TDL.ORG TEXAS DIGITAL LIBRARY

Journal of Regulatory Science

http:\\journalofregulatoryscience.org

Journal of Regulatory Science 7 (2019) 1–14

Regulatory Science

Comparison of Southern-by-Sequencing (SbS™) Technology and Southern Blot Analysis for Molecular Characterization of Genetically Modified Crops

Kent Brink[a,*], S. C. Anitha[b], Mary K. Beatty[a], Jennifer Anderson[a], Megan Lyon[a], Janine Weaver[a,c], Nina Dietrich[a,d]

[a]*Corteva Agriscience™, Agriculture Division of DowDuPont, Johnston, IA 50131*
[b]*Corteva Agriscience™, Agriculture Division of DowDuPont, Hyderabad, TS 502336, India*
[c]*Stine Research Center, Bldg 320/333A, 1090 Elkton Rd., PO Box 30, Newark, DE 19714*
[d]*Delaware Criminal Justice Information System (DELJIS), 800 Silver Lake Blvd Suite 101, Dover, DE 19904*

Abstract

Southern blot analysis is typically used for molecular characterization of genetically modified (GM) crops. Southern-by-Sequencing (SbS™ technology, hereafter referred to as SbS) is a high-throughput, sequence-based alternative technique utilizing targeted sequence capture coupled with next-generation sequencing (NGS) and bioinformatics tools to achieve the same molecular endpoints. To demonstrate that both SbS and Southern blot analysis reach the same conclusions about insertion copy number and intactness of the inserted DNA, both techniques were used to characterize four GM soybean [*Glycine max* (L.) Merr.] events containing simple or complex DNA insertions. To demonstrate that both techniques reach the same conclusions about the presence of unintended DNA, GM maize (*Zea mays* L.) events containing *Agrobacterium* plasmid backbone fragments were characterized. Additionally, oligonucleotides containing varying lengths of target sequence were analyzed to compare both techniques' sensitivity for detecting small insertions. SbS and Southern blots had similar sensitivity and provided comparable results for copy number and intactness of simple and complex DNA insertions. Both techniques also had comparable results for detection of unintended plasmid backbone DNA sequences and small DNA fragments. Thus, SbS can deliver the same endpoints as Southern blot analysis for key molecular characterization aspects of GM crops and gene edited varieties, providing important information to inform regulatory decisions.

Keywords: Molecular characterization, genetically modified (GM) crops, Southern-by- Sequencing (SbS™) technology, Southern blot, next-generation sequencing

Abbreviations: GM, genetically modified; SbS, Southern-by-Sequencing (SbS™ technology); NGS, next-generation sequencing; bp, base pair; SI, single intact insertion; DI, tandem double insertion; DR, single insertion with deleted region; CI, complex insertion with multiple joined fragments; DIG, digoxigenin; JSA, junction sequence analysis; PCR, polymerase chain reaction

1. Introduction

The safety assessment framework for genetically modified (GM) crops is well established [9, 5, 6]. Typically, Southern blot hybridization analysis [16], polymerase chain reaction (PCR), and Sanger sequencing are used to characterize GM crops to support regulatory data requirements. Southern blot analysis is used to determine copy number and intactness of the inserted DNA, which includes determining the number of separate insertions within the plant genome and the number of copies of the intended DNA within each insertion, as well as confirming the DNA was inserted as expected (i.e., detecting if fragmented, rearranged, or unintended plasmid backbone DNA was incorporated along with the intended DNA insertion). PCR is typically used to confirm the presence of the insertion, and Sanger sequencing of the entire insertion and flanking genomic regions is used to detect small truncations at the ends of the intended DNA insertion or any changes within the inserted DNA. Paired with PCR and Sanger-based sequencing analysis, Southern blot analysis can provide the required information on the structure of the inserted DNA and flanking regions. While these molecular techniques are robust, provide high quality, reproducible data, and are well understood by regulators and scientific experts, Southern blot analysis can be labor intensive, and is technically challenging for complex insertions [8]. For example, Southern blot techniques require the use of multiple restriction enzymes and numerous different probes to analyze the inserted DNA. Additionally, because Southern blot analysis

*Corresponding author: Kent Brink, Phone: (515) 535-2048, Email: kent.brink@corteva.com, ORCID iD: 0000-0003-4395-9740

is image-based, interpretation often requires considerable experience, and variability at multiple steps can lead to subjective evaluation.

Next-generation sequencing (NGS) technologies have been developed to provide comprehensive molecular characterization data with high efficiency and data quality [3, 8, 12, 19]. NGS is an approach that analyzes millions of short overlapping sequence reads across the DNA sample and analyzes them in a bioinformatics pipeline, which enables detection of copy number of inserted DNA, determination of insertion intactness, absence of plasmid backbone DNA sequences, and other important molecular characterization endpoints [12]. Furthermore, NGS can detect more complex rearrangements and other structural variants, when compared to Southern blots. NGS technology coupled with junction sequence analysis (JSA) has been used to verify the intactness and stability of the inserted DNA in several GM lines, including MON 87411, MON 87419, and MON 87403, by comparing the comprehensive coverage of sequence fragments across the genome of non-modified plants (i.e., near isogenic controls) and the GM plant [18].

Southern-by-Sequencing (SbS™ technology, hereafter referred to as SbS) has recently been highlighted as a robust sequence-level application that utilizes sequence capture coupled with NGS technology and bioinformatics tools for high-throughput event selection [19]. The advantage of SbS analysis for molecular characterization of GM lines is the targeted capture step prior to NGS, which allows for very high depth of sequence reads at any sequence junction involving DNA derived from the transformation plasmid [8, 19]. This approach provides high confidence that all such junctions are accurately identified. The combination of junction number and the assembled sequence adjacent to the junctions, when compared to the known transformation DNA sequence, allow a detailed insertion map to be developed. This map includes all the major elements needed for molecular characterization: number of separate insertions, copy number of each genetic element found in the insertion(s), any rearrangements or truncations that occurred in the inserted DNA, and whether any unintended plasmid DNA was incorporated into the plant genome.

A comprehensive direct comparison of SbS and Southern blot for sensitivity and accuracy has not yet been published. Herein, we compare the sensitivity of SbS to Southern blot analysis, with respect to their ability to characterize insertions ranging from simple to complex, to detect short sequences of inserted DNA, and to determine the presence of plasmid backbone DNA sequences in soybean and GM maize lines. The latter is also directly applicable to gene edited varieties where the absence of unintentionally incorporated plasmid DNA is one of the key objectives of molecular characterization.

2. Methods

2.1. Southern-by-Sequencing

SbS analysis was conducted following methods previously described [19]. Briefly, a DNA probe library was designed and constructed to capture the transformation plasmid DNA sequence (Roche NimbleGen, Madison, WI, USA). Leaf tissue was collected from each transformed GM line and the control line, and genomic DNA was extracted (Omega Bioteck E-Z 96 Plant DNA Kit, Norcross, GA, USA). Sequencing libraries were constructed and sequencing fragments were enriched with target capture probes. Enriched library pools of each transformed line and the control line were sequenced on the Illumina HiSeq2500 system (Illumina Inc., San Diego, CA, USA), following manufacturer's protocols. Alignment-based transformation plasmid backbone analysis was followed by junction detection and filtering to detect plasmid backbone junction sequences, and was subsequently followed by detection and removal of endogenous junctions (i.e., junctions that exist within the plant genome that are captured by the SbS process). To generate physical maps of the insertions, the final junctions were mapped to characterize insertion site and intactness of the inserted DNA using BLAT, version 35x1 [10].

During the bioinformatics analysis following NGS, sequence reads that showed partial homology to the plasmid DNA sequence (while the rest of the read did not match the contiguous plasmid sequence) were identified as junctions between inserted DNA and genomic DNA, or between insertions of two plasmid DNA sequences that were not contiguous in the original plasmid. Multiple sequence reads were generated for each junction and these reads were compiled into a consensus sequence for the junction. A unique junction was defined as one in which the plasmid-derived sequence and the adjacent sequence were the same, although the overall length of the multiple reads for that junction varied due to the sequencing process. The number of unique junctions was related to the number of plasmid insertions present in the genome (for example, a single DNA insertion was expected to have two unique junctions). Detection of additional unique junctions beyond the two plasmid-genome junctions expected for a single insertion indicated the presence of additional plasmid insertions.

2.2. Southern blot

Southern blot analysis was conducted following methods previously described [2], with some modification. Briefly, genomic DNA was isolated from leaf tissues using the urea-based procedure [4] or using a high-salt extraction buffer (2.0 M NaCl, 100 mM Tris-HCl pH-8.0, 50 mM of sodium salt of EDTA and 100 mM sodium metabisulphite) procedure. DNA was quantified on a spectrofluorometer using the Quant-iT™ PicoGreen®dsDNA Assay Kit (Molecular Probes, Inc., Eugene, OR, USA) and visualized on an agarose gel to determine the DNA quality. Following restriction enzyme digestion (New England BioLabs, Ipswich, MA or Thermo Fisher Scientific., Waltham, MA, USA), fragments were separated, transferred to a nylon membrane, crosslinked by ultraviolet light, and detected as discrete bands when hybridized with a digoxigenin (DIG)-labeled probe, following methods similar to [2] and using DIG-labeled molecular weight markers (DIG II, DIG VI and DIG VII; Roche, Indianapolis, IN, USA). A CDP-Star Chemiluminescent Nucleic Acid Detection System with DIG Wash and Block Buffer Set (Roche) and an appropriate image analyzer (for example, ImageQuant LAS 4000, GE Healthcare

Bio-Sciences) or x-ray film were used to visualize and digitally capture images.

2.3. Analysis of simple and complex insertions

To determine if both SbS and Southern blot methods provide comparable results for the characterization of simple and complex inserts, a population of soybean GM lines was created from embryogenic cultures following microprojectile bombardment transformation protocol [7, 11, 17]. Briefly, soybean embryogenic suspension cultures were generated, as described previously [13] and were maintained in 250 ml flasks containing 50 ml of liquid media on rotary shakers at 26 °C under cool white fluorescent lights with a 16:8 h day:night photoperiod. Freshly sub-cultured cultures were bombarded with 0.6 μ gold particles coated with the linear DNA fragment PHP63750A, derived from plasmid PHP63750 (Supplement 1: Figure S1), using a biolistic instrument PDS1000/HE (Bio-Rad, Hercules, CA, USA). All regenerated soybean GM lines were analyzed by SbS, as described above. SbS utilized capture probes homologous to the transformation plasmid to isolate genomic DNA that hybridized to the probe sequences. Captured DNA was then sequenced using a NGS procedure and analyzed using bioinformatics tools [19].

Four soybean GM lines that contained DNA insertions of varying complexity [SI (single intact insertion), DI (tandem double insertion), DR (single insertion with deleted region), and CI (complex insertion with multiple joined fragments)] were selected for Southern blot analysis. Soybean plants from each of the four soybean GM lines and a control line (untransformed soybean line of the same genetic background) were grown under greenhouse conditions, and leaf tissue was harvested and frozen prior to DNA extraction for Southern blot analysis. Genomic DNA samples extracted from transformed and control plants were digested with *Bgl*II or a double digest of *Spe*I and *Avr*II (Supplement 1: Figure S1) for characterization of the PHP63750A insertions. The CI line was analyzed further by digestion with *Bam*HI, *Bcl*I, *Eco*RV, or *Nco*I and hybridization with the same probes for a more complete analysis (Supplement 1: Figure S2). Plasmid PHP63750 DNA was also included to verify probe hybridization and to serve as a size reference for fragments internal to the DNA insertion. Following transfer, bound DNA fragments were detected as discrete bands when hybridized with a labeled probe (Supplement 1: Table S1). The probes were designed to cover all nucleotides of the PHP63750A transformation fragment and included some adjacent nucleotides from the PHP63750 plasmid backbone. Southern blot analysis was used to characterize the DNA insertions and create physical maps of the insertions in these four soybean lines to compare to the maps derived from SbS analysis.

2.4. Detection of small DNA fragments

To determine if both SbS and Southern blot methods have comparable ability to detect short sequences, eleven 250 bp double stranded oligonucleotides were designed and synthesized by a commercial vendor (Life Technologies Corporation, Carlsbad, CA, USA or Integrated DNA Technologies, Coralville, IA, USA). Methods used to create the oligonucleotide sequences and to detect them by SbS were previously described [19]. Each 250 bp oligonucleotide contained a variable length fragment of *Agrobacterium* plasmid backbone DNA (size ranged from 35 to 100 bp) inserted between maize genomic sequences (Supplement 2: Table S1). A 250 bp oligonucleotide entirely composed of sequence from a different part of the *Agrobacterium* plasmid backbone was used as a positive control for the process. The same set of oligonucleotides was added to maize genomic DNA and analyzed by Southern blot to determine method sensitivity.

Leaf tissue from untransformed maize plants was harvested and maintained frozen (< -50 °C) until processing. Maize genomic DNA was extracted with a high-salt buffer and was sequentially precipitated using one-sixth volume of 5.0 M potassium acetate and 0.6 volume of isopropyl alcohol. DNA was treated with ribonuclease enzyme and was precipitated using one-tenth volume of 3.0 M sodium acetate and double volume of chilled ethanol, and was subsequently purified prior to being digested with *Hind*III, as described above. Following digestion, the fragments produced were spiked with the synthesized oligonucleotides at a concentration to provide one copy of oligonucleotide per copy of genomic DNA equivalent (calculated by comparing the fragment size to the genome size and accounting for the amount of DNA loaded into the gel), and the mixture was separated on an agarose gel. The DNA fragments were denatured *in situ*, transferred to a nylon membrane and fixed to the membrane by UV crosslinking (UVP, LLC, Cambridge, UK). Two DIG-labeled DNA probes within the *Agrobacterium* backbone region were hybridized to the target DNA on the nylon membranes, where plasmid backbone probe 45 hybridizes to test oligonucleotide sequences and plasmid backbone probe 7 hybridizes to the control oligonucleotide sequence. Images were captured by detection with Syngene G-Box Chemi XT16 and XX6 (Syngene, Inc., Cambridge, UK).

2.5. Detection of plasmid backbone DNA

To determine if both SbS and Southern blot methods provide comparable results for detection of unintentionally incorporated plasmid backbone sequence, a population of GM maize lines was created in mid-2012 using standard *Agrobacterium*-mediated transformation [20] with plasmid PHP59391. Following SbS analysis [19], six GM maize lines (A, B, C, D, E and F) containing various fragments from the *Agrobacterium* transformation plasmid backbone DNA were selected for Southern Blot analysis. To analyze the GM maize events by Southern blot, genomic DNA samples extracted from both the GM and control plants were digested with *Eco*RV, as described above. An untransformed maize line with the same genetic background was used as a control and plasmid PHP59391 was used as a positive control to verify probe hybridization. The DNA fragments bound to the nylon membrane were hybridized using a series of 45 probes. These probes were designed to test for the presence of the entire plasmid backbone DNA sequence (i.e., outside of the T-DNA Right and Left Borders).

3. Results

3.1. Analysis of simple and complex insertions

SbS and Southern blot analysis were used to characterize the DNA insertions and create physical maps of the insertions in four soybean GM lines with DNA insertions of variable complexity. The analysis of the SI (single intact insertion) and DR (single insertion with deleted region) events are described below, while detailed information about the DI (tandem double insertion) and CI (complex insertion with multiple joined fragments) events is provided in Supplemental Materials (refer to sections 10.1.1. Supplement 1: Analysis of DI event, and 10.1.2. Supplement 1: Analysis of CI event).

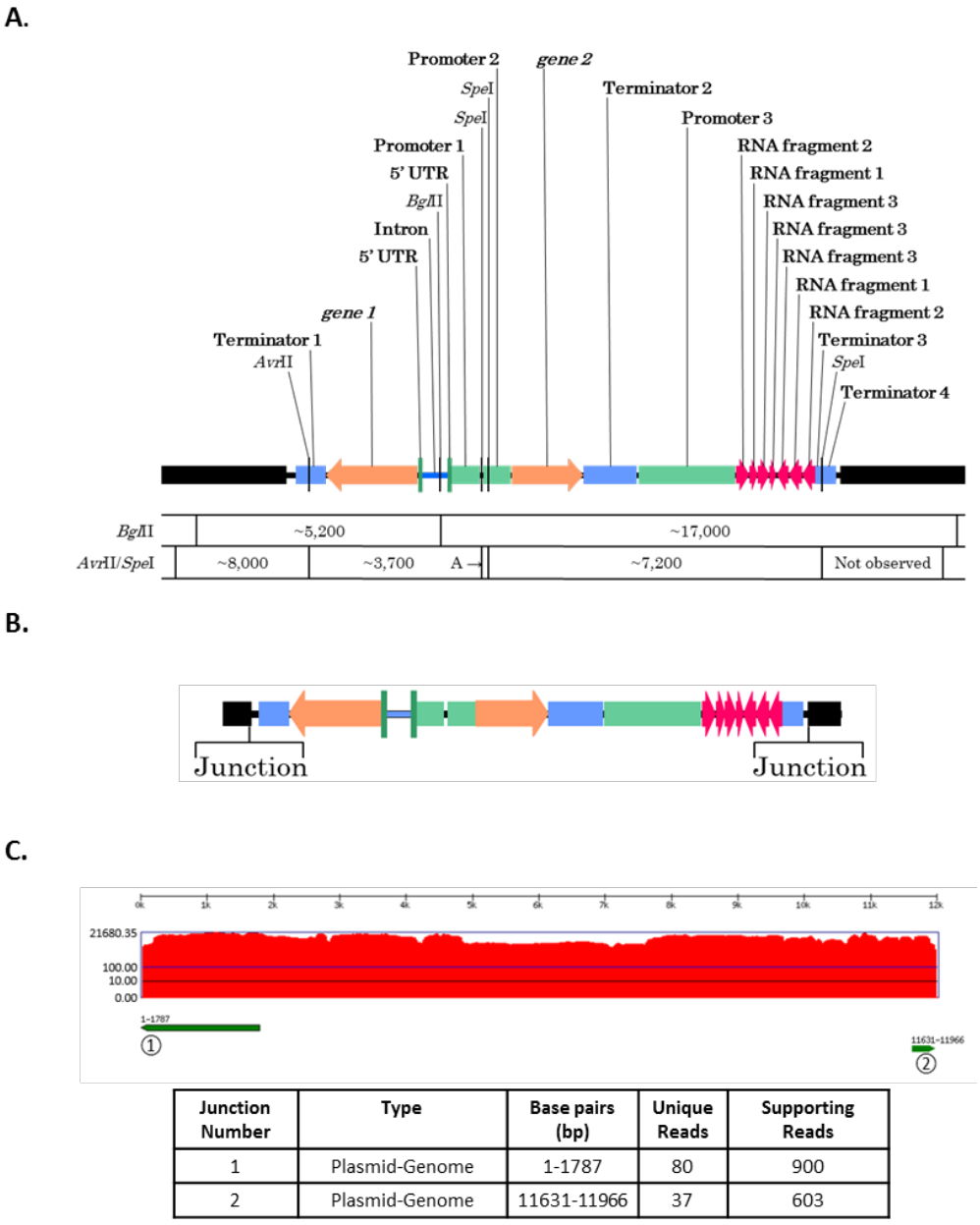
SbS analysis of the SI event resulted in high levels of coverage across nearly the entire transformation fragment and yielded a total of two plasmid-genome junctions (junction 1 and 2 in Figure 1C). The 5′ junction began at base pair (bp) 1 of the PHP63750A transformation fragment, while the 3′ junction occurred at bp 11,966 out of the 11,976 total bp in the fragment, indicating that 10 bp were truncated at the 3′ end during integration into the soy genome. The presence of only two plasmid-genome junctions demonstrates that there is a single insertion in the SI genome, while their locations near the ends of the transformation fragment, and the absence of any other junctions, indicate that the fragment was integrated into the genome intact except for the missing 10 bp at the 3′ end. A map of the inserted DNA in the SI event was created using this information (Figure 1B).

For Southern blot analysis, genomic DNA from the SI event, untransformed control soybean DNA, and plasmid PHP63750 were digested with restriction enzyme *Bgl*II or double digestion with *Avr*II/*Spe*I and hybridized to probes for all elements within the PHP63750A fragment (Supplement 1: Figure S1). Two types of genomic fragments are expected to be observed from these digests and hybridizations: 1) border fragments (indicated by "border" in Supplement 1: Table S2 and Table S3) where an enzyme site is located at one end of the fragment hybridizing to the probe and a second site is expected in the soy genome, and 2) internal fragments where known enzyme sites flank the probe region and the fragments are completely contained within PHP63750A. Border fragment sizes are unique for each insertion due to the varying location of the restriction sites within the surrounding soy genome. The number of bands produced from a given enzyme digestion is directly related to the number of inserted copies. One hybridizing band produced from an enzyme that cleaves once in the insert, outside of the probe region, indicates the presence of one copy of the inserted DNA at a single locus in the genome. Border fragments formed from the insertion of a full-length PHP63750A are expected to be larger than the size predicted from the PHP63750A sequence due to the inclusion of genomic DNA in the fragment. The exact size of border fragments cannot be predicted in advance due to the unknown location of the cleavage site in the soy genome. Internal fragments provide a means to assess the intactness of the inserted DNA and whether it has changed from the intended arrangement.

Restriction enzyme *Bgl*II was selected for copy number analysis of the SI soybean event, as there is a single site for *Bgl*II, located at bp 3,280 of PHP63750A (Supplement 1: Figure S1) and is predicted to yield border fragments of greater than 3,300 bp and greater than 8,700 bp for a single inserted copy of PHP63750A (Supplement 1: Table S2). For a single insertion, hybridization with all probes (Supplement 1: Table S1 and Figure S1) would result in a single insertion-derived band for each probe, except for the Intron probe, which will have two bands due to the location of the restriction site within the probe region. For the SI event, all probes located 5′ of the *Bgl*II site hybridized to a single band of approximately 5,200 bp, while the probes 3′ to the cut site resulted in a band of about 17,000 bp (Figure 1A, Supplement 1: Table S2 and Figures S3 - S8). For example, the results of hybridization with the Terminator 2 probe and the RNA fragments probe are shown in Figures 2A and 2B, respectively (*Bgl*II panel, lane 7, band ~17,000 bp). The Intron probe hybridized to both bands, as expected (Supplement 1: Table S2 and Figure S4). These fragments are consistent with the presence of a single PHP63750A insertion in the SI event.

A double digest with the restriction enzymes *Avr*II and *Spe*I was used to analyze the intactness of the PHP63750A insertion in the SI soybean event. *Avr*II has a single site in PHP63750A at bp position 426, while *Spe*I has three sites located at bp positions 4,196; 4,338; and 11,560 (Supplement 1: Figure S1). Digestion of an intact PHP63750A insertion with *Avr*II/*Spe*I is predicted to yield a 5′ border band of greater than 400 bp, three internal bands of 3,770 bp, 142 bp, and 7,222 bp, and a 3′ border band of greater than 400 bp (Supplement 1: Table S3 and Figure S1). The presence of the expected bands, and absence of any insert-derived bands other than the predicted bands, would provide a strong indication that the inserted PHP63750A is complete and was not truncated upon insertion. Hybridization of the *Avr*II/*Spe*I digest with all probes resulted in the expected internal bands matching the corresponding plasmid bands, and single border bands located at each end of the inserted DNA, outside the restriction sites (Figure 1A, Supplement 1: Table S3 and Figures S9 - S14). For example, the results of hybridization with the Terminator 2 probe and the RNA fragments probe are shown in Figures 2A and 2B, respectively (*Avr*II/*Spe*I panel, lane 7, band ~7,200 bp). The 142 bp band expected with the Promoter 1 probe was not observed, as its small size meant it was run off the gel during electrophoresis. The presence of all expected internal bands shows that the restriction enzyme sites within the PHP63750A transformation fragment are all intact, and therefore the inserted DNA is also intact, with the caveat that the restriction enzymes are not located exactly at the ends of the inserted DNA (as shown in Figure 1A).

SbS analysis of the DR event resulted in high levels of coverage across much of the transformation fragment, with the exception of a 523-bp region located in the gene 2/terminator 2 region that showed very low coverage, and yielded two plasmid-genome junctions (junctions 1 and 4 in Figure 3C) and two plasmid-plasmid junctions (junctions 2 and 3 in Figure 3C). The 5′ genomic junction began at bp 11 of the PHP63750A transformation fragment, while the 3′ genomic junction occurred at



Junction Number	Type	Base pairs (bp)	Unique Reads	Supporting Reads
1	Plasmid-Genome	1-1787	80	900
2	Plasmid-Genome	11631-11966	37	603

Figure 1: (A) Schematic map of the SI soybean line based on Southern blot analysis, showing the *Bgl*II, *Avr*II, and *Spe*I restriction sites and the sizes of observed fragments approximated to the nearest 100 bp (not to scale). The letter A designates the 142 bp *Spe*I fragment that ran off the bottom of the gel during electrophoresis. Not observed signifies that the fragment was not detected with the digests and hybridizations described. The flanking genome is represented by the horizontal black bar. (B) Schematic map of the SI soybean line based on SbS analysis, where the genetic elements correspond to panel A, and the plasmid-genome junctions are indicated. (C) SbS coverage graph mapped against the transformation fragment depicted as a "ruler" across the top, showing number of reads using a logarithmic scale. Junction locations are indicated by numbered arrows below the coverage graph with details below the coverage graph. "Unique reads" are the compiled reads that contain both plasmid and genomic sequences and thus define the junction; Supporting reads are the total number of reads across the junction (sequencing depth). Multiple identical Supporting reads are included in each Unique Read.

bp 11,976 out of the 11,976 total bp in the fragment, indicating that 10 bp were truncated at the 5′ end during integration into the soy genome. The presence of only two plasmid-genome junctions demonstrates that there is a single insertion in the DR genome. Detection of two plasmid-plasmid junctions demonstrates that there was a 523-bp deletion between bp 6,371 and bp 6,895 of the PHP63750A transformation fragment. The very low coverage across the deleted region of the construct (~10 reads) is likely due to background amplification of environmental bacterial sequences [19]. A map of the inserted DNA in the SI event was created using this information (Figure 3B).

Southern blot analysis with *Bgl*II demonstrated that there

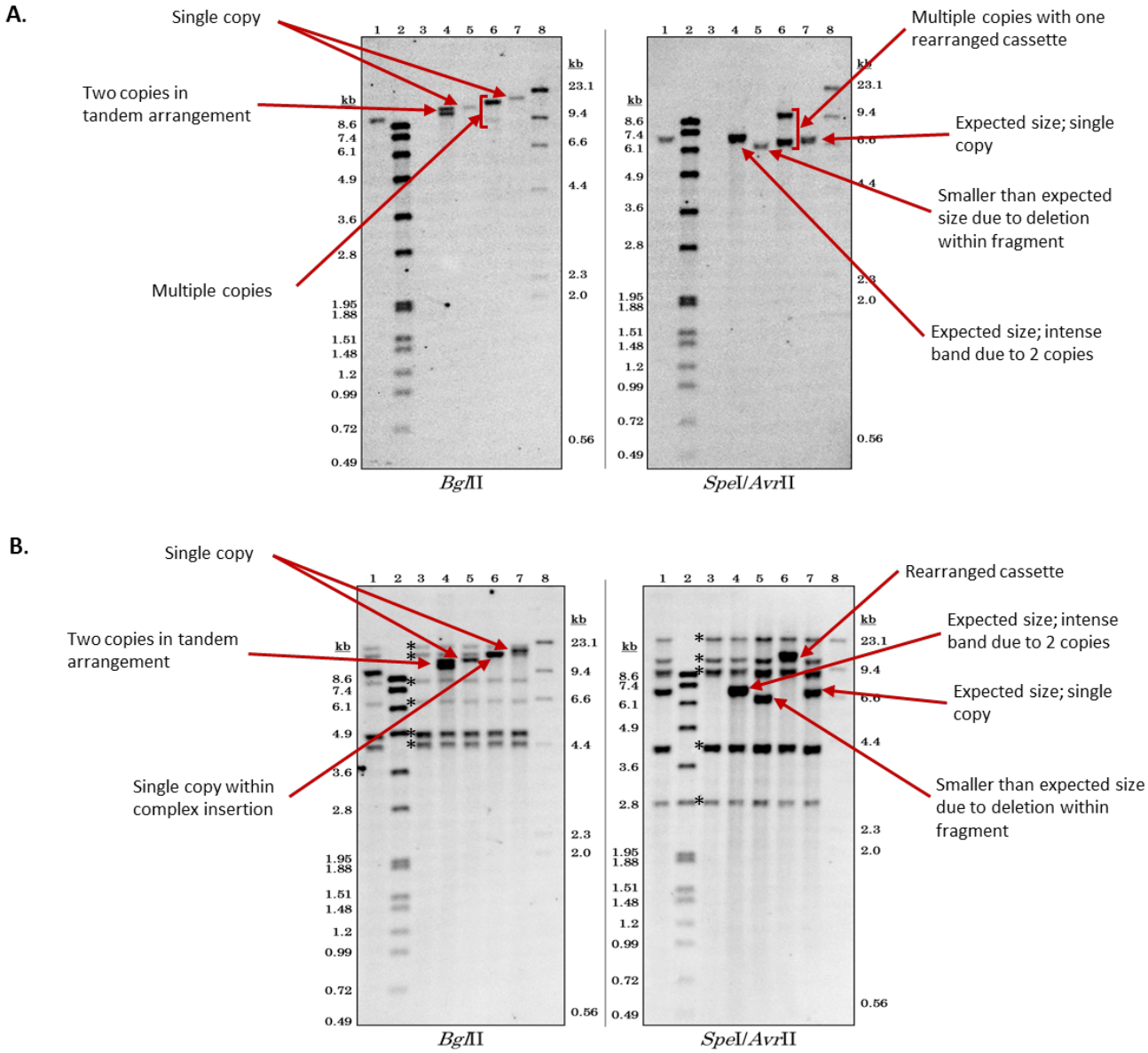
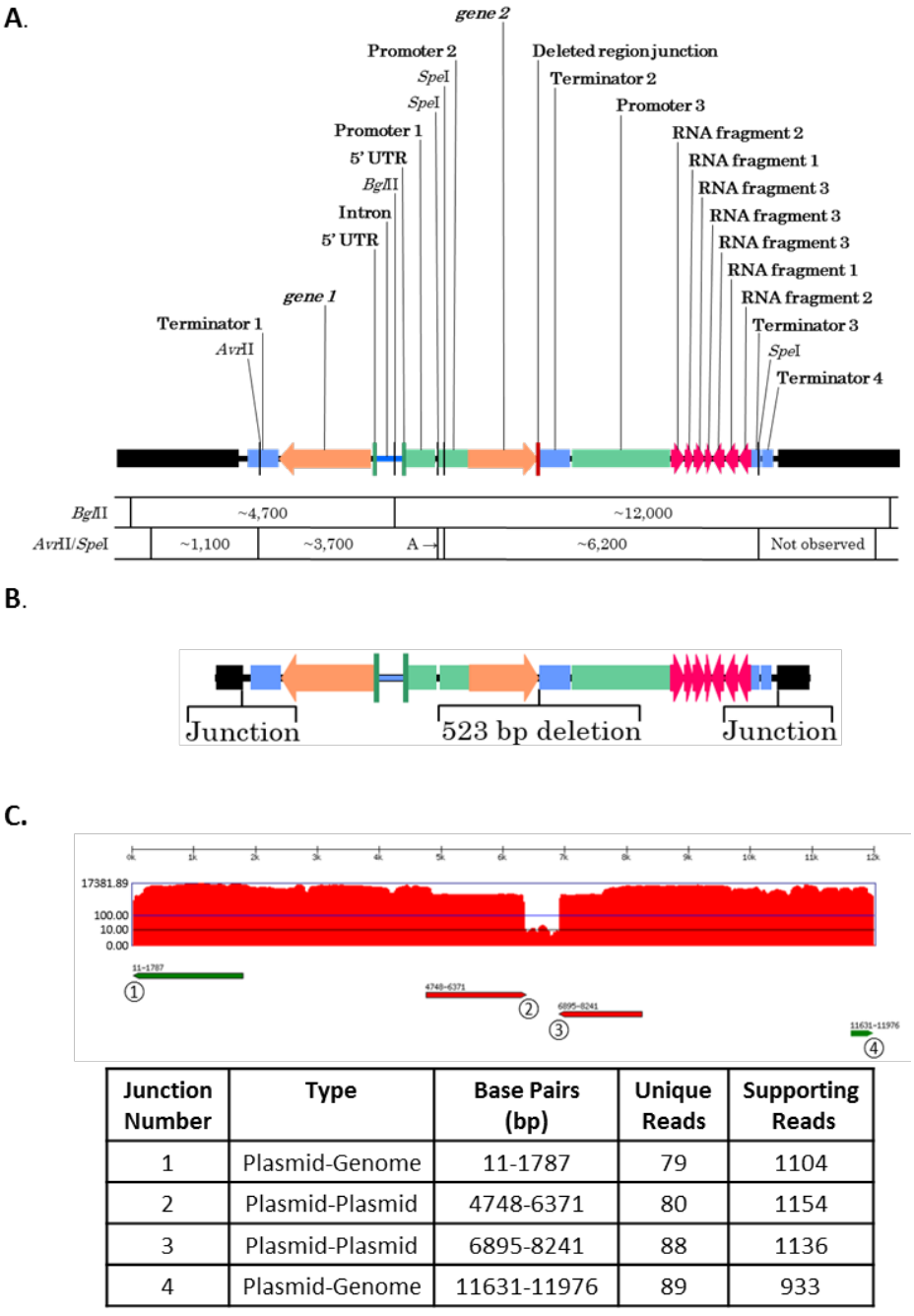


Figure 2: Southern blot analysis of the four soybean GM lines containing simple to complex insertions. Molecular weight markers are in Lane 2 and Lane 8. The two control DNAs (control soybean DNA spiked with plasmid and control soybean DNA) are in Lane 1 and Lane 3, respectively. Lanes 4-7 are the four transformed soybean GM lines (Lane 4 - DI; Lane 5 - DR; Lane 6 - CI; and Lane 7 - SI). DNA was digested with *Bgl*II (left panel) for copy number or *Avr*II/*Spe*I (right panel) to determine insertion intactness. Blots were probed with Terminator 2 probe (2A) or RNA fragments probe (2B). Soybean endogenous genomic bands are depicted with an asterisk (*) in Lane 3 of each blot for the RNA Fragments probe (2B) and otherwise visible, but unmarked in lanes 4-7.

is a single PHP63750A insertion in the DR event. All probes located 5′ of the *Bgl*II site hybridized to a single band of approximately 4,700 bp, while the probes 3′ to the cut site resulted in a band of about 12,000 bp (Figure 3A, Supplement 1: Table S2 and Figures S3 - S8). For example, the results of hybridization with the Terminator 2 probe and the RNA fragments probe are shown in Figures 2A and 2B, respectively (*Bgl*II panel, lane 5, band ~12,000 bp). The Intron probe hybridized to both bands, as expected (Supplement 1: Table S2 and Figure S4). These fragments are consistent with the presence of a single PHP63750A insertion in the DR event. Hybridization of the *Avr*II/*Spe*I digest resulted in the expected internal band of ~3,700 bp with the probes located between the *Avr*II site and the first *Spe*I site, indicating that this region of the DR event is intact (Supplement 1: Table S3). The probes located between the second and third *Spe*I sites all yielded a band of ~6,200 bp, which is smaller than the predicted size of 7,222 bp (Figure 3A, Supplement 1: Table S3 and Figures S9 - S14). Since restriction sites are present as expected at the ends of this fragment, the smaller than expected size of the band indicates that a region was deleted within the fragment. The results of hybridization with the Terminator 2 probe and the RNA fragments probe are



Junction Number	Type	Base Pairs (bp)	Unique Reads	Supporting Reads
1	Plasmid-Genome	11-1787	79	1104
2	Plasmid-Plasmid	4748-6371	80	1154
3	Plasmid-Plasmid	6895-8241	88	1136
4	Plasmid-Genome	11631-11976	89	933

Figure 3: Schematic map of the DR soybean line based on Southern blot analysis (A) and SbS (B); and the flanking genome is represented by the horizontal black bar. (A) For Southern blot analysis, the *Bgl*II, *Avr*II, and *Spe*I restriction sites are indicated with the sizes of observed fragments approximated to the nearest 100 bp (not to scale). The letter A designates the 142 bp *Spe*I fragment that ran off the bottom of the gel during electrophoresis. Not observed signifies that the fragment was not detected with the digests and hybridizations described. (B) Schematic of the inserted DNA based on SbS analysis. The genetic elements correspond to panel A, and the plasmid-genome junctions and deleted region are indicated. (C) SbS coverage graph mapped against the transformation fragment, as in Figure 1. Junctions 1 and 4 indicate the plasmid-genome junctions. Junctions 2 and 3 indicate the ends of the retained plasmid sequences that are joined together due to the deletion of the 523 bp region.

shown in Figures 2A and 2B, respectively (*Avr*II/*Spe*I panel, lane 5, band ~6,200 bp). The presence of the internal bands shows that the restriction enzyme sites within the PHP63750A transformation fragment are all intact, and since the band derived from the 5′ end of the insertion matches the predicted size, this end is intact. However, since the band derived from the 3′ end of the insertion is smaller than predicted, it demonstrates that a deletion within this region occurred during transformation.

Comparison of the SbS results and the Southern blot results

shows that both types of analysis yielded the same conclusion regarding the copy number and intactness of the inserted DNA in the SI event (Figures 1A and 1B) and the DR event, (Figures 3A and 3B). Southern blot analysis and SbS also yielded the same conclusions regarding the copy number and intactness of the inserted DNA in the DI line (Supplement 1: Analysis of DI event) and the CI line (Supplement 1: Analysis of CI event).

3.2. Detection of small DNA fragments

Results of SbS analysis of oligonucleotide sequences for small fragment detection were previously described [19]. To directly compare Southern blot results to the SbS results, genomic DNA from untransformed maize plants was digested with *Hind*III and spiked with oligonucleotides containing variable length of the *Agrobacterium* plasmid backbone (35 to 100 bp), corresponding to those oligonucleotides used for SbS analysis. Following electrophoresis and transfer, the Southern blots were hybridized to plasmid backbone probe 45 (Figure 5). A band of approximately 250 bp was observed in lanes corresponding to 40, 45, 50, 55, 60, 65, 70, 75, 80 and 100 bp backbone fragments (Figure 4). A very faint band was detected with the oligonucleotide containing the 35 bp backbone DNA fragment, and no bands were seen in either the control oligonucleotide or unspiked maize DNA (Figure 4). Hybridization with plasmid backbone probe 7 showed a band of approximately 250 bp for the positive control oligonucleotide fragment (250 bp from backbone region) (Supplement 2: Figure S1). Detection of 250 bp band in control oligonucleotide sequence served as positive control for this Southern blot analysis. As expected, no bands were observed in test oligonucleotide fragment or unspiked maize DNA lanes (Supplement 2: Figure S1). Results from this study demonstrate that Southern blot analysis can detect fragments as small as 40 bp. SbS analysis could detect 35-100 bp insertions in the oligonucleotides [19], which indicates that sensitivity of both methods to detect small inserted DNA fragments is comparable.

3.3. Detection of plasmid backbone DNA

Six GM maize lines (denoted A-F) were generated by *Agrobacterium*-mediated transformation with plasmid PHP59391 (Figure 5) and analyzed by SbS to detect presence of PHP59391 plasmid backbone DNA fragments. For maize lines A, C, D, E, and F, SbS detected insertion of the backbone DNA fragments of 4,300 bp, 821 bp, 1,883 bp, 214 bp, and 2,315 bp, respectively (Figure 5 and Table 1). For maize line B, the entire backbone region was detected (Figure 5 and Table 1).

Southern blot analysis with *Eco*RV restriction enzyme was used to confirm the plasmid backbone fragments detected by SbS. The locations of *Eco*RV restriction sites within the PHP59391 plasmid backbone are shown in Figure 5. Digestion of PHP59391 with *Eco*RV is predicted to yield backbone fragments of 10,743 bp, 13,680 bp, 8,055 bp, 7,767 bp and 10,646 bp (Figure 5 and Supplement 3: Table S1). Genomic DNA samples from the six GM maize lines, untransformed control maize, and PHP59391 plasmid DNA were digested with *Eco*RV and hybridized with 45 probes that cover the entire plasmid backbone region. Locations of the probes are shown on the plasmid backbone map (Figure 5). Four of the backbone probes (11, 25, 35 and 43) spanned one of the *Eco*RV restriction sites within the backbone and therefore should hybridize to two plasmid fragments (Figure 5). Another four of the backbone probes have sequences that are repeated in the plasmid (4, 5, 6 and 7) and therefore should also hybridize to two plasmid fragments (Figure 5).

The results of the hybridization of selected backbone probes to the *Eco*RV digested genomic DNA from the six GM maize lines are shown in Figure 6, and the remainder are shown in Supplement 3: Figures S1 - S45 and are summarized in Supplement 3: Table S1. A single plasmid backbone-derived band was detected at approximately 19,000 bp for maize line A with probes 37-42 (Supplement 3: Table S1 and Figures S36 - S41), 14,000 bp for maize line C with probes 26 and 27 (Supplement 3: Table S1 and Figures S25 - S26), and approximately 6,800 bp for maize line D with probes 21-23 (Supplement 3: Table S1 and Figures S20 - S22). For maize line C, the hybridization with backbone probe 27 is faint due to an overlap of only 137 bp and can only be seen on the X-ray film. Similarly, for maize line D, backbone probe 23 hybridization is faint due to an overlap of only 66 bp.

The SbS results indicate that maize lines E and F would hybridize to backbone probes 13-14 and 12-15, respectively (Figure 5). With Southern blot analysis, a single band was detected at approximately 4,000 bp for maize line E with probes 13-14 (Supplement 3: Table S1 and Figures S12 - S13); however, the hybridization with backbone probe 13 is faint due to an overlap of only 113 bp. For maize line F, a single band was detected at approximately 10,000 bp with probes 12-15 (Supplement 3: Table S1 and Figures S11 - S14); however, backbone probe 15 hybridization is fainter due to an overlap of only 149 bp.

For maize line B, containing the entire plasmid backbone, all of the backbone probes are expected to hybridize to the corresponding *Eco*RV generated fragment (Figure 5). A single band was detected at approximately 7,000 bp with backbone probes 1 - 10 (Supplement 3: Table S1 and Figures S1 - S9). Hybridization with backbone probes 4 - 7 each also had an additional band of about 11,000 bp due to the repetitive sequences contained within the plasmid (Supplement 3: Table S1 and Figures S4 - S7). Backbone probe 11, due to the sequence spanning an *Eco*RV restriction site, produced two bands of approximately 7,000 bp and 14,000 bp (Supplement 3: Table S1 and Figure S10); however, the approximately 7,000 bp band is faint due to having only an 88 bp overlap to the fragment. A single band of about 14,000 bp was detected with backbone probes 11 - 24 (Supplement 3: Table S1 and Figures S10 - S23). Backbone probe 25 also spans an *Eco*RV restriction site and produced two bands of about 14,000 bp and 8,000 bp, the second of which is faint due to having only a 117 bp overlap to the fragment (Supplement 3: Table S1 and Figure S24). A band of about 8,000 bp was detected with backbone probes 26 - 34 (Supplement 3: Table S1 and Figures S25 - S33). Spanning another *Eco*RV site, backbone probe 35 shows only a single band of about 8,000 bp due to the overlapping fragment sizes of 8,055 bp and 7,767 bp (Supplement 3: Table S1 and Figure S34). Hybridization with backbone probes 36 - 42 resulted in a single band of approx-

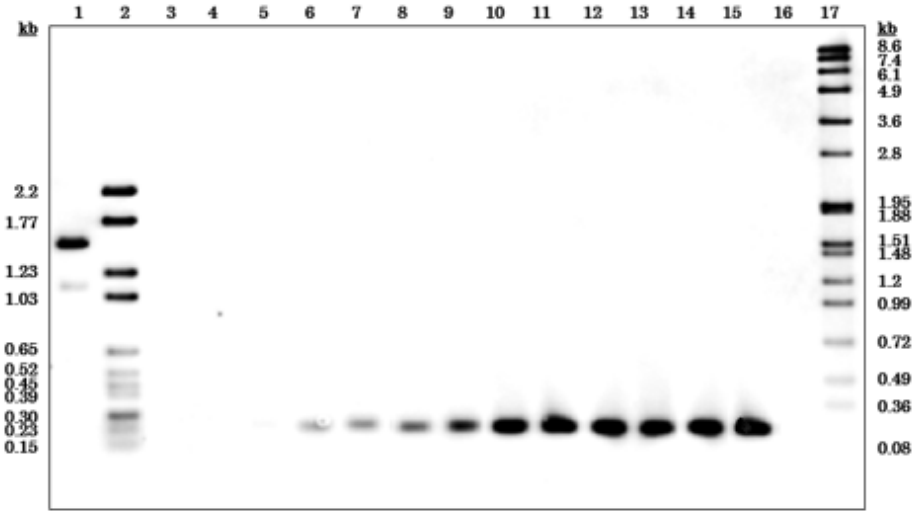


Figure 4: Southern blot sensitivity analysis using synthetic oligonucleotides. Molecular weight markers are in Lane 2 and Lane 17. The positive controls (maize DNA spiked with *Agrobacterium* plasmid, and maize DNA with 250 bp positive control oligonucleotide) are in Lanes 1 and 3, respectively. Negative control DNA from untransformed maize is in Lanes 4 and 16. Lanes 5-14 contain increasing lengths of plasmid backbone fragments (starting at 35 bp in Lane 5, and increasing by 5 bp for a final amount of 80 bp in Lane 14) and maize genomic sequence to a total of 250 bp. Lane 15 contains 100 bp of plasmid backbone and maize genomic sequence to a total of 250 bp.

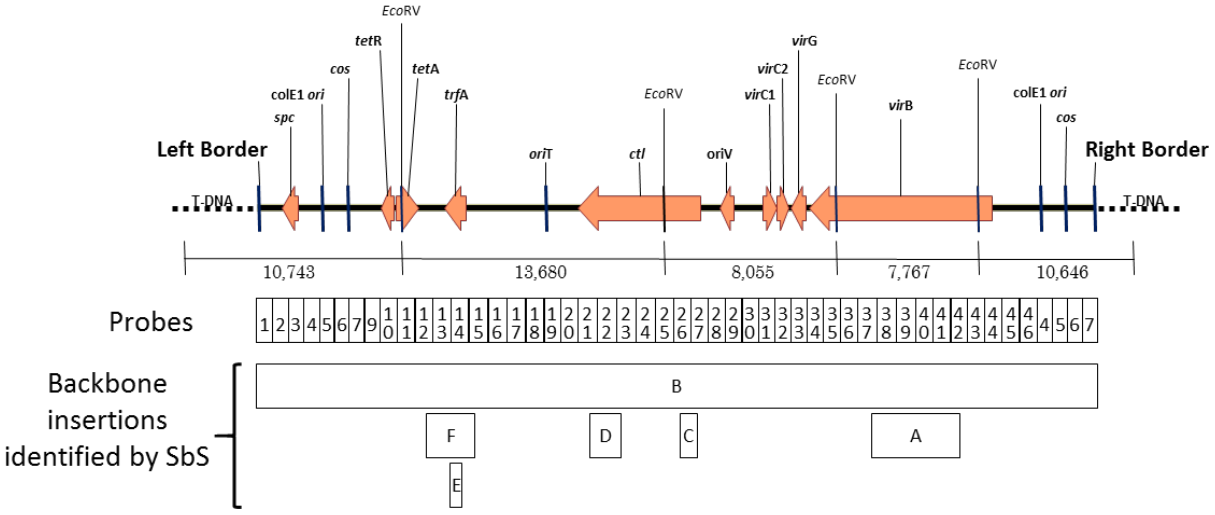


Figure 5: *Agrobacterium* plasmid backbone map and probes. The upper portion of the figure shows a schematic of the *Agrobacterium* plasmid backbone region (not including T-DNA) with genetic elements and *Eco*RV restriction enzyme sites. Below it is a representation of the fragments obtained by digestion of the plasmid with *Eco*RV, and the 45 Southern probes for backbone sequences (probes 1-46; note there is no probe 8; some probes are in repeated regions and thus appear twice on the map). The lower part of the figure shows the plasmid backbone regions identified in each line. Maize line A contains a 4,300 bp backbone DNA fragment and is aligned with probes 37-42, line B contains the entire backbone DNA (all probes), line C consists of a 821 bp fragment (probes 26-27), line D consists of a 1,883 bp fragment (probes 21-23), line E consists of a 214 bp fragment (probes 13-14), line F consists of a 2,315 bp fragment (probes 12-15). All expected DNA backbone fragments were detected with the corresponding probes by Southern blot analysis (Figure 6 and Supplement 3: Figure S1 - S45).

imately 8,000 bp (Supplement 3: Table S1 and Figures S35 - S41). Backbone probe 43, containing the final *Eco*RV restriction site, produced two bands of about 8,000 bp and 11,000 bp (Supplement 3: Table S1 and Figure S42). The final three backbone probes 44, 45, and 46 resulted in a single band of about 11,000 bp (Supplement 3: Table S1 and Figures S43 - S45).

For all six maize lines, only the Southern Blot probes corresponding to the backbone regions detected by SbS hybridized to the corresponding genomic DNA samples (Figure 5 and Supplement 3: Table S1), indicating that the probes were effective in detecting the backbone sequences when they are present. The backbone fragment sequences detected by Southern blot analysis agree with the results obtained by SbS.

Maize Lines	Fragment Size Detected by SbS (bp)	Backbone Probes
A	4,300	T37 - 42 (145 bp overlap to probe 37)
B	Entire backbone	All
C	821	26 - 27 (137 bp overlap to probe 27)
D	1,883	21 - 23 (66 bp overlap to probe 23)
E	214	13 - 14 (113 bp overlap to probe 13)
F	2,315	12 - 15 (149 bp overlap to probe 15)

Table 1: Fragment size detected by Southern-by-Sequencing (SbS) in six maize lines containing PHP59391 plasmid backbone fragments.

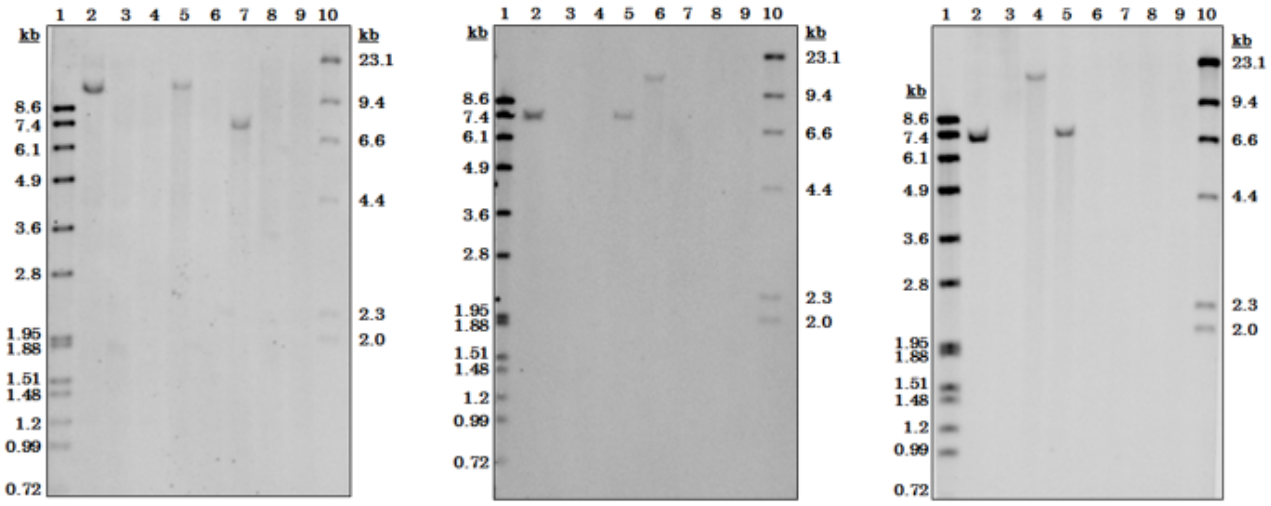


Figure 6: Southern blot analysis of maize containing *Agrobacterium* plasmid backbone DNA fragments. Results of three probes are depicted (probes 22, 26, and 28, from left to right). The molecular weight markers are in Lane 1 and Lane 10. Control DNAs (untransformed maize DNA with plasmid and untransformed maize DNA) are in Lanes 2 and 3, respectively. Lanes 4-9 contain maize lines A-F (described in Figure 5), respectively. See Supplement 3: Figure S1 - S45 for additional Southern blot analysis results.

4. Discussion

Detailed molecular characterization of GM crops informs the safety assessment and is required for global regulatory approvals prior to commercialization [9, 5, 6]. Traditionally, Southern blot analysis and Sanger sequencing are used to characterize the structure of the inserted DNA and the nucleotide sequence of the DNA insertion. These techniques can accurately characterize insertion structure and nucleotide sequence; however, they are often accompanied by several technical challenges. Not only is Southern blot analysis laborious and time-consuming, data analysis of Southern blot images is subjective, requiring multiple restriction enzyme digests and hybridizations and interpretation of band size. In the cases of rearranged or truncated insertions, additional restriction enzyme digestions are likely to be required to clarify the presence of unexpected hybridization bands and create a map of a complex insertion, further complicating and prolonging the analysis. An additional technical challenge arises when hybridization probes closely or exactly match endogenous genomic sequences and result in bands that are not derived from the inserted DNA. These endogenous bands can be numerous, potentially obscure insertion bands, and must be identified with absolute certainty by comparison to the untransformed control DNA samples before they can be excluded from the set of insertion-related bands [1, 15].

NGS is an alternative approach for the molecular characterization of DNA insertion structure [3, 8, 12, 14, 19]. NGS relies on either whole genome sequencing or hybridization enrichment and sequencing of the targeted fragments, followed by bioinformatics pipelines that identify junctions between the endogenous genomic DNA and any inserted DNA. NGS also detects unexpected sequence junctions within the inserted DNA that indicate rearrangements or other changes from the expected insertion structure. Analysis of the junction sequences and comparison to the intended transformation sequence allow creation of an insertion map similar to that resulting from Southern blot analysis. For example, the detection of only two genomic-inserted DNA junctions at the ends of the DNA intended to be inserted would indicate the presence of a single intact inser-

tion that matches the expected insertion. Additional junctions, either genomic DNA-inserted DNA or inserted DNA-inserted DNA, would indicate the presence of more than a single insertion as well as rearrangements or truncations of the intended DNA insertion. Also, since NGS typically aligns the sequence reads to the entire transformation plasmid, both the intended inserted DNA and unintentionally inserted backbone DNA can be detected.

SbS utilizes targeted capture of plasmid-related sequences coupled with NGS and bioinformatics data analysis to identify sequence junctions and characterize the same molecular endpoints as the traditional methods. An advantage of SbS is the targeted capture step prior to NGS, which allows for substantially increased sequence coverage at any junction involving DNA derived from the transformation plasmid [8, 19]. This results in a high degree of confidence that all plasmid DNA junctions are accurately identified and allows for the development of a detailed map of the insertion structure.

To demonstrate that Southern blots and SbS reach the same conclusions about copy number and DNA intactness, four soybean GM lines containing simple (SI, single insertion) and complex (DI, tandem double insertion; DR, single insertion with a deleted region; and CI, multiple joined fragment) insertions were analyzed. To demonstrate that both techniques reach the same conclusions about the presence of unintended plasmid backbone DNA, six GM maize lines containing *Agrobacterium* plasmid backbone fragments of varying sizes were characterized. Additionally, the sensitivity of Southern blot analysis and SbS was compared by their ability to detect small (35-100 bp) sequence fragments. Results from these studies demonstrate that both Southern blot analysis and SbS reach the same conclusions for these molecular characterization endpoints.

The insertion maps of the four soybean GM lines created using SbS junction analysis matched the maps generated by Southern blot analysis, which demonstrates that the two techniques are equivalent for determination of copy number and arrangement of the physical structure of simple to complex insertions. As highlighted by this study, the biggest differences between the Southern blots and SbS approach are the complexity of the Southern blot experimental phase (amount of time, labor, and number of blots needed for complete analysis) and subjective interpretation of the Southern blot output (interpretation of the observed bands which may include unexpected size bands, faint bands, or endogenous genomic DNA bands). For example, to analyze the soybean GM lines by Southern blot, two different restriction digests were required for the simple insertions, and six restriction digests were required for the most complex multiple joined fragment insertion. Additionally, Southern blot analysis required a total of twelve hybridizations of each digest with probes that covered the entire intended transformation fragment to generate restriction maps of the insertions based on interpretation of bands. Conversely, the SbS results were obtained in a single targeted capture and sequencing run, which was much simpler in terms of time and resources, and data interpretation was much less subjective. Adding to the complexity for Southern blot analysis of the soybean GM lines was the presence of endogenous soybean bands. For example, most of the genetic elements in the transformation fragment were derived from endogenous soy genomic DNA sequences, and thus probes derived from these elements and used in this study were homologous to soy endogenous genomic sequences. Therefore, Southern blot analysis of the SI, DI, DR and CI soybean lines with these probes exhibited a number of bands that were due to hybridization to the endogenous sequences. Using SbS, endogenous elements were accounted for by utilizing an untransformed soybean DNA that was captured with the same probe library and analyzed against the transformation fragment. Sequence junctions identified through this process could then be removed from all downstream analyses.

Both Southern blot analysis and SbS analysis of the six GM maize lines for the presence of unintended plasmid DNA also generated comparable results. In all cases, the Southern probes that hybridized to the maize DNA were those expected based on the SbS results, while no hybridization was observed for probes that did not correspond to the plasmid backbone DNA regions identified by SbS for each maize plant. The GM maize line containing the full plasmid backbone hybridized to all 45 probes and showed two bands for each of the probes to the duplicated sequences. The exact correlation of the Southern blot results for backbone probes with the backbone fragments identified by SbS in these six GM maize lines shows that both techniques are effective at detecting the presence of unintended plasmid DNA sequences in transformed plants. However, SbS analysis of these GM maize events was more efficient and less subject to interpretation than Southern blot. Analysis by Southern blot required the use of 45 different hybridization probes to cover the entire 43 kb of the *Agrobacterium* plasmid backbone DNA, including four probes to sequences that are duplicated in the backbone, whereas SbS tested for the presence of the same backbone sequences in a single experiment. As in the Southern blot analysis of the soybean GM lines, probes hybridizing to the endogenous genetic elements added complexity to the interpretation of results in the experiment with GM maize lines. Although the probes were designed to target plasmid DNA sequences, there was enough similarity between some of the probes and maize DNA to allow endogenous hybridization bands to be observed on some of the Southern blots, complicating data analysis and making it more difficult to verify that no backbone DNA sequences were incorporated into the plant genome.

A further advantage of SbS for this type of analysis is that the capture probe library can contain as many plasmid DNA backbone sequences as desired, so that a single capture experiment can verify the absence of sequences from multiple plasmids. This is desirable for transformations that utilize more than a single plasmid; for example, co-transformation experiments to generate GM plants or in gene editing experiments in which several plasmids are used to deliver different components of the gene editing system. Southern blot analysis would necessitate specific digest, probe, and hybridization designs for each different plasmid, with the accompanying increase in time and resources needed for such customization, while SbS is high-throughput and allows for analysis of multiple constructs in a single probe library.

While SbS has been shown to be capable of detecting small

fragments [19], a direct comparison of the sensitivity of SbS and Southern blot methods have not been made. Previously, the ability of SbS to detect small fragments was evaluated using 250 bp oligonucleotides containing increasing amounts of plasmid backbone (ranging in size of 35-100 bp) [19]. SbS could consistently detect plasmid fragments of 50 bp or larger, with variable ability to detect fragments as small as 35 bp [19]. In this study, Southern blot analysis clearly detected the 40 bp fragment and showed faint detection of the 35 bp fragment. Given that the sensitivity of Southern blot detection can vary from experiment to experiment depending on efficiency of DNA separation and transfer to the membrane, hybridization conditions, and probe sequence, it can be concluded that the two techniques have similar sensitivity for small fragments. In certain cases, SbS may have an advantage, as Southern blotting will not detect identical duplicated sequences separately if they are on the same restriction fragment, while SbS, being sequence-based, can identify junctions for multiple fragments and report the presence of each one in the insertion.

5. Conclusion

SbS is an efficient alternative tool to traditional Southern blot analysis for the regulatory molecular characterization of transformation-derived crops. Both techniques show comparable results for determining the necessary endpoints of DNA insertion structure, as they give the same results for both simple and complex DNA insertions and are equivalent for detecting the presence of unintended plasmid backbone DNA sequences in plant genomes. The two methods also demonstrate similar sensitivity for the detection of small inserted DNA fragments. SbS presents advantages over Southern blotting in simplicity and consistency of overall experimental design across different transformation processes, amount of labor and time necessary to complete the analysis, and reduction in potential experimental variation due to fewer manual steps. Since SbS output is sequence-based rather than image-based, as in Southern blotting, data analysis, interpretation, and reporting are simplified. The combination of SbS and Sanger-based sequencing of inserted DNA and adjacent genomic regions can provide all the information needed for molecular characterization of insertion structure and sequence for any new GM event. Furthermore, SbS is effective for any transformation method and can be used for any crop variety for which a reference genome sequence is available. SbS application is not limited to characterization of GM crops, as it can also be used for gene edited varieties where confirmation of absence of unintended plasmid DNA is one of the key objectives of molecular characterization. SbS offers an efficient and reliable tool for molecular characterization of crop varieties created using genetic transformation techniques.

6. Disclaimer

All authors were employed at Corteva Agriscience™ Agriculture Division of DowDuPont at the time of preparation of this manuscript.

7. Acknowledgement

We thank Mary Locke, Tracey Fisher, Ajith Anand, Masha Fedorova, Shveta Bagga, and many others for critical scientific review of the manuscript.

8. Article Information

This article was received November 16, 2018, in revised form March 19, 2019, and made available online July 1, 2019.

9. References

[1] Altpeter, F., Vasil, V., Srivastava, V., & Vasil, I. K. (1996). Integration and expression of the high-molecular-weight glutenin subunit 1Ax1 gene into wheat. *Nature Biotechnology, 14(9)*, 1155.

[2] Brink, K., Chui, C.-F., Cressman, R. F., Garcia, P., Henderson, N., Hong, B., . . . Stecca, K. L. (2014). Molecular characterization, compositional analysis, and germination evaluation of a high-oleic soybean generated by the suppression of FAD2-1 expression. *Crop Science, 54(5)*, 2160-2174.

[3] Cade, R., Burgin, K., Schilling, K., Lee, T.-J., Ngam, P., Devitt, N., & Fajardo, D. (2018). Evaluation of whole genome sequencing and an insertion site characterization method for molecular characterization of GM maize. *Journal of Regulatory Science, 6(1)*, 1-14.

[4] Chen, J., & Dellaporta, S. (1994). Urea-based Plant DNA Miniprep. In M. Freeling, V. Walbot (Eds.), *The Maize Handbook* (pp. 526-527). New York: Springer.

[5] EFSA. (2006). Guidance document of the Scientific Panel on Genetically Modified Organisms for the risk assessment of genetically modified plants and derived food and feed. *EFSA Journal, 4*(4), 1-100.

[6] FAO/WHO. (1991). Strategies for assessing the safety of foods produced by biotechnology: report of a joint FAO/WHO consultation. Geneva, Switzerland: World Health Organization.

[7] Finer, J. J., & McMullen, M. D. (1991). Transformation of soybean via particle bombardment of embryogenic suspension culture tissue. *In Vitro Cell & Developmental Biology - Plant, 27*(4), 175-182.

[8] Guttikonda, S. K., Marri, P., Mammadov, J., Ye, L., Soe, K., Richey, K., . . . Kumpatla, S. P. (2016). Molecular Characterization of Transgenic Events Using Next Generation Sequencing Approach. *PLOS ONE, 11*(2), 1-17.

[9] Joint FAO/WHO Codex Alimentarius Commission, Food and Agriculture Organization of the United Nations. (2009). *Foods Derived from Modern Biotechnology* (2nd ed., pp. 85). Rome, Italy: World Health Organization.

[10] Kent, W. J. (2002). BLAT – the BLAST-like alignment tool. *Genome Research, 12*(4), 656-664.

[11] Klein, T. M., Wolf, E. D., Wu, R. & Sanford, J. C. (1987). High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature 327*, 70-73.

[12] Kovalic, D., Garnaat, C., Guo, L., Yan, Y., Groat, J., Silvanovich, A., . . . Christian, A. (2012). The Use of Next Generation Sequencing and Junction Sequence Analysis Bioinformatics to Achieve Molecular Characterization of Crops Improved Through Modern Biotechnology. *The Plant Genome, 5*(3), 149-163.

[13] Samoylov, V. M., Tucker, D. M., Thibaud-Nissen, F., & Parrott, W. A. (1988). A liquid-medium-based protocol for rapid regeneration from embryogenic soybean cultures. *Plant Cell Reports, 18*, 49-54.

[14] Schouten, H. J., Schijlen, E., Schaart, J., van de Geest, H., Papadimitriou, S., Smulders, M. J. M., . . . Sanchez Perez, G. (2016). GM plants compared to baseline; a whole genome sequencing approach. Retrieved from https://www.cogem.net/index.cfm/en/publications/publication/gm-plants-compared-to-the-baseline-a-whole-genome-sequencing-approach?order=relevance&q=van+Tienderen&category=&from=30-09-1998&to=06-04-2019&sc=fullcontent

[15] Smith, C. J., Watson, C. F., Morris, P. C., Bird, C. R., Seymour, G. B., Gray, J. E., . . . Grierson, D. (1990). Inheritance and effect on ripening of antisense polygalacturonase genes in transgenic tomatoes. *Plant Molecular Biology, 14*(3), 369-379.

[16] Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology, 98*(3), 503-517.
[17] Stewart, C. N., Jr., Adang, M. J., All, J. N., Boerma, H. R., Cardineau, G., Tucker, D., & Parrott, W. A. (1996). Genetic transformation, recovery, and characterization of fertile soybean transgenic for a synthetic Bacillus thuringiensis cryIAc gene. *Plant Physiology 112*, 121-129.
[18] USDA APHIS. (n.d.). Petitions for Determination of Nonregulated Status. Retrieved October 2018 from https://www.aphis.usda.gov/aphis/ourfocus/biotechnology/permits-notifications-petitions/petitions/petition-status
[19] Zastrow-Hayes, G. M., Lin, H., Sigmund, A. L., Hoffman, J. L., Alarcon, C. M., Hayes, K. R., . . . Beatty, M. K. (2015). Southern-by-sequencing: A Robust Screening Approach for Molecular Characterization of Genetically Modified Crops. *The Plant Genome, 8*(1), 1-15.
[20] Zhao, Z.-y., Gu, W., Cai, T., Tagliani, L., Hondred, D., Bond, D., . . . Pierce, D. (2002). High throughput genetic transformation mediated by *Agrobacterium tumefaciens* in maize. *Molecular Breeding, 8*(4), 323-333.

10. Supplemental Materials

10.1. Supplement 1

Supplement 1 materials can be found on pages 1 - 25 located at http://www.feedhaccp.org/distance/elearning/JRS/2019 /jrs-v07brink_appendix.pdf.

10.1.1. Supplement 1: Analysis of DI event

SbS analysis of the DI event resulted in high levels of coverage across the transformation construct, and yielded two plasmid-genome junctions (junctions 1 and 4 in Supplement 1: Figure S15C) and two plasmid-plasmid junctions (junctions 2 and 3). The presence of only two plasmid-genome junctions demonstrates that there is a single insertion in the DI genome, however, detection of two plasmid-plasmid junctions indicates the presence of multiple copies in the insertion. Junction 1 begins at bp 19 of the PHP63750A transformation fragment, and junction 2 indicates the end of the first copy at bp 11,976. The second copy of PHP63750A begins at bp 1 of the fragment (junction 3) and ends at bp 11,973 with plasmid-genome junction 4. These data show there are two nearly-complete copies of PHP63750A inserted at a single location in the soybean genome, with the first copy truncated by 18 bp at the 5′ end but otherwise intact, while the second copy is intact except for 3 bp deleted at the 3′ end. A map of the inserted DNA in the DI event was created using this information (Supplement 1: Figure S15B).

Southern blot analysis of the DI event with *Bgl*II resulted in a total of three bands: ~3,700 bp with the probes located 5′ to the *Bgl*II site, ~11,000 bp with the probes 3′ to the *Bgl*II site, and ~12,000 bp with all probes (Supplement 1: Figure S15A). The results of hybridization with the Terminator 2 probe and the RNA fragments probe are shown in Figures 2A and 2B, respectively (*Bgl*II panel, lane 4, bands at ~11,000 bp and ~12,000 bp). The ~12,000 bp band detected with all probes indicates a tandem arrangement of two copies of PHP63750A; as this is the size of the transformation fragment, two copies in tandem would produce a band of ~12,000 bp from the sequences between the two *Bgl*II sites in such an arrangement. The ~3,700 bp and ~11,000 bp bands result from the genomic border bands at the 5′ and 3′ ends, respectively, of the tandem insertion in this event.

Hybridization of the *Avr*II/*Spe*I digest resulted in the expected internal bands of ~3,700 bp and ~7,200, indicating that the restriction sites in the PHP63750A fragment are intact (Supplement 1: Figures S9 - S14, Figure S15A, and Table S3). The 5′ probe and Terminator 1 probe also yielded a band of ~11,000 bp from the 5′ genomic border, and an additional band of ~850 bp. The band of ~850 bp was also detected with the 3′ probe, indicating that this band is derived from the region of the insertion containing the junction between the two copies of PHP63750A (Supplement 1: Figures S9 - S14, Figure S15A, and Table S3). The results of hybridization with the Terminator 2 probe and the RNA fragments probe are shown in Figures 2A and 2B, respectively (*Avr*II/*Spe*I panel, lane 4). The presence of the internal bands shows that the restriction enzyme sites within both copies of the PHP63750A transformation fragment are all intact, and the presence of the ~850 bp band with both the 5′ probe and 3′ probe supports the proposed tandem arrangement of two copies of PHP63750A in the DI event.

The Southern blot analysis of the DI event for copy number and intactness demonstrated that there is a single DNA insertion comprising two copies of the PHP63750A transformation fragment in a tandem arrangement. A map of the insertion was created using the Southern blot results (Supplement 1: Figure S15A). Comparison of the SbS results and the Southern blot results shows that both types of analysis yielded the same conclusion regarding the copy number and intactness of the inserted DNA in this event.

10.1.2. Supplement 1: Analysis of CI event

SbS analysis of the CI event resulted in high levels of coverage across the transformation construct, and yielded two plasmid-genome junctions (junctions 1 and 8 in Supplement 1: Figure S16D) and six plasmid-plasmid junctions (junctions 2-7). The presence of only two plasmid-genome junctions demonstrates that there is a single insertion in the CI genome; however, detection of multiple plasmid-plasmid junctions indicates the presence of several fragments of PHP63750A within this one insertion. Junctions 1 and 8 are both located in the middle of PHP63750A; therefore, both ends of the CI insertion resulted from truncated fragments. The six plasmid-plasmid junctions detected by SbS indicate that there are three fragments within the insertion, as each physical junction yields two SbS junctions (one SbS junction from each side of the physical junction). Using the location of the junctions within PHP63750A, and the sequences generated by SbS on either side of each junction, allows the creation of a putative map of the inserted DNA in the CI event (Supplement 1: Figure S16C). This map shows the presence of several truncated PHP63750A fragments within the CI insertion, with one or two copies of each genetic element.

Southern blot analysis of the CI event with *Bgl*II resulted in a total of three bands of ~15,000 bp, ~9,000 bp, and ~8,000 bp, with the specific band(s) detected depending on the probe used (Supplement 1: Figure S3 - S8, Figure S16A, and Table S2). The band at ~8,000 bp was very faint and was determined

to be an artifact due to restriction enzyme digestion or electrophoresis. The remaining Southern blot analysis all support the presence of only two copies of each of these elements. The results of hybridization with the Terminator 2 probe and the RNA fragments probe are shown in Figures 2A and 2B, respectively (*Bgl*II panel, lane 6). The Terminator 2 probe detected all three of these bands, while the RNA fragments probe showed only the ~15,000 bp band. These Southern blot results indicate multiple fragments within the insertion, but *Bgl*II alone is not sufficient to determine exact copy number of each genetic element due to the insertion's complexity.

Hybridization of the *Avr*II/*Spe*I digest did not result in the expected internal bands with any of the probes, indicating that no part of the PHP63750A transformation fragment inserted intact (Supplement 1: Figure S9 - S14, Figure S16A, and Table S3). The results of hybridization with the Terminator 2 probe and the RNA fragments probe are shown in Figures 2A and 2B, respectively (*Avr*II/*Spe*I panel, lane 6). Rather than the expected internal band of 7,222 bp, the Terminator 2 probe yielded two bands of ~10,000 bp and ~6,600 bp, while the RNA fragments probe gave a band of ~16,000 bp, confirming that the PHP63750A fragments were truncated and rearranged upon insertion into the genome.

Due to the apparent complexity of the CI insertion, the *Bgl*II and *Avr*II/*Spe*I digests did not provide enough information to allow construction of an insertion map. Therefore, additional restriction digests were used to generate additional Southern blot data: *Bam*HI, *Bcl*I, *Eco*RV, and *Nco*I (Supplement 1: Figure S16B). The results of hybridizations with these digests are provided in Supplement 1: Tables S4 and S5 and the Southern blots are shown in Supplement 1: Figure S17 - S28. Combination of the data from all six Southern blot digests allowed for the development of an insertion map (Supplement 1: Figure S16A and B), showing the presence of several truncated fragments of PHP63750A, with one or two copies of each genetic element. This map is in complete accordance with the map generated from the SbS results (Supplement 1: Figure S16C), indicating that even in the case of a highly complex insertion, SbS and Southern blots provide equivalent information about copy number and insertion structure.

10.2. Supplement 2

Supplement 2 materials can be found on pages 26 - 27 located at http://www.feedhaccp.org/distance/elearning/JRS/2019/jrs-v07brink_appendix.pdf.

10.3. Supplement 3

Supplement 3 materials can be found on pages 28 - 41 located at http://www.feedhaccp.org/distance/elearning/JRS/2019/jrs-v07brink_appendix.pdf.