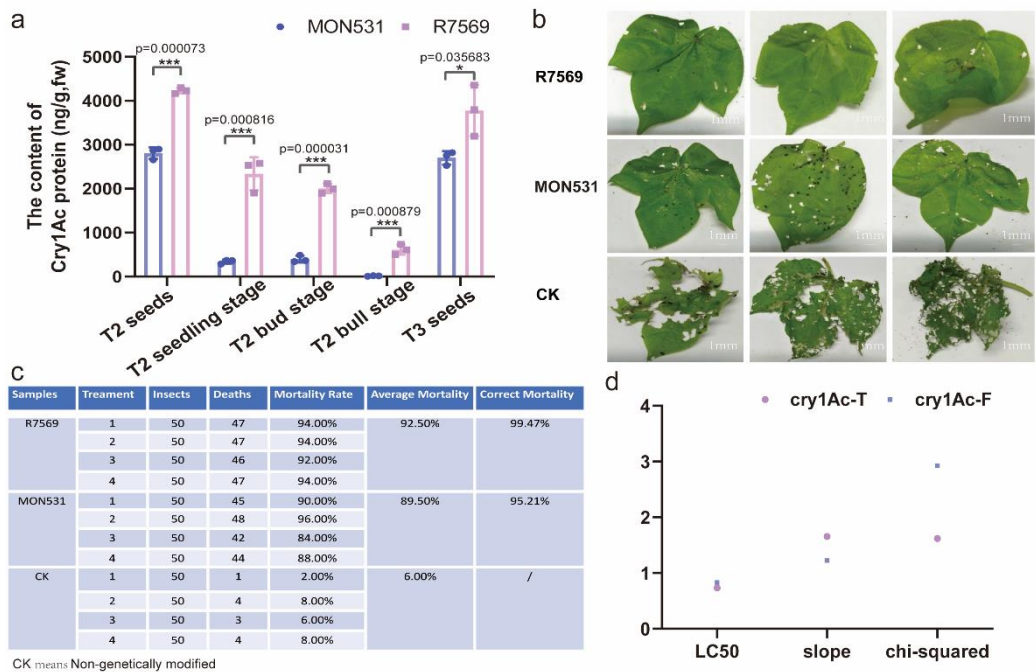
Figure 3. Determination of protein content and insect resistance of R7569. (a). The content of truncated Cry1Ac protein was determined using ELISA. The stars on the horizontal line represented significant differences (t-test). The blue dot represented MON531 and the rose square represented R7569. **(b,c)**. The resistance levels of Cry1Ac proteins with different lengths were determined by leaf feeding in vitro according to previous method. **(d)**. The LC$_{50}$ values of the truncated and full-length Cry1Ac protein were determined. "F" indicated the full length of *cry1Ac* gene, and "T" indicated the segmental length of *cry1Ac* gene.

The Truncated Insecticidal Protein Cry1Ac Had a Higher Level of Biological Resistance than the Full-Length Insecticidal Protein Cry1Ac

The area of bollworm feeding on the leaves of insect-resistant cotton can reflect the insect resistance of *cry1Ac* gene cotton R7569, MON531 and non-transgenic cotton (Figure 3b). Bollworms on R7569 cotton leaves to a lesser degree of damage and the least area of feeding; bollworms on MON531 cotton leaves to the second degree of damage, leaving a slightly larger damage area; non-transgenic cotton damage was the most serious, leaving the largest damage area (Figure 3b). The insect resistance of the seedling leaves of the incubator-grown transgenic insect-resistant cotton MON531 and R7569 was determined by the isolated leaf method. From the statistical data, it can be seen that the corrected mortality rate of insect-resistant cotton containing truncated Cry1Ac in R7569 against bollworm larvae was 99.47%, and the resistance level of it was high; the corrected mortality rate of MON531 was 95.21%, and the resistance level of it was the same as R7569 (Figure 3c).

Compared with the non-transgenic counterpart, the resistance level of both to the target insect, cotton bollworm, was at a high level. The insect resistance of the truncated Cry1Ac insecticidal protein was determined by the Bt protein mixing method, and the results showed that the LC50 value, slope and chi-square value of the truncated Cry1Ac protein of R7569 was 0.732 ng/g (dw), 1.654 and 1.617, respectively. For the content of full-length of Cry1Ac protein, there were 0.827 ng/g (dw), 1.225 and 2.926, respectively. The insecticidal activity of the full-length and truncated Cry1Ac proteins was better, and the insecticidal activity of the truncated Cry1Ac protein was better than that of the full-length Cry1Ac protein (Figure 3d).

4. Discussion

The application of genetically modified organisms (GMOs) in the field of crops has realized multiple benefits and is expected to alleviate the current challenges associated with the commercial

and agricultural sectors [13]. The *cry1Ac* gene from Bacillus thuringiensis was artificially modified and introduced into the cotton genome, where the cry1Ac gene was transcriptionally translated into insecticidal proteins that could specifically control target pests such as Lepidoptera and Coleoptera [14]. In this study, a deletion of the exogenous insecticidal protein Cry1Ac was found for the first time in a commercially available transgenic crop; however, this deletion of the truncated *cry1Ac* insecticidal gene possessed better performance for biological applications. The transcription and expression levels of the truncated *cry1Ac* gene in leaves were significantly higher than those of the full-length *cry1Ac* gene during the critical period of cotton growth (t-test, $P < 0.05$ or $P < 0.0001$), suggesting that it may have better biological resistance. Subsequently, the results of indoor bioassays in the cotton isolated leaf method and feed mix method showed that both full-length and truncated Cry1Ac insecticidal protein resistance grades were at high levels compared with the non-transgenic cotton control, but the corrected mortality of bollworm larvae was higher in the R7569 insecticide-resistant cotton with truncated Cry1Ac insecticidal protein than that of the MON531 insecticide-resistant cotton with full-length Cry1Ac insecticidal protein (t-test, $P < 0.05$). Meanwhile, the LC50 value of the truncated Cry1Ac protein in R7569 was smaller than that of the full-length Cry1Ac protein in MON531 (t-test, $P > 0.05$). Both slopes were close to 1 indicating reliable results. The results showed that the insecticidal activity of the truncated Cry1Ac protein was superior to that of the full-length Cry1Ac protein. This phenomenon that the natural variation of exogenous genes occurring in commercially applied transgenic crops showed superior biological resistance can provide a multifaceted reference for the cultivation and evaluation of transgenic crops that will be widely applied in China.

The introduction of exogenous genes into the genome of recipient plants in transgenic crops can cause abnormal recombination and deletion of exogenous insertion structures, resulting in incomplete single-copy or multi-copy insertion of exogenous genes. The exogenous insertion structure of R7569 was verified, and it was identical to that of the commercially applied transgenic insecticide cotton event (MON531), with the same flanking sequences of the 5' and 3' ends of the exogenous gene and the only difference, a recombination deletion of the 3' end of the exogenous insecticide gene and the terminator sequence of R7569. The presence of incomplete reading frames in exogenous genes has been reported previously [5], and the insertion structure of a generation of insect-resistant cotton MON757 event developed by Monsanto contains an incomplete section of the *cry1Ac* gene, including the promoter and part of the insecticidal gene sequence [12]. In addition, the insect-resistant transgenic maize MON810 was found to have sequences of the 3' end coding region of the cry1Ab gene in the exogenous insertion structure and the NOS terminator lost during transformation [15,16], and analysis of the study showed that its cry1Ab gene passages to the incomplete reading frame of the exogenous gene in MON757 and MON810 events was accompanied by an intact exogenous insertion that steadily the presence of the exogenous gene in the events [5,17]. Accompanying intact exogenous insertions was stably present in both cotton and maize genomes, as demonstrated in material exposed early in their developmental stage, and were therefore likely to have arisen during the integration of exogenous genes into the cotton genome. However, the reason for the generation of the truncated *cry1Ac* gene in R7569 has not yet been clarified, but since R7569 and MON531 have identical flanking sequences at both ends, and based on the extremely high specificity of the genome integrated by Agrobacterium transformation, the study hypothesised that R7569 was a recombination deletion mutant event originating from MON531, and the specific reasons for the occurrence of the recombination deletion were speculated as follows There are two ways: the first one is that the two integrated structures were generated during the early breeding process, but R7569 has always existed in the early breeding materials, and had good trait performance during the later application process, so it was not screened out, and has always existed along with the MON531 population. The second possibility is that the truncated *cry1Ac* gene arose as a result of chromosomal and gene fragment recombination in the cotton genome itself during multiple generations of breeding in the MON531 population. Since R7569 was found to be "heterozygous" for both the full-length and the truncated *cry1Ac* gene during the identification of R7569, it was concluded that R7569 was most likely derived from the MON531 event. In addition, since R7569 and MON531 have the same flanking

sequence, the existing event-specific detection methods cannot distinguish between the two, this study designed a specific quantitative PCR method for the recombination region of R7569, and found that there was a certain proportion of R7569 in the transgenic cotton material in China's market. For the recombination-deficient events found in this study, the recombination-based specific detection method may provide a more scientific and reasonable judgement standard for the event-specific detection method, and help the detection and monitoring of the transgenic components at the same time.

The variation and expression of natural full-length Cry1Ac protein, naturally occurring truncated Cry1Ac protein, artificially modified full-length Cry1Ac protein, and artificially modified truncated Cry1Ac protein based on the transgenic plants have been less reported [6]. It was found that the truncated *cry1Ac* gene was able to be transcribed normally and that the transcript levels of the *cry1Ac* gene were significantly higher in both seedling and bud stage leaves of the insect-resistant cotton event of the pure R7569 event than those of MON531 (t-test, $P<0.05$). Yang analyzed the transcription of the truncated *cry1Ac* gene in the exogenous *cry1Ac* gene high-expressing and low-expressing types in the MON757 event clarified that the transcription level of the *cry1Ac* gene in the high-expressing type was seven times higher than that in the low-expressing type [12], confirming that the truncated gene was able to be transcribed normally. The higher transcription level of this truncated gene may be caused by the higher transcription efficiency of the short fragment, and the specific reason needs to be further explored. In addition, the truncated *cry1Ab* gene in the transgenic insect-resistant maize MON810 was also able to transcribe and express insecticidal proteins normally [17]. Koul et al. introduced the full-length (3534bp) and truncated (1845bp) *cry1Ac* genes into tomato to obtain the insect-resistant transgenic tomatoes called pNBR1-1 and pRD400, respectively [6]. The results of Cry1Ac insecticidal protein content showed that the soluble insecticidal protein content in pNBR1-1 was 0.0001-0.0026%, and the soluble insecticidal protein content in pRD400 was 0.0020-0.0128%, and the insecticidal protein content of the transgenic tomato containing the full-length *cry1Ac* gene was lower than that of the transgenic tomato containing the truncated *cry1Ac* gene [6].The exogenous insecticidal protein content of R7569 was higher than that of MON531 during the same reproductive period (t-test, $P<0.0001$) indicating that the expression of truncated Cry1Ac protein was higher than that of full-length Cry1Ac protein. Genes become mature proteins through transcription and translation to exert their functional activities, and it is not the case that genes with high transcription levels also have high protein expression, but the higher content of insecticidal proteins may be related to the higher transcription level of their *cry1Ac* genes. The higher transcript level of the truncated *cry1Ac* gene than that of the full-length *cry1Ac* gene may be one of the reasons why the Cry1Ac insecticidal protein content of R7569 was higher than that of MON531. However, genes with higher transcription levels do not necessarily exhibit the same level of protein expression, so the fact that the truncated gene can be transcribed normally and has a high transcription level does not directly infer the excellent performance of its exogenous insecticidal protein content and biological functional activity [18]. It has been shown that different *cry1Ac* protein contents were correlated with bollworm survival, and that as the cotton growing season lengthens and plant maturity increases, the Cry1Ac insecticidal protein content decreases, accompanied by an increase in bollworm survival [19,20]. Koul conducted a study on insect resistance of transgenic tomatoes containing truncated cry1Ab genes. transgenic tomatoes containing the truncated *cry1Ab* gene were assayed for insect resistance, and the results showed that the corrected mortality of cotton bollworms to transgenic tomatoes containing the truncated *cry1Ab* gene was 100% [21]. In the present study, the corrected mortality rate of R7569 transgenic insect-resistant cotton against bollworm larvae was 99.47%, suggesting that truncated insecticidal proteins have better insecticidal activity. The protein content of full-length and truncated insecticidal proteins differed significantly, but both had good insecticidal activity. Since heterozygous transgenic cotton monocots for both full-length and truncated *cry1Ac* genes were found in R7569 in this study, the possibility of mutualism between full-length and truncated Cry1Ac proteins, and the resulting altered resistance to the target pests need to be further explored.

5. Conclusions

This study was the first to identify a deletion of an exogenous gene present in a commercially applied transgenic crop that nevertheless exhibits improved levels of transcription, expression, and biotic resistance. The resulting detection methods, market monitoring, and biological management strategies for target pests due to interactions between full-length and truncated proteins need to be addressed, and this study provides a reference for the detection, monitoring, and safety evaluation of transgenic crops for wider market applications.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Figure S1:Standard curve of *cry1Ac* gene using RT-qPCR; Table S1:Primers used in this study.

Author Contributions: Data curation, Chenhui Hou and Haoyuan Guan; Funding acquisition, Jiajian Xie; Methodology, Yongjun Zhang; Supervision, Xiuping Chen; Validation, Huilin Yu; Writing – original draft, Chunmeng Huang. All authors have read and agreed to the published version of the manuscript.
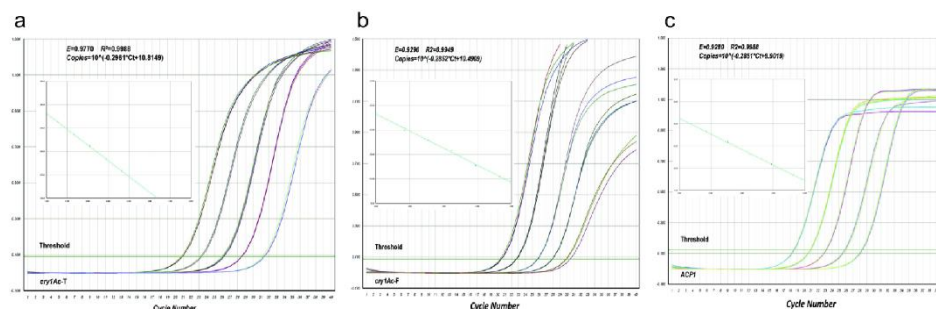
Data Availability Statement: All data generated or analyzed during this study are included in this published article.

Conflicts of Interest: The authors have not declared a conflict of interest.

Appendix A

Figure A1. Standard curve of *cry1Ac* gene using RT-qPCR. (a) Standard curve of truncated *cry1Ac* gene using RT-qPCR. (**b**) Standard curve of full *cry1Ac* gene using RT-qPCR. (**c**) Standard curve of *ACP1* gene using RT-qPCR.

References

1. Hashmi, J.A.; Zafar, Y.; Arshad, M.; Mansoor, S.; Asad, S. Engineering Cotton (Gossypium Hirsutum L.) for Resistance to Cotton Leaf Curl Disease Using Viral Truncated AC1 DNA Sequences. *Virus Genes* **2011**, *42*, 286–296, doi:10.1007/s11262-011-0569-9.
2. Zhou, Y.; Huang, C.; Chen, Y.; Han, L.; Xie, J.; Chen, X. Sensitivies of Fall Armyworm (Spodoptera Frugiperda) Populations in Different Regions of China to Four Bt Proteins. *Agronomy* **2023**, *13*, doi:10.3390/agronomy13092415.
3. Székács, A.; Lauber, É.; Juracsek, J.; Darvas, B. Cry1Ab Toxin Production of MON 810 Transgenic Maize. *Environ. Toxicol. Chem.* **2010**, *29*, 182–190, doi:10.1002/etc.5.
4. la Paz, J.L.; Pla, M.; Papazova, N.; Puigdomènech, P.; Vicient, C.M. Stability of the MON 810 Transgene in Maize. *Plant Mol. Biol.* **2010**, *74*, 563–571, doi:10.1007/s11103-010-9696-2.
5. Rosati, A.; Bogani, P.; Santarlasci, A.; Buiatti, M. Characterisation of 3' Transgene Insertion Site and Derived MRNAs in MON810 YieldGard® Maize. *Plant Mol. Biol.* **2008**, *67*, 271–281, doi:10.1007/s11103-008-9315-7.
6. Koul, B.; Yadav, R.; Sanyal, I.; Amla, D.V. Comparative Performance of Modified Full-Length and Truncated Bacillus Thuringiensis-Cry1Ac Genes in Transgenic Tomato. *Springerplus* **2015**, *4*, doi:10.1186/s40064-015-0991-x..

7.   Meng, Y.; Wang, S.; Guo, J.; Yang, L. Collaborative Ring Trial of the Applicability of a Reference Plasmid DNA Calibrant in the Quantitative Analysis of GM Maize Event MON810. *Foods* **2022**, *11*, doi:10.3390/foods11111538.

8.   Nielsen, C.R.; Berdal, K.G.; Bakke-Mckellep, A.M.; Holst-Jensen, A. Dietary DNA in Blood and Organs of Atlantic Salmon (Salmo Salar L.). *Eur. Food Res. Technol.* **2005**, *221*, 1–8, doi:10.1007/s00217-005-1160-1.

9.   Holst-Jensen, A.; Bertheau, Y.; de Loose, M.; Grohmann, L.; Hamels, S.; Hougs, L.; Morisset, D.; Pecoraro, S.; Pla, M.; den Bulcke, M. Van; et al. Detecting Un-Authorized Genetically Modified Organisms (GMOs) and Derived Materials. *Biotechnol. Adv.* **2012**, *30*, 1318–1335, doi:10.1016/j.biotechadv.2012.01.024.

10. Manzanares-Palenzuela, C.L.; Mafra, I.; Costa, J.; Barroso, M.F.; De-Los-Santos-Álvarez, N.; Delerue-Matos, C.; Oliveira, M.B.P.P.; Lobo-Castañón, M.J.; López-Ruiz, B. Electrochemical Magnetoassay Coupled to PCR as a Quantitative Approach to Detect the Soybean Transgenic Event GTS40-3-2 in Foods. *Sensors Actuators, B Chem.* **2016**, *222*, 1050–1057, doi:10.1016/j.snb.2015.09.013.

11. Tahar, S. Ben; Salva, I.; Brants, I.O. Genetic Stability in Two Commercialized Transgenic Lines (MON810). *Nat. Biotechnol.* **2010**, *28*, 779, doi:10.1038/nbt0810-779b.

12. Yang, Y.; Ye, W.; Fan, J.J. Event-Specific PCR Detection Methods of Genetically Modified Cotton (Gossypium Hirsutum) MON757 and Their Application. **2016**, *24*, 908–918, doi:10.3969/j.issn.1674-7968.2016.06.014.

13. Raman, R. The Impact of Genetically Modified (GM) Crops in Modern Agriculture: A Review. *GM Crop. Food* **2017**, *8*, 195–208, doi:10.1080/21645698.2017.1413522.

14. Naranjo, S.E.; Ruberson, J.R.; Sharma, H.C.; Wilson, L. *Integration of Insect-Resistant Genetically Modified Crops within IPM Programs*; 2008; ISBN 9781402083730.

15. VAN RIE, J.; JANSENS, S.; HÖFTE, H.; DEGHEELE, D.; VAN MELLAERT, H. Specificity of Bacillus Thuringiensisδ-endotoxins: Importance of Specific Receptors on the Brush Border Membrane of the Mid-gut of Target Insects. *Eur. J. Biochem.* **1989**, *186*, 239–247, doi:10.1111/j.1432-1033.1989.tb15201.x.

16. Hernández, M.; Pla, M.; Esteve, T.; Prat, S.; Puigdomènech, P.; Ferrando, A. A Specific Real-Time Quantitative PCR Detection System for Event MON810 in Maize YieldGard® Based on the 3′-Transgene Integration Sequence. *Transgenic Res.* **2003**, *12*, 179–189, doi:10.1023/A:1022979624333.

17. La Paz, J.L.; Vicient, C.; Puigdomènech, P.; Pla, M. Characterization of Polyadenylated CryIA(b) Transcripts in Maize MON810 Commercial Varieties. *Anal. Bioanal. Chem.* **2010**, *396*, 2125–2133, doi:10.1007/s00216-009-3176-z.

18. Hamamoto, K.; Fukaya, T. Dynamic Interplay between Non-Coding Enhancer Transcription and Gene Activity in Development. *bioRxiv* **2022**, 2022.08.02.502573, doi:10.1038/s41467-023-36485-1.

19. Adamczyk, J.J.; Perera, O.; Meredith, W.R. Production of MRNA from the Cry1Ac Transgene Differs among Bollgard® Lines Which Correlates to the Level of Subsequent Protein. *Transgenic Res.* **2009**, *18*, 143–149, doi:10.1007/s11248-008-9198-z.

20. Park, D.; Choi, I.Y.; Kim, N.S. Detection of MPing Mobilization in Transgenic Rice Plants. *Genes and Genomics* **2020**, *42*, 47–54, doi:10.1007/s13258-019-00877-9.

21. Koul, B.; Srivastava, S.; Sanyal, I.; Tripathi, B.; Sharma, V.; Amla, D.V. Transgenic Tomato Line Expressing Modified Bacillus Thuringiensis Cry1Ab Gene Showing Complete Resistance to Two Lepidopteran Pests. *Springerplus* **2014**, *3*, 1–13, doi:10.1186/2193-1801-3-84.