

Event 5307 Maize:

**Genome to Insert Junction Analysis for Translated Open Reading Frames With
a Minimum Size of 30 Amino Acids: Assessment of Amino Acid Sequence
Similarity to Known or Putative Toxins**

Data Requirement:	Not applicable
Author:	Stephen New
Study Completion Date:	June 23, 2010
Performing Laboratory:	Syngenta Biotechnology, Inc. Regulatory Science and Trait Safety 3054 East Cornwallis Road PO Box 12257 Research Triangle Park, NC 27709-2257, USA
Syngenta Study No.:	Not applicable
Report No.:	SSB-187-10

STATEMENTS OF DATA CONFIDENTIALITY CLAIMS

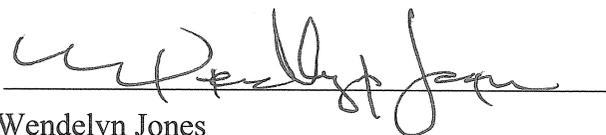
The following statement applies to submissions to the United States Environmental Protection Agency (US EPA).

Statement of No Data Confidentiality Claim

No claim of confidentiality is made for any 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).

Company: *Syngenta Seeds, Inc.*

Company Representative:



Wendelyn Jones
Regulatory Affairs Manager

6/23/10
Date

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any other regulatory authority for the support of regulatory approval of this product or any other product, and (ii) not be published or disclosed to any third party including, but not limited to, any authority for the support of regulatory approval of any products.

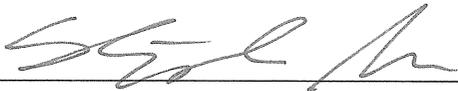
Its submission does not constitute a waiver of any right to confidentiality that may exist in any other country.

© 2010. Syngenta. All Rights Reserved.

STATEMENT CONCERNING GOOD LABORATORY PRACTICES STANDARDS

The provisions of Good Laboratory Practices Standards (GLPS) (40 CFR Part 160, US EPA 1989) pursuant to the Federal Insecticide, Fungicide, and Rodenticide Act do not apply to this report because it summarizes the analyses of peptide sequence data. These analyses were performed according to accepted scientific practices, and relevant study records (including raw data) have been retained.

Study Director:



Stephen New
Regulatory Scientist, Molecular Analysis Team
Regulatory Science and Trait Safety
Syngenta Biotechnology, Inc.

June 23, 2010

Date

Submitted by:



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

6/23/10

Date

Sponsor:



Hope Hart
Team Leader, Molecular Analysis Team
Regulatory Science and Trait Safety
Syngenta Biotechnology, Inc.

June 23, 2010

Date

TABLE OF CONTENTS

STATEMENTS OF DATA CONFIDENTIALITY CLAIMS2

STATEMENT CONCERNING GOOD LABORATORY PRACTICES STANDARDS3

LIST OF FIGURES.....5

LIST OF TABLES5

LIST OF ACRONYMS AND ABBREVIATIONS6

SUMMARY7

INTRODUCTION.....8

MATERIALS AND METHODS8

ORF Determination..... 8

Sequence Similarity Search 8

RESULTS9

Data Quality and Integrity 11

CONCLUSION 11

RECORDS RETENTION..... 12

CONTRIBUTING SCIENTISTS 12

REFERENCES..... 13

Published 13

Unpublished..... 13

LIST OF FIGURES

Figure 1. Location of the 5307 maize putative ORF	10
Figure 2. Alignment of the translated 5307 maize putative ORF with NP_053411.1	10

LIST OF TABLES

Table 1. Putative ORFs spanning either junction of the maize genomic DNA and the 5307 maize insert.....	9
Table 2. Description of alignment associated with the 5307 maize putative ORF	10

LIST OF ACRONYMS AND ABBREVIATIONS

3'	three prime
5'	five prime
BLASTP	Basic Local Alignment Search Tool for Proteins
BLOSUM62	Blocks Substitution Matrix62
Cry1Ab	Cry1Ab protein
Cry3A	Cry3A protein
DNA	deoxyribonucleic acid
<i>ecry3.1Ab</i>	eCry3.1Ab gene
eCry3.1Ab	eCry3.1Ab protein
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
GLPS	Good Laboratory Practices Standards
<i>manA</i>	phosphomannose isomerase gene
NCBI	National Center for Biotechnology Information
ORF	open reading frame
<i>pmi</i>	phosphomannose isomerase gene
PMI	phosphomannose isomerase protein
US EPA	United States Environmental Protection Agency
®	registered trademark
™	trademark

SUMMARY

Maize plants derived from transformation Event 5307 (“5307 maize”) contain the gene *ecry3.1Ab* encoding an eCry3.1Ab protein and the gene *pmi* (also known as *manA*) encoding the enzyme phosphomannose isomerase (PMI).

Bioinformatic analysis of the deoxyribonucleic acid (DNA) sequences spanning the junctions between the maize genomic sequence and the 5307 maize insert identified one putative open reading frame (ORF) with a minimum putative translation size of 30 amino acids. For this analysis, ORFs were defined as any DNA sequence between a putative start codon (ATG) and a putative stop codon (TAG, TAA, or TGA). The Basic Local Alignment Search Tool for Proteins (BLASTP) program was used to search the National Center for Biotechnology Information (NCBI) Entrez® Protein Database to determine whether the translations of this putative ORF showed significant similarity to known and putative toxins. The results of this analysis suggest that the amino acid translation of the putative ORF spanning the junction between the maize genomic sequence and the 3' region of the 5307 maize insert shows no sequence similarity to any known or putative toxins.

INTRODUCTION

Using the techniques of modern molecular biology, Syngenta has transformed maize (*Zea mays*) to produce Event 5307 maize, a new cultivar that has insecticidal activity against certain corn rootworm (*Diabrotica*) species. Maize plants derived from transformation Event 5307 ("5307 maize") contain the gene *ecry3.1Ab* encoding an eCry3.1Ab protein and the gene *pmi* (also known as *manA*) encoding the enzyme phosphomannose isomerase (PMI). The eCry3.1Ab protein is an engineered chimera of modified Cry3A (mCry3A) and Cry1Ab proteins. The gene *pmi* was obtained from *Escherichia coli* strain K-12 and the protein it encodes was utilized as a plant selectable marker during development of 5307 maize.

Bioinformatic analysis was used to determine any putative open reading frames (ORFs) that span either junction between the maize genomic sequence and the 5307 maize insert. For this analysis, ORFs were defined as any deoxyribonucleic acid (DNA) sequence between a putative start codon (ATG) and a putative stop codon (TAG, TAA, or TGA) with a minimum putative translation size of 30 amino acids. The purpose of this analysis is to determine whether the amino acid translations of any putative ORFs spanning either junction of the maize genomic DNA and the 5307 maize insert have similarity with proteins that are known or putative toxins.

The translation of any putative ORFs spanning either junction of the maize genomic sequence and the 5307 maize insert was systematically compared with the latest posting of the National Center for Biotechnology Information (NCBI) Entrez® Protein Database (NCBI 2010). This procedure identified (1) whether any translated putative ORFs showed similarity to sequences in the NCBI Entrez® Protein Database (*i.e.*, alignments with Basic Local Alignment Search Tool for Proteins [BLASTP] Expectation values [*E*-values] of ten or lower), and (2) whether any proteins showing sequence similarity to any of the translated putative ORFs were known or putative toxins.

MATERIALS AND METHODS

ORF Determination

The insert and flanking sequences for 5307 maize were previously reported (New 2010a, New 2010b). The Vector NTI Advance™ program, version 10.3.0, was used to determine putative ORFs that span the junctions between the maize genomic sequence and the 5307 maize insert. For this analysis, putative ORFs were defined as DNA sequences that are contained between a putative start codon (ATG) and a putative stop codon (TAG, TAA, or TGA), and have a minimum translation size of 30 amino acids; these putative ORFs were determined and translated *in silico* into amino acid sequences. The translations of any putative ORFs spanning either junction of the maize genomic DNA and the 5307 maize insert were screened for sequence similarity to known or putative toxins.

Sequence Similarity Search

The BLASTP program (Altschul *et al.* 1997) was used to search the NCBI Entrez® Protein Database with translations of any putative ORFs spanning the junction between

the maize genomic sequence and the 5307 maize insert as the query sequences. The NCBI Entrez® Protein Database was used to identify the potential similarity between amino acid sequences of the query proteins and known and putative toxins. For this analysis, the NCBI Entrez® Protein Database (containing over 10 million sequences) was accessed on April 12, 2010. The BLASTP is a program used to search protein sequences for sequence similarities. Most proteins are modular in nature and contain repeating functional domains within the protein. Similarly, functional domains are conserved across different proteins from different species. The BLASTP algorithm is optimized to identify these domains or shorter sequence similarities present within the full length query sequence; as a result, this approach detects more similarities than would a search started by aligning two sequences over their entire length. The BLASTP program, version 2.2.19, was used to perform the similarity searches with the following default parameters:

- No complexity filter
- E*-value = 10
- Word size = 3
- Gap costs: existence = 11 and extension = 1
- Similarity matrix: Blocks Substitution Matrix62 (BLOSUM62)

The *E*-value is a measure of the probability that matches between sequences occurred by chance. Search results involving comparisons between proteins 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. These sequences were evaluated for source and biological function.

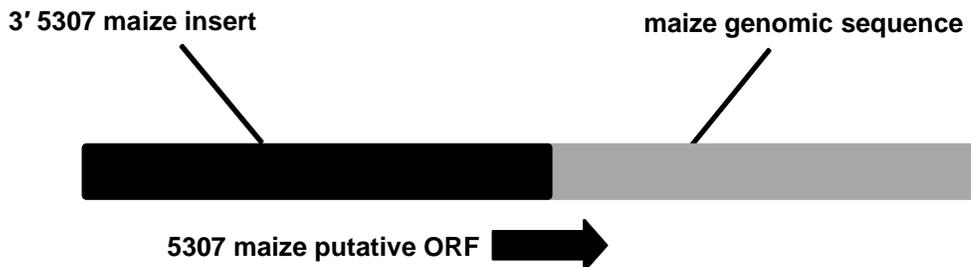
RESULTS

The reading frame analysis identified one putative ORF spanning the junction between the maize genomic sequence and the 3' region of the 5307 maize insert (5307 maize putative ORF). Table 1 shows the size of the putative ORF in base pairs, the translation of this sequence, and the size of the translation in amino acids. Figure 1 illustrates the location of the 5307 maize putative ORF.

Table 1. Putative ORFs spanning either junction of the maize genomic DNA and the 5307 maize insert

Sequence name (Figure 1)	Size of sequence analyzed (base pairs)	Size of putative translation (amino acid)	Putative translation
5307 maize putative ORF	243	81	MLLDLLALQEIYRCPPGGQHGRIRNVLL SCLSVNLFTPQYTLFPGPGWAHWQR VHRTVRCPKARNPSFCFVLFFQFGFC SNL

Figure 1. Location of the 5307 maize putative ORF



For the translated 5307 maize putative ORF, the BLASTP analysis identified one alignment (Table 2). This alignment was to a hypothetical protein from *Agrobacterium tumefaciens* (Entrez® Database Accession No. NP_053411.1 [NCBI 2010]) (Figure 2). The region of the translated 5307 maize putative ORF that has similarity with NP_053411.1 is located entirely within the 5307 maize insert. The region of similarity is to the left border sequence from *A. tumefaciens* and insert sequence adjacent to the left border. Because this portion of the translated junction sequence originates from sequence within the borders of pSYN12274, the 5307 maize transformation plasmid, it is not surprising to see similarity with a putative *A. tumefaciens* protein. The *E*-value for the alignment between this protein sequence and the translated 5307 maize putative ORF was 3.2.

The hypothetical protein identified is not a known or putative toxin.

Table 2. Description of alignment associated with the 5307 maize putative ORF

Accession number	Alignment description	Score	E-value
NP_053411.1	NR gi 10954991 ref NP_053411.1 hypothetical protein pTi-SAKURA_p173 [Agrobacterium tumefaciens]tiorf171 [Agrobacterium tumefaciens]	35.0	3.2

Figure 2. Alignment of the translated 5307 maize putative ORF with NP_053411.1

```
>NR|gi|10954991|ref|NP_053411.1| hypothetical protein
      pTi-SAKURA_p173 [Agrobacterium tumefaciens]
      tiorf171 [Agrobacterium tumefaciens]
      Length = 77
```

```
Score = 34.7 bits (78), Expect = 3.2, Method: Compositional matrix
adjust.
```

```
Identities = 16/20 (80%), Positives = 17/20 (85%)
```

```
Query: 25 VLLSCLSVNLFTPQYTLFPG 44
      +LLSCLSVNLFTPQY L P
Sbjct: 1 MLLSCLSVNLFTPQYILPPA 20
```

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.

CONCLUSION

One putative ORF spanning the junction between the maize genomic sequence and the 3' region of the 5307 maize insert was determined. The results of a comprehensive amino acid similarity search of the NCBI Entrez® Protein Database suggest that the translation of this putative ORF shows no similarity with any known or putative toxins.

RECORDS RETENTION

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

CONTRIBUTING SCIENTISTS

The analytical work reported herein was conducted by Stephen New, B.S. This work was conducted at Syngenta Biotechnology, Inc.

Reported by:  June 23, 2010
Date
Stephen New
Regulatory Scientist, Molecular Analysis Team
Regulatory Science and Trait Safety
Syngenta Biotechnology, Inc.

Approved by:  June 23, 2010
Date
Hope Hart
Team Leader, Molecular Analysis Team
Regulatory Science and Trait Safety
Syngenta Biotechnology, Inc.

REFERENCES**Published**

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.

NCBI. 2010. Entrez Protein 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=Protein> (accessed April 12, 2010).

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

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

Unpublished

New S. 2010a. *Event 5307 Maize: Flanking Sequence Determination*. Report No. SSB-160-10 (unpublished). Research Triangle Park, NC: Syngenta Biotechnology, Inc.

New S. 2010b. *Event 5307 Maize: Insert Sequence Analysis*. Report No. SSB-159-10 (unpublished). Research Triangle Park, NC: Syngenta Biotechnology, Inc.