Event SYHT0H2 Soybean

Basic Local Alignment Search Tool for Nucleotides (BLASTN) Analysis of Soybean Genomic Sequences Flanking the Insert Using the Nucleotide Collection Database and a Plant Expressed Sequence Tags (EST) Database

Assessment

DATA REQUIREMENTS: Not Applicable

AUTHOR: [redacted]

COMPLETION DATE: April 23, 2012

PERFORMING LABORATORY: Syngenta Crop Protection, LLC
3054 East Cornwallis Road
Research Triangle Park, NC 27709-2257 USA

LABORATORY PROJECT ID: Report Number: TK0059667
Task Number: TK0059667

SUBMITTER:
Syngenta Seeds, Inc.
3054 East Cornwallis Road
Post Office Box 12257
Research Triangle Park, NC 27709-2257 USA

SPONSOR:
Syngenta Crop Protection, LLC
410 Swing Road
Post Office Box 18300
Greensboro, NC 27419-8300 USA

VOLUME 1 OF 2 OF REPORT

STATEMENTS OF DATA CONFIDENTIALITY CLAIMS

The following statement applies to submissions to the United States Environmental Protection Agency (US EPA).

Statement of Data Confidentiality Claim

A claim of confidentiality is made for information contained in this report on the basis of its falling within the scope of Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Section 10 (d) (1) (A), (B), or (C). The information claimed confidential has been removed to a confidential volume and is cited by cross-reference number in the body of this volume.

Company: *Syngenta Seeds, Inc.*

Company Representative:

[redacted] April 19, 2012

Date

Manager, Regulatory Affairs

These data are the property of Syngenta Seeds, Inc. and, as such, are considered to be confidential for all purposes other than compliance with the regulations implementing FIFRA Section 10. Submission of these data in compliance with FIFRA does not constitute a waiver of any right to confidentiality that may exist under any other provision of common law or statute or in any other country.

The following statement applies to submissions to regulatory agencies and other competent authorities other than the US EPA and all other viewers.

This Document Contains Confidential Business Information

This document contains information that is proprietary to Syngenta and, as such, is considered to be confidential for all purposes other than compliance with the relevant registration procedures.

Without the prior written consent of Syngenta, this information may (i) not be used by any third party including, but not limited to, any other regulatory authority for the support of regulatory approval of this product or any other product, and (ii) not be published or disclosed to any third party including, but not limited to, any authority for the support of regulatory approval of any products.

Its submission does not constitute a waiver of any right to confidentiality that may exist in any other country.

GOOD LABORATORY PRACTICE COMPLIANCE STATEMENT

The information provided in this volume is not subject to the US EPA Good Laboratory Practice Standards (GLPS) (40 CFR Part 160, US EPA 1989). Therefore, a GLPS Compliance Statement is not applicable.

Study Director:

[redacted] April 23, 2012

Date

Technical Expert, Product Safety
Syngenta Crop Protection, LLC

Submitted by:

[redacted] April 19, 2012

Date

Manager, Regulatory Affairs
Syngenta Seeds, Inc.
3054 East Cornwallis Road
Post Office Box 12257
Research Triangle Park, NC 27709-2257 USA

Sponsor:

[redacted] April 23, 2012

Date

Technical Leader, Product Safety
Syngenta Crop Protection, LLC
410 Swing Road
Post Office Box 18300
Greensboro, NC 27419-8300 USA

GENERAL INFORMATION

Contributors

The following contributed to this report in the capacities indicated:

Name	Title
[redacted]	Study Director, Syngenta Crop Protection, LLC

Records Retention

Raw data, the original copy of this report, and other relevant records are archived at Syngenta, 3054 East Cornwallis Road, Research Triangle Park, NC 27709-2257 USA.

TABLE OF CONTENTS

STATEMENTS OF DATA CONFIDENTIALITY CLAIMS **2**

GOOD LABORATORY PRACTICE COMPLIANCE STATEMENT **3**

GENERAL INFORMATION **4**

TABLE OF CONTENTS **5**

LIST OF FIGURES **7**

LIST OF ACRONYMS AND ABBREVIATIONS **8**

1.0 EXECUTIVE SUMMARY **9**

2.0 INTRODUCTION **10**

3.0 MATERIALS AND METHODS **10**

3.1 Database Descriptions 11

3.2 BLASTN Parameters 11

3.3 Control of Bias 12

3.4 Statistical Analysis 12

4.0 RESULTS AND DISCUSSION **12**

4.1 BLASTN Analyses Using the NCBI nr/nt Database 12

4.2 BLASTN Analysis Using the PlantGDB Viridiplantae EST Database 13

4.3 Data Quality and Integrity 13

5.0 CONCLUSION **13**

6.0 REFERENCES **14**

APPENDICES SECTION **15**

APPENDIX A Top 10 results of BLASTN analysis of 1000 bp of DNA sequence flanking the 5' region of the SYHT0H2 insert against the NCBI nr/nt database 16

APPENDIX B Alignments corresponding to the top 10 results of BLASTN analysis of 1000 bp of DNA sequence flanking the 5' region of the SYHT0H2 insert against the NCBI nr/nt database 17

APPENDIX C Top 10 results of BLASTN analysis of 1000 bp of DNA sequence flanking the 3' region of the SYHT0H2 insert against the NCBI nr/nt database 18

APPENDIX D Alignments corresponding to the top 10 results of BLASTN analysis of 1000 bp of DNA sequence flanking the 3' region of the SYHT0H2 insert against the NCBI nr/nt database 19

APPENDIX E Top 10 results of BLASTN analysis of 1000 bp of DNA sequence flanking the 5' region of the SYHT0H2 insert against the PlantGDB Viridiplantae EST database 20

APPENDIX F Alignments corresponding to the top 10 results of BLASTN analysis of 1000 bp of DNA sequence flanking the 5' region of the SYHT0H2 insert against the PlantGDB Viridiplantae EST database 21
APPENDIX G Top 10 results of BLASTN analysis of 1000 bp of DNA sequence flanking the 3' region of the SYHT0H2 insert against the PlantGDB Viridiplantae EST database 22
APPENDIX H Alignments corresponding to the top 10 results of BLASTN analysis of 1000 bp of DNA sequence flanking the 3' region of the SYHT0H2 insert against the PlantGDB Viridiplantae EST database 23

LIST OF FIGURES

FIGURE 1 Soybean genomic sequence (1000 bp) flanking the 5' region of the SYHT0H2 insert 10
FIGURE 2 Soybean genomic sequence (1000 bp) flanking the 3' region of the SYHT0H2 insert 10

LIST OF ACRONYMS AND ABBREVIATIONS

AvHPPD-03	*p*-hydroxyphenylpyruvate dioxygenase enzyme encoded by *avhppd-03*
avhppd-03	*p*-hydroxyphenylpyruvate dioxygenase gene derived from oat
BLASTN	Basic Local Alignment Search Tool for Nucleotides
bp	base pair
CBI	confidential business information
DNA	deoxyribonucleic acid
E-value	expectation or expect value
EST	expressed sequence tags
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
GenBank	National Institutes of Health genetic sequence database
GLPS	Good Laboratory Practice Standards
HPPD	*p*-hydroxyphenylpyruvate dioxygenase enzyme
mRNAs	messenger ribonucleic acids
NCBI	National Center for Biotechnology Information
nr/nt	non-redundant nucleotide
PAT	phosphinothricin acetyltransferase enzyme
pat	phosphinothricin acetyltransferase gene
RefSeq	NCBI Reference Sequence Collection representing DNA, RNA, and protein
RNAs	ribonucleic acids
US EPA	United States Environmental Protection Agency

1.0 EXECUTIVE SUMMARY

Soybean (*Glycine max* [L.] Merrill) has been genetically modified to express the novel genes *avhppd-03* derived from oat (*Avena sativa* L.) and *pat* from *Streptomyces viridochromogenes*. The gene *avhppd-03* encodes a *p*-hydroxyphenylpyruvate dioxygenase (HPPD) enzyme, designated AvHPPD-03, that catalyzes the formation of homogentisic acid, the aromatic precursor in plastoquinone and vitamin E biosynthesis. In comparison with the native soybean HPPD, the AvHPPD-03 isozyme from oat has lower binding affinity for mesotrione, an herbicide that inhibits HPPD. Expression of *avhppd-03* in transgenic Event SYHT0H2 soybean plants confers a mesotrione-tolerance phenotype. The gene *pat* encodes the enzyme phosphinothricin acetyltransferase (PAT), which inactivates the herbicide glufosinate, an inhibitor of glutamine synthetase, an enzyme in the nitrogen assimilation pathway. Expression of *pat* confers a glufosinate-tolerance phenotype, which was used as a selectable marker in the development of SYHT0H2 soybeans.

The purpose of this study was to ascertain whether the SYHT0H2 insert disrupts any known endogenous soybean gene. The Basic Local Alignment Search Tool for Nucleotides (BLASTN) was used to screen the soybean genomic sequences flanking the SYHT0H2 insert for similarity to sequences in a non-redundant nucleotide database and an expressed sequence tags database. The results of these analyses indicate that the SYHT0H2 insert does not disrupt any known endogenous soybean gene.

2.0 INTRODUCTION

Soybean (*Glycine max* [L.] Merrill) has been genetically modified to express the novel genes *avhppd-03* derived from oat (*Avena sativa* L.) and *pat* from *Streptomyces viridochromogenes*. The gene *avhppd-03* encodes a *p*-hydroxyphenylpyruvate dioxygenase (HPPD) enzyme, designated AvHPPD-03, that catalyzes the formation of homogentisic acid, the aromatic precursor in plastoquinone and vitamin E biosynthesis. In comparison with the native soybean HPPD, the AvHPPD-03 isozyme from oat has lower binding affinity for mesotrione, an herbicide that inhibits HPPD. Expression of *avhppd-03* in transgenic Event SYHT0H2 soybean plants confers a mesotrione-tolerance phenotype. The gene *pat* encodes the enzyme phosphinothricin acetyltransferase (PAT), which inactivates the herbicide glufosinate, an inhibitor of glutamine synthetase, an enzyme in the nitrogen assimilation pathway. Expression of *pat* confers a glufosinate-tolerance phenotype, which was used as a selectable marker in the development of SYHT0H2 soybeans.

The purpose of this study was to ascertain whether the SYHT0H2 insert disrupts any known endogenous soybean gene. The Basic Local Alignment Search Tool for Nucleotides (BLASTN) was used to screen the soybean genomic sequences flanking the SYHT0H2 insert for similarity to sequences in a non-redundant nucleotide (nr/nt) database and an expressed sequence tags (EST) database.

3.0 MATERIALS AND METHODS

A previous study determined the 1000 base pair (bp) soybean genomic sequences that flank the SYHT0H2 insert (de Framond 2012). These sequences (shown in Figures 1 and 2) were screened for similarity with deoxyribonucleic acid (DNA) sequences found in public databases. Sequence similarity analyses were performed with the BLASTN program, version 2.2.19 (Altschul *et al.* 1997). The sequences were compared with the DNA sequences in the latest version of the National Center for Biotechnology Information (NCBI) nr/nt database (NCBI 2012) and with sequences in the latest version of the PlantGBD Viridiplantae EST database (PlantGDB 2012).

FIGURE 1 Soybean genomic sequence (1000 bp) flanking the 5' region of the SYHT0H2 insert

{Volume 2: Confidential Business Information (CBI) Cross-reference Number 1}

FIGURE 2 Soybean genomic sequence (1000 bp) flanking the 3' region of the SYHT0H2 insert

{Volume 2: CBI Cross-reference Number 2}

3.1 Database Descriptions

The nr/nt database contains all nucleotide sequences from the National Institutes of Health genetic sequence database (GenBank®), the NCBI Reference Sequence Collection (RefSeq), the European Molecular Biology Laboratory, and the DNA Database of Japan, together with nucleotide sequences derived from the three-dimensional structures described in the Brookhaven Protein Data Bank. GenBank® is an archival repository of all sequences, whereas the RefSeq database is a non-redundant set of reference standards that includes chromosomes, complete genomic molecules (organelle genomes, viruses, and plasmids), intermediate assembled genomic contigs, curated genomic regions, ribonucleic acids (RNAs), messenger RNAs (mRNAs), and proteins. The nr/nt database does not contain any EST sequences, sequence tagged sites, genome sequence survey sequences, or phase 0, 1, or 2 high-throughput genome sequences. The nr/nt database was updated on March 14, 2012 and contained over 15 million unique sequences.

The PlantGBD Viridiplantae EST database contains the GenBank® EST sequences for the Viridiplantae taxon (green plants), generated from sequences extracted from the NCBI EST database (Boguski *et al*. 1993) and classified as belonging to the Viridiplantae taxon through parsing conducted by PlantGDB (Duvick *et al.* 2008). The PlantGBD Viridiplantae EST database was updated on January 26, 2012 and contained over 21 million sequences.

3.2 BLASTN Parameters

The following parameters were used for the BLASTN analyses:

- Expect = 10. The expectation value (*E*-value) is a measure of the probability that matches between sequences occurred by chance. Search results involving comparisons between nucleotides with highly similar sequences yield *E*-values approaching zero; the probability that sequence similarities occurred by chance increases with higher *E*-values (Ponting 2001). The search identified all sequences in the database with search results yielding an *E*-value of 10 or lower.
- Match/mismatch scoring = default for nucleotide searches: +1 for a match and –3 for a mismatch.
- Gap costs: existence = 5; extension = 2. A gap is a space introduced into an alignment to compensate for insertions or deletions in one sequence relative to another. The introduction of a gap causes the deduction of a fixed value from the alignment score to prevent the accumulation of excessive gaps in an alignment. Extension of the gap to encompass additional nucleotides is also penalized in determining the score of an alignment. The resultant score is derived from the number of identical matches between the query sequence and the database entry, with higher scores indicating greater similarity between the two sequences.
- Filter = low complexity. The low complexity filter masks regions of low compositional complexity that could cause spurious or misleading results.

3.3 Control of Bias

Any rejected data, and the documented reasons for the rejection of those data, are retained in the study file.

3.4 Statistical Analysis

No statistical analysis was conducted during this study other than the BLASTN analyses.

4.0 RESULTS AND DISCUSSION

4.1 BLASTN Analyses Using the NCBI nr/nt Database

The BLASTN analyses using the NCBI nr/nt database resulted in numerous alignments for the soybean genomic sequences flanking the SYHT0H2 insert. Alignments to known endogenous soybean genes would be expected to have the lowest *E*-values in the BLASTN analyses. The ten results with the lowest *E*-values (and therefore the greatest sequence similarity) are listed in Appendix A for the 5′ flanking sequence and Appendix C for the 3′ flanking sequence, and the corresponding alignments are shown in Appendices B and D, respectively. All alignments were examined, and none of them provided evidence that a known soybean gene was interrupted by the SYHT0H2 insert. Analysis of the remaining alignments with higher *E*-values did not change the interpretation of the data; for this reason, remaining alignments were not included in this report.

In BLASTN analysis of the soybean genomic sequence flanking the 5′ region of the insert, two of the ten alignments with the highest *E*-values were to soybean sequences. The two alignments (*E*-value = 2.1) were to unannotated soybean genomic sequence (database accession numbers AC235173.1 and EF533702.1). Alignments with the lowest *E*-values (0.13) were nucleotide sequence from *Oryza sativa* (rice) (database accession numbers AC136229.3 and AC130609.2) and nucleotide sequence from *Mus musculus* (mouse), (database accession number AC093445.4). The five remaining sequences align to unannotated genomic sequences from pig, rat, human, and mouse.

In BLASTN analysis of the soybean genomic sequence flanking the 3′ region of the insert, the ten lowest *E*-values were for alignments to genomic sequences from *G. max* and *Glycine soja* (wild soybean) (*E*-value = 1×10^{-66} to 3×10^{-64}). One alignment (*E*-value = 5×10^{-66}) was to intron sequence from a O-acetylserine(thiol)lyase-encoding gene in *G. soja* (database accession number EF535995.1) (Zhang *et al.* 2008). Another alignment (*E*-value = 7×10^{-65}) was to sequence predicted to be a protein complex responsible for biogenesis of lysosome-related organelles (database accession number XM_003544873.1) (Falcón-Pérez *et al.* 2002). The remaining eight of the top ten alignments were to unannotated genomic sequence from the *G. max.* Most of the alignments analyzed were to the same region of the 3′ genomic sequence flanking the insert, suggesting that the sequence in this region is repetitive. None of the alignments to soybean genomic sequences were immediately adjacent to the genome to insert junction. Furthermore, none of these alignments corresponded to genes for which alignments were found in the 5' flanking sequence.

4.2 BLASTN Analysis Using the PlantGDB Viridiplantae EST Database

BLASTN analyses using the PlantGDB Viridiplantae EST database resulted in numerous alignments for the soybean genomic sequences flanking the SYHT0H2 insert. The ten results with the lowest *E*-values are listed in Appendix E for the 5' flanking sequence and Appendix G for the 3' flanking sequence, and the corresponding alignments are shown in Appendices F and H, respectively. All alignments were examined, and none of them provided evidence that a known soybean gene was interrupted by the SYHT0H2 insert. Analysis of the remaining alignments with higher *E*-values did not change the interpretation of the data; for this reason, remaining alignments were not included in this report.

In BLASTN analysis of the soybean genomic sequence flanking the 5′ region of the insert, none of the alignments were to soybean sequences. The lowest *E*-value was for mRNA sequence from *Zingiber officinale* (ginger) (0.009). The remaining top alignments (*E*-value = 2.3) were to mRNA sequences from citrimelo, cowpea, and meadow fescue.

In BLASTN analysis of the soybean genomic sequence flanking the 3′ region of the insert, all of the ten lowest *E*-values were for alignments to soybean mRNA sequences. All alignments were to the same region of the 3′ genomic sequence flanking the insert, suggesting that the sequence in this region is repetitive. It has been noted in literature that gene prediction programs erroneously label repetitive regions as genes (Bennetzen *et al*. 2004) resulting in EST databases containing repetitive regions that have been incorrectly labeled as genes. None of the alignments to soybean mRNA sequences were immediately adjacent to the genome to insert junction, and no corresponding alignments were found in the 5′ flanking sequence.

4.3 Data Quality and Integrity

No circumstances occurred during the conduct of this analysis that would have adversely affected the quality or integrity of the data generated.

5.0 CONCLUSION

The results of BLASTN analyses of the genomic sequences flanking the SYHT0H2 insert indicate that the insert does not disrupt any known endogenous soybean gene.

6.0 REFERENCES

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.

Bennetzen JL, Coleman C, Liu R, Jianxin M, Ramakrishna W. 2004. Consistent over-estimation of gene number in complex plant genomes. *Curr Opin Plant Biol* 7:732–736.

Boguski MS, Lowe TMJ, Tolstoshev CM. 1993. dbEST – database for "expressed sequence tags." *Nat Genet* 4:332–333.

de Framond A. 2012. *Event SYHT0H2 Soybean: Flanking Sequence Determination*. Study No. TK0059646 (unpublished). Research Triangle Park, NC: Syngenta Crop Protection, LLC.

Duvick J, Fu A, Muppirala U, Sabharwal M, Wilkerson MD, Lawrence CJ, Lushbough C, Brendel V. 2008. PlantGDB: a resource for comparative plant genomics. *Nucleic Acids Res* 36:D959–D965.

Falcón-Pérez JM, Starcevic M, Gautam M, Dell'Angelica EC. 2002. BLOC-1, a novel complex containing the pallidin and muted proteins involved in the biogenesis of melanosomes and platelet-dense granules. *J Biol Chem* 277:28191–28199.

NCBI. 2012. NCBI Nucleotide Database. Bethesda, MD: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health. http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide (updated March 14, 2012).

PlantGDB. 2012. PlantGDB EST Assemblies. http://www.plantgdb.org/prj/ESTCluster (updated January 26, 2012).

Ponting C. 2001. Issues in predicting protein function from sequence. *Brief Bioinform* 2:19–29.

US EPA. 1989. Good Laboratory Practice Standards. 40 CFR Part 160.

Zhang C, Meng Q, Gai J, Yu D. 2008. Cloning and functional characterization of an O-acetylserine(thiol)lyase-encoding gene in wild soybean (*Glycine soja*). *Mol Biol Rep* 35:527–534.

APPENDICES SECTION

APPENDIX A Top 10 results of BLASTN analysis of 1000 bp of DNA sequence flanking the 5' region of the SYHT0H2 insert against the NCBI nr/nt database

Accession Number	Description	*E*-value	Score (bits)
AC136229.3	*Oryza sativa* (japonica cultivar-group) chromosome 5 PAC clone P0690B12, complete sequence	0.13	48.1
AC130609.2	*Oryza sativa* Japonica Group chromosome 5 clone P0048F12, complete sequence	0.13	48.1
AC093445.4	*Mus musculus* strain C57BL/6J chromosome 1 clone rp23-267j18, complete sequence	0.13	48.1
FP102622.4	Pig DNA sequence from clone CH242-210G17 on chromosome X, complete sequence	0.52	46.1
AC242413.4	*Rattus norvegicus* Y Chr BAC RNECO-2P24 (Amplicon Express Rat SHR-Akr (*Eco*R1 Digest) BAC library) complete sequence	0.52	46.1
AC240974.3	*Rattus norvegicus* Y Chr BAC RNAEX-276G09 (Amplicon Express Rat SHR-Akr BAC library) complete sequence	0.52	46.1
AJ334050.1	*Homo sapiens* genomic sequence surrounding NotI site, clone HSJ6-AN5C	0.52	46.1
AC235173.1	*Glycine max* strain Williams 82 clone GM_WBb0002O19, complete sequence	2.1	44.1
EF533702.1	*Glycine max* clone BAC GM_WBb098N11, complete sequence	2.1	44.1
AC102315.9	*Mus musculus* chromosome 3, clone RP24-441C20, complete sequence	2.1	44.1

APPENDIX B Alignments corresponding to the top 10 results of BLASTN analysis of 1000 bp of DNA sequence flanking the 5' region of the SYHT0H2 insert against the NCBI nr/nt database

{Volume 2: CBI Cross-Reference Number 3}

APPENDIX C Top 10 results of BLASTN analysis of 1000 bp of DNA sequence flanking the 3' region of the SYHT0H2 insert against the NCBI nr/nt database

Accession Number	Description	*E*-value	Score (bits)
AC235423.1	*Glycine max* strain Williams 82 clone GM_WBb0137C18, complete sequence	1.00E-66	264
EF535995.1	*Glycine soja* O-acetylserine(thiol)lyase gene, complete cds	5.00E-66	262
AC235257.1	*Glycine max* strain Williams 82 clone GM_WBb0038A08, complete sequence	2.00E-65	260
EF533695.1	*Glycine max* clone BAC GM_WBb098N15, complete sequence	2.00E-65	260
XM_003544873.1	PREDICTED: *Glycine max* biogenesis of lysosome-related organelles complex 1 subunit 1-like (LOC100811869), mRNA	7.00E-65	258
AP011822.1	*Glycine max* DNA, BAC clone: WBb35C13, complete sequence	3.00E-64	256
AP011811.1	*Glycine max* DNA, BAC clone: MiB300H01, complete sequence	3.00E-64	256
AC235696.2	*Glycine max* clone GM_WBb0070F18, complete sequence	3.00E-64	256
AC235301.1	*Glycine max* strain Williams 82 clone GM_WBb0063F13, complete sequence	3.00E-64	256
AC235288.1	*Glycine max* strain Williams 82 clone GM_WBb0056C20, complete sequence	3.00E-64	256

APPENDIX D Alignments corresponding to the top 10 results of BLASTN analysis of 1000 bp of DNA sequence flanking the 3' region of the SYHT0H2 insert against the NCBI nr/nt database

{Volume 2: CBI Cross-Reference Number 4}

APPENDIX E Top 10 results of BLASTN analysis of 1000 bp of DNA sequence flanking the 5' region of the SYHT0H2 insert against the PlantGDB Viridiplantae EST database

Accession Number	Description	*E*-value	Score (bits)
DY367775.1	EST ZO__Ee0008I21 f ZO__Ee *Zingiber officinale* cDNA clone ZO__Ee0008I21 5', mRNA sequence	0.009	50.1
CX669619.1	EST UCRCP01_046_A11_T7 *Swingle citrumelo* nematode-challenged root cDNA library - UCRCP01 Citrus x paradisi x Citrus trifoliata cDNA clone UCRCP01_046_T7_A11, mRNA sequence	2.3	42.1
CX669618.1	EST UCRCP01_046_A11_T3 *Swingle citrumelo* nematode-challenged root cDNA library - UCRCP01 Citrus x paradisi x Citrus trifoliata cDNA clone UCRCP01_046_T3_A11, mRNA sequence	2.3	42.1
CX669467.1	EST UCRCP01_045_B02_T7 *Swingle citrumelo* nematode-challenged root cDNA library - UCRCP01 Citrus x paradisi x Citrus trifoliata cDNA clone UCRCP01_045_T7_B02, mRNA sequence	2.3	42.1
CX669466.1	EST UCRCP01_045_B02_T3 *Swingle citrumelo* nematode-challenged root cDNA library - UCRCP01 Citrus x paradisi x Citrus trifoliata cDNA clone UCRCP01_045_T3_B02, mRNA sequence	2.3	42.1
CX665397.1	EST UCRCP01_016_C03_T7 *Swingle citrumelo* nematode-challenged root cDNA library - UCRCP01 Citrus x paradisi x Citrus trifoliata cDNA clone UCRCP01_016_T7_C03, mRNA sequence	2.3	42.1
FF553918.1	EST MOPFF09TF MOP *Vigna unguiculata* cDNA 5', mRNA sequence	2.3	42.1
FF549311.1	EST MOPFF09TR MOP *Vigna unguiculata* cDNA 3', mRNA sequence	2.3	42.1
GO896788.1	EST RDD_cLiFpEfI12nb_9x1_4_H18_cLibkit5_D09 RDD_cLiFpEfI12nb_plant *Festuca pratensis* cDNA clone 9H18 5', mRNA sequence	2.3	42.1
GO885113.1	EST CLS_CLiFpEfSpn_95a1_1_k21cLibkit5LD_F11 CLS_CLiFpEfSPn_plant *Festuca pratensis* cDNA clone 95k21 5', mRNA sequence	2.3	42.1

APPENDIX F Alignments corresponding to the top 10 results of BLASTN analysis of 1000 bp of DNA sequence flanking the 5' region of the SYHT0H2 insert against the PlantGDB Viridiplantae EST database

{Volume 2: CBI Cross-Reference Number 5}

APPENDIX G Top 10 results of BLASTN analysis of 1000 bp of DNA sequence flanking the 3' region of the SYHT0H2 insert against the PlantGDB Viridiplantae EST database

Accession Number	Description	*E*-value	Score (bits)
BG550747.1	EST sad23e01.y1 Gm-c1074 *Glycine max* cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1074-986 5', mRNA sequence	2.00E-68	268
FK012993.1	EST GLQAO36TF JCVI-SOY2 *Glycine max* cDNA 5', mRNA sequence	3.00E-67	264
BI969156.1	EST GM830007B10C06 Gm-r1083 *Glycine max* cDNA clone Gm-r1083-2435 3', mRNA sequence	2.00E-65	258
BI426794.1	EST sag07e12.y1 Gm-c1080 *Glycine max* cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1080-599 5', mRNA sequence	2.00E-65	258
AW317170.1	EST sf38b10.x1 Gm-c1028 *Glycine max* cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1028-2228 3' similar to SW:GC5L_ARATH O22929 GCN5-LIKE PROTEIN 1, mRNA sequence	2.00E-65	258
BI701458.1	EST sag47a11.y1 Gm-c1081 *Glycine max* cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1081-1941 5', mRNA sequence	2.00E-65	258
CO978999.1	EST GM89003A2G03.r1 Gm-r1089 *Glycine max* cDNA clone Gm-r1089-1062 3, mRNA sequence	8.00E-65	256
BI470345.1	EST sah90g06.y1 Gm-c1050 *Glycine max* cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1050-3395 5', mRNA sequence	1.00E-63	252
EH039713.1	EST GMSeed0535 Soybean endosperm tissue in developing seeds *Glycine max* cDNA clone HYI28.xg_B01_013, mRNA sequence	1.00E-63	252
CA852475.1	EST E08B11_D11_04.ab1 cDNA Peking library 6, 8 day SCN3 *Glycine max* cDNA clone E08B11 5', mRNA sequence	5.00E-63	250

APPENDIX H Alignments corresponding to the top 10 results of BLASTN analysis of 1000 bp of DNA sequence flanking the 3' region of the SYHT0H2 insert against the PlantGDB Viridiplantae EST database

{Volume 2: CBI Cross-Reference Number 6}