This report has personal information and commercial confidential information (CCI) redacted

STUDY TITLE

Allergen and Toxin Evaluation of Open Reading Frames in Z6

AUTHORS

[Personal information redacted]

REPORT DATE

March 12, 2019

PERFORMING LABORATORY

SPS Regulatory 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	3/12/2019
[Personal information redacted]	Date
Regulatory Affairs Manager	

TABLE OF CONTENTS

SUMMARY

Objective: To determine whether pSIM1278 and pSIM1678 open reading frames (ORFs) in Z6 are homologs of known or putative toxins or allergens based on sequence identity.

Methods: ORFs were defined as contiguous, ≥30-amino acid sequences between translation start and stop codons. The ORFs were identified in all six reading frames from both pSIM1278 and pSIM1678 inserts, and from the flanking regions of the genome to include junction ORFs.

Allergen homology searches were performed using online tools available through the Food Allergy Research and Resource Program (FARRP; http://www.allergenonline.org/databasefasta.shtml). The AllergenOnline database (version 19, released February 10, 2019) contained 2,129 protein sequences. Full-length, 80-mer, and 8-mer searches were performed. The full-length search identified matches with at least 50% sequence identity between ORF and allergen protein sequences with an E-value cutoff of 10^{-4}. The 80-mer search identified matches with at least 35% sequence identity between contiguous 80 amino acid ORF sequences and allergen sequences. The 8-mer search identified matches with 100% sequence identity between 8 amino acid ORF sequences and allergen sequences.

Toxin homology searches were performed against the non-redundant protein sequence database (nr) using BLAST (blastp; https://blast.ncbi.nlm.nih.gov/Blast.cgi). Searches were performed on March 7, 2019 using the Entrez query “toxin” and an E-value cutoff of 10^{-2}.

Results: Sixty (60) ORFs were identified in Z6 that were associated with either the pSIM1278 and pSIM1678 inserts, the junction between the insert and genomic DNA, or the *Rpi-vnt1* gene sequence from pSIM1678.

Allergen Search

Two ORFs associated with the VInv sequence from the pSIM1678 insert were identified as having homology to a minor allergen found in tomato (*Solanum lycopersicum*). Alignment with the tomato invertase protein was not unexpected as the sequence of the VInv inverted repeat in pSIM1678 is derived from potato and shares 95% sequence identity with the tomato protein. The VInv sequence naturally exists in potato, and Z6 potatoes are not expected to contain higher levels of invertase than conventional potatoes as the inverted repeat is designed to lower expression of the vacuolar invertase protein.

Toxin Search

A search of the NCBI, non-redundant protein database did not identify homology with any toxins. Some matches to asparagine synthetase and other non-toxic sucrose degrading enzymes were identified in toxic or pathogenic bacteria. The proteins are homologs of the potato asparagine synthetase and vacuolar invertase proteins, they are not themselves toxic. The matches were identified because the accession record was annotated with the word “toxin” reflecting the proteins microbial source.

The toxin search matched the VNT1 coding sequence with four proteins that contained the keyword “toxin” in their accession records. These proteins, however, are not toxins, but function to protect

against toxins. LOV1, RP3-like, and Tsn1, are R-proteins that recognize pathogen-expressed effector proteins (i.e. victorin, Pc toxin, and ToxA, respectively) and elicit a protective immune response. RGA2-like, from apple (*Malus domestica*), is an R-protein that shares a short, specific region of sequence identity with VNT1. This short region may not be correctly annotated in the NCBI database as it is also found in over 200 other eukaryotic proteins, none of which have similarly annotated accession records.

Conclusion: The allergen and toxin homology searches did not identify any safety concerns for the ORFs in Z6. The vacuolar invertase protein from potato is similar to the vacuolar invertase protein from tomato and is not expected to be translated into protein because transcripts of the inverted repeat enter the RNAi pathway as processed siRNA. These siRNA are intended to decrease levels of the invertase protein in potato. Furthermore, as invertase naturally exists in potatoes, Z6 potatoes are no more likely than conventional varieties to cause an allergic reaction in individuals already sensitive to the tomato vacuolar invertase protein.

INTRODUCTION

Z6 was developed by transforming Snowden potatoes with pSIM1278 and then with pSIM1678. The T-DNA region of pSIM1278 consists of two cassettes designed to reduce *Asn*, *Ppo*, *R1*, and *PhL* transcripts in tubers. The cassettes are designed to express inverted repeats, which form double-stranded structures that are processed into small interfering RNA (siRNA) by the cell's RNA interference machinery. Processing of the dsRNA makes it unlikely that transcripts from the pSIM1278 insert would be translated into protein.

The T-DNA region of pSIM1678 also consists of two cassettes. The first is designed to express the *Rpi-vnt1* late blight resistance gene from the wild *Solanum* species, *S. venturii*, and the second to reduce expression of the potato vacuolar invertase gene in tubers. Like with the inverted repeats in pSIM1278, processing of the dsRNA into siRNA makes it unlikely that transcripts are translated into protein.

An allergic reaction is an immune-like response that occurs when antibodies in the serum of allergic individuals interact with specific antigens (typically proteins) present in pollen or food. Bioinformatics is used to search for sequence identity between a protein introduced into a biotech food crop and proteins known or suspected to be allergens. Homology between queried sequences and known or suspected allergens could indicate potential antibody cross-reactivity that could cause an allergic response in individuals with allergy to the identified protein.

The allergen and toxin potential of open reading frames (ORFs) associated with each insert was assessed using well-established bioinformatics techniques (Goodman et al., 2008; Ladics et al., 2007; Terrat and Ducancel, 2013). The ORF sequences queried in this study included the VNT1 protein sequence (891 amino acids; ACJ66594) expressed from *Rpi-vnt1*, and all ORFs associated with the inserts including junction ORFs. Some insert-associated ORFs in Z6 are duplicated between pSIM1278 and pSIM1678 because the T-DNA elements were constructed using similar DNA sequences from potato. Duplicate ORFs were removed before performing the searches.

STUDY OBJECTIVE

To determine whether pSIM1278 and pSIM1678 ORFs in Z6 are homologs of known or putative toxins or allergens based on sequence identity.

STUDY DATES

12/2018-03/2019

KEY STUDY PERSONNEL

[Personal information redacted]

MATERIALS AND METHODS

Identification of ORFs Associated with pSIM1278 and pSIM1678

An ORF was defined as a contiguous, ≥30 amino acid sequence between start- and subsequent in-frame stop-codons. Nucleotide sequences were translated in three reading frames from two directions. All six reading frames within the pSIM1278 and pSIM1678 inserts and flanking regions were analyzed for ORFs (shown schematically in Figure 1). The results were converted into FASTA-formatted files using CLC Genomic Workbench software (Qiagen). An ORF was considered novel if it would not otherwise exist in the potato genome.

All ORFs were identified that were contained within the Z6 inserts, including those adjacent to flanking genomic sequence (junction ORFs). Duplicate ORFs that resulted from having repetitive elements in the insert (e.g. pAgp and pGbss) were removed.

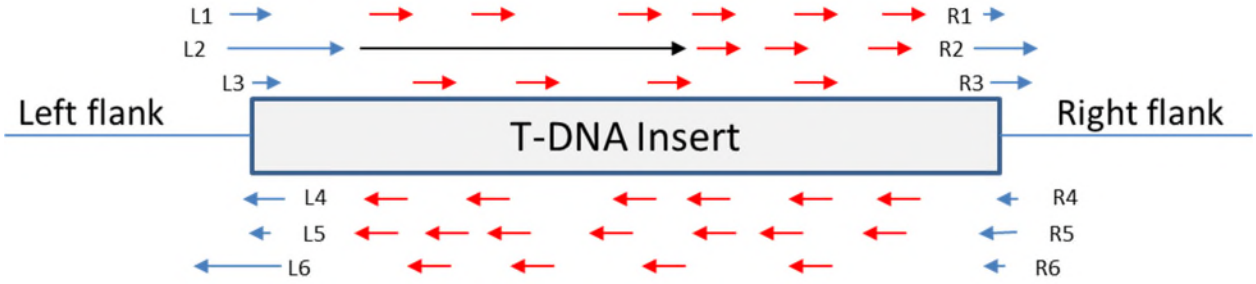

Figure 1. Complete ORF Analysis Scheme
A representative T-DNA insertion site in the plant genome. All ORFs ≥30 amino acids, contained within the DNA insert (red lines), including an introduced protein coding region (**black** line) or overlapping the junctions between the insert and plant genome (blue lines), were identified and used in subsequent analyses. All lines are representative and do not indicate actual ORFs.

Allergenicity Searches

Searches were performed to identify homology of ORF sequences, from the pSIM1278 and pSIM1678 inserts in Z6, to known allergens in the AllergenOnline.org database. AllergenOnline is made available by the Food Allergy Research and Resource Program (FARRP) at the University of Nebraska (http://www.allergenonline.org/databasefasta.shtml). Sequences deposited in the database are reviewed annually by a panel of qualified food allergenicity experts. Version 19 of the AllergenOnline database (available on February 10, 2019) contains 2,129 sequences, and was searched using FASTA to identify sequence identity between input and database entries (Pearson and Lipman, 1988). The search methods used are supported by recent published guidance for protein allergenicity prediction in food products, and are based on sequence identity as well as structural aspects of the interaction between antibodies and protein targets (cross-reactivity potential) (Goodman et al., 2008; Ladics, 2008; Ladics and Selgrade, 2009).

Full-Length Sequence Search

A full-length search used FASTA alignment to identify matches between full-length queries and the allergens contained in the AllergenOnline database. Homology was determined using the following criteria: (1) sequence identity was greater than 50% between the query protein and database entry, and (2) an E-value of less than 10^{-4}. The E-value describes the number of hits one can expect to see by chance when querying a database. The E-value is dependent on database size. Low E-values (e.g. $<10^{-4}$) indicate a higher degree of sequence similarity and that the hit is less likely to be due to chance.

80-mer Sliding Window Search

The 80-mer sliding window search identifies localized regions of similarity between the ORFs and known allergens by comparing all contiguous 80 amino acid sequences within an ORF to sequences in the AllergenOnline database. Matches were defined as sequences having greater than 35% identity to known allergens and an E-value of 10 was used by the AllergenOnline search algorithm to filter alignments containing large sequence gaps. Sequences in the database with less than 35% identity are not considered likely candidates for cross-reactivity (Ladics, 2008; Ladics and Selgrade, 2009) and are therefore not detailed in the search output.

8-mer Exact Match Search

The 8-mer exact-match search identifies small, localized regions consisting of eight amino acids of identity between the queried ORF sequence and known or suspected allergens in the database. Although the occurrence of matching 8-mer sequences can be used to assess the potential for cross-reactivity, FARRP warns that the 8-mer search can also lead to identification of false positive results, and suggests caution be used when interpreting findings based on this search alone. There is presently no scientific rationale for a small region of sequence with high identity to allergens being a safety concern without substantiation by other search methods (Goodman et al., 2008).

Toxicity Searches

Methods familiar to regulators and established in some countries include an approach modeled after the allergenicity studies where bioinformatics is used to inform on the potential of sequence similarity between protein sequences and known toxins.

The NCBI database was queried on March 7, 2019 using all protein sequences annotated with the keyword “toxin” (Entrez query: “toxin”; E-value $< 10^{-2}$). All matches are reported, although many proteins that are not actual toxins may be identified using this search criteria. Matches may include actual toxins, proteins involved in the synthesis of toxins in a host, proteins that interact with toxins, proteins involved in toxin-induced defense responses, or non-toxic proteins from organisms that produce other known toxins. Only proteins that match to known toxins indicate a potential safety concern. For all others, a brief explanation and rationale for safety is included.

RESULTS

The ORFs identified from the pSIM1278 and pSIM1678 inserts were assessed for homology to potential toxins and known allergens. The sequences associated with each ORF are shown in Appendix A. The ORF sequences were separated into three categories:

1. Gene sequences
    - The pSIM1678 insert contains the *Rpi-vnt1* late blight resistance gene, which expresses the VNT1 protein. There are no other protein expressing genes in the inserts from pSIM1278 or pSIM1678.
2. Junction sequences
    - These include ORFs spanning the insert and flanking genomic DNA.
    - A single ORF was identified at the left junction of each insert. There were no ORFs associated with the right junction from either insert.
3. Insert sequences
    - Most of these ORFs are contained within the individual elements in the inserts and are comprised of potato DNA.

Assessment of ORFs for Allergen Homology

Gene Sequences

No homologous allergens were identified for the coding sequence of *Rpi-vnt1* (VNT1), using the full-length, 80-mer, or 8-mer searches.

Junction Sequences

No homologous allergens were identified for any ORFs associated with junctions between inserts and genomic DNA using the full-length, 80-mer, or 8-mer searches.

Insert Sequences

Two ORFs associated with the VInv sequence from the pSIM1678 insert matched with a minor allergen, beta-fructofuranosidase precursor-vacuolar invertase, from tomato (Table 1). The ORFs are complementary as a result of the VInv inverted repeat and have similar sequences (Table 1, bold sequence). This similarity results in a common match being identified from the bioinformatics analysis.

Table 1. Allergen Homology of pSIM1678 ORFs in Z6

ORF[1]	Query Match	Organism	Accession
>ORF35 **MLSWQRTAYHFQPQKNWMNDPNGPLYHKGWYH LFYQYNPDSAIWGNITWGHAVSKDLIHWLYLPFAM VPDQWYDINGVWTGSA**TILPDGQIMMLYTGDTDDY VQVQNLAYPTNLSDPLLLDWVKYKGNPVLVPPPGIGV KDFRGIQRTQSRHR >ORF54 **MLSWQRTAYHFQPQKNWMNDPNGPLYHKGWYH LFYQYNPDSAIWGNITWGHAVSKDLIHWLYLPFAM VPDQWYDINGVWTGSA**EFL	Minor allergen beta-fructofuranosidase precursor–vacuolar invertase (Foetisch et al., 2003)	*Solanum lycopersicum*	AAL75449, AAL75450

[1]Bold sequence indicates region of identity between ORFs

Although homology was identified between these ORFs and the tomato allergen, it is not a safety concern. Because the ORF sequences exist naturally in potato as part of the vacuolar invertase protein, Z6 potatoes are no more likely than conventional potatoes to cause an allergic reaction in individuals sensitive to the tomato vacuolar invertase. Additionally, translation of the ORF into protein is unlikely because the transcript from the inverted repeat forms a dsRNA that gets processed by the RNAi machinery into small-interfering RNA (siRNA) to prevent ribosomal translation through RNA interference. The VInv cassette leads to production of siRNA that down-regulate the potato vacuolar invertase gene resulting in reduced amounts of vacuolar invertase in Z6 compared to conventional potatoes and eliminating any allergenicity concerns that may be associated with these ORFs.

Assessment of ORFs for Toxin Homology

ORFs were assessed for homology to toxins by alignment with proteins in the NCBI database limited to accession records annotated with the Entrez keyword, “toxin”.

Gene Sequences

No homologous toxins were identified from the coding sequence for VNT1. Some proteins that align with VNT1 were identified because they contain the keyword “toxin” in their accession records. These proteins function to protect plants against toxins or toxin-producing pathogens (Table 2).

Table 2. Summary of Toxin Homology for VNT1

ORF	Query Match	Organism	Accession
>ORF32 (VNT1) MNYCVYKTWAVDSYFPFLILTFRKKKFNEKLKEMA EILLTAVINKSIEIAGNVLFQEGTRLYWLKEDIDWLQ REMRHIRSYVDNAKAKEVGGDSRVKNLLKDIQQL AGDVEDLLDEFLPKIQQSNKFICCLKTVSFADEFAM EIEKIKRRVADIDRVRTTYSITDTSNNNDDCIPLDRR RLFLHADETEVIGLEDDFNTLQAKLLDHDLPYGVVS IVGMPGLGKTTLAKKLYRHVCHQFECSGLVYVSQQ PRAGEILHDIAKQVGLTEEERKENLENNLRSLLKIKR YVILLDDIWDVEIWDDLKLVLPECDSKIGSRIIITSRN SNVGRYIGGDFSIHVLQPLDSEKSFELFTKKIFNFVN DNWANASPDLVNIGRCIVERCGGIPLAIVVTAGML RARGRTEHAWNRVLESMAHKIQDGCGKVLALSYN DLPIALRPCFLYFGLYPEDHEIRAFDLTNMWIAEKLI VVNTGNGREAESLADDVLNDLVSRNLIQVAKRTYD GRISSCRIHDLLHSLCVDLAKESNFFHTEHNAFGDP SNVARVRRITFYSDDNAMNEFFHLNPKPMKLRSLF CFTKDRCIFSQMAHLNFKLLQVLVVVMSQKGYQH VTFPKKIGNMSCLRYVRLEGAIRVKLPNSIVKLKCLE TLDIFHSSSKLPFGVWESKILRHLCYTEECYCVSFASP FCRIMPPNNLQTLMWVDDKFCEPRLLHRLINLRTL CIMDVSGSTIKILSALSPVPRALEVLKLRFFKNTSEQI NLSSHPNIVELGLVGFSAMLLNIEAFPPNLVKLNLV GLMVDGHLLAVLKKLPKLRILILLWCRHDAEKMDL SGDSFPQLEVLYIEDAQGLSEVTCMDDMSMPKLK KLFLVQGPNISPISLRVSERLAKLRISQVL	LOV1: Confers susceptibility to the fungus *Cochliobolus victoriae* by conditioning victorin-dependent induction of defense-associated proteins. Victorin is a toxin synthesized by *C. victoriae.*	*Arabidopsis thaliana*	A7XGN8, A9QGV6
	RP3-like: Confers resistance to Pc toxin.	*Sorghum bicolor*	ACE86400, ACE86402, ACE86396
	Tsn1: Confers sensitivity to the wheat fungal pathogen ToxA.	*Aegilops speltoides*	ADG84875, ADG84876
	PREDICTED—RGA2-like: Region 107-171 annotated as "Ribonuclease toxin, BrnT, of type II toxin-antitoxin system".	*Malus domestica*	XP_008342325

LOV1, RP3-like, and Tsn1 are R-protein homologs that function in the sensitivity of their host to fungal pathogens through recognition of effector molecules, i.e. victorin, Pc toxin, and ToxA. A literature review confirmed that these R-proteins are in fact not toxins or substances with toxic properties (Faris et al., 2010; Lorang et al., 2007; Nagy and Bennetzen, 2008; Walton, 1996).

RGA2-like is a predicted disease resistance protein from apple (*Malus domestica*). The alignment between RGA2-like and VNT1 is shown in Figure 22. A 65 amino acid region annotated as BrnT toxin is highlighted in yellow. The sequence similarity between VNT1 and RGA2-like in this region is low, so VNT1 would not be considered homologous to the region annotated as BrnT toxin.

PREDICTED: disease resistance protein RGA2-like [Malus domestica]
Sequence ID: **XP_008342325.1** Length: 353 Number of Matches: 1
Range 1: 57 to 349

Score	Expect	Method	Identities	Positives	Gaps	Frame
81.6 bits(200)	3e-15()	Compositional matrix adjust.	90/314(29%)	146/314(46%)	39/314(12%)	

Features:

```
Query  93   DSRVKNLLKDIQQLAGDVEDLLDEFLPKIQQSNKFICCLKTVSFADEFAMEIEKIKRRVA  152
            D  + + L  ++ ++ D++D+LDEF  + Q+    +  L T S          + IK +V
Sbjct  57   DHLLTDWLGKLKDVSYDIDDVLDEF--EFQKLRMQVLGLGTGS---------DTIKGKV-  104

Query  153  DIDRVRTTYSITDTSNNNDDCIPLDRRRLFLHADETEVIGLEDDFNTLQAKLLD--HDLP  210
             I   +  +++ + S +              F+HA   +VIG E +   +  +L    H  P
Sbjct  105  -IAAAKAQFNLAERSVDWHGMHMERETHSFVHA--PDVIGRESEKEEIVVQLFKDTHTGP  161

Query  211  -----YGVVSIVGMPGLGKTTLAKKLY--RHVCHQFECSGLVYVSQQPRAGEILHDIAKQ  263
                 V+SI G+ GLGKTTLAK +Y    V   FE    V VS    +  +L +I
Sbjct  162  GDEENVSVISINGLGGLGKTTLAKLVYNDNRVVTNFELRIWVCVSDDFDSKRLLSEIVTA  221

Query  264  VGLTE--EERKENLENNLRSLLKIKRYVILLDDIWD-------VEIWDDLKLVLPECDSK  314
                +  +E  E ++  LR  L  K+ +++LDD+WD       V+ W DLK +L    +
Sbjct  222  ATSQKCGDESIEQMQMKLRRALTGKKLLLVLDDVWDKGPMGITVKKWIDLKSLLNV--AA  279

Query  315  IGSRIIITSRNSNVGRYIGGDFSIHVLQPLDSEKSFELFTKKIFNFVNDNWANASPDLVN  374
             G +II+T+RN +V   + GD  +H+L+ L        +F K  F    +      P+L+
Sbjct  280  CGRKIIVTTRNESVALLM-GDAHMHLLKVLPLSDCMTIFVKVAFARREE---QNHPNLMK  335

Query  375  IGRCIVERCGGIPL  388
            IG  IV+   G PL
Sbjct  336  IGEDIVKSVEGFPL  349
```

Figure 2. Alignment of VNT1 with Disease Resistance Protein, RGA2-like, from Apple
The BLAST (blastp) alignment of the VNT1 (Query) with the RGA2-like protein (Sbjct) from apple, which has a region annotated as "BrnT toxin" (highlighted in yellow). The identity between the two proteins is low across the highlighted region.

Additionally, this region of the RGA2-like protein appears to have been incorrectly annotated as a toxin. The annotation was made by a predictive software algorithm (Gnomon) in June of 2016, which has not been confirmed by human QC. A BLAST query of the sequence against the NCBI protein database identified over 200 matches (E-value ≤ 1) to eukaryotic proteins, none of which were annotated as toxins or the BrnT toxin. The same BLAST search did not identify this region in prokaryotes.

Junction Sequences

No homologous toxins were identified for any junction ORFs aligned to the entries in the NCBI database.

Insert Sequences

No homologous toxins were identified for ORFs, associated with the pSIM1278 or pSIM1678 inserts, when aligned to the entries in the NCBI database.

Alignment of pSIM1278 ORFs with the NCBI protein database identified asparagine synthetase homologs that contain the keyword, "toxin", in their accession record (Table 3). Asparagine synthetase homologs are expressed in many organisms but are not toxins. The pSIM1278 insert contains an inverted repeat

containing sequence from the potato asparagine synthetase (*Asn1*) gene. Each side of the inverted repeat produces a unique ORF. The two ORFs are similar in sequence resulting in a common set of protein matches when assessed using bioinformatics (Table 3, bold sequence).

Table 3. Toxin Search Results for pSIM1278 ORFs

ORF[1]	Query Match	Organism	Accession
>ORF4 **MFLMSRKIKSLG** **VKMVISGEGAD** **EIFGGYLYFHKAP** **NKEEFHTETCRKI** **KALHQYDCLRA** **NKATSAWGLEA** **RVPFLDKEF**EFV MCGLQKGESTKL QMNKNKTEIDF >ORF29 **MFLMSRKIKSLG** **VKMVISGEGAD** **EIFGGYLYFHKAP** **NKEEFHTETCRKI** **KALHQYDCLRA** **NKATSAWGLEA** **RVPFLDKEF**VNK LVIN	Asparagine Synthetase B or Asparagine Synthase	*Escherichia coli*	AIZ81697, KFD76705, NP_308731, EIL13740, EYZ18706, EYY50888, EZH10307, EIL02389, EYY57333, EIL08426, KDV14966, EYU77209, EIL30960, EZA17347, EYV14508, KNZ12093, EYW20264, EZA37061, EZB27777, EYV89799, EYZ96359, EJF05567, ANO76830, PJI61967, OTB64025, KYR19924, KYT66274, KYU11696, PBQ68604, OTB49553, KYS77729, KYR79236, KYU23155, KYR36119, PAU22677, PBR97907, KYR73496, OTB69004, PBR58416, KYR16525
	Asparagine Synthetase B	*Enterobacter cloacae*	ASQ17067
	Asparagine Synthetase B	*Shigella flexneri, Shigella boydii, Shigella dysenteriae*	PQO16610, PQM97814, PQM84983, PQN36548, PQN35957, PQN20224
	Asparagine Synthetase	*Vibrio cholerae*	EEY43083
	Asparagine Synthetase or Asparagine Synthase	*Clostridium botulinum*	KIL09560

[1]Bold is used to highlight regions of identity between the two ORFs

The alignment of pSIM1678 ORFs with the NCBI protein database identified sucrose-degrading enzymes that are homologs of potato vacuolar invertase and ubiquitously expressed in bacteria, including both pathogenic and non-pathogenic strains (Table 4). The accession records for these proteins contained the keyword “toxin” due to the pathogenicity of the host organism, which is not related to the sucrose-degrading enzymes. The pSIM1678 insert contains an inverted repeat containing sequence from the potato vacuolar invertase gene. Each side of the inverted repeat produces a unique ORF. The two ORFs are similar (Table 4, bold sequence) resulting in a common set of protein matches when assessed using bioinformatics.

Table 4. Toxin Search Results for pSIM1678 ORFs

ORF[1]	Query Match	Organism	Accession
>ORF35 **MLSWQRT AYHFQPQK NWMNDP NGPLYHKG WYHLFYQY NPDSAIWG NITWGHA VSKDLIHW LYLPFAMV PDQWYDI NGVWTGS A**TILPDGQI MMLYTGD TDDYVQVQ NLAYPTNLS DPLLLDWV KYKGNPVL VPPPGIGVK DFRGIQRT QSRHR	Glycosyl Hydrolase Family 32, Sucrose-6-Phosphate Hydrolase	*Escherichia coli*	KYU57217, KYU29761, KYU30531, KOZ65786, NP_311270, EYZ93717, EYV91737, KOZ09537, EZB27384, EYZ68862, EDV64124, EYY18841, EYZ04892, KDV15364, KFD75665, EZQ28926, EYV93762, PBQ65362, PBQ55252, PBQ65657, KYR67720, KYR68655, KYR82912, KYR32265, KYR67990, KYR41985, OTB64516, OTB37363, OTB74045, OTB58160, OTB47925, OTB40657, OTB85828, EYW80343, OTE04890, PBR16851, PBR55824, KYS71623, KYS89136, KYS68613, KYS70798, OTC48590, OTC43741, OTD44623, OTD31325, ARA06893, PAU30426, PAU23617, PJI56720
	Glycosyl Hydrolase Family 32	*Shigella dysenteriae, Shigella flexneri*	PQN16446, PQM92544
	Sucrose-6-Phosphate Hydrolase	*Enterobacter cloacae*	ASQ18782, ASQ18131
	Sucrose-6-Phosphate Hydrolase, Beta-Fructofuranosidase	*Enterococcus faecalis*	EFK77719, ELA02140
>ORF54 **MLSWQRT AYHFQPQK NWMNDP NGPLYHKG WYHLFYQY NPDSAIWG NITWGHA VSKDLIHW LYLPFAMV PDQWYDI NGVWTGS A**EFL	Glycosyl Hydrolase Family 32, Sucrose-6-Phosphate Hydrolase, Levanase, Beta-Fructofuranosidase	*Bacillus cereus, Bacillus anthracis, Bacillus amyloliquefaciens*	BAL16531, EDX61098, PDO99061, AWM53161, AWM45605, AWM49398, AWM45947, AWM49739, AWM53492, AWM53917, AWM46166, AWM49952
	Sucrose-6-Phosphate Hydrolase	*Vibrio cholera*	EEY52113, EEY41740, EEY48061
	Sucrose-6-Phosphate Hydrolase	*Enterococcus faecium*	EFF60448, EFF59908, EFF60441
	Glycosyl Hydrolase Family 32	*Clostridioides difficile*	OYO90556
	Sucrose-6-Phosphate Hydrolase	*Clostridium botulinum, Clostridium butyricum*	KIL08234, APF21264, APF21879
	Sucrose-6-Phosphate Hydrolase	*Staphylococcus hyicus*	AJC95603
	Sucrose-6-Phosphate Hydrolase	*Staphylococcus aureus*	KPE20402, KPE18748
	Hypothetical protein	*Ensifer aridi*	WP_026617525

Alignments of ORFs containing sequence from the potato asparagine synthetase and vacuolar invertase genes did identify non-toxic proteins whose annotation includes the keyword “toxin”. Both families of proteins are broadly expressed in pathogenic and non-pathogenic organisms but are not related to the toxicity of those organisms. Accession records for proteins are often annotated with the keyword “toxin” as a result of the source organism’s toxicity. The search criteria used (see Methods and Materials) can potentially result in identification of false-positives, as seen here.

CONCLUSION

This study used bioinformatics to assess the allergen and toxin potential of ORFs associated with the pSIM1278 and pSIM1678 inserts in Z6. The ORFs associated with the pSIM1278 and pSIM1678 inserts are unlikely to be translated into protein. Transcription of the inverted repeats within the inserts is driven by opposing promoters (pAgp and pGbss), which limits transcriptional read through. Transcription initiation is expected to be limited to the promoter elements contained within the insert, which except for the pVNT1 element, are not designed for expressing protein. The inverted repeats are designed to produce dsRNA that are processed into siRNA, which are not recognized or translated by the ribosome. The siRNA associated with the PHL/R1 and VInv inverted repeats would prevent accumulation of transcripts due to RNA interference mechanisms. In addition, any unexpected expression would require productive transcription, including capping, splicing, and polyadenylation of messenger RNA (mRNA), nuclear export, and translation into stable protein. Thus, only the pVNT1 element is anticipated to lead to production of a protein, i.e. VNT1. As was shown above, based on bioinformatics, none of the ORFs associated with the inserts in Z6 are homologous to known toxins or allergens.

Homology between the pSIM1278 and pSIM1678 insert-associated ORFs and known allergens was limited to an expected minor allergen in tomatoes, whose potato homolog is targeted for reduced expression in Z6. The presence of a potato vacuolar invertase sequence within the pSIM1678 inverted repeat led to ORFs with homology to a minor tomato allergen. As the *VInv* gene and sequences comprising the VInv inverted repeat are found in conventional potatoes, Z6 potatoes are no more likely than conventional potatoes to cause an allergic reaction in individuals sensitive to the tomato vacuolar invertase protein. The VInv cassette produces dsRNA that reduce expression of invertase in Z6, reducing levels of the potato protein homologous to the minor tomato allergen.

No toxins were identified with homology to the ORFs associated with the inserts in Z6. Proteins with homology to VNT1 were not toxins, but plant disease resistance proteins that protect against toxins and toxin-producing pathogens. Asparagine synthetase and sucrose degrading enzymes identified as homologs to ORFs associated with the asparagine synthetase and vacuolar invertase inverted repeats in Z6 potatoes are expressed ubiquitously in bacteria and are not toxins.

REFERENCES

Faris, J.D., Zhang, Z., Lu, H., Lu, S., Reddy, L., Cloutier, S., Fellers, J.P., Meinhardt, S.W., Rasmussen, J.B., Xu, S.S., et al. (2010). A Unique Wheat Disease Resistance-like Gene Governs Effector-Triggered Susceptibility to Necrotrophic Pathogens. PNAS *107*, 13544–13549.

Foetisch, K., Westphal, S., Lauer, I., Retzek, M., Altmann, F., Kolarich, D., Scheurer, S., and Vieths, S. (2003). Biological Activity of IgE Specific for Cross-Reactive Carbohydrate Determinants. The Journal of Allergy and Clinical Immunology *111*, 889–896.

Foster, S.J., Park, T.-H., Pel, M., Brigneti, G., Śliwka, J., Jagger, L., van der Vossen, E., and Jones, J.D.G. (2009). *Rpi-vnt1.1*, a *Tm-2*^{2}K Homolog from *Solanum Venturii*, Confers Resistance to Potato Late Blight. Molecular Plant-Microbe Interactions *22*, 589–600.

Goodman, R.E., Vieths, S., Sampson, H.A., Hill, D., Ebisawa, M., Taylor, S.L., Ree, R. Van, and van Ree, R. (2008). Allergenicity Assessment of Genetically Modified Crops--What Makes Sense? Nature Biotechnology *26*, 73–81.

Ladics, G.S. (2008). Current Codex Guidelines for Assessment of Potential Protein Allergenicity. Food and Chemical Toxicology *46*, S20-3.

Ladics, G.S., and Selgrade, M.K. (2009). Identifying Food Proteins with Allergenic Potential: Evolution of Approaches to Safety Assessment and Research to Provide Additional Tools. Regulatory Toxicology and Pharmacology *54*, S2-6.

Ladics, G.S., Bannon, G.A., Silvanovich, A., and Cressman, R.F. (2007). Comparison of Conventional FASTA Identity Searches with the 80 Amino Acid Sliding Window FASTA Search for the Elucidation of Potential Identities to Known Allergens. Molecular Nutrition and Food Research *51*, 985–998.

Lorang, J.M., Sweat, T.A., and Wolpert, T.J. (2007). Plant Disease Susceptibility Conferred by a "Resistance" gene. PNAS *104*, 14861–14866.

Nagy, E.D., and Bennetzen, J.L. (2008). Pathogen Corruption and Site-Directed Recombination at a Plant Disease Resistance Gene Cluster. Genome Research 1918–1923.

Pearson, W.R., and Lipman, D.J. (1988). Improved Tools for Biological Sequence Comparison. PNAS *85*, 2444–2448.

Terrat, Y., and Ducancel, F. (2013). Are There Unequivocal Criteria to Label a given Protein as a Toxin? Permissive versus Conservative Annotation Processes. Genome Biology *14*, 406.

Walton, J.D. (1996). Host-Selective Toxins: Agents of Compatibility. The Plant Cell *8*, 1723–1733.

APPENDIX A. ORFs in Z6

[CCI]