STUDY TITLE

Sequence Characterization of the Inserts in Z6

AUTHOR

[personal information redacted]

REPORT DATE

March 11, 2019

PERFORMING LABORATORY

SPS Molecular Lab

STUDY NUMBER

[CCI]

CERTIFICATION PAGE

I, the undersigned, declare that, to the best of my knowledge, this report provides an accurate evaluation of data in this study.

Signed	3/12/19
[personal information redacted]	Date
Senior Molecular Scientist	

TABLE OF CONTENTS

INTRODUCTION

Z6 was developed by transforming Snowden potatoes with pSIM1278 and then with pSIM1678. The pSIM1278 T-DNA consists of two inverted repeats designed to reduce the expression of asparagine synthetase, polyphenol oxidase, R1 glucan water dikinase, and phosphorylase L in tubers. The pSIM1678 T-DNA consists of the late blight resistance gene Rpi-v*nt1* and an inverted repeat to reduce expression of vacuolar invertase in tubers.

This study provides a molecular characterization of the pSIM1278 and pSIM1678 inserts in Z6. Locus number, insert structure, flanking sequence, absence of backbone, and insertion site characterizations were conducted with a combination of methods including droplet digital PCR [CCI], captured mate-pair Illumina sequencing, Southern blotting [CCI], PCR, and Sanger sequencing.

STUDY OBJECTIVES

Describe T-DNA insertion sites in Z6 using multiple forms of DNA fragment and sequence analysis. Determine locus number, flanking sequence, absence of backbone, and insertion site characterization. Validate number of inserts and insert structure determined previously by ddPCR and Southern blotting.

STUDY DATES
August 2015-March 2019

KEY STUDY PERSONNEL

[personal information redacted]

MATERIALS AND METHODS

Plant Material

Tissue culture (G0) plants were grown in Sunshine mix-1 (www.sungro.com) in two-gallon pots in a greenhouse controlled for temperature (18 °C minimum/27 °C maximum) and light (16 h photoperiod with an intensity of about 1,500 µmol/m^2/s). After one to two months of growth, leaf tissue was collected for genomic DNA isolation.

DNA Isolation

A 1.0 g sample of young potato leaves was ground into a fine powder using a mortar and pestle under liquid nitrogen. The ground tissue was transferred to a pre-cooled 15 mL conical tube with a pre-cooled spatula and stored at -80 °C. Powdered tissue was thoroughly mixed with 10 mL extraction buffer (0.35 M sorbitol, 0.1 M Tris-HCl, pH 8.0, 0.05 M EDTA) and the suspension was centrifuged at 3,000 rpm for 15 min at room temperature. The pellet was resuspended in 2 mL extraction buffer containing 200 µg RNase A. After incubating the suspended DNA at 65 °C for 20 min with 2 mL nuclear lysis buffer (0.2 M Tris-HCl pH 7.5, 0.005 M EDTA pH 8.0, 20 mg/mL CTAB, 800 µL 5% sarcosyl) it was mixed with an equal volume of chloroform: isoamyl alcohol (24:1), vortexed for 1 min, and centrifuged at 3,000 rpm for 5 min at room temperature. To precipitate the DNA the aqueous layer was mixed with an equal volume of isopropyl alcohol, washed with 70% ethanol, air dried, and dissolved in 400-700 µL 1X Tris/EDTA buffer (TE). DNA concentration was measured using Qubit Fluorometric Quantitation (Life Technologies).

Polymerase Chain Reaction (PCR) confirming flanking sequences and characterizing integration site
Primers were designed to confirm the junction regions between the insert and genomic DNA. To characterize the integration site, primers were designed using the Snowden flanking genomic sequences obtained for each side of the insert. A forward primer from the left flank was combined with a reverse primer from the right flank to amplify each integration site (Table 1). PCR reaction conditions are described in Table 2. PCR products were visualized following electrophoresis on a 1.0% agarose gel, then extracted using a [CCI].

Table 1. PCR primers

[CCI]		

Table 2. PCR reaction conditions

Temperature (^{0}C)	Time	Cycle Number
95	2 min	1
95	20 seconds	35
55	20 seconds	
72	3 min	
72	2 min	1

Sanger Sequencing

Plasmids purified using QIAprep Spin Miniprep Kit (Qiagen) were Sanger sequenced at the Sequetech Corporation using the BigDye® Terminator v3.1 Cycle Sequencing Kit and a PRISM 3730xl DNA Analyzer (Applied Biosystems).

Databases

Databases for the potato reference genome (Potato Genome Sequencing Consortium, 2011; Sharma SK, Bolser D, de Boer J, Sønderkær M et al, 2013)were utilized for sequence alignments and BLAST searches, obtained from the PGSC Data download site (http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml), specifically "PGSC *S. tuberosum* group Phureja DM1-3" Pseudomolecules (v4.03) and Transcripts (v3.4).

Illumina Library Preparation, Capture, and Sequencing

[CCI]

Illumina Sequence Analysis

[CCI]

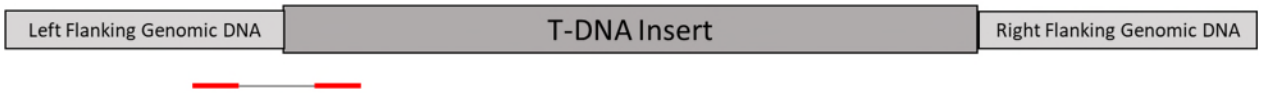

Figure 1. Junction Finding Method using Illumina Data
Sequence pairs are collected where one read aligns to the T-DNA and the other read aligns to the reference genome. Loci with >20 reads unique to a given transformation event indicate a T-DNA insertion. Reads mapping to additional loci can arise from potato-derived DNA in the T-DNA cassette, flanking sequence that is repetitive, and insertion sites. Loci identified by junction finding scripts are inspected. Junctions are confirmed with PCR and Sanger sequencing.

Read pairs with one read mapping to the T-DNA cassette of either pSIM1278 or pSIM1678 and the other read mapping to a locus in the genome may not be uniquely associated with insertion sites due to the presence of potato DNA in the T-DNA cassettes. To distinguish between these two possibilities, [CCI]

[CCI]

After flanking sequences were confirmed by PCR and Sanger sequencing, and the structure confirmed by Southern blotting, the insert and flanking sequences were assembled. Illumina mate pairs were mapped to the assembled sequences with [CCI]

.

Results showing next generation sequence alignments display reads by orientation, with forward reads (green), reverse reads (red), and pairs (light and dark blue). Blue read pairs indicate pairing has been detected taking into account the reverse/forward orientation associated with mate pair type libraries, and light and dark blue indicate the orientation of the random fragment.

To detect possible vector backbone in the potato genome, reads mapping with low stringency (85% identity over 50% of the read) to the binary vector backbone were extracted. These reads were re-mapped, requiring 99.5% identity over 90% of the read, and the results were inspected. Any reads mapping to the backbone were further analyzed to see whether the mate pair maps to the potato genome or if they might be derived from common laboratory high copy plasmid DNA.

RESULTS

Number of Insert Validation

Research using Illumina Next Generation Sequencing (NGS) data to characterize transgene inserts suggests that this method can be used to replace Southern blots (Zastrow-Hayes et al., 2015). Here, a similar approach was used in conjunction with Southern blot analysis. Mate pair libraries and multiplexed targeted sequencing were used to adapt to the challenges of identifying inserts containing native, potato sequences and inverted repeats. Insertion locus identification with bioinformatic scripts indicated one insert each from pSIM1278 and pSIM1678 in the Z6 genome.

Insert Flanking Sequences

The pSIM1278 T-DNA insert and the pSIM1678 T-DNA insert in Z6 each consists of a single locus as indicated by ddPCR, next generation sequencing analysis, and Southern blotting [CCI]. PCR generated products for Sanger sequencing across the junctions near the left border (LB) and right border (RB) of each insert including at least 1 kb of flanking DNA (Table 1). After Sanger sequencing of the left and right junction regions, a reference sequence for the pSIM1278 insert and flanking sequence (depicted in Figure 2A) and a reference sequence for the pSIM1678 insert and flanking sequence (depicted in Figure 2B) were assembled for validation by Illumina data.

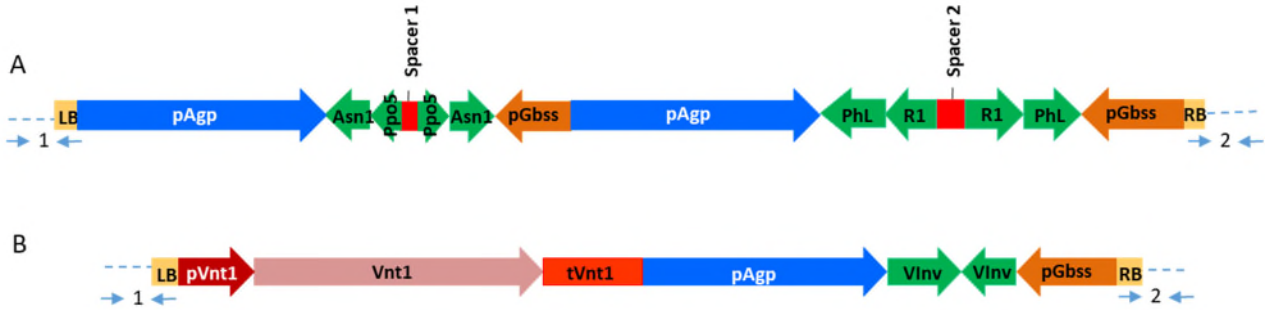

Figure 2. Insert Structures in Z6 with Junctions Labeled
The Z6 event contains a single, nearly full-length T-DNA from (A) pSIM1278 insert and (B) pSIM1678 insert. The plant genomic flanking sequence is indicated with a dashed line. Arrows denote primer pairs (Table 1) used to amplify the junction near the left border (LB, 1) and the right border (RB, 2) for each insert.

Insert Sequence Validation

Illumina data were used to validate the insert structure that was determined by Southern blotting [CCI]. The Z6 Illumina library was generated by sequencing mate pair DNA fragments sharing sequence identity to either pSIM1278 or pSIM1678 after targeted capture. As described above, the reference pSIM1278 insert (event V11) and the reference pSIM1678 insert were assembled from Sanger sequenced flanking genomic junctions and the T-DNA sequence. Illumina mate pair sequencing reads were aligned to both references as well as the genome. Targeted capture and resulting enrichment yielded thorough depth of coverage of read pairs across the entire pSIM1278 insert, including the junctions and adjacent flanking genomic DNA (Figure 3A) and the entire pSIM1678 insert, including the junctions and adjacent flanking genomic DNA (Figure 3B). The alignments were inspected to confirm accuracy by depth of coverage and sequence quality scores across the entire pSIM1278 insert and pSIM1678 insert.

Sequence data (Appendix A) showed that the pSIM1278 insert contains a nearly full-length T-DNA with a 14 bp deletion from the left border annotation and a 23 bp deletion from the right border annotation. Sequence data (Appendix B) showed that the pSIM1678 insert contains a nearly full length T-DNA with a 9 bp deletion from the left border annotation and a 36 bp deletion from the right border annotation.

The sequences described are consistent with the structures of the pSIM1278 and pSIM1678 inserts determined by Southern blot analyses [CCI].

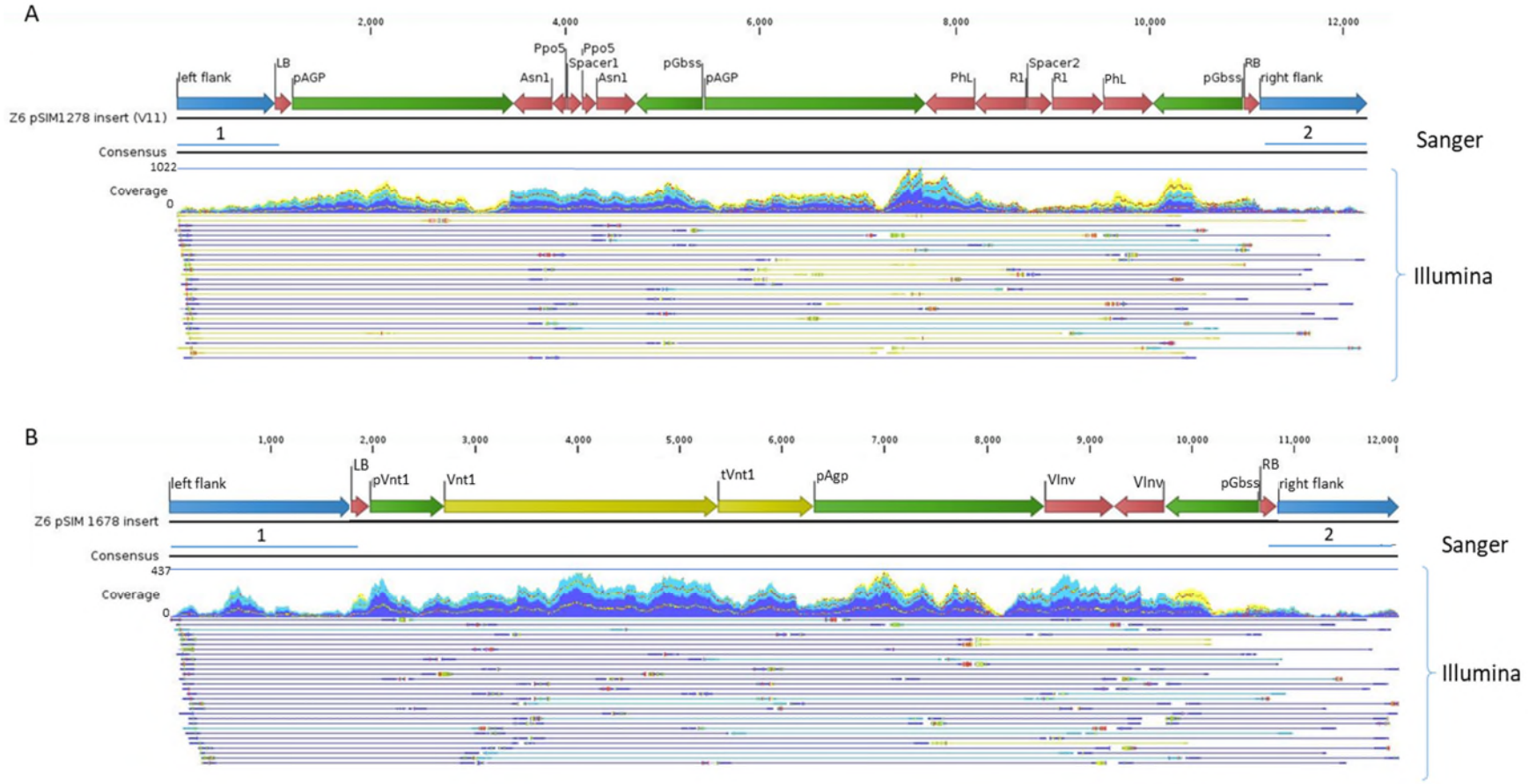

Figure 3. Summary of the Sequence Characterization of the pSIM1278 and pSIM1678 Inserts

Maps of the (A) pSIM1278 insert (V11) and (B) pSIM1678 insert are shown with blue lines used to represent regions Sanger sequenced, where 1 is the left junction and 2 is the right junction. Illumina sequence coverage depth of paired reads is displayed, indicating that there are more reads than shown. Yellow corresponds to reads that align to multiple places on the inserts (i.e. inverted repeats or duplicated promoters).

Insertion Site Characterization

Agrobacterium-mediated transformation results in T-DNA being introduced at one or more random locations in the genome. Snowden potatoes are tetraploid, containing four homologous sets of chromosomes, and each insert is expected to occur at a unique locus on one of the four homologues. Therefore, Z6 retains three chromosomes with a native locus homologous to the insertion site following transformation.

The sequence of the genomic integration site for the insert from pSIM1278 was determined using primers that hybridize to the flanking genomic regions of the insert (Figure 4). PCR with these primers amplified the homologous, native loci in untransformed Snowden, resulting in an amplicon of approximately 1.4 kb (Table 1). Comparison of the Sanger sequenced insertion site (Appendix C) and the insert described above (Appendix A) indicated that 72 bp of genomic DNA was deleted as a result of the T-DNA insertion from pSIM1278. BLAST of the 1.4 kb sequence aligns to a region of the potato reference genome chromosome 4 (region 63937636-63939054) with 97% ID. The insertion site (72 bp deleted sequence highlighted purple, Figure 4B) does not indicate interruption of known genes.

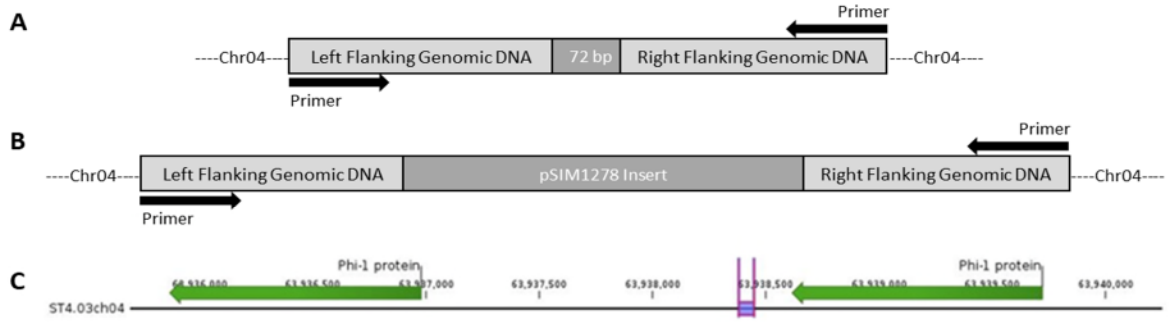

Figure 4. Characterization of the pSIM1278 Insertion Site on Chromosome 4
(A) Native locus showing 72-bp deletion due to insertion of the pSIM1278 insert. PCR primer locations are indicated on the left and right flanking regions of the native locus. (B). Z6 locus showing the pSIM1278 insertion and left and right flanking genomic regions of chromosome 4 and PCR primer locations. (C) The region deleted by the T-DNA insertion (purple highlight) occurs between two gene annotations in the reference genome.

The sequence of the genomic integration site for the insert from pSIM1678 was determined using primers that hybridize to the flanking genomic regions of the insert (Figure 5). PCR with these primers amplified the homologous, native loci in Snowden. Due to a duplication of 957 bp at the left and right flanking sequence, primers were designed outside using sequence unique to the left and right flank. After Sanger sequencing, identity of the insertion site (Appendix D) was annotated relative to the insert described above (Appendix B). A BLAST search of the 957 nt duplicated sequence at each flank aligns to a region of the potato reference genome chromosome 9 with 97% ID (region 3678695-3677323), which does not indicate interruption of known genes (Figure 5B).

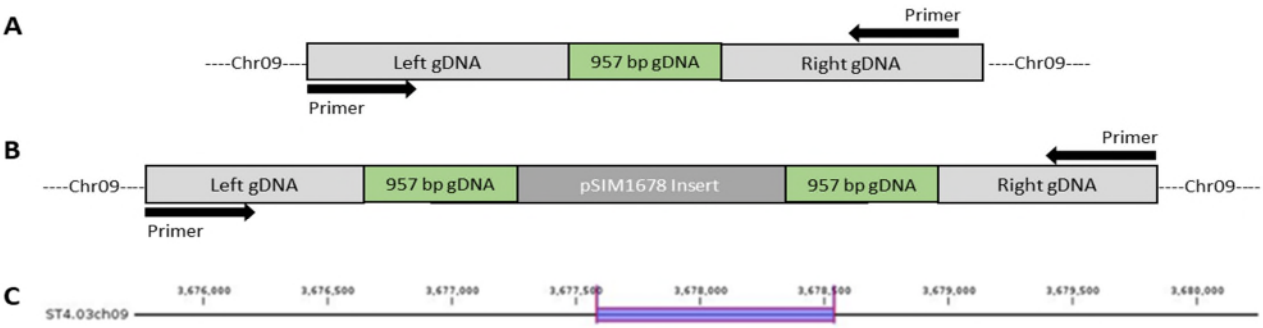

Figure 5. Characterization of the pSIM1678 Insertion Site on Chromosome 9
(A) Native locus showing 957 bp genomic DNA (gDNA) region duplicated upon insertion of the pSIM1678 insert. PCR primer locations are indicated on the left and right flanking regions of the native locus. (B) Z6 locus showing the pSIM1678 insertion and left and right flanking genomic regions of chromosome 9. The duplicated 957 bp gDNA is shown in green and PCR primer locations are present on the left and right flanking gDNA. (C) The region of the reference genome duplicated by the T-DNA insertion (purple highlight).

Absence of Vector Backbone DNA

Illumina mate pair sequencing libraries were captured to enrich for high coverage of sequences derived from pSIM1278 or pSIM1678, including the backbone (which is shared between the two vectors). The backbone contains two elements comprised of potato DNA, the *Ubi7* promoter and the *Ubi3* terminator, which were the only regions of backbone that contained mapped reads (Figure 6). The mate pairs of these reads were inspected and do not connect to either of the pSIM1278 or pSIM1678 insertion site loci. The two reads that map to the backbone are most likely derived from common laboratory high copy plasmid DNA. The results are comparable to the Snowden conventional control. These results indicate that no backbone is inserted in Z6, corroborating Southern blot data [CCI].

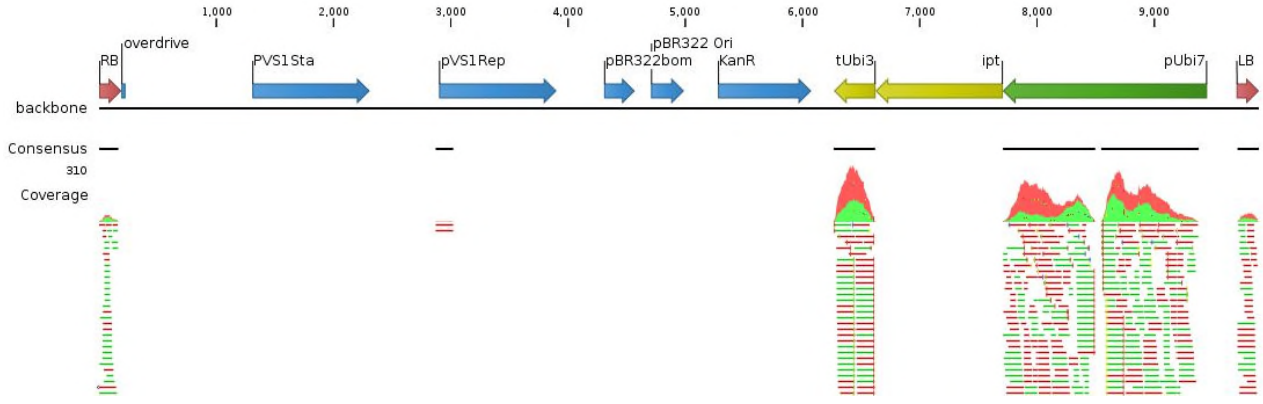

Figure 6. Illumina sequencing reads aligned to the binary vector backbone

Potato DNA derived elements Ubi7pro, Ubi3T, and the T-DNA borders show significant representation in the Z6 enriched sequencing library, as well as in the untransformed Snowden control (not shown). Red and green lines are forward and reverse reads, respectively.

CONCLUSION

A combination of Sanger and Illumina NGS sequencing corroborated studies using ddPCR and Southern blots, showing the presence of single inserts associated with transformation of Z6 using pSIM1278 and pSIM1678. The structure and sequences of the two inserts in Z6 are provided, with flanking DNA sequence. No backbone DNA was integrated into the Snowden genome. No annotated genes were disrupted by the insertion of these T-DNAs.

REFERENCES

Illumina Inc. (2012). Data Processing of Nextera® Mate Pair Reads on Illumina Sequencing Platforms. 1–12.

Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. Bioinformatics *25*, 1754–1760.

Potato Genome Sequencing Consortium (2011). Genome Sequence and Analysis of the Tuber Crop Potato. Nature *475*, 189–195.

Sharma SK, Bolser D, de Boer J, Sønderkær M, et al. (2013). Construction of Reference Chromosome-Scale Pseudomolecules for Potato: Integrating the Potato Genome with Genetic and Physical Maps. Genes Genomes Genetics *3*, 2031–2047.

Zastrow-Hayes, G.M., Lin, H., Sigmund, A.L., Hoffman, J.L., Alarcon, C.M., Hayes, K.R., Richmond, T.A., Jeddeloh, J.A., May, G.D., and Beatty, M.K. (2015). Southern-by-Sequencing: A Robust Screening Approach for Molecular Characterization of Genetically Modified Crops. The Plant Genome *8*, 1–15.

APPENDIX A

pSIM1278 T-DNA insert in V11 including 1018 nt of left flanking sequence (lowercase), T-DNA insert (uppercase), and 1121 nt right flanking sequence (lowercase). Primer regions are in yellow.

[CIC]

APPENDIX B

pSIM1678 T-DNA insert in Z6 including 1111 nt of left flanking sequence (lowercase), T-DNA insertion (uppercase), and 1189 nt of right flanking sequence (lowercase). Underlined sequence indicates 957bp genomic DNA duplicated during insertion. Primer regions are in yellow.

[CCI]

APPENDIX C

Chromosome 4 insertion site sequence of 1443 nt including 768 nt sequence homologous to the V11 left flanking sequence (lowercase), 72 nt deleted during insertion (uppercase and underlined), and 603 nt sequence homologous to the V11 flanking sequence (lowercase). Primer regions are in yellow.

[CCI]

APPENDIX D

1285 nucleotides of chromosome 9 insertion site sequence including left and right flanking regions (lowercase) and 957 nt sequence duplicated at both flanks in uppercase. Primer regions are in yellow.

[CCI]