B.T.

Industria y Comercio

SUPERINTENDENCIA

Delegatura de propiedad Industrial

División de Nuevas Creaciones

PCT

SOLICITUD

PATENTE DE INVENCIÓN

- PCT ENTRADA EN FASE NACIONAL -

21. **EXPEDIENTE No.** 06-1259

54. **TÍTULO** PROTEINAS INSECTICIDAS SECRETADAS DE BACILLUS THURINGIENSIS Y USOS DE LAS MISMAS

51. **CLASIFICACIÓN INTERNACIONAL** C12N15/32, 1/21, C07K14/325, C12P21/02, A01N63/02, A01H1/137/18

71. **SOLICITANTE** MONSANTO TECHNOLOGY, LLC

DOMICILIO 800 North Lindbergh Boulevard, E2NA - St. Louis, Missouri 63167 - ESTADOS UNIDOS DE AMERICA

74. **APODERADO** DRA. MARGARITA CASTELLANOS ABONDANO

22. **BOGOTÁ, D. C.,** 06 ENE. 2006

(FORMA P-10)

PETITORIO

Industria y Comercio
SUPERINTENDENCIA

SUPERINDUSTRIA Y COMERCIO
Radicacion : 06001259 00000000 Folios: 314
Fecha (AMD): 2006-01-06 15:32:18
Tramite : 011 PCT-PATENT 379 FASE NACI 411 PRESENTAC
Dependencia: 2020 DIVISION DE NUEVAS CREACIONES

FORMULARIO UNICO DE SOLICITUD DE PATENTE
PCT
ENTRADA EN FASE NACIONAL

07-02-2006

SOLICITUD DE:

①

[X] Patente de Invención. [ ] Patente de Modelo de Utilidad.

[ ] Capítulo I [X] Capítulo II

② **SOLICITANTE (71)**	Nombre **MONSANTO TECHNOLOGY, LLC** Dirección **800 North Lindbergh Boulevard E2NA - St. Louis, Missouri 63167, ESTADOS UNIDOS DE AMERICA** Nacionalidad o Domicilio **St. Louis, MO, ESTADOS UNIDOS DE AMERICA** Lugar de Constitución **ESTADOS UNIDOS DE AMERICA** Teléfono Fax E-mail	**IDENTIFICACIÓN** C C. [ ] NIT [ ] C E [ ] Otro [ ] Número Cual
③ **REPRESENTANTE O APODERADO (74)**	Nombre **MARGARITA CASTELLANOS A.** Dirección **CALLE 94A No. 7A-09** *Teléfono* **610 7420** *Fax* **610 0706** E-mail mail@alvarocastellanos.com	**IDENTIFICACIÓN** C C [X] NIT [ ] C E [ ] Otro [ ] Cual Numero **39'779.654** TPA **70 819**
④ **INVENTOR (ES) (72)**	Nombre: **VER ANEXO # 1** Dirección **VER ANEXO # 1** Nacionalidad o Domicilio **VER ANEXO # 1**	

⑤ **PCT (86)**

Solicitud Internacional No **PCT/US2004/021692** Fecha **06/07/2004**

Publicación Internacional No **WO 2005/019414** Fecha **03/03/2005**

⑥ **Título (54)** **PROTEINAS INSECTICIDAS SECRETADAS DE BACILLUS THURINGIENSIS Y USO DE LAS MISMAS**

⑦ **Clasificación Internacional (51)** C12N 15/32, 1/21, C07K 14/325, C12P/21/02, A01N63/00, A01H 37/18.

⑧ **Prioridad** SI [X] NO [ ]

(33) País de Origen	(32) Fecha	(31) Número de Solicitud
US	07/07/2003	60/485,483

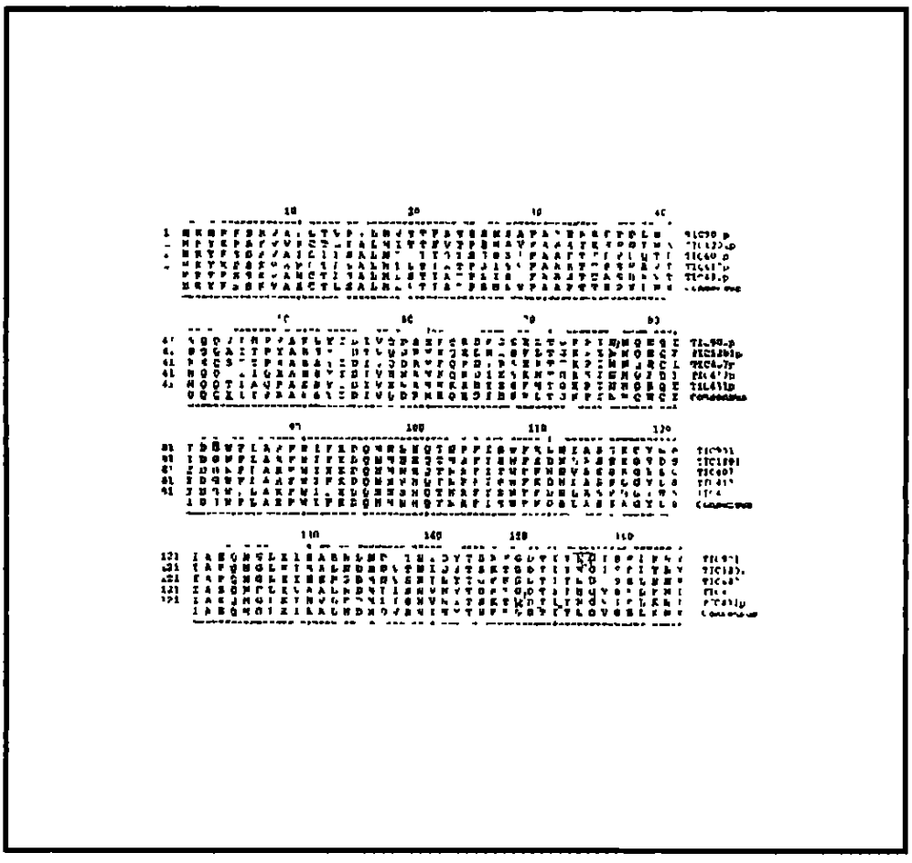
⑨ **Comprobantes de pago Nos.** **05-96,475 y 06-1,173** **Fecha** **Diciembre 20 de 2005 y Enero 5 de 2006**

Instrucciones para el diligenciamiento del presente formulario, ver reverso de esta página.

(10) Anexos

- [X] Comprobante de pago de la tasa de presentación de la solicitud
- [ ] Documento que acredita la existencia y representación legal cuando el solicitante sea persona jurídica
- [ ] Poderes, si fuere el caso, junto con la sustitución del mismo a mi favor
- [ ] Documento de cesión del inventor al solicitante o a su causante
- [ ] Declaraciones
- [X] Copia de la solicitud internacional (Resumen, descripción, reivindicaciones, dibujos)
- [X] Traducción de la solicitud internacional al castellano (Resumen, descripción, reivindicaciones, dibujos)
- [ ] Copia del examen preliminar internacional efectuado por la Autoridad Internacional de Examen Preliminar (IPEA)
- [ ] Traducción del examen preliminar internacional efectuado por la IPEA
- [ ] De ser el caso, copia del contrato de acceso
- [ ] De ser el caso, documento que acredite la licencia o autorización de uso de conocimientos tradicionales de las comunidades indígenas
- [ ] De ser el caso, certificado de depósito del material biológico
- [X] Arte final 12x12
- [ ] De ser el caso, información sobre otras solicitudes de patente o títulos obtenidos en el extranjero por el mismo titular o su causante, relacionadas parcial o totalmente con la invención de esta solicitud
- [ ] De ser el caso, modificaciones efectuadas a la solicitud internacional
- [X] Otros **Tarjeta de Propietario, Tarjeta Temática y Extracto para Publicación.-**

(11) FIGURA CARACTERISTICA:



(12) Solicito la concesión de la patente,

NOMBRE **MARGARITA CASTELLANOS ABONDANO**

FIRMA

C C No **39.779.654 Usaquén** - TPA **70.819 C.S.J.**

06 ENE. 2006

Instrucciones para el diligenciamiento del presente formulario, ver reverso de esta página. - MCA/gcb - 6027-FNPCT

ALVARO CASTELLANOS M. & CIA.
ABOGADOS
Calle 94 A No. 7 A-09
Apartado 6349
Bogotá, D.C. – COLOMBIA

ANEXO # 1

INVENTOR(ES) (72)

	NOMBRE	DIRECCION	NACIONALIDAD O DOMICILIO
1	JAMES A. BAUM	321 South Elm Ave , Webster Groves, MO 63119, ESTADOS UNIDOS DE AMERICA	Estadounidense
2	JUDITH C. DONOVAN	817 Phaeton Drive, Manchester, MO 63021, ESTADOS UNIDOS DE AMERICA	Estadounidense
3	WILLIAM P. DONOVAN	817 Phaeton Drive, Manchester, MO 63021, ESTADOS UNIDOS DE AMERICA	Estadounidense
4	JAMES T. ENGLEMAN	429 Ridge Avenue, Ephrata, PA 17522, ESTADOS UNIDOS DE AMERICA	Estadounidense
5	KARINA KRASOMIL-OSTERFELD	237 Clayton Oaks Drive, Ellisville, MO 63011, ESTADOS UNIDOS DE AMERICA	Estadounidense
6	JOHN W. PITKIN	16526 Winterleaf Drive, Wildwood, MO 63011, ESTADOS UNIDOS DE AMERICA	Estadounidense
7	JAMES K. ROBERTS	15422 Clover Ridge Drive, Chesterfield, MO 63017, ESTADOS UNIDOS DE AMERICA	Estadounidense

6027.
4

SUPERINDUSTRIA Y COMERCIO
Radicacion : 05001259 00000000 Folios: 314
Fecha (AMD): 2006-01-06 15:32:10
Tramite : 011 PCT-PATENT 379 FASE NACI 411 PRESENTAC
Dependencia: 2020 DIVISION DE NUEVAS CREACIONES

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

NIT : 800176089-2

RECIBO OFICIAL DE CAJA : 05 - 96,475
FECHA : DICIEMBRE 20 DE 2005

***** C O N S I G N A C I O N *****

DEPOSITANTE	TIPO PAGO	BANCO	CUENTA	No. PAGO	FECHA PGO	Vr. PAGO
ALVARO CASTELLANOS	CONSIGNACION	BANCO POPULAR	050-00110-6	49043983	20/12/2005	2,393,550.00

***** C O N C E P T O *****

CANT.	RENTISTICO	CONCEPTO	TOTAL CONCEPTO
1	50005-01-01 SOLICITUDES	1 TRAMITES DE SOL. DE PATENTE DE IN	443,000.00
		TOTAL : $	443,000.00

SON: CUATROCIENTOS CUARENTA Y TRES MIL PESOS

RESPONSABLE : _____

RECIBO DE CAJA APLICADO AL EXPEDIENTE No._____

6027
5

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

NIT : 800176089-2

SUPERINDUSTRIA Y COMERCIO
Radicacion : 06001259 00000000 Folios: 314
Fecha (AMD): 2006-01-06 15:32:18
Tramite : 011 PCT-PATENT 379 FASE NACI 411 PRESENTAC
Dependencia: 2020 DIVISION DE NUEVAS CREACIONES

RECIBO OFICIAL DE CAJA : 06 - 1.173
FECHA : ENERO 5 DE 2006

C O N S I G N A C I O N

DEPOSITANTE	TIPO PAGO	BANCO	CUENTA	No. PAGO	FECHA PGO	Vr. PAGO
ALVARO CASTELLANOS	CONSIGNACION	BANCO POPULAR	050-00110-6	49043995	05/01/2006	22,630,400.00

C O N C E P T O

CANT.	RENTISTICO	CONCEPTO		TOTAL CONCEPTO
1	50005-01-09 OTROS SERVICIOS DE PROPIEDAD I	998	VALOR COMPLEMENTO TARIFA A#O ANTR	20,000.00
			TOTAL : $	20,000.00

SON: VEINTE MIL PESOS

RESPONSABLE : _____

RECIBO DE CAJA APLICADO AL EXPEDIENTE No._____

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 March 2005 (03.03.2005)

PCT

(10) International Publication Number
WO 2005/019414 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US2004/021692

(22) International Filing Date: 6 July 2004 (06.07.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/485,483 7 July 2003 (07.07.2003) US

(71) Applicant *(for all designated States except US)*: MONSANTO TECHNOLOGY, LLC [US/US], Gail Wuellner, E2NA, 800 North Lindbergh Blvd, Creve Coeur, MO 63198 (US)

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: BAUM, James, A. [US/US], 321 South Elm Ave, Webster Groves, MO 63119 (US) DONOVAN, Judith, C. [US/US], 817 Phaeton Drive, Manchester, MO 63021 (US) DONOVAN, William, P. [US/US], 817 Phaeton Drive, Manchester, MO 63021 (US). ENGLEMAN, James, T. [US/US], 429 Ridge Avenue, Ephrata, PA 17522 (US). KRASOMIL-OSTERFELD, Karina [US/US], 237 Clayton Oaks Drive, Ellisville, MO 63011 (US) PITKIN, John, W. [US/US], 16526 Winterleaf Drive, Wildwood, MO 63011 (US) ROBERTS, James, K. [US/US], 15422 Clover Ridge Drive, Chesterfield, MO 63017 (US).

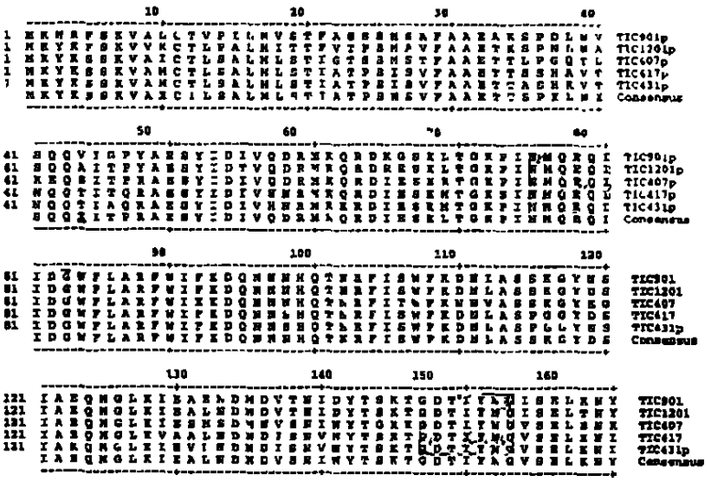
(74) Agent: BALL, Timothy, K., E2NA, Monsanto Company, 800 North Lindbergh Blvd, Creve Coeur, MO 63198 (US)

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

[Continued on next page]

(54) Title: INSECTICIDAL PROTEINS SECRETED FROM BACILLUS THURINGIENSIS AND USES THEREFOR



(57) Abstract: The present invention relates to the isolation and characterization of nucleotide sequences encoding novel insecticidal proteins secreted into the extracellular space from Bacillus thuringiensis and related strains The proteins are isolated from culture supernatants of Bacillus thuringiensis and related strains and display insecticidal activity against coleopteran insects including Colorado potato beetle (*Lymantria dispar*) and Southern Corn Rootworm (*Diabrotica undecempunctata*) Insecticidal proteins encoded by nucleotide sequences that hybridize to the isolated and characterized nucleotide sequences are disclosed Also disclosed are methods of making and using transgenic cells and plants comprising the novel nucleotide sequence of the invention

WO 2005/019414 A2

European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette

APPLICATION FOR PATENT

For

INSECTICIDAL PROTEINS SECRETED FROM BACILLUS SPECIES AND USES THEREFOR

Co-Inventors.

James A. Baum
Judith C. Donovan
William P. Donovan
James T. Engleman
Karina Krasomil-Osterfeld
John W. Pitkin
and
James K. Roberts

Each *B. thuringiensis* subspecies often produces unique types of insecticidal crystal proteins For example, *B thuringiensis* subspecies *kurstaki* produces crystal proteins of approximately 130 kilo Daltons (kD) and 70 kD in size that are toxic to caterpillars, whereas *B. thuringiensis* subspecies *tenebrionis* produces a crystal protein of about 72 kD which is toxic to beetles.

A characteristic of crystal proteins is their ability to coalesce to form crystals inside the *B. thuringiensis* mother cell Upon lysis of the mother cell the proteins are released as crystals into the external environment. In addition, *B. thuringiensis* also produces non-crystal proteins that, in contrast to crystal proteins, are secreted by *B. thuringiensis* cells as soluble proteins into the culture medium. Secreted non-crystal proteins of *B. thuringiensis* include phospholipases, proteases, and β-lactamase that have little, if any, insecticidal activity. However, three secreted non-crystal proteins of *B. thuringiensis* designated Vip1, Vip2 and Vip3 have been reported to be toxic to coleopteran or lepidopteran insects (Estruch et al., 1996; U. S. Patent No. 5,866,326; WO94/21795; WO96/10083) A non-crystal protein of *B. thuringiensis* designated CryV is reported to be toxic to lepidopteran insects (Kostichka et al , 1996) A number of *Bacillus thuringiensis* isolates producing extracellular secreted insecticidal protein toxins have been previously identified (US Patent Serial No 5,840,868, US Patent Serial No 5,849,870; US Patent Serial No 5,866,326; US Patent Serial No 5,872,212, US Patent Serial No. 5,877,012; US Patent Serial No. 5,888,801; US Patent Serial No. 6,204,435; US Patent Serial No. 6,242,669; US Patent Serial No 6,279,369). Such strains have all been shown to produce one or more of these VIP or CryV toxin proteins or closely related homologs. Surprisingly, the inventors herein disclose a new class of extracellular secreted insecticidal protein toxins that do not exhibit homology to the known VIP or CryV classes of proteins.

Comparisons of amino acid sequences indicate that the Vip1, Vip2, Vip3, WAR, MIS and CryV protein classes are not related to the proteins of the present invention. Further comparison shows that none of the one hundred thirty-seven, more or less, known insect-toxic proteins of *B. thuringiensis* (Crickmore et al., 1998), are related to the proteins of the present invention. In fact, no significant homology was found between the sequences of the proteins of the present invention and any of the thousands of protein sequences contained in the National Center for Genome Resources (GenBank), Santa Fe, NM. A BLAST search identified only two proteins in the GenBank database that suggested a possible homology to TIC901. The *Bacillus sphaericus* Mtx2 insecticidal protein exhibited only a 21% amino acid sequence identity over a contiguous 135 amino acid sequence length when aligned with TIC901. A putative amino acid sequence that may be expressed from a Fowlpox virus genome exhibited only a 27% amino acid sequence identity over a contiguous 147 amino acid sequence length when aligned with TIC901.

SUMMARY OF THE INVENTION

In one embodiment, the present invention relates to an isolated and purified insecticidal protein, exhibiting an amino acid sequence substantially as set forth in SEQ ID NO: 4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, and SEQ ID NO:33 whether as precursor amino acid sequences or mature and/or processed and secreted forms of these amino acid sequences, or related amino acid sequences and homologs thereof. Insecticidal activity of TIC901 and related proteins has been

extracellular space surrounding the bacterial strain during fermentation. An exemplary strain EG12450 has been deposited in a permanent culture collection pursuant to the Budapest Treaty and has been assigned the accession No. NRRL B-30357.

In a further embodiment, the invention also relates to a biologically pure culture of a *B. thuringiensis* bacterium designated as strain EG2158 exhibiting insecticidal activity against coleopteran insects. *B. thuringiensis* strain EG2158 represents a wild type *B. thuringiensis* strain from which a *tic*901 coding sequence was isolated and has been deposited in a permanent culture collection pursuant to the Budapest Treaty and has been assigned the accession No. NRRL B-18213. EG2158 is shown herein to produce at least two insecticidal proteins comprising the amino acid sequences selected from the group consisting of SEQ ID NO 4 and SEQ ID NO·10.

In a further embodiment, the present invention provides a vector comprising a nucleotide sequence as set forth in SEQ ID NO:3 encoding a TIC901 amino acid sequence as set forth in SEQ ID NO:4. An *Escherichia coli* strain containing a vector comprising SEQ ID NO.3 has been deposited on February 6, 2002 in the Northern Regional Research Lab of Agricultural Research Service Center Collection (NRRL), USDA, under the provisions of the "Budapest Treaty on the International Recognition of the Deposit of Microorganism for the Purposes of Patent Procedure" and given the Accession No. NRRL B-30549. One plasmid containing said nucleotide sequence is set forth herein as pEG1381.

In a further embodiment, the present invention provides a nucleotide sequence as set forth in SEQ ID NO:3 encoding a TIC901 amino acid sequence, and an oligonucleotide portion that can be labeled and used as a hybridization probe for identifying additional related genes encoding related insecticidal proteins or homologues thereof. Other related nucleotide sequences specifically exemplified herein comprise sequences as set forth in SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO.9, and SEQ ID NO.32, each of which encode insecticidal protein toxins as set forth in SEQ ID NO: 4, SEQ ID NO 6, SEQ ID NO 8, SEQ ID NO 10, and SEQ ID NO 33, respectively.

In yet a further embodiment, the invention provides plant cells and plants that have been transformed with a nucleotide sequence encoding a TIC901 protein as set forth in SEQ ID NO 4 or an insecticidal fragment thereof. The nucleotide sequence can be translated and expressed by plant cells and in plant tissues at levels sufficient to inhibit or kill Coleopteran insect pests. Both monocot and dicot plants are within the scope of the invention. Modification of the sequence may be required in order to effect the maximum level of expression and to enhance the ability of the plant containing the sequence to produce insecticidal levels of the TIC901 protein.

In yet a further embodiment, the present invention also provides a method for producing a transgenic plant that exhibits increased expression levels of a nucleotide sequence encoding TIC901, and thereafter increased levels of the insecticidal protein TIC901. Thus plants transformed with nucleotide sequences modified from those disclosed herein exhibit improved and increased levels of coleopteran pest resistance abilities in comparison to a plant lacking a nucleotide sequence encoding a TIC901 or related protein.

In accomplishing the foregoing, a method for expressing a nucleotide sequence encoding a TIC901 protein in a plant is provided comprising the steps of inserting into the genome of a plant cell a

nucleic acid sequence comprising in the 5' to 3' direction, a plant functional promoter operably linked to a structural DNA sequence optimized for plant expression that causes production of an RNA sequence encoding a TIC901 polypeptide sequence as set forth in SEQ ID NO: 4, or a sequence having at least from about 80%, or from at least about 85%, or from at least about 90%, or from at least about 95%, or from at least about 99% sequence identity to the amino acid sequence as set forth in SEQ ID NO: 4; and a 3' non-translated DNA sequence that functions in the cells of the plant to cause transcription termination and polyadenylation, obtaining transformed plant cells containing said nucleic acid sequence; and regenerating from the transformed plant cells genetically transformed plants that express the nucleotide sequence encoding the TIC901 protein, wherein the transformed plants are morphologically normal and exhibit elevated or improved levels of coleopteran pest resistance compared to a plant not transformed to express said protein.

Another embodiment of the present invention is the provision for antibodies that bind specifically to epitopes presented only by the TIC901 protein or homologs Antibodies can be used for identifying the presence of a TIC901 protein or homolog, for purifying said protein or homolog, for identifying a nucleotide sequence from which a TIC901 protein or homolog is being expressed, and for use in kits designed to allow the detection of a TIC901 protein or homolog or the detection of a nucleotide sequence expressing said protein or homolog.

A particular advantage of the present invention comprises an improvement in insect resistance management. The ability to combine two or more insecticidal agents, each toxic to the same insect pest species, into a single composition, and each agent exhibiting a mode of action different from the other insecticidal agents with which it is combined, present a means for more effectively controlling a particular insect pest species by substantially reducing the likelihood that resistance to the insecticidal composition will develop in a population. The TIC901 protein of the present invention can be combined with any number of known insectidical agents to achieve the level of resistance management in a particular composition, preferably by expression of the combination of insecticidal agents in plants In particular TIC901 protein compositions can be combined with Cry3 or Cry3 amino acid sequence variants to achieve control of various coleopteran plant pest species, or with other appropriate Cry proteins such as PS149B1, CryET33/34, CryET80/76, CryET70, Cry22, CryET39, CryET76, Cry5Ba, Cry6a, and Cry12a, and the like, and with VIP, WAR, or MIS proteins and the like, and with various insecticidal compositions derived from *Xenorhabdus* and *Photorhabdus* bacterium species that have been shown to exhibit insecticidal bioactivity directed to Coleopteran plant pest species. Preferably the *in planta* use of these compositions would be directed to enhanced expression of the proteins in the parts of the plant that exhibit the greatest vulnerability to coleopteran insect predation. For protection of potato against CPB, it would be preferable to achieve the highest levels of expression in the leaves and stems of the plant. For maize species susceptible to wireworm and to rootworms, it would be preferable to achieve the highest levels of expression in the subterranean parts of the plant, i.e., within the root systems of the plant.

Another embodiment comprises an isolated polynucleotide that encodes a *Bacillus thuringiensis* insecticidal toxin or insecticidal fragment thereof, active against an insect pest, wherein the toxin or insecticidal fragment has a molecular weight between approximately 36,000 Daltons and

approximately 42,500 Daltons. In addition, the nucleotide sequence encoding the toxin, or the complement thereof, hybridizes under stringent conditions to SEQ ID NO:3. The toxin preferably exhibits biological activity in controlling or killing a coleopteran insect pest, preferably Colorado potato beetle and/or corn rootworms. In one embodiment the nucleotide sequence encoding the toxin is optimized for expression in plants, yet encodes substantially the toxin or an insecticidal fragment thereof, i e., encodes the same or substantially the same amino acid sequence as present in the native amino acid sequence.

Another embodiment of the present invention provides for host cells transformed to contain a polynucleotide encoding an insecticidal protein of the present invention or an insecticidal fragment thereof Preferably the nucleotide sequences of the present invention are modified to improve expression of the proteins of the present invention in a preferred host cell The host cell of the present invention is selected from the group consisting of a bacterial cell, a fungal cell, and a plant cell Expression in a plant cell can comprise expression to achieve accumulation of the insecticidal protein in the cytoplasm, or can result in the insecticidal protein being accumulated into a subcellular organelle such as a plastid, chloroplast, or mitochondria. Alternatively the insecticidal protein of the present invention or insecticidal fragments thereof could be localized to the protein secretion machinery of the particular host cell and result in an accumulation of the protein product out side of the cell and into the extracellular spaces surrounding the cell

An additional embodiment of the present invention provides a method for controlling infestation of a plant by a coleopteran insect species. Preferably a pesticidal amount of an insecticidal protein of the present invention or insectidal fragment thereof is provided for consumption by the insect pest in the diet of the insect. The diet can consist of a plant part that the insect normally feeds upon, such as a plant tissue or plant cell. The insecticidal protein or insecticidal fragment thereof can be provided in a composition that is applied to the surface of the plant tissue, plant part, or plant cell or more preferably can be produced by the protein synthesis machinery of the cell and, as described above, accumulated within the plant cell or secreted outside of the plant cell, so long as the amount of the protein toxin provided is an insecticidal amount sufficient to inhibit the insect pest from further feeding, or to inhibit the further growth and development of the insect pest, or to cause mortality to the insect pest. The diet provided to the insect can also be an artificial diet that contains the toxin protein uniformly distributed within or topically applied to the exposed surface(s) of the diet substrate, or included as a concentration gradient within or topically applied to the exposed surface(s) of the diet substrate. The insecticidal toxin or fragment thereof is derived from a nucleotide sequence that is encoded in *Bacillus thuringiensis* by a nucleotide sequence that hybridizes under stringent conditions to the nucleotide sequence substantially complementary to SEQ ID NO:3.

The present invention also provides a method for detecting a first nucleotide sequence that hybridizes to a second nucleotide sequence as set forth in SEQ ID NO:3, wherein the first nucleotide sequence encodes an insecticidal protein or insecticidal fragment thereof and hybridizes under stringent hybridization conditions to the second nucleotide sequence. Exemplary sequences are SEQ ID NO:2 and SEQ ID NO:3.

The present invention also provides non-naturally occurring or synthetic nucleotide sequences that encode a TIC901 insecticidal protein or insecticidal fragment thereof or homolog thereof, wherein said TIC901 protein or insecticidal fragment thereof or homolog thereof is selected from the group of sequences consisting of SEQ ID NO:5, and SEQ ID NO.7. Preferably the non-naturally occurring nucleotide sequence or sequences provided for herein that encode an insecticidal protein or insecticidal fragment thereof are provided for expression of a TIC901 or related protein in plant cells. Therein, plant cells transformed with such sequences are provided for herein. Plants grown from the transformed plant cells are provided by the instant inventions. Seeds from the transformed plants of the present invention are also provided so long as the seeds contain at least the sequences encoding the insecticidal proteins or insecticidal protein fragments thereof.

Exemplary sequences of the present invention, in addition to those related to SEQ ID NO:3 and SEQ ID NO:4 include at least: (1) the nucleotide sequence as set forth in SEQ ID NO:5, and the amino acid sequence encoded by SEQ ID NO 5 as set forth in SEQ ID NO 6, also referred to herein as insecticidal protein TIC1201; (2) the nucleotide sequence as set forth in SEQ ID NO.7, and the amino acid sequence encoded by SEQ ID NO:7 as set forth in SEQ ID NO:8, also referred to herein as insecticidal protein TIC407; (3) the nucleotide sequence as set forth in SEQ ID NO:9, and the amino acid sequence encoded by SEQ ID NO:9 as set forth in SEQ ID NO:10, also referred to herein as insecticidal protein TIC417, and (4) the nucleotide sequence as set forth in SEQ ID NO:32, and the amino acid sequence encoded by SEQ ID NO:32 as set forth in SEQ ID NO.33, also referred to herein as insecticidal protein TIC431. Each of these proteins and the native *B. t.* nucleotide sequences encoding these proteins are related to TIC901 as defined herein. For example, and respectively, SEQ ID NO 5 is a nucleotide sequence encoding a TIC1201 insecticidal protein as set forth in SEQ ID NO:6 SEQ ID NO 5 as shown herein is identifiable by hybridization to SEQ ID NO:3 under stringent conditions. SEQ ID NO:5 encodes a protein that exhibits coleopteran toxic biological activity, exhibiting toxicity to corn rootworms and to Colorado potato beetles. SEQ ID NO.5, SEQ ID NO.3, SEQ ID NO.7, SEQ ID NO:9, and SEQ ID NO:32 are each capable of hybridizing to each other under hybridization conditions selected from the group consisting of stringent hybridization conditions and specific hybridization conditions. Each sequence can also be identified by hybridization to SEQ ID NO:2 under conditions selected from the group consisting of stringent hybridization conditions and specific hybridization conditions. Each sequence can also be identified by amplification using, for example, an oligonucleotide primer pair as set forth in SEQ ID NO:11 and SEQ ID NO:12, and an oligonucleotide primer pair as set forth in SEQ ID NO 23 and SEQ ID NO:27. The primer pair as set forth in SEQ ID NO:11 and SEQ ID NO:12, and the primer pair as set forth in SEQ ID NO:23 and SEQ ID NO:27 are exemplary and diagnostic for identifying the presence of a nucleotide sequence encoding a TIC901 or related insecticidal protein in a sample. These oligonucleotide pairs, when used alone or together under defined amplification conditions and in the presence of a suitable nucleotide sequence substrate, produce an amplicon consisting of from about 540 to about 640 base pairs. Thermal amplification reactions using these primer sets are useful for detecting the presence of a *B.t.* gene encoding an insecticidal protein corresponding to a TIC901 or related protein in a sample, and greatly simplifies the search for and identification of such related sequences Other amplicons derived

from the use of other primer pairs are also envisioned based on the nucleotide sequence alignment of, for example, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, and SEQ ID NO:32. Regions of substantial amino acid sequence identity of the proteins encoded by these nucleotide sequences correspond to nucleotide sequences that can be used for preparing complementary or substantially complementary sequences for use as probes or primers for use in thermal amplification reactions that allow for the detection of sequences related to TIC901, TIC1201, TIC407, TIC417, and TIC431.

Degenerate oligonucleotide probes and primers as set forth in SEQ ID NO:23 through SEQ ID NO:29 are additionally provided as a means for identifying any nucleotide sequence encoding a secreted insecticidal protein from at least a *Bacillus thuringiensis* species in which the nucleotide sequence identified with the degenerate oligonucleotide probes hybridizes under stringent conditions to one or more of the sequences selected from the group consisting of SEQ ID NO.2, SEQ ID NO 3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO 11, SEQ ID NO·12, SEQ ID NO·15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO·18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23,.SEQ ID NO 24, SEQ ID NO.25, SEQ ID NO:26, SEQ ID NO.27, SEQ ID NO:28, SEQ ID NO 29, and SEQ ID NO 32. Exemplary sequences identifiable using such oligonucleotides include sequences selected from the group consisting of SEQ ID NO 3, SEQ ID NO:5, SEQ ID NO 7, SEQ ID NO 9, SEQ ID NO 30, and SEQ ID NO 32, each encoding respectively peptides as set forth in SEQ ID NO·4, SEQ ID NO 6, SEQ ID NO 8, SEQ ID NO 10, SEQ ID NO 31, and SEQ ID NO.33.

Another embodiment comprises a method of detecting *tic*901 and related protein coding sequences in *Bacillus* strains comprising the steps of culturing a *Bacillus* species for from about 16 to about 45 hours in rich broth under aerobic conditions, detecting a protein in the culture supernatant that exhibits TIC901 related antigen identity and/or cross reactivity to an antibody that binds specifically to one or more TIC901 protein peptides or antigens, identifying and purifying a nucleotide sequence that encodes the detected protein, expressing the protein from the nucleotide sequence, and demonstrating insecticidal activity using the expressed protein.

Kits for detecting the presence of the nucleotide sequences of the present invention, as well as probes, primers, analogues and derivatives of the same, are also contemplated. Such kits contain one or more nucleotide sequences each for use either as a probe for detecting the presence of a nucleotide sequence encoding an insecticidal protein of the present invention or fragment thereof or related nucleotide sequences, or for use in combination with one or more other probes or primers included in such kit for amplifying one or more sequences of the present invention or a related nucleotide sequence. Such kits could also or alternatively contain antibody specific for binding to one or more peptides or the proteins of the present invention, as well as reagents for use with the probe or antibody, and the kits would also contain control samples for use in ensuring that the nucleotides or peptides identified with the probe and or antibody and reagents were functioning according to the manufacturers' instructions. All of the reagents necessary for carrying out the methods of identification of either nucleotide sequences or peptides would be packaged together in a kit along with instructions for use. An exemplary kit will contain a nucleotide sequence derived from a TIC901, TIC1201,

TIC407, TIC417, and/or TIC431 coding sequence along with a sample of nucleotide sequence amplification primers, for example, as set forth in SEQ ID NO:11 and SEQ ID NO.12, or various combinations of SEQ ID NO:23-26 and SEQ ID NO.27-29 together with the reagents necessary for carrying out an amplification reaction, all packaged together in said kit

It is therefore contemplated that the compositions and methods disclosed by the present invention will provide many advantages over the prior art including those specifically outlined above. In addition, the present invention provides an entirely new class of insecticidal proteins and nucleotide sequences encoding these proteins that were not previously known in the art.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 is an amino acid sequence alignment of the precursor proteins TIC901p, TIC407p, TIC417p, TIC1201p, and TIC431p; each amino acid sequence contains a predicted thirty (30) amino terminal amino acid sequence characteristic of a type II signal peptide followed by thirteen additional amino acids from amino acid position thirty-one (31) through amino acid position forty-three (43) of their respective sequences that is not present in the mature protein isolated from spent fermentation media. The underlined amino acid in the consensus sequence at position 44 represents the mature protein amino terminal amino acid. The native nucleotide sequence encoding the corresponding shaded amino acids at positions 75-83, 147-153, and 275-283 were used as the basis for constructing redundant nucleotide probes and primers used for identifying sequences encoding these and other related insecticidal proteins from Bacillus species nucleotide sequences.

BRIEF DESCRIPTION OF THE SEQUENCES

SEQ ID NO. 1 represents an amino acid sequence deduced from the results of Edman degradation of a gel purified about 38 kDa insecticidal protein secreted into the media by *B. thuringiensis* strain EG2158 cells, and corresponds substantially to the amino acid sequence as set forth in SEQ ID NO:4 from amino acid sequence position 44 through 58.

SEQ ID NO: 2 represents a synthetic nucleotide sequence for use as a probe for detecting a *tic*901 or related nucleotide sequence, or for use as one of a pair of thermal amplification primers to amplify all or a part of a *tic*901 or related nucleotide sequence, and corresponds to codon triplets preferred for use by *B thuringiensis* and other *Bacillus* species, in particular exhibiting codon usage that is biased towards containing an A or a T at the third base pair position within each codon.

SEQ ID NO: 3 represents a native (also referred to herein as wild-type) *Bacillus thuringiensis* nucleotide sequence encoding a TIC901 protein. A predicted Pribnow box or Shine & Dalgarno sequence is located at about nucleotides 141-147. The predicted ORF encoding the predicted precursor TIC901 protein corresponds to nucleotides from position 153 through position 1,253. Nucleotides from position 282-325 correspond substantially to the sequence of the oligonucleotide probe as set forth in SEQ ID NO 2, which hybridizes to the complement of nucleotides 282-325 as

set forth in SEQ ID NO:3 The GTA valine codon at nucleotide position 282-284 corresponds to the amino terminal amino acid in the secreted form of TIC901

SEQ ID NO: 4 represents a 367 residue TIC901 amino acid sequence deduced from the open reading frame as set forth in SEQ ID NO:3 from nucleotide position 153 through nucleotide position 1,253. The full length 367 residue amino acid sequence corresponds to the predicted precursor protein amino acid sequence expressed from the native/wild type coding sequence in *B. thuringiensis* The amino acid sequence from residue number 1 through residue 30 as set forth in SEQ ID NO:4 corresponds to the predicted amino terminal signal peptide or secretory signal peptide that is produced in *B. thuringiensis* from the expression of the nucleotide sequence as set forth in SEQ ID NO:3, is followed by thirteen (13) amino acids that are not present in the mature/secreted form of the 324 amino acid residue mature insecticidal protein sequence upon expression in *B. thuringiensis*. The 324 amino acid residue mature insecticidal protein sequence corresponds to the insecticidally effective TIC901 mature and secreted protein sequence. The 43 residue amino terminal amino acid sequence is predicted to be proteolytically cleaved from the precursor protein, either in part during translocation across the bacterial cytoplasmic membrane, or in part by an as yet undefined signal peptidase or other protease that recognizes the consensus sequence comprising the amino acid sequence residues XAA1-XAA2-GLN immediately before the scissile breakpoint, releasing the mature insecticidal protein into the extracellular milieu, where XAA1 corresponds to serine (SER), lysine (LYS),or asparagine (ASN), and XAA2 corresponds to glutamate (GLU) or glutamine (GLN).

SEQ ID NO 5 represents a native *B. thuringiensis* nucleotide sequence encoding an insecticidal protein designated herein as TIC1201. The sequence includes 529 nucleotides of sequence upstream of the predicted ATG initiation codon positioned at nucleotide position 530-532 A predicted consensus Pribnow box or Shine & Dalgarno sequence is positioned upstream of the predicted ATG initiation codon from nucleotide position 518 through 524. The open reading frame encoding the predicted precursor TIC1201 protein, like TIC901, comprises an amino terminal amino acid sequence corresponding to a predicted signal peptide or secretory targeting peptide The sequence encoding the TIC1201 signal peptide is predicted to encode thirty (30) amino acids, followed by thirteen (13) additional amino acids that are not present in the mature/secreted form of the insecticidal protein. These thirteen additional amino acids terminate in a sequence encoding the SER-GLN-GLN peptidase recognition sequence identical to the sequence present in the TIC901 precursor protein sequence. The amino acid sequence of the insecticidal TIC1201 protein released into the media from *B thuringiensis* strains expressing this sequence is predicted from the coding sequence to comprise 321 amino acid residues, being encoded by the nucleotide sequence from position 659 through 1621 as set forth in SEQ ID NO:5. Even though the predicted ORF encoding the TIC1201 precursor protein is identified herein as being within nucleotides 530-1621, the ORF could possibly extend from nucleotide 437 through 1621. This is predicted to be unlikely because of the similarity of the signal peptide to that of TIC901, and the lack of any consensus Pribnow box or Shine & Dalgarno sequence within an reasonable proximity to any ATG initiation codon upstream of that positioned at nucleotides 518 through 524.

SEQ ID NO:6 represents a deduced 395 amino acid sequence of the TIC1201precursor protein as encoded by the nucleotide sequence from nucleotide 530 through nucleotide 1621as set forth in SEQ ID NO:5.

SEQ ID NO 7 represents a nucleotide sequence encoding the insecticidal protein designated herein as TIC407.

SEQ ID NO.8 represents a deduced 368 amino acid sequence for TIC407 as encoded by the nucleotide sequence from nucleotide position 196 through 1299 as set forth in SEQ ID NO:7.

SEQ ID NO:9 represents a nucleotide sequence encoding the insecticidal protein designated herein as TIC417.

SEQ ID NO:10 represents a deduced 364 amino acid sequence for TIC417 as encoded by the nucleotide sequence from nucleotide position 92 through 1173 as set forth in SEQ ID NO:9

SEQ ID NO:11 represents a forward amplification thermal primer sequence, or a probe sequence, designated herein as prJWP139, corresponding to the coding sequence as set forth in SEQ ID NO 3 from nucleotide position 438 through 458, and further corresponding to the codons encoding the amino acid sequence ASN-ASN-ASN-HIS-GLN-THR-ASN-ARG from amino acid sequence position 96-103 as set forth in SEQ ID NO.4, biased towards codons preferred for use in gene sequences derived from *Bacillus thuringuensis* or other *Bacillus* species strains, in which the codons contain A and or T in the third position.

SEQ ID NO.12 represents a reverse amplification thermal primer sequence, or a probe sequence, designated herein as prJWP143, corresponding to the reverse complement of the coding sequence as set forth in SEQ ID NO 3 from nucleotide position 978 through 998, and further corresponding to the codons encoding the amino acid sequence GLN-LYS-PHE-ILE-TYR-PRO-ASN from amino acid sequence position 276-282 as set forth in SEQ ID NO:4, biased towards codons preferred for use in gene sequences derived from *Bacillus thuringiensis* or other *Bacillus* species strains, in which the codons contain A and or T in the third position.

SEQ ID NO:13 represents a synthetic, artificial, or non-naturally occurring nucleotide sequence encoding a TIC901 amino acid sequence variant.

SEQ ID NO:14 represents an amino acid sequence deduced from the coding sequence as set forth in SEQ ID NO:13 from nucleotide position 1 through nucleotide position 1104.

SEQ ID NO:15 represents an artificial nucleotide sequence for use as a probe or primer, described herein as prJWP151.

SEQ ID NO:16 represents an artificial nucleotide sequence for use as a probe or primer, described herein as prJWP152.

SEQ ID NO:17 represents an artificial nucleotide sequence for use as a probe or primer, described herein as prJWP186.

SEQ ID NO.18 represents an artificial nucleotide sequence for use as a probe or primer, described herein as prJWP183.

SEQ ID NO:19 represents an artificial nucleotide sequence for use as a probe or primer, described herein as prJWP155.

SEQ ID NO.19 represents an artificial nucleotide sequence for use as a probe or primer, described herein as prJWP155.

SEQ ID NO.20 represents an artificial nucleotide sequence for use as a probe or primer, described herein as prJWP156.

SEQ ID NO·21 represents an artificial nucleotide sequence for use as a probe or primer, described herein as prJWP168.

SEQ ID NO·22 represents an artificial nucleotide sequence for use as a probe or primer, described herein as prJWP170.

SEQ ID NO.23 represents a degenerate artificial oligonucleotide sequence for use as a probe or primer, described herein as prJWP200.

SEQ IDNO:24 represents a degenerate artificial oligonucleotide sequence for use as a probe or primer, described herein as prJWP201.

SEQ IDNO.25 represents a degenerate artificial oligonucleotide sequence for use as a probe or primer, described herein as prJWP202

SEQ IDNO:26 represents a degenerate artificial oligonucleotide sequence for use as a probe or primer, described herein as prJWP203

SEQ IDNO:27 represents a degenerate artificial oligonucleotide sequence for use as a probe or primer, described herein as prJWP204.

SEQ IDNO·28 represents a degenerate artificial oligonucleotide sequence for use as a probe or primer, described herein as prJWP205.

SEQ IDNO.29 represents a degenerate artificial oligonucleotide sequence for use as a probe or primer, described herein as prJWP206.

SEQ ID NO:30 represents a fragment of a nucleotide coding sequence derived from thermal amplification of the genome of EG2158 with oligonucleotides prJWP200 and prJWP204.

SEQ ID NO 31 represents the amino acid sequence from the primary open reading frame set forth in SEQ ID NO 30

SEQ ID NO 32 represents a nucleotide sequence encoding the insecticidal protein designated herein as TIC431.

SEQ ID NO:33 represents a deduced 364 amino acid sequence for TIC431 as encoded by the nucleotide sequence from nucleotide position 1 through 1092 as set forth in SEQ ID NO:32.

DETAILED DESCRIPTION OF THE INVENTION

In accordance with the present invention, a new genus of nucleotide sequences encoding a new genus of insecticidal proteins derived from *Bacillus thuringiensis* and related *Bacillus* strains has been discovered As defined elsewhere herein, these nucleotide coding sequences hybridize to one another under appropriate hybridization conditions and the proteins encoded by these nucleotide sequences cross react with antiserum raised against any one of the other proteins. An alignment of the nucleotide sequences encoding the mature/secreted forms of the TIC1201, TIC901, TIC407, TIC417, and TIC431 proteins reveals that the coding sequence encoding the mature TIC901 protein fragment

(secreted form of TIC901 from amino acid position 44 through amino acid position 367 as set forth in SEQ ID NO:4) is from about 79 to about 91 percent identical to each of the sequences encoding the other four mature protein fragments disclosed herein; the sequence encoding the predicted mature/secreted TIC417 protein fragment (amino acid position 44 through amino acid position 364 as set forth in SEQ ID NO 10) is from about 75 to about 95 percent identical to each of the sequences encoding the other three mature protein fragments disclosed herein; the sequence encoding the predicted mature/secreted TIC407 protein fragment (amino acid position 44 through amino acid position 368 as set forth in SEQ ID NO:8) is from about 75 to about 82 percent identical to each of the sequences encoding the other four mature protein fragments disclosed herein; the sequence encoding the mature/secreted TIC1201 fragment (amino acid position 44 through amino acid position 395 as set forth in SEQ ID NO.6) is from about 80 to about 91 percent identical to each of the sequences encoding the other four mature protein fragments disclosed herein, and the sequence encoding the mature/secreted TIC431 fragment (amino acid position 44 through amino acid position 364 as set forth in SEQ ID NO:33) is from about 75 to about 95 percent identical to each of the sequences encoding the other four mature protein fragments disclosed herein. The proteins encoded by each of these nucleotide coding sequences exhibit coleopteran species inhibitory biological activity, exhibit substantial amino acid sequence identity in part and substantial amino acid sequence similarity in part, and therefore are considered to be related insecticidal proteins The predicted mature/secreted form amino acid sequence for the TIC417 insecticidal protein (TIC417m) is about 78 9 percent identical to the corresponding mature/secreted form amino acid sequence for TIC901 (TIC901m) The predicted mature/secreted form amino acid sequence for the TIC1201 insecticidal protein (TIC1201m) is about 90.1 percent identical to the corresponding amino acid sequence for TIC901m, and is about 80.7 percent identical to the corresponding amino acid sequence for TIC417m and TIC431m. The predicted mature/secreted form amino acid sequence for the TIC407 insecticidal protein (TIC407m) is about 80% identical to the corresponding TIC901m, about 75% identical to the corresponding TIC417m, and about 82% identical to the corresponding TIC1201m The predicted mature/secreted form amino acid sequence for the TIC431 insecticidal protein (TIC431m) is about 75% identical to the mature TIC407, about 79% identical to the mature TIC901, about 80% identical to the TIC1201, and about 95% identical to the TIC417 mature protein amino acid sequence. Each of the proteins encoded by the nucleotide sequences disclosed herein can be expressed in plants alone or in various combinations with each other or with other coleopteran inhibitory insecticidal agents such as proteins, crystal proteins, δ-endotoxins, lectins, patatins, and other toxins and the like to achieve a means of insect resistance management in the field that has not been feasible before by merely using the known coleopteran insecticidal proteins derived from *Bacillus thuringiensis* strains, such as Cry3 proteins, VIP and/or WAR and/or MIS proteins, and various coleopteran inhibitory insecticidal proteins derived from *Bacillus latersoporous* species, *Bacillus sphaericus* species, and *Xenorhabdus* and *Photorhabdus* bacterial species. The proteins of the present invention can also be used in plants in combination with other types of insecticidal agents and or insecticidal toxins for achieving plants transformed to contain at least one means for controlling one or more of each of the common plant pests selected from the groups consisting of coleopteran insect pests, lepidopteran insect pests, piercing and sucking insect pests, and

the like. Other proteins and or insect controlling agents that can be expressed in a plant in combination with the proteins of the present invention include but are not limited to lepidopteran insecticidal proteins from *Bacillus* species, such as Cry proteins derived from *Bacillus thuringiensis*, *Bacillus laterosporous*, and *Bacillus sphaericus* species, and WAR, MIS, and/or VIP proteins isolatable from various *Bacillus* species, insecticidal proteins derived from *Xenorhabdus* and *Photorhabdus* bacterial species, and compositions such as transgenic dsRNA's expressly directed to suppression of one or more genes in one or more target insect pests. As used herein, "insecticidal polypeptide" or "insecticidal protein" or "insecticidal fragment thereof" refers to a polypeptide exhibiting insecticidal properties, *e.g.*, a polypeptide that inhibits the growth, development, viability or fecundity of target insect pests, and an insecticidal agent including all of these as well as double stranded RNA's directed to suppression of one or more genes in one or more target pests.

Surprisingly, the proteins of the present invention appear to be unrelated to any of the *Bacillus thuringiensis* insecticidal proteins heretofore discovered in the art. The proteins of the present invention are shown herein to be excreted into the extracellular space surrounding the *Bacillus* species from which they are derived. These proteins are shown herein to be significantly smaller than the known Cry, VIP, WAR and MIS proteins previously known in the art, and may be expressed during the vegetative stage of growth of isolated and purified bacterial cell cultures. This is unlike the expression of Cry proteins which are expressed generally in the sporulation phase of growth and which form various crystalline bodies within the forespore of the cell.

As will become apparent to those of skill in the art, the inventors herein disclose the isolation and purification of a nucleotide sequence, *tic*901, encoding a precursor TIC901 protein (TIC901p) that is subsequently proteolytically processed to release a mature TIC901 protein (TIC901m) that exhibits coleopteran species inhibitory biological activity. The inventors herein disclose the use of the *tic*901 sequence as a means for identifying a *multitude* of other related sequences, which each also encode insecticidal proteins related to TIC901, TIC901p, and TIC901m, and disclose the use of antibodies raised against TIC901m in an ELISA method for detecting strains of Bt that produce TIC901 related proteins expressed from coding sequences related to that encoding TIC901

Nucleotide sequences disclosed herein and encoding TIC901 were derived from strains of *Bacillus thuringiensis*, including strains EG2158, EG6489, EG6561, EG12450, and EG4653. These strains have been deposited under the provisions of the Budapest Treaty with the permanent collection of the Agricultural Research Service Culture Collection, Northern Regional Research Laboratory (NRRL), U.S. Department of Agriculture (USDA), 1815 North University Street, Peoria, IL 61604. The relevant strains were deposited with the NRRL between April 29, 1987 and February 6, 2002 *B. thuringiensis* strain EG2158 was provided with the NRRL accession No NRRL B-18213; and *B thuringiensis* strain EG12450 was provided with the NRRL accession No. NRRL B-30357. Nucleotide sequences related to *tic*901, and amino acid sequences related to TIC901 (including precursor and mature species of TIC901) which are disclosed herein include but are not limited to *tic*1201 and the encoded insecticidal protein TIC1201 isolated from and produced at least by *B.t.* strains EG3618 and 86833, *tic*407 and the encoded insecticidal protein TIC407 isolated from and produced at least by *B.t.* strain EG6618, and *tic*417 and the encoded insecticidal protein TIC417 isolated from and produced at

least by *B.t.* strains EG2158, EG6489, and EG6561. and TIC431 encoded at lease by *B t.* strain EG4653.

A broth culture derived from the purified *B.t.* strain EG2158 was tested for insecticidal activity, and was determined to exhibit coleopteran insect inhibitory biological activity directed against Colorado Potato Beetle (CPB). A protein exhibiting a mass of about 38 kDa as judged by SDS-PAGE was purified from the broth culture and was observed to contain the indicated coleopteran toxicity. This protein was subjected to automated Edmund degradation, and the results yielded an amino acid sequence believed to be the 15 amino terminal amino acid sequence (SEQ ID NO:1) of the about 38 kDa protein, i.e., the TIC901m protein. A semi-redundant synthetic oligonucleotide sequence (WD444; SEQ ID NO 2) corresponding to naturally occurring codons preferred for use in protein coding sequences isolated from *Bacillus thuringiensis* or related *Bacillus* species bacteria, i.e. exhibiting a preference for A or T in the third position of each codon, was constructed for use as a probe for detecting sequences of homology that could conceivably encode the about 38 kDa insecticidal protein in *Bacillus thuringiensis* A nucleotide sequence library was constructed from DNA purified from *Bt* strain EG2158. The purified DNA was digested to completion with *Hind*III, and fragments were inserted into a *Hind*III digested, calf intestine phosphatase treated pUC18 plasmid vector to construct a *Bt* strain EG2158 genomic library. The DNA library was transformed into an *E. coli* strain and the transformation mixture was plated onto solid selection media. The colonies that arose were probed with a sample of alkaline phosphatase conjugated synthetic nucleotide sequence probe WD444 (SEQ ID NO·2). A recombinant *E. coli* strain designated as EG12447 containing plasmid pEG1379 (also known as pMON74007), a pUC18 derivative that contained an 8 kb *Hind*III fragment isolated from *B. thuringiensis* strain EG2158, hybridized to the alkaline phosphatase conjugated oligonucleotide probe The 8 kb *Hind*III fragment in plasmid pEG1379was determined by nucleotide sequence analysis to contain the entire nucleotide sequence encoding the TIC901p protein. NRRL received a viable deposit of strain EG12447 and designated the deposited sample with the NRRL accession No. NRRL B-30549 on February 6, 2002

The nucleotide sequence of the open reading frame encoding the about 38 kDa insecticidal protein was determined. The open reading frame encoding the TIC901p protein was designated as *tic*901. The open reading frame consists of a nucleotide sequence of 1101 nucleotides (nucleotides 153-1,253 as set forth in SEQ ID NO·3), and is predicted to encode a precursor protein consisting of 367 amino acids (SEQ ID NO 4). The predicted molecular weight of the amino acid sequence deduced from the open reading frame is 41,492 Daltons, which is within reasonable expectations of the mass of the secreted protein estimated by SDS-PAGE analysis as about 38 kDa considering the loss of mass of a signal peptide of between three and four kDa. The precursor protein (TIC901p) is predicted to exhibit an isoelectric point of 6.368 and a net charge of –2.102 at pH 7.0. The total composition of the nucleotide sequence encoding the precursor protein is comprised of 69% AT, which is consistent with other coding sequences identified from *B. thuringiensis* and other *Bacillus* strains. The nucleotide composition of the coding sequence is also consistent with other genes characterized from *Bacillus* species, containing about 39% adenosine, about 18% guanosine, about 30% thymidine and about 13% cytosine.

The native *tic*901 coding sequence in pEG1379 appeared to be incapable of producing a measurable amount of the TIC901m protein from recombinant *E. coli* cultures containing this plasmid. This was not unexpected given the known lack of functionality of *Bacillus* promoters in *E. coli*, and the differences in codon preference between the two organisms. No CPB inhibition was observed with strain EG12447 culture supernatants or cells containing this plasmid. The 8 kB insert was excised and placed into an *E coli / B. thuringiensis* shuttle vector to form plasmid pEG12450. pEG12450 was transformed into an acrystalliferous strain of *Bacillus thuringiensis*, EG10650 (US Patent No. 6,468,523), to produce strain EG12450. EG10650 was derived from an acrystalliferous *B. thuringiensis* strain EG10368 (identified in US Patent NO. 5,322,687) by replacing the *npr* and *apr* (neutral protease and acidic protease genes, respectively) with deletion mutant alleles of these two protease genes, *npr*3 and *apr*1 respectively (US Patent No. 5,759,538) Culture supernatants derived from EG12450 tested positive for Colorado Potato Beetle (CPB) inhibitory activity Protein purified from the culture supernatants from EG12450 also tested positive for CPB inhibitory activity. No crystal structures were observed in sporulated cultures, but the cell pellets/sporulated culture biomass also tested positive for CPB inhibitory bioactivity suggesting that some portion of the TIC901 protein remained associated with the spore/culture or that spores consumed by the test species of insect germinated within the insect and produced sufficient TIC901 insecticidal protein to cause an observable inhibitory effect.

The culture supernatants and purified protein from strain EG12450 were also tested for biological activity against corn rootworms. Inhibitory bioactivity was observed for both Southern and Western corn rootworms (*Diabrotica undecempunctata howardii* and *Diabrotica virgifera virgifera* respectively).

A diverse collection of *Bacillus* strains was examined as disclosed herein by the inventors in order to determine whether these *B thuringiensis* and/or *B. sphaericus* strains also produced extracellular proteins related to TIC901. In particular, cell paste and spent media were processed from 279 strains and provided in bioassay to southern corn rootworm larvae. About one third of the strains produced secreted proteins into the growth media that tested positive for rootworm inhibitory activity. Priority was given to thirty six (36) strains that produced extracellular proteins that exhibited the greatest effective rootworm inhibition These strains were screened further by determining whether the TIC901 coding sequence hybridized to sequences present in each strains' genome, and comparing the results for these strains with the results obtained by probing strain EG2158 with the native TIC901 coding sequence. The results are shown in Table 1.

Table 1. RFLP Isotypes of *Bacillus* Strains Producing Secreted CRW Insecticidal Proteins

Strain	tic901 RFLP[1]	TIC Protein Homolog ID
EG2158[2]	A	901/417
EG2211	B	*1201*
EG2874	B	*1201*
EG2904	B	*1201*
EG3109	B	*1201*
EG 3111	B	*1201*
EG 3116	B	*1201*
EG 3117	B	*1201*
EG 3119	B	*1201*
EG 3120	B	*1201*
EG 3171	B	*1201*
EG 3173	B	*1201*
EG 3177	B	*1201*
EG 3354	B	*1201*
EG 3458	B	*1201*
EG 3461	B	*1201*
EG 3618	B	1201[3]
EG 3619	B	*1201*
EG 3620	B	*1201*
EG 3753	B	*1201*
EG 3787	B	*1201*
EG 4189	B	*1201*
EG 4191	B	*1201*
EG 4193	B	*1201*
EG 4332[4]	-	
EG 4834	B	*1201*
EG 5194	B	*1201*
EG 5552	-	
EG 5858[4]	-	
EG 6489	A	901/417
EG 6555	B	*1201*[3]
EG 6561	A	901/417
EG 6564	B	*1201*
EG 6618	C	*407*[3]
EG 6890	B	*1201*
EG 10650	-	
86833	B	1201

1-Total DNA isolated from each strain was digested to completion with *HindIII* and analyzed by Southern blot. The letter A, B, C or D corresponds to the size of a particular restriction fragment (polymorphism) illuminated by tic901 or SEQ ID NO.2 probe.

2- EG2158 also contains a 3rd nucleotide sequence homolog related to *tic*901, *tic*1201, *tic*407, and *tic*417 as set forth in SEQ ID NO.30

3- indicates only partial nucleotide sequence identification obtained, but all or portion of sequence obtained encodes protein sequence 100% identical to indicated TIC protein

4- indicates that these strains have been shown to produce extracellular substances that exhibit Lygus inhibitory biological activity

Four (4) of the thirty-six (36) strains failed to produce a nucleotide fragment that hybridizes with the VIP1 or VIP 2 probes (strains EG4332, EG5858, EG6618, and 86833) under conditions of high stringency, i.e., washes at 0.5X SSC and at 65°C. All other strains produced a nucleotide fragment that hybridized to a TIC901 probe, however, interesting hybridization signal intensity variations were observed, specifically with reference to the *Hind*III restriction fragment(s) that were illuminated by the TIC901 probe. There were essentially three different sized restriction fragments (polymorphisms) that were identified by hybridization to a *tic*901 probe. It was at first believed that only one of the three different fragments capable of hybridizing under these conditions to a *tic*901 probe was shown to be present in any one strain. Two restriction fragment polymorphisms exhibited a different signal intensity when hybridized with the *tic*901 probe compared to the signal produced by hybridization of this probe to the *tic*901 gene fragment from strain EG2158. This result suggests that there are at least three alleles of the *tic*901 coding sequence present in these thirty six (36) strains. Sequence analysis, as described herein, of each of these three restriction fragment length polymorphisms has allowed the identification of each of these related ORF's (open reading frames) encoding a TIC901 related protein contained within each sequence A comparison of these nucleotide sequences has been made as shown herein by aligning the sequences to the native TIC901 coding sequence to determine the extent of identity. In addition, the proteins encoded by the TIC901 and TIC1201 ORF's identified in these restriction fragments have been tested for insecticidal properties, and each exhibits coleopteran pest toxicity. It is therefore believed that the TIC407 and TIC417 proteins will also exhibit coleopteran insecticidal biological activity based on their high degree of relationship to the TIC901 and TIC1201 proteins. As a consequence of the significant identity of amino acid sequence relationship between the TIC407, TIC417, and TIC1201 proteins in comparison to TIC901, the proteins are described herein as amino acid sequence variants of each other TIC1201, for example, contains three fewer amino acids than TIC901 and contains 31 amino acid sequence variations in comparison to TIC901. Therefore, when compared to TIC901, the other TIC amino acid sequences contain amino acid variations that may contribute to different insecticidal spectrum and/or virulence and potency. Subsequent analysis of these and other strains using thermal amplification methods with primer pairs that have degeneracy incorporated into their sequences based on nucleotide sequence alignments of the coding sequences for the TIC901, TIC407, TIC1201, and TIC417 proteins resulted in amplicons that could correspond to a TIC407 coding sequence present in the genome of the strains EG5858, EG5552, and EG4332. These sequences may not have appeared using a TIC901 specific blot because it was determined that the TIC407 sequence was present on a large, approximately 18-19 kb, *Hind*III fragment which may not have transferred effectively to the blot membrane.

It is intended that the proteins of the present invention be used for agricultural purposes, i.e., for protecting plants from insect pest infestation, and more particularly for protecting plants from coleopteran insect pest infestation. As exemplified herein, the proteins of the present invention are useful for protecting plants at least from Colorado Potato Beetle infestation and at least from Corn Rootworm infestation. Plant protection can be achieved by topical application of a plant or plant parts such as by applying to the surface of the plant, i.e., the leaves, flowers, stems, stalks, and roots, a

composition in the form of a dust, spray, powder, or emulsion or other agriculturally acceptable excipient that contains an insecticidally effective amount of one or more of the proteins of the present invention. Alternatively, the agricultural excipient can contain, in addition to one or more of the proteins of the present invention, one or more additional insecticidal proteins effective for inhibiting the same spectrum of insect pests believed to be controlled by the proteins of the present invention such as Cry3 proteins, CryET33/34, CryET80/76, PS149B1 and other coleopteran inhibitory binary toxin proteins, VIP/MIS/WAR proteins, and the like, and/or proteins that are effective in controlling an altogether different spectrum of plant insect pests such as Cry1's, Cry2's, Cry9's, and the like. It is also within the scope of the present invention for an agricultural excipient as described above to contain other types of pesticidal compositions such as fungicides and/or acaricides and the like. Alternatively, and preferably, the plant itself will be transformed to contain one or more nucleotide sequences modified for improved expression of one or more of the proteins of the present invention *in planta* or expression of an insecticidal portion thereof, alone or in combination with other insecticidal agents such as those capable of being produced *in planta* using methods in molecular biology, including double stranded RNA mediated methods for suppressing genes in target pest cells

The TIC901, TIC1201, TIC407, TIC417, and TIC431 proteins are insecticidal compounds active against coleopteran insects such as CPB and rootworms. These proteins as set forth in SEQ ID NO:4, SEQ ID NO.6, SEQ ID NO:8, SEQ ID NO:10, and SEQ ID NO 33 respectively, and insecticidal fragments thereof, and related insecticidal proteins may be used as the active ingredient in insecticidal formulations useful for controlling coleopteran insects As used herein and with reference to insecticidal proteins that are related to these proteins, it is intended that related insecticidal proteins are those that are identified as homologs of these proteins or those that are identified as being encoded by a nucleotide sequence that hybridizes either under stringent hybridization conditions or specific hybridization conditions to all or a part of the native *Bacillus thuringiensis* sequence encoding the TIC901 protein, the TIC1201 protein, the TIC417 protein, the TIC407 protein, the TIC431 protein or an insecticidal portion thereof. Stringent conditions, as defined herein, comprise at least hybridization at 42°C followed by two washes for five minutes each at room temperature with 2X SSC, 0 1% SDS, followed by two washes for thirty minutes each at 65°C in 0 5X SSC, 0.1% SDS. Of course, one skilled in the art will recognize that, due to the redundancy of the genetic code, many other sequences are capable of encoding such related proteins, and those sequences, to the extent that they function to express insecticidal proteins either in *Bacillus* strains or in plant cells, are intended to be encompassed by the present invention, recognizing of course that many such redundant coding sequences will not hybridize under these conditions to the native sequences encoding TIC901, TIC1201, TIC407, TIC417, and/or TIC431. It should be understood that when referring to a TIC901 or related insecticidal protein or insecticidal fragment thereof, or when referring to a nucleotide sequence encoding a TIC901 or related insecticidal protein or insecticidal fragment thereof, that TIC901 is interchangeable and indistinguishable from reference to TIC407, TIC417, TIC431, and TIC1201 and the like, including the amino acid sequence SEQ ID NO:31 encoded by the nucleotide sequence as set forth in SEQ ID NO:30, which includes the full length insecticidal protein encoded therefrom by the full length coding

sequence for the protein, a part of which is exemplified by the amino acid sequence as set forth in SEQ ID NO:31.

Some nucleotide sequences that may be related to the nucleotide sequences of the present invention may not hybridize under stringent conditions, but may in fact hybridize to a *tic*901, *tic*1201, *tic*407, *tic*431, and/or *tic*417, or a related sequence using specific hybridization conditions. Such sequences may encode a protein that has at least about 30% amino acid sequence identity to the proteins of the present invention. Proteins exhibiting at least about 30% sequence identity may also exhibit very similar tertiary structures and so may also exhibit similar or related biological activity. With reference to the instant invention, such similarity in tertiary structure would include insecticidal biological activity. Specific hybridization conditions that enable the identification of more distantly related nucleotide sequences include a first hybridization at a low temperature, typically about 40°C or so, followed by washes as indicated above at room temperature to remove non-specifically bound probe, followed by exposure to film (in instances in which an investigator uses isotopic labeling means) or exposure to immunological reagents and chemical developing reagents to identify nucleotide fragments that hybridize to a specific gene probe. An indication that the hybridization is non-specific is one in which many hybridizing fragments are observed. In instances in which a number of hybridizing fragments are observed, the blot is washed one or more times, each time at a slightly higher temperature than the previous wash (for example, each wash could be accomplished at a temperature of about 5°C more than the previous wash) until only one or a few (two or three) hybridizing fragments are observed, this one or few fragments being exhibiting specific hybridization, and can then be cloned and sequences to determine the extent of homology and/or identity to the original probe sequence. Such sequences would be specifically related in that they encode proteins that have a related function, for example, insecticidal activity, to the protein encoded by the original sequence derived from the nucleotide probe.

Coding sequences are conceivable that function to encode all or an insecticidal portion of a TIC901 or related protein that do not hybridize under stringent conditions However, such sequences are derived from the native nucleotide sequence on the basis that the native nucleotide sequence is capable of being modified to exhibit a non-native sequence that still encodes the same or substantially the same native amino acid sequence, or that the native amino acid sequence is capable of being used along with a codon table to back-translate from the amino acid sequence, allowing the skilled artisan to arrive at a nucleotide sequence that encodes all or an insecticidal portion of a TIC901 or related protein All such sequences are intended to be within the scope of the present invention.

Insecticidal compositions can be produced from bacterial strains expressing the proteins of the present invention. A *B. thuringiensis* strain containing one or more nucleotide sequences encoding one or more TIC901 or related proteins and/or substantial equivalents thereof, can be cultured using known standard media and fermentation techniques. Because the proteins of the present invention are preferably secreted into the extracellular milieu, upon completion of the fermentation cycle, the bacteria expressing TIC901 or a homolog thereof can be harvested by first separating the *B. thuringiensis* along with any spores and crystals produced therein, from the spent fermentation broth by means well known in the art. The recovered *B. thuringiensis* spores and crystals can be formulated into

a wettable powder, a liquid concentrate, granules, solution, emulsion, spray, suspension, powder, foam, paste, aerosol, capsule or other finely or coarsely divided material or impregnant for natural or synthetic material, or other formulation, in admixture with suitable carriers, diluents, adjuvants, preservatives, dispersants, solvents, emulsifying agents, inert carriers and other components suitable for physically or chemically associating with plants or their locus, for oral uptake by target plant pathogens, and to facilitate handling and application for particular target pests. The formulation and application procedures are all well known in the art.

Formulated bait granules containing an attractant and spores and crystals of the *B. thuringiensis* isolates or concentrated spent fermentation media or insecticidal proteins purified from the spores or spent fermentation media, or recombinant microbes comprising the nucleotide sequences encoding TIC901 or related insecticidal proteins obtainable from the *B. thuringiensis* isolates disclosed herein, can be applied to the environment of the pest. The bait may be applied liberally since the toxin does not affect animals or humans. Product may also be formulated as a spray or powder. Pests pick the product up on their feet or abdomen and carry it back to the nest where other pests will be exposed to the toxin. The *B. thuringiensis* isolate or recombinant host expressing a nucleotide sequence or gene encoding a TIC901 or related protein of the present invention may also be incorporated into a bait or food source for the pest.

As would be appreciated by a person skilled in the art, the pesticidal concentration will vary widely depending upon the nature of the particular formulation, particularly whether it is a concentrate or to be used directly. The pesticide will be present in at least 1% by weight and may be 100% by weight. The dry formulations will have from about 1-95% by weight of the pesticide while the liquid formulations will generally be from about 1-60% by weight of the solids suspended or capable of being suspended in the liquid phase. The formulations will generally have from about 10^2 to about 10^4 cells/mg or from about 5 to about 100 parts per million of the active component insecticidal protein, i.e., the TIC901 protein, amino acid sequence variant thereof, insecticidal portion or fragment thereof, or homolog thereof such as TIC431, TIC1201, TIC417, and TIC407 These formulations will be administered at from about 50 mg (liquid or dry) to about 1 kg or more per hectare. The formulations can be applied to the environment of the coleopteran pests, e.g., plants, soil, or water by spraying, dusting, sprinkling, or the like, and can also be applied to the surfaces of seeds as a seed treatment or seed coating and can be permeated into the seed coat and/or cotyledon(s). One skilled in the art will also recognize that combinations of the proteins of the present invention when combined together in a composition or formulation, may also have particularly useful and beneficial effects, for example, providing a broader host range for controlling insect infestation, or increasing the virulence and potency of a composition intended for use as an insecticidal agent.

It is well within the skill of the art to construct a variant or modified nucleotide sequence that encodes the insecticidal protein of the present invention, or an insecticidal fragment thereof, or an insectidical amino acid sequence variant thereof that exhibits improved insecticidal activity compared to the native amino acid sequence, and place that nucleotide sequence into an expression cassette that functions in plants to cause the transcription of the coding sequence into a messenger RNA that is subsequently translated in the cells of the plant such that an insecticidally effective amount of the

insecticidal protein is produced within the plant tissues. It also within the skill of the art to transform a plant cell, preferably a corn, cotton, soybean, canola, rice, wheat, oat, milo, grass, forage plant, fruit tree, ornamental flower, tomato, potato, carrot, kale, and tobacco plant cell and the like with a nucleotide sequence embedded within a plant functional expression cassette, to select for cells that contain the sequence and are expressing insecticidally effective amounts of a TIC901 protein, (and/or amino acid sequence variant thereof, insecticidal portion or fragment thereof, or homolog thereof such as TIC431, TIC1201, TIC417, and TIC407) and to produce plants from such transformed cells. One skilled in the art would know to use electroporation, infusion, ballistic methods, or *Agrobacterium tumefaciens* mediated methods and the like for introducing the nucleotide sequences of the present invention or modifications thereof into a plant cell. Such methods are well known in the art.

The term "variant or modified" with reference to nucleotide sequences is intended to refer to nucleotide sequences which encode the same toxins or which encode equivalent toxins having similar insecticidal activity, the term "equivalent toxin" referring to a toxin exhibiting the same, essentially the same, or improved biological activity against the target pests as the claimed native or referent toxin. A variant or modified nucleotide sequence intended for use in dicot plants would encode substantially the same amino acid sequence as the native coding sequence, i e , the coding sequence found in nature, but would comprise a total combined GC composition from about 49 to about 58 percent, and would avoid utilizing the least preferred codon used by the intended dicot plant, determined by compiling such preference and usage frequencies from a consortium of coding sequences derived from one or more individual dicot plant species intended to be transformed with the variant or modified nucleotide sequence. A variant or modified nucleotide sequence intended for use in a monocot plant would also encode substantially the same amino acid sequence as the native coding sequence, and would comprise a total combined GC composition from about 52 to about 64 percent or more, and would also avoid utilizing the least preferred codon for encoding any amino acid as determined by compiling such preference and usage frequencies from a consortium of coding sequences derived from one or more individual monocot plant species intended to be transformed with the variant or modified nucleotide sequence. Codon usage frequency is intended to refer to the number of times, on average, that a particular codon is used in a coding sequence. For a particular plant species, a codon that is intended to cause the incorporation of a particular amino acid into a nascent amino acid sequence will be utilized on average with some relative fixed frequency. For amino acids that utilize only two codons, this frequency is generally about fifty-fifty, i.e., each codon being used about half the time, unless one of the codons utilizes a substantially greater number of purines or pyrimidines that are not typically representative of the GC content of the particular plant species. For *Bacillus* species, for example, coding sequences generally are from about 60 to about almost 70 per cent AT. Codon usage in *Bacillus* species is biased toward the use of codons that are enriched for the presence of A or T in a particular codon, and more particularly with A or T in the third base position of any particular codon. Therefore, codons that primarily utilize G or C are used in a native and/or naturally occurring *Bacillus* coding sequence with a much lower frequency than codons that contain A's or T's. Therefore, when producing a variant or modified nucleotide sequence intended for use in a particular plant, monocot or dicot, it is important to ensure that appropriate attention is given to avoiding using the least preferred

codon with any great frequency for that particular plant In fact, for monocots, it is preferred that the coding sequence mimic the GC distribution found in most native or naturally occurring monocot plants; that being a preferred about 65% GC for about the first 10% of the coding sequence, tapering down to about 60%GC for the second 10-15% of the coding sequence, and leveling off to about a 50-55% GC for the middle one half or more of the coding sequence, and then gradually increasing the GC % of the coding sequence up to about 60-64% through the last 15-20% of the coding sequence. This distribution of GC% seems to mimic as closely as possible the GC% distribution of a naturally occurring monocot gene and therefore it is believed that a synthetic or artificially produced gene or coding sequence should resemble this architecture as well

As used herein, "synthetic coding sequences" or "non-naturally occurring coding sequences" encoding the *B. thuringiensis* TIC901 proteins or homologs or derivatives thereof as insecticidal toxins of the present invention are those prepared in a manner involving any sort of genetic isolation or manipulation that results in the preparation of a coding sequence that encodes a TIC901 insecticidal protein or related amino acid sequence or homolog or variant or the substantial equivalent thereof including coding sequences that encode at least an insecticidal portion of a TIC901 protein, a TIC431 protein, a TIC1201 protein, a TIC407 protein, or a TIC417 protein. This includes isolation of the coding sequence from its naturally occurring state, manipulation of the coding sequence as by codon modification (as described herein), chemical synthesis such as phosphoramidite chemistry and the like, or site-specific mutagenesis (as described herein), truncation of the coding sequence or any other manipulative or isolative method.

As used herein, the phrase "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity. A sequence which is identical at every position in comparison to a reference sequence is said to be identical to the reference sequence and vice-versa A first nucleotide sequence when observed in the 5' to 3' direction is said to be a "complement" of a second or reference nucleotide sequence observed in the 3' to 5' direction if the first nucleotide sequence exhibits complete complementarity with the second or reference sequence. As used herein, nucleic acid sequence molecules are said to exhibit "complete complementarity" when every nucleotide of one of the sequences read 5' to 3' is complementary to every nucleotide of the other sequence when read 3' to 5'. A nucleotide sequence that is identical at every position when read 5' to 3' in comparison to a reference nucleotide sequence read 5' to 3' is said to be identical to the reference sequence and vice-versa. A nucleotide sequence that is complementary to a reference nucleotide sequence will exhibit a sequence identical to the reverse complement sequence of the reference nucleotide sequence. These terms and descriptions are well defined in the art and are easily understood by those of ordinary skill in the art.

As used herein, "substantial homology", with reference to nucleic acid sequences, refers to nucleotide sequences that hybridize under stringent conditions to the coding sequences as set forth in SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO.9, or SEQ ID NO:32 or the complements thereof. Sequences that hybridize under stringent conditions to SEQ ID NO:3, SEQ ID NO.5, SEQ ID NO:7, SEQ ID NO:9, or SEQ ID NO:32 or the complements thereof, in particular from about nucleotide position 153 to about nucleotide position 1,253 of SEQ ID NO:3, and more particularly from about nucleotide position 282 to about nucleotide position 1,253 of SEQ ID NO:3, or from about nucleotide position 437 to about nucleotide position 1621 of SEQ ID NO:5, more particularly from about nucleotide position 530 to about nucleotide position 1621 of SEQ ID NO:5, and even more particularly from about nucleotide position 659 to about nucleotide position 1621 of SEQ ID NO:5; or from about nucleotide position 196 to about nucleotide position 1299 of SEQ ID NO.7, or more particularly from about nucleotide position 325 to about nucleotide position 1299 of SEQ ID NO:7; or from about nucleotide position 215 to about nucleotide position 1306 of SEQ ID NO.9, or more particularly from about nucleotide position 344 to about nucleotide position 1306 of SEQ ID NO:9, or from about nucleotide position 1 through about nucleotide position 1092 as set forth in SEQ ID NO:32, or more particularly from about nucleotide position 130 to about nucleotide position 1092 as set forth in SEQ ID NO:32, contain one or more contiguous nucleotide sequences that are sufficiently identical to one or more contiguous nucleotide sequences of SEQ ID NO:3, SEQ ID NO.5, SEQ ID NO:7, SEQ ID NO:32, or SEQ ID NO:9, such that an alignment is able to take place between the two sequences, and the two sequences are then able, under stringent conditions, to form hydrogen bonds with corresponding bases on the opposite strand to form a duplex molecule that is sufficiently stable under the stringent conditions for a long enough period of time to be detectable using methods well known in the art. Such substantially homologous sequences are preferably from about 67% to about 70% identical, or more preferably from about 80% to about 85% identical, or most preferable from about 90% to about 95% identical, to about 99% identical or greater to the referent nucleotide sequences as set forth in SEQ ID NO.32, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO 7, or SEQ ID NO:9, or the complements thereof. In addition, nucleotide sequences that encode insecticidal proteins isolatable from *Bacillus thuringiensis* or other *Bacillus* species strains and the like, that hybridize under stringent conditions to SEQ ID NO.2 are also envisioned to exhibit substantial homology with the above listed referent nucleotide sequences that hybridize under stringent conditions to the *tic*901, *tic*1201, *tic*407, *tic*417, and *tic*431 coding sequence as set forth in SEQ ID NO:3 or SEQ ID NO:5, SEQ ID NO 7, SEQ ID NO:9, and SEQ ID NO 32 respectively or the complements thereof. Such nucleotide sequences are referred to herein as homologs of SEQ ID NO:3 and the like and comprise SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, and SEQ ID NO:32 and related sequences and homologues thereof.

With reference to polypeptide sequences, the terms "substantial identity" or "substantial similarity" refers to polypeptides which exhibit a substantial amino acid sequence identity or a substantial amino acid sequence similarity to a referent amino acid sequence, particularly in view of the fact, as described herein, that certain amino acids may be substituted by other amino acids based on hydropathicity or hydrophilicity indices and still result in a protein with similar biological activity, *i.e.* still obtain a biological functionally equivalent protein. Therefore, an amino acid sequence exhibiting a

substantial identity or a substantial similarity to a referent polypeptide or amino acid sequence would exhibit from about 30% to about 50% amino acid sequence identity to the referent sequence, and more preferably exhibit from about 70% to about 80% amino acid sequence identity to the referent sequence, more preferably from about 86% to about 90% amino acid sequence identity to the referent sequence, and even more preferably from about 95% to about 99% amino acid sequence identity to the referent polypeptide sequence. More specifically, the inventors envision peptides exhibiting insecticidal activity that are related to the peptides of the present invention exhibit substantial peptide identity or substantial peptide similarity to the peptides of the present invention and exhibit at least from about 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, and 99 percent identity or similarity to the referent polypeptide sequences as set forth herein and selected from the group consisting of SEQ ID NO:1, SEQ ID NO.4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, and SEQ ID NO:33.

With reference to the proteins of the instant application, the terms "variant amino acid sequence", or "amino acid sequence variant", or "modified amino acid sequence variant" are intended to refer to amino acid sequences that are substantially equivalent to the amino acid sequences of the present invention. For example, a protein produced by the introduction of a restriction site for convenience of molecular manipulations into a coding sequence of the present invention that results in the addition or subtraction of one or more codons without otherwise (1) disrupting the native coding sequence, (2) disrupting the native open reading frame, and (3) disrupting the insecticidal biological activity of the protein, would constitute (a) a variant amino acid sequence compared to the native insecticidal toxin, (b) an amino acid sequence variant compared to the native insecticidal toxin, or (c) a modified amino acid sequence variant compared to the native insecticidal toxin. One skilled in the art would recognize that there are other types of modifications that can be made to the amino acid sequence of the present invention without disrupting the biological activity of the protein. The use of the term "disrupting", with reference to biological activity, is intended to refer to modifications of the native amino acid sequence by insertion or deletion of one or more amino acids, or exchange or substitution of one amino acid for another, which do not decrease the insecticidal biological activity of the protein Insertions, deletions, and substitutions are within the scope of the present disclosure to the extent that the resulting amino acid sequence variant exhibits insecticidal activity no less than that of the native insecticidal protein. Chimeras of the proteins disclosed herein, fusions of the proteins or parts of the proteins disclosed herein, and permuteins of the proteins disclosed herein are specifically contemplated.

Proteins that are substantially equivalent to the proteins of the instant application are intended to be biologically functionally equivalent. As used herein, the phrase "biological functional equivalents", with respect to the insecticidal proteins of the present invention, are peptides, polypeptides and proteins that contain a sequence or moiety exhibiting sequence similarity to the novel peptides of the present invention, such as a TIC901 protein or insecticidal fragment thereof, or a TIC1201, TIC410, TIC407, or TIC431 protein or insecticidal fragments thereof, and that exhibit the same or similar functional properties as that of the polypeptides disclosed herein, including insecticidal activity. Biological equivalents also include peptides, polypeptides and proteins that react with, *i.e.*

specifically bind to antibodies raised against epitopes present on or within TIC901 and related proteins such as TIC1201, TIC417, TIC407, and TIC431and insecticidal fragments thereof, and that exhibit the same or similar binding or reactive activity, including to both monoclonal and polyclonal antibodies.

It is also contemplated that the proteins of the present invention could be useful for protecting dicot plants from insect infestation. Such infestations could be the result of coleopteran, dipteran, lepidopteran, or even infestation by mites, mealworms, grubs, or a wide variety of insects that injure the plant by piercing the plant tissues and extracting the nutrients intended for plant growth and development. Modifications to the primary amino acid sequence of the proteins of the present invention could result in a protein that exhibits a host range different from that of the native protein.

The proteins of the present invention, because of their localization into the extracellular space when expressed by *Bacillus* strains, may be useful for targeting other proteins for localization into the extracellular space. For example, the skilled artisan would know to link a first protein that is not normally secreted into the extracellular space to a second protein that is normally secreted into the extracellular space in order to achieve the localization of the first protein into the extracellular space. The proteins of the present invention could be fused by any number of means well known in the art to one or more insecticidal toxins such as crystalline delta-endotoxins to form a chimeric protein that is targeted for secretion into the extracellular space surrounding a particular host cell. It is even envisioned that the secretion event itself could lead to the separation of the two protein parts such that two separate and distinct insecticidal proteins are released into the extracellular space surrounding a particular host cell The two proteins could either (1) both be toxic to the same insect species but effectuate their insecticidal activity using different modes of action, or (2) each be toxic to different insect species. It is conceivable that any number of insecticidal proteins could be linked end-to-end to the proteins of the present invention to form multimeric chimeras that are targeted to the extracellular space surrounding a particular host cell. Such "other" proteins conceivably could be green fluorescent and related proteins and variants, kinases and phosphatases for modulating cell signaling processes, nucleases, lipases, herbicide tolerance proteins expressed from genes such as *gox*, various *epsps* homologues, *bar* and homologues and the like, PhnO, NptII, Aad, and the like. All of these proteins could be used as selectable markers as well, particularly when linked to a gene encoding one or more of the proteins of the present invention, to track the presence of the genes encoding one or more of the proteins of the present invention in a plant or other host cell

The proteins of the present invention could be targeted for import into a subcellular organelle. For example, a first nucleotide sequence encoding a chloroplast or plastid targeting sequence could be operably linked or fused to a second nucleotide sequence encoding an insecticidal protein of the present invention to produce a chimeric precursor protein that is targeted for insertion into the chloroplast or plastid within a plant cell. Expression of such chimeric proteins would result in the import of the proteins of the present invention into the plant chloroplast or plastid, resulting in the localization of the insecticidal toxin or insecticidal fragment thereof into the chloroplast or plastid. Additionally, a nucleotide sequence encoding one or more proteins of the present invention could be localized to the chloroplast or plastid for expression. The localization of the nucleotide sequences to the plastid or chloroplast could result in the incorporation of the nucleotide sequences into the chloroplast or plastid

genome, or could result in the presence of an autonomously replicating nucleic acid sequence encoding the protein of the present invention In either sense, the proteins of the present invention would be localized to the chloroplast or plastid. As used herein therefore, the phrase "chloroplast or plastid localized" refers to a biological molecule, either polynucleotide or polypeptide, which is positioned within the chloroplast or plastid such that the molecule is isolated from the cellular cytoplasmic milieu, and functions within the chloroplast or plastid cytoplasm to provide the effects claimed in the instant invention. Localization of a biological molecule to the chloroplast or plastid can occur, with reference to polynucleotides, by artificial mechanical means such as electroporation, mechanical microinjection, or by polynucleotide coated microprojectile bombardment, or with reference to polypeptides, by secretory or import means wherein a natural, synthetic, or heterologous plastid or chloroplast targeting peptide sequence is used which functions to target, insert, assist, or localize a linked polypeptide into a chloroplast or plastid.

As used herein, the phrase "operatively linked" or "operably linked" refers to nucleic acid coding segments connected in frame so that the properties of one influence the expression of the other. These phrases and groups of words can also be used to refer to amino acid sequences which exhibit some function when linked to another amino acid sequence, for example, a signal peptide when linked to a protein of interest is referred to as being operably linked to the protein of interest for the purpose of targeting the protein of interest to the secretory apparatus of the host cell in which the protein is produced, or to a subcellular compartment such as an endoplasmic reticulum, a chloroplast or a plastic, a mitochondrion, a vacuole, the nucleus or nucleolus, or other subcellular compartment and the like

For the purposes of the present invention, the word "gene" refers to a nucleotide sequence that contains an open reading frame encoding a TIC901 protein, a TIC1201 protein, a TIC417 protein, a TIC407 protein, a TIC431 protein, or an insecticidal fragment thereof, or an amino acid sequence variant thereof, or a related protein homolog or insecticidal fragment thereof or amino acid sequence variant thereof that is at least operably linked to a promoter sequence and a transcription termination sequence, wherein the promoter and transcription termination sequences are functional in the host cell in which the protein is produced. As used herein, "structural gene" refers to a gene that is expressed to produce a polypeptide. A structural gene of the present invention can contain, in addition to promoter and transcription termination sequences, five prime untranslated sequences, intronic sequences, and enhancer elements that function in plants in particular, and preferably those that are derived from maize or other monocotyledonous plants that, when linked together in proper sequence with one or more coding sequences of the present invention result in improved levels of expression in particular plant tissues, and preferably result in enhanced expression in root tissues of maize plants.

Nucleotide sequence information provided by the present invention allows for the preparation of relatively short DNA sequences, referred to herein as probes or primers, having the ability to specifically hybridize to sequences of the selected polynucleotides disclosed herein. Such nucleic acid probes of an appropriate length are prepared based on a consideration of selected polypeptide sequences encoding the insecticidal polypeptides of the present invention, *e.g.*, a sequence such as that shown in SEQ ID NO:2, all or a probe specific part of SEQ ID NO:3 from about nucleotide 153 to about nucleotide 1,253; all or a probe specific part of SEQ ID NO.5 from about nucleotide 530 to about

nucleotide 1621, all or a probe specific part of SEQ ID NO:7, all or a probe specific part of SEQ ID NO:9, all or a probe specific part of SEQ ID NO:32, or all or a probe specific or primer specific part of the sequences selected from the group consisting of SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO.20, SEQ ID NO:21, and SEQ ID NO:22, and complements thereof and the like. Reference to the phrase "all or a probe specific part of" is intended to refer to a probe comprising at least a 15 to 50, more or less, contiguous nucleotide sequence selected from the group of nucleotides set forth in a particular referent sequence such as SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO.7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO.18, SEQ ID NO.19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO.22, and SEQ ID NO·32 and complements thereof and the like. The ability of such nucleic acid probes to specifically hybridize to a nucleotide sequence encoding an insecticidal polypeptide sequence lends to them particular utility in a variety of embodiments. Most importantly, the probes may be used in a variety of assays for detecting the presence of complementary sequences in a given sample.

In certain embodiments, it is advantageous to use oligonucleotide primers. The sequence of such primers is designed using a polynucleotide of the present invention for use in detecting, amplifying or modifying a defined segment of an insecticidal protein coding sequence from *B. thuringiensis* or from *Bacillus sphaericus* and the like using thermal amplification technology. Segments of nucleotide sequences related to the polynucleotides encoding the insecticidal polypeptides of the present invention may also be isolated and characterized using thermal amplification technology and such primers.

To provide certain of the advantages in accordance with the present invention, a preferred nucleic acid sequence employed for hybridization studies or assays or as a primer includes sequences that are complementary to at least a 14 to 30 or more contiguous stretch of nucleotides of a polynucleotide sequence encoding all or a part of an insecticidal protein of the present invention, such as that shown in SEQ ID NO:2, SEQ ID NO 3, SEQ ID NO:5, SEQ ID NO 7, SEQ ID NO:9, SEQ ID NO·11, SEQ ID NO:12, SEQ ID NO·13, SEQ ID NO·15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO·19, SEQ ID NO:20, SEQ ID NO.21, SEQ ID NO:22, and SEQ ID NO:32 and complements thereof and the like

A primer or probe size of at least 14 nucleotides in length helps to ensure that the fragment will be of sufficient length to form a duplex molecule that is both stable and selective. Molecules having complementary sequences over segments greater than 14 bases in length are generally preferred. In order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained, one will generally prefer to design nucleic acid molecules having *tic*901-complementary sequences and the like of 14 to 20 nucleotides, or even longer where desired; or having *tic*1201-complementary sequences and the like of 14 to 20 nucleotides, or even longer where desired; *tic*417-complementary sequences and the like of 14 to 20 nucleotides, or even longer where desired; *tic*407-complementary sequences and the like of 14 to 20 nucleotides, or even longer where desired; and *tic*431-complementary sequences and the like of 14 to 20 nucleotides, or even longer where desired. Such fragments may be readily prepared by, for example, directly

synthesizing the fragment by chemical means, by application of nucleic acid reproduction technology, or by excising selected DNA fragments from recombinant sequences localized in plasmids or other vectors containing appropriate inserts and suitable restriction sites.

The inventors herein have also designed degenerate universal probes and primers for use in identifying naturally occurring nucleotide sequences encoding amino acid sequences derived from insecticidal proteins that are homologues of the proteins of the present invention. The nucleotide sequences identified using the exemplified probes and primers set forth in SEQ ID NO:23-SEQ ID NO:29 hybridize under stringent conditions to the nucleotide sequences encoding secreted insecticidal proteins as set forth herein.

The amino acid sequences alignment as shown in Figure 1 provided a basis for identifying amino acid sequences that are highly conserved between the four aligned insecticidal precursor proteins. For example, the amino acid sequence as set forth in SEQ ID NO 4 from about amino acid seventy-five (75) through about amino acid eighty-three (83) is a sequence that is conserved in both sequence and position within the primary sequences of the TIC901, TIC1201, TIC407, TIC417, and TIC431 proteins. This is described herein as a 'first conserved amino acid sequence'. The amino acid sequence from about amino acid one-hundred-forty-seven (147) through about one-hundred-fifty-three (153) as set forth in SEQ ID NO:4 is also conserved in both sequence and position within the primary sequences of the TIC901, TIC1201, TIC407, TIC431, and TIC417 proteins This is described herein as a 'second conserved amino acid sequence'. Similarly, the amino acid sequence from about amino acid two-hundred-seventy-five (275) through about amino acid two-hundred-eighty-three (283) as set forth in SEQ ID NO.4 is also conserved in both sequence and position within the primary sequences of the TIC901, TIC1201, TIC407, TIC431, and TIC417 proteins. This is described herein as a 'third conserved amino acid sequence'. These sequences each correspond to substantially conserved but slightly degenerate nucleotide sequences in the respective coding sequence for each protein that can be used either as a probe sequence or as a primer sequence for identifying the presence of a nucleotide segment homologous to a sequence comprising at least a fourteen base sequence selected from the group consisting of SEQ ID NO.3, SEQ ID NO.5, SEQ ID NO.7, SEQ ID NO:32, and SEQ ID NO:9 or the complement thereof. For example, a thermal amplification reaction that uses a degenerate primer sequence, the synthesis of which is based on the compiled sequences selected from SEQ ID NO:3, SEQ ID NO·5, SEQ ID NO:7, SEQ ID NO:32, and SEQ ID NO.9, and the complements thereof when inverse thermal amplification is contemplated, corresponding to the sequence coding for the 'first conserved amino acid sequence' as described above, comprises a twenty-six-mer (26-mer) oligonucleotide corresponding to the first conserved nucleotide sequence, for example as set forth in SEQ ID NO:3 from about nucleotide position three-hundred-seventy-five (375) to about nucleotide position four-hundred-one (401), would be one of the degenerate sequences that could be used for probing a sample for the presence of a nucleotide sequence homologue corresponding to a sequence that hybridizes to the corresponding sequence within SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO·7, SEQ ID NO:32, and SEQ ID NO:9. Alternatively, the oligonucleotide sequence could be used as one of a pair of oligonucleotide primers in a thermal amplification reaction for producing an amplicon sequence that would hybridize to the corresponding sequences within one or more of the sequences as

set forth herein, for example, within a sequence comprising SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO.7, SEQ ID NO:32, and SEQ ID NO:9, under stringent hybridization conditions.

The inventors herein have therefore constructed sets of primers and probes that can be used alone or in combination with each other for identifying sequences that are related to the proteins of the present invention, including TIC901, TIC1201, TIC407, TIC431, and TIC417, that encode secreted insecticidal proteins, and that hybridize to one or more of the sequences disclosed herein under stringent hybridization conditions. One set of primers consists of degenerate oligonucleotide sequences that correspond to sequences as set forth in SEQ ID NO.23 through SEQ ID NO.25. SEQ ID NO:23 corresponds to a set of degenerate oligonucleotide sequences corresponding to from about nucleotide position three-hundred-seventy-five (375) to about nucleotide position four-hundred-one (401) as set forth in SEQ ID NO.3, from about nucleotide position seven-hundred-fifty-two (752) to about nucleotide position seven-hundred-seventy-eight (778) as set forth in SEQ ID NO:5, from about nucleotide position three-hundred-ninety-one (391) to about nucleotide position four-hundred-seventeen (417) as set forth in SEQ ID NO:7, from about nucleotide position four-hundred-thirty-seven (437) to about nucleotide position four-hundred-sixty-three (463) as set forth in SEQ ID NO 9, and from about nucleotide position two-hundred-twenty-three (223) to about nucleotide position two-hundred-forty-nine (249) as set forth in SEQ ID NO:32. All possible combinations of nucleotide sequences encoding the corresponding 'first conserved amino acid sequence' described herein above that would reasonably be expected to be present within a *Bacillus* species are contemplated therein as set forth in SEQ ID NO:23 SEQ ID NO 24 and SEQ ID NO:25 are degenerate oligonucleotide sequences comprising subsets of the sequences corresponding to SEQ ID NO:23, and like SEQ ID NO:23, contain codons that are biased toward the codon usage preference of *Bacillus* coding sequences.

A second set of primers and probes that can be used for identifying sequences as described herein that are related to SEQ ID NO 3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO.9, and SEQ ID NO:32 and that encode secreted insecticidal proteins is contemplated by the degenerate sequences as set forth in SEQ ID NO:26. These oligonucleotide sequences correspond to the range of anticipated nucleotide sequences that would be preferred by a *Bacillus* species for encoding the amino acid sequence described herein above as the 'second conserved amino acid sequence', and further correspond to from about nucleotide position five-hundred-ninety-one (591) through about nucleotide position six-hundred-eleven (611) as set forth in SEQ ID NO:3, from about nucleotide position nine-hundred-sixty-eight (968) to about nucleotide position nine-hundred-eighty-eight (988) as set forth in SEQ ID NO:5, from about nucleotide position six-hundred-sever (607) to about nucleotide position six-hundred-twenty-seven (627) as set forth in SEQ ID NO:7, from about nucleotide position six-hundred-fifty-three (653) to about nucleotide position six-hundred-seventy-three (673) as set forth in SEQ ID NO.9, and from about nucleotide position four-hundred-thirty-nine (439) to about nucleotide position four-hundred-fifty-six (456) as set forth in SEQ ID NO:32 All possible combinations of nucleotide sequences encoding the corresponding 'second conserved amino acid sequence' described herein above that would reasonably be expected to be present within a *Bacillus* species are contemplated therein as

set forth in SEQ ID NO:26 and the codons selected for incorporation into the degenerate oligonucleotide sequence are biased toward the codon usage preference of *Bacillus* coding sequences.

Still, a third set of primers and probes that can be used for identifying sequences as described herein that are related to SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, and SEQ ID NO:32 and that encode secreted insecticidal proteins are contemplated by the degenerate sequences as set forth in SEQ ID NO 27 - SEQ ID NO.29. These oligonucleotide sequences correspond to the range of anticipated nucleotide sequences that would be preferred by a *Bacillus* species for encoding the amino acid sequence described herein above as the 'third conserved amino acid sequence', and further correspond to the reverse complement of the nucleotide sequence from about nucleotide position nine-hundred-seventy-five (975) to about nucleotide position one-thousand-one (1,001) as set forth in SEQ ID NO:3, the reverse complement of the nucleotide sequence from about nucleotide position one-thousand-three-hundred-fifty-two (1,352) to about nucleotide position one-thousand-three-hundred-seventy-eight (1,378) as set forth in SEQ ID NO 5, the reverse complement of the nucleotide sequence from about nucleotide position nine-hundred-ninety-one (991) to about nucleotide position one-thousand-seventeen (1,017) as set forth in SEQ ID NO.7, the reverse complement of the nucleotide sequence from about nucleotide position one-thousand-thirty-seven (1,037) to about nucleotide position one-thousand- sixty-three (1,063) as set forth in SEQ ID NO 9, and the reverse complement of the nucleotide sequence from about nucleotide position eight-hundred-twenty-three (823) to about nucleotide position eight-hundred-forty-six (846) as set forth in SEQ ID NO 32. All possible combinations of oligonucleotide sequences encoding the corresponding 'third conserved amino acid sequence' described herein above that would reasonably be expected to be present within a *Bacillus* species are contemplated therein as set forth in SEQ ID NO:27. SEQ ID NO:28 and SEQ ID NO 29 are also degenerate nucleotide sequences comprising subsets of the sequences corresponding to SEQ ID NO:27, and like SEQ ID NO 27, contain codons that are biased toward the codon usage preference of *Bacillus* coding sequences.

Any of the sequences contemplated by the sequences as set forth in SEQ ID NO 23 through SEQ ID NO:29 can be used alone as a probe for identifying the presence of a nucleotide sequence in a sample that encodes a secreted insecticidal protein related to any of the proteins exemplified herein, in particular a nucleotide sequence that hybridizes to one or more of the nucleotide sequences of the present invention, including but not limited to SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO.5, SEQ ID NO:7, SEQ ID NO.9, SEQ ID NO:32 and any of the nucleotide sequences set forth herein as probes and/or primers selected from the group consisting of SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO.22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO 25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, and SEQ ID NO:29.

Alternatively, and preferably, various combinations of the probes and primers set forth herein can be used together in a thermal amplification reaction to produce one or more amplicons that are diagnostic for the presence of a nucleotide sequence in a sample that encodes a secreted insecticidal protein related to the proteins of the present invention in that the nucleotide sequence encoding the insecticidal protein hybridizes to one or more of the nucleotide sequences exemplified herein

comprising SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO.5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO.11, SEQ ID NO:12, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO.24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO.28, SEQ ID NO:29, and SEQ ID NO:32 or the complements thereof, under stringent hybridization conditions. For example, combining one or more of the nucleotides as set forth in SEQ ID NO·23-25, such as primer prJWP200 (SEQ ID NO:23), with one or more of the nucleotides as set forth in SEQ ID NO:27-29, such as primer prJWP204 (SEQ ID NO:27), each at a concentration of at least about 1 pico-mole per micro-liter in a thermal amplification reaction containing 1X TAQ amplification buffer, 0.2 molar each deoxy-nucleotide tri-phosphate (dATP, dTTP, dCTP, and dGTP), 2 millimolar $MgCl_2$, 2 units TAQ polymerase, and from about ten (10) to about one hundred (100) nano-grams of a sample containing DNA template in which one or more of the sequences of the present invention encoding a secreted insecticidal protein or fragment thereof is present, results in the synthesis of a double-stranded DNA fragment that is an amplicon comprising from about 600 to about 650 base pairs, more preferably from about 615 to about 630 base pairs, and even more preferably from about 623 to about 626 base pairs An amplicon of this size is diagnostic for the presence of a nucleotide sequence in a sample encoding all or part of an secreted insecticidal protein that is related to one or more of the proteins of the present invention, and is of a size that can reasonably be anticipated, in particular if the thermal amplification cycle conditions consist of an initial denaturation of about 2 minutes at 94°C followed by 35 cycles of a denaturation step of 30 seconds at 94°C, an annealing step of 30 seconds at 50°C, and an extension step of 45 seconds at 72°C followed by a final extension step of 7 minutes at 72°C. The temperature of the annealing step should be decreased by 0.3°C for each successive cycle so that the final annealing temperature is about 39.8°C. Also, combining one or more of the nucleotides as set forth in SEQ ID NO:26 such as prJWP203 with one or more primers as set forth in SEQ ID NO:27-29 in a similar thermal amplification reaction would result in the synthesis of an amplicon of from about 400 to about 415 base pairs, or an amplicon of from about 400 to about 415 base pairs, or from about 405 to about 410 base pairs, that is diagnostic for the presence of a nucleotide sequence in a sample that encodes a secreted insecticidal protein related the proteins of the present invention.

Following these conditions, and using the primers prJWP200 (SEQ ID NO:23) and prJWP204 (SEQ ID NO:27) along with genomic DNA from strain EG2158, resulted in the production of an amplicon segment corresponding to about 620 base pairs Several individual clones representing this segment were isolated at random from the amplification reaction and the nucleotide sequence of each individual clone was obtained. As expected, a first sequence identical to the corresponding nucleotide sequence for TIC901 was identified in the clone population (not including the primer sequences at either end of the clone, from about nucleotide position four-hundred-two (402) through about nucleotide position nine-hundred-seventy-four (974) as set forth in SEQ ID NO:3). Also, and as expected, a second sequence identical to the corresponding nucleotide sequence for TIC417 was identified in the clone population (not including the primer sequences at either end of the clone, from about nucleotide position four-hundred-sixty-four (464) through about nucleotide position one-thousand-thirty-six (1,036) as set forth in SEQ ID NO:9).

Surprisingly, a third sequence (SEQ ID NO:30) was also identified in the clone population that did not correspond identically to any of the sequences set forth herein but which was substantially similar in nucleotide sequence to each of the sequences of the present invention, including the corresponding coding sequence within the *tic*901, *tic*1201, *tic*407, *tic*417, and *tic*431 coding sequences. As exemplified herein, it was quite unexpected to find the *tic*417 coding sequence in the strain EG2158 genomic DNA. However, it was even more surprising to identify yet a third nucleotide sequence from the EG2158 genome that likely corresponds to a nucleotide segment that encodes yet a third secreted insecticidal protein different from TIC901, TIC1201, TIC407, TIC417, and TIC431, but which is sufficiently similar in sequence to the proteins of the present invention to be classified as one of the species within the genus of secreted insecticidal proteins encoded by a nucleotide sequence that hybridizes to one or more of the sequences set forth herein, and is exemplary of the novelty and utility of the degenerate oligonucleotide probes and primers exemplified herein for use in identifying sequences that encode secreted insecticidal proteins and that hybridize under stringent conditions to the related *tic*901, *tic*1201 *tic*07, *tic*417, and *ttc*431 coding sequences as set forth herein.

The amino acid sequence encoded by the uninterrupted open reading frame as set forth in SEQ ID NO:30, which has had the twenty-six-mer degenerate oligonucleotide sequences deleted from both the 5' and 3' ends, is set forth in SEQ ID NO 31. The amino acid sequence set forth in SEQ ID NO:31 is substantially similar to the amino acid sequence of TIC901 from about amino acid position eighty-five (85) through about amino acid position two-hundred-seventy-four (274) as set forth in SEQ ID NO 4, containing only two (2) amino acids that are different from the analogous sequence in SEQ ID NO.4, corresponding to an about 98.9% identity between SEQ ID NO.31 and the corresponding sequence in SEQ ID NO.4. The amino acid sequence set forth in SEQ ID NO.31 is substantially similar to the amino acid sequence of TIC1201 from about amino acid position eighty-five (85) through about amino acid position two-hundred-seventy-four (274) as set forth in SEQ ID NO 6, containing only thirteen (13) amino acids that are different from the analogous sequence in SEQ ID NO:6, corresponding to an about 93.2% identity between SEQ ID NO:31 and the corresponding sequence in SEQ ID NO.6. The amino acid sequence set forth in SEQ ID NO:31 is substantially similar to the amino acid sequence of TIC417 from about amino acid position eighty-five (85) through about amino acid position two-hundred-seventy-four (274) as set forth in SEQ ID NO:10, containing only thirty (30) amino acids that are different from the analogous sequence in SEQ ID NO.10, corresponding to an about 83.7% identity between SEQ ID NO:31 and the corresponding sequence in SEQ ID NO:10. The amino acid sequence set forth in SEQ ID NO:31 is also substantially similar to the amino acid sequence of TIC401 from about amino acid position eighty-five (85) through about amino acid position two-hundred-seventy-four (274) as set forth in SEQ ID NO:8, containing forty-one (41) amino acids that are different from the analogous sequence in SEQ ID NO:8, corresponding to an about 78.4% identity between SEQ ID NO:31 and the corresponding sequence in SEQ ID NO:8 The amino acid sequence as set forth in SEQ ID NO:31 is also substantially similar to the amino acid sequence as set forth in SEQ ID NO:33 (TIC431).

The present invention also contemplates an expression vector comprising a polynucleotide of the present invention Thus, in one embodiment an expression vector is an isolated and purified DNA

molecule comprising a promoter operatively linked to a coding region that encodes a polypeptide of the present invention, which coding region is operatively linked to a transcription-terminating region, whereby the promoter drives the transcription of the coding region. The coding region may include a segment encoding a *B. thuringiensis* insecticidal toxin of the present invention and a segment encoding a chloroplast or plastid targeting peptide. The DNA molecule comprising the expression vector may also contain a functional intron sequence positioned either upstream of the coding sequence or even within the coding sequence, and may also contain a five prime (5') untranslated leader sequence (i.e., a UTR or 5'-UTR) positioned between the promoter and the point of translational initiation.

As used herein and with reference to promoter elements, the terms "operatively linked" or "operably linked" are intended to indicate that a nucleotide sequence that contains a promoter, i e. a genetic element that functions in a particular host cell to drive the initiation of transcription, is connected to a coding region in such a way that the transcription of that coding region is controlled and substantially regulated by that promoter. Means for operatively linking a promoter to a coding region are well known in the art. Promoters that function in bacteria are well known in the art. Exemplary and preferred promoters for the *B. thuringiensis* crystal proteins include the *sigA*, *sigE*, and *sigK* gene promoters. Alternatively, native, mutagenized, heterologous, or recombinant promoters derived from *Bacillus thuringiensis* or other *Bacillus* species can be used for achieving expression of the proteins of the present invention in a *Bacillus* species strain.

Where a nucleotide sequence encoding all or an insecticidal part of a protein of the present invention is to be used to transform a plant, a promoter is selected that has the ability to drive expression of the coding sequence in that particular species of plant. Promoters that function in different plant species are also well known in the art. Promoters useful for expression of polypeptides in plants are those which are inducible, viral, synthetic, or constitutive as described in Odell *et al.* (Nature 313·810-812, 1985), and/or promoters that are temporally regulated, spatially regulated, and spatio-temporally regulated. Preferred promoters include the enhanced CaMV35S promoters, and the FMV35S promoter For optimum control of rootworm species by expression of the proteins of the present invention in plants, it is preferable to achieve the highest levels of expression of these proteins within the roots of maize plants. A number of root enhanced promoters have been identified and are known in the art (Lu et al., J. Plant Phys., 2000, 156(2):277-283; US Patent No. 5,837,848; US Patent No. 6,489,542). Substantial temporal or spatial regulation refers to the expression of a gene within a plant or plant tissue from a plant operable promoter. With reference to temporal regulation, a promoter may be regulated for expression only during specific times during plant cell or tissue or even whole plant growth and development. A promoter which is actively expressing one or more genes only during seed germination would be one example of temporal regulation. Other examples could include promoters which are actively expressing one or more genes only during times when the plant, plant cell or plant tissue is exposed to certain light intensities or during total darkness. Substantial temporal regulation refers to a promoter which is actively expressed at a certain time but which may or may not be completely suppressed at other times, such that expression may still be detected by monitoring for the presence of some indicator such as an enzyme produced from a coding sequence linked to such promoter, or as measured by the increase or decrease in some gene product such as an mRNA produced

at various times throughout plant growth, differentiation, and development and/or in response to various environmental stimuli. Substantial spatial regulation refers to the expression of a gene linked to a promoter from which expression proceeds only during growth and development of certain cells or tissues within a plant. For example, a tapetal promoter would only be expected to be substantially spatially expressed during flower growth and development Similarly, a root specific or root enhanced promoter would only be expected to be substantially spatially expressed from within root cells or root tissues. Substantially spatially regulated also refers to the level of expression from a particular tissue specific promoter in that particular tissue and as related to levels of expression from that or a similar promoter in other tissues, wherein expression may also be detected in tissues other than the particular tissue in which the promoter expression is preferred, but at significantly lower expression levels as measured by the production of an enzyme produced from a coding sequence linked to the promoter or by the appearance of some detectable gene product Promoters can also be both substantially temporally and substantially spatially regulated together and simultaneously in a coordinately regulated manner. Other promoters specifically intended to be within the scope of the present invention include but are not limited to the ubiquitin promoter, the sugarcane bacilliform DNA virus promoter, the ribulose bis-phosphate carboxylase large subunit promoter, among others.

Preferred intron sequences for achieving optimum expression of non-naturally occurring nucleotide sequences in monocotyledonous plants may also be included in the DNA expression construct Such an intron is typically placed near the 5'-end of the mRNA within or immediately downstream of an untranslated sequence. The intron could be obtained from, but not limited to, a set of introns consisting of the maize Heat Shock Protein (HSP) 70 intron (U. S. Patent 5,424,412; 1995), the rice *Act1* intron (McElroy *et al.*, Plant Cell 2:163-171, 1990), the Adh intron 1 (Callis *et al* , Genes & Develop. 1:1183-1200, 1987), or the sucrose synthase intron (Vasil *et al.*, Plant Phys. 91.1575-1579, 1989).

Another element that functions to regulate or to modulate gene expression is the DNA sequence between the transcription initiation site and the start of the coding sequence, termed the untranslated leader sequence (UTL). Compilations of leader sequences have been made to predict optimum or sub-optimum sequences and generate "consensus" and preferred leader sequences (Joshi, Nucl. Acids Res. 15.9627-9640, 1987). Preferred leader sequences are contemplated to include those that comprise sequences predicted to direct optimum expression of the linked structural gene, *i.e.* to include a preferred consensus leader sequence that increases or maintains mRNA stability and prevents inappropriate initiation of translation. The choice of such sequences will be known to those of skill in the art in light of the present disclosure. Sequences that are derived from genes that are highly expressed in plants, and in particular in maize will be most preferred. One particularly useful leader is the petunia HSP70 leader.

Transcription enhancers or duplications of enhancers could be used to increase expression. These enhancers often are found 5' to the start of transcription in a promoter that functions in eukaryotic cells, but can often be inserted in the forward or reverse orientation 5' or 3' to the coding sequence. Examples of enhancers include elements from the CaMV 35S promoter, octopine synthase

genes (Ellis *et al.*, EMBO Journal 6:11-16, 1987), the rice actin gene, and promoter from non-plant eukaryotes (*e.g.*, yeast; Ma *et al.*, Nature 334:631-633, 1988).

RNA polymerase transcribes a nuclear genome DNA coding sequence through a site where polyadenylation occurs. Typically, DNA sequences located a few hundred base pairs downstream of the polyadenylation site serve to terminate transcription. Those DNA sequences are referred to herein as transcription-termination regions. Those regions are required for efficient polyadenylation of nuclear transcribed messenger RNA (mRNA). For coding sequences introduced into a chloroplast or plastid, or into a chloroplast or plastid genome, mRNA transcription termination is similar to methods well known in the bacterial gene expression art. For example, either in a polycistronic or a monocistronic sequence, transcription can be terminated by stem and loop structures or structures similar to bacterial *rho* dependent sequences.

Expression constructs will typically include a coding sequence exemplified in the present invention or a derivative thereof along with a 3′ end DNA sequence that functions as a signal to terminate transcription and, in constructs intended for expression from the plant nuclear genome, allow for the 3' end polyadenylation of the resultant RNA transcript The most preferred 3′ elements are contemplated to be those from the nopaline synthase gene of *A tumefaciens* (*nos* 3'end), the terminator for the T7 transcript from the octopine synthase gene of *A. tumefaciens*, and the pea RUBISCO synthase E9 gene (E9 3') 3' non-translated transcription termination and polyadenylation sequence. These and other 3′ end regulatory sequences are well known in the art.

Preferred plant transformation vectors include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those disclosed, *e.g* , by Herrera-Estrella (Nature 303·209-213, 1983), Bevan (Nature 304:184-187,1983), Klee (Bio/Technol. 3:637-642, 1985) and Eur. Pat Appl. No EP 0120516 (each specifically incorporated herein by reference).

The present invention discloses isolated and purified nucleotide sequences encoding insecticidal proteins derived from *Bacillus* species, and particularly from *Bacillus thuringiensis* species. In particular, the *B. thuringiensis* strains 86833, EG2158, EG3618, EG6489, EG6561, EG6618, and EG4653 are each shown herein to produce one or more soluble insecticidal proteins that are localized to culture supernatants (see Table 1 except for EG4653 which is described in detail in Example 11).

The *B. thuringiensis* strains and other bacterial strains described herein may be cultured using conventional growth media and standard fermentation techniques. The *B. thuringiensis* strains harboring one or more *tic*901, *tic*1201, *tic*407, *tic*417, *tic*431or related genes may be fermented as described herein until the cultured *B. thuringiensis* cells reach the stage of their growth cycle when the TIC901, TIC1201, TIC407, TIC417, TIC431 and/or related proteins are produced.

Subject cultures have been deposited under conditions that assure that access to the culture will be available to authorized parties during the pendency of this patent application or patents issued. However, it should be understood that the availability of a deposit does not constitute a license to practice the subject invention in derogation of patent rights granted by governmental action.

Further, these microorganism deposits have been made under the provisions of the "Budapest Treaty on the International Recognition of the Deposit of Microorganism for the Purposes of Patent

Procedure." The subject culture deposits will be stored and made available to the public in accord with the provisions of the Budapest Treaty, i.e., it will be stored with all the care necessary to keep it viable and uncontaminated for a period of at least five years after the most recent request for the furnishing of a sample of the deposits, and in any case, for a period of at least 30 (thirty) years after the date of deposit or for the enforceable life of any patent which may issue disclosing the culture. The depositor acknowledges the duty to replace the deposits should the depository be unable to furnish a sample when requested, due to the conditions of the deposits. All restrictions on the availability to the public of the subject culture deposit will be irrevocably removed upon the granting of a patent disclosing it.

TIC901, TIC1201, TIC407, TIC417, TIC431 and related proteins of the present invention are shown herein to be produced and secreted into the growth media by several strains of *Bacillus thuringiensis*. Fermentations using the strains of the present invention may be continued through the sporulation stage when crystal proteins, if any, are formed along with the spores. The spores and cell debris can be separated from the supernatant by centrifugation, and the spent culture medium can be used to isolate the insecticidal proteins of the present invention. The inventors herein illustrate the method of ammonium sulfate precipitation as one means for concentrating and collecting all or most of the proteins present in the spent and clarified culture medium. However, one skilled in the art will recognize that there are a number of other means available for purifying and isolating the proteins of the present invention. Gel filtration and size exclusion chromatography are two readily available means for extracting proteins directly from the spent media. Spent media can also be desalted and the filtrate used to extract protein using ion exchange columns. Also, affinity columns, containing antibodies that bind specifically to TIC901, TIC1201, TIC407, TIC417, TIC431 or related proteins can be used to purify the proteins of the present invention directly from the media.

The amino acid sequences of the present invention have been compared to the amino acid sequences present in commercially available protein sequence databases, and no significant homologies or similarities have been identified. Based on this analysis, the TIC901, TIC1201, TIC407, TIC417, and TIC431 proteins and related sequences appear to be unique and form the basis for the establishment of a new and separate class of *Bacillus* insecticidal proteins because the proteins of the present invention have not been observed to exhibit any significant relationship to other known insecticidal proteins.

Modification and changes may be made in the structure of the peptides of the present invention and DNA segments that encode them and still obtain a functional molecule that encodes a protein or peptide with desirable characteristics. The biologically functional equivalent peptides, polypeptides, and proteins contemplated herein should possess from about 65 to about 70% or greater amino acid sequence similarity, or from about 80% or greater amino acid sequence similarity, or from about 90% or greater amino acid sequence similarity, to the sequence of, or corresponding moiety within, the fundamental TIC901 amino acid sequence as set forth in SEQ ID NO:4, or the corresponding moiety within the amino acid sequences of TIC1201, TIC407, TIC417, and TIC431 as set forth respectively in SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, and SEQ ID NO.32 and related sequences.

According to the present invention reference to the *tic*901 gene and encoded protein toxin, includes not only the full length sequences disclosed herein but also fragments of these sequences, natural variants, mutants, and recombinant or genetically engineered derivatives of the *tic*901 gene comprising SEQ ID NO: 3. Proteins encoded by these sequences should retain essentially the same as or greater characteristic insecticidal properties than those of the TIC901 protein comprising SEQ ID NO:4. The proteins useful in the present invention may also include fusion proteins that retain the characteristic insecticidal properties essentially the same as or greater than those of the TIC901 protein. In some instances, the fusion protein may contain, in addition to the characteristic insecticidal properties of the proteins specifically exemplified herein, another insecticidal activity contributed by the amino acid sequence of the fusion partner. Alternatively, crystallographic analysis of the TIC901 protein or insecticidal variants thereof may provide a means for determining whether the protein would be a candidate for the construction of a permutein that exhibits the same or preferably greater insecticidal activity than the native TIC901 or related protein, and which preferably exhibits improved characteristics related to expression in a preferred host cell such as a plant cell. The same qualities and characteristics apply as well to *tic*1201, *tic*407, and *tic*417; to not only the full length sequences disclosed herein but also fragments of these sequences, natural variants, mutants, and recombinant or genetically engineered derivatives of these genes comprising respectively SEQ ID NO:5, SEQ ID NO.7, SEQ ID NO:9, and SEQ ID NO:32. The proteins encoded by these sequences should retain essentially the same as or greater characteristic insecticidal properties than those of the TIC1201, TIC407, TIC417, and TIC431 proteins comprising respectively SEQ ID NO:6, SEQ ID NO.8, SEQ ID NO:10, and SEQ ID NO:32

It should be apparent to a person skilled in the art that nucleotide sequences encoding insect inhibitory toxins, and particularly coleopteran inhibitory toxins, can be identified and obtained through several means as disclosed herein. The specific sequences and related sequences as exemplified herein may be obtained from the isolates deposited at a culture depository as described above. These sequences, or insecticidal portions or variants thereof, may also be constructed synthetically, for example, by use of a nucleotide sequence synthesizer. Variations of coding sequences may be readily constructed using standard techniques for making point mutations Also, fragments of these sequences can be made using commercially available exonucleases or endonucleases according to standard procedures. For example, enzymes such as *Bal*31 or site-directed mutageneais may be used to systematically excise nucleotides from the ends of such sequences as exemplified herein or from within the protein coding sequence. Also, nucleotide sequences that encode insecticidally active protein fragments may be obtained using a variety of restriction enzymes, endonucleases, thermal amplification methods, and the like Proteases such as proteinase K, trypsin, chymotrypsin, pepsin, and the like may be used to directly obtain active fragments of these toxins.

Other toxins and nucleotide sequences encoding such toxins related to the toxins and coding sequences of the present invention can be derived from DNA obtained from *B. thuringiensis*, *B. laterosperous*, *B. sphaericus*, and related *Bacillus* species isolates using the teachings provided in the art in combination with the nucleotide sequences disclosed herein. Such toxins and nucleotides sequences that are related to the toxins and coding sequences of the present invention are deemed

herein to be equivalent to the toxins and nucleotide sequences of the present invention. By "equivalent" it is meant that a protein exhibits the characteristics of one or more of the proteins described herein, including but not limited to similar insecticidal inhibitory bioactivity, host range of insecticidal bioactivity, exhibits similar antigenic epitopes that cross react with antibodies raised against TIC901, TIC1201, TIC407, and TIC417 and the like, and including related proteins, exhibit a similar size relative to TIC901 and related proteins, exhibit similar expression profiles and characteristics, exhibit a propensity for seclusion to the extracellular environment when expressed in *Bacillus thuringiensis* or related bacterial species, and the like. The phrase "exhibit a propensity for seclusion to the extracellular environment" is intended to include TIC901 and related proteins including but not limited to TIC1201, TIC407, TIC417, and TIC431 and the like that are produced by the bacterium or host cell as a precursor protein that contains an amino acid sequence inked to the insecticidal protein that functions to target the insecticidal protein to a bacterial or host cell secretory apparatus and which, upon contact with the secretory apparatus, is proteolytically cleaved by a signal peptidase, releasing the mature or insecticidal protein into the extracellular environment in the case of a gram positive microbe, at least into the periplasm in the case of a gram negative microbe, and into the endoplasmic reticulum or secretory vesicle or into a subcellular organelle such as a mitochondria or chloroplast or plastic in the case of a fungal or plant or other eukaryotic host cell. Cryptic nucleotide coding sequences are also contemplated to be within the scope of the invention herein.

There are a number of methods for identifying the presence of and obtaining equivalent insecticidal toxins related to the peptides disclosed herein. For example, antibodies to the insecticidal toxins disclosed and claimed herein can be used to identify and isolate other toxins from a mixture of proteins. Specifically, antibodies may be raised to the portions of the toxins that are most constant within the new class of proteins and most distinct from other *B. thuringiensis* toxins. These antibodies can then be used to specifically identify equivalent toxins with the characteristic activity by immuno-precipitation, enzyme linked immuno-sorbent assay (ELISA), or Western blotting. Antibodies to the toxins disclosed herein, or to equivalent toxins, or fragments of these toxins, can readily be prepared using standard procedures in the art. The nucleotide sequences that encode these toxins can then be obtained from the microorganism or other various sources.

A further method for identifying the toxins and genes of the present invention is through the use of oligonucleotide probes. These probes are essentially nucleotide sequences that hybridize under stringent hybridization conditions to the TIC901 coding sequence or a sequence related to a TIC901 coding sequence. As is well known in the art, if a probe molecule and nucleic acid sequence molecule in a sample hybridize by forming a strong enough bond between the two molecules, it can be reasonably assumed that the two molecules exhibit substantial homology. Probe binding is detected using any number of means known in the art including but not limited to fluorescence, luminescence, isotopic, immunological, surface plasmon resonance spectroscopy, and the like. Such probe analysis provides a rapid method for identifying toxin-encoding genes of the present invention. The nucleotide segments that are used as probes according to the invention can be synthesized by use of DNA synthesizers using standard procedures or by other means known in the art. These nucleotide

sequences can also be used as PCR primers to amplify nucleotide sequences of the present invention or portions thereof.

Fragments and equivalents of related proteins that retain at least the insecticidal activity of the exemplified toxins are within the scope of the present invention. Also, because of the redundancy of the genetic code, a variety of different DNA sequences can encode the amino acid sequences disclosed herein. It is well within the skill of a person trained in the art to create these alternative DNA sequences encoding the same, or essentially the same, toxins These variant DNA sequences are within the scope of the present invention

It is well known in the art that certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence, and, of course, its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. It is thus contemplated by the inventors that various changes may be made in the peptide sequences of the disclosed compositions, or corresponding DNA sequences which encode said peptides without appreciable loss of their biological utility or activity. Such substitutions are also known in the art as conservative substitutions.

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like.

Each amino acid has been assigned a hydropathic index on the basis of their hydrophobicity and charge characteristics (Kyte and Doolittle, 1982), these are: isoleucine (+4.5); valine (+4 2); leucine (+3 8); phenylalanine (+2 8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1 8); glycine (–0.4); threonine (–0.7); serine (–0.8); tryptophan (–0.9), tyrosine (–1.3); proline (–1 6); histidine (–3 2); glutamate (–3.5), glutamine (–3.5); aspartate (–3.5); asparagine (–3.5); lysine (–3 9); and arginine (–4.5).

It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydropathic index or score and still result in a protein with similar biological activity, *i.e.* still obtain a biological functionally equivalent protein. In making such changes, the substitution of amino acids whose hydropathic indices are within ±2 is preferred, those that are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. The greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein (U. S. Patent 4,554,101).

As detailed in U. S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0), lysine (+3 0); aspartate (+3 0 ± 1); glutamate (+3.0 ± 1); serine (+0.3), asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (–0.4); proline (–0.5 ± 1); alanine (–0 5); histidine (–0 5); cysteine (–1.0); methionine (–1.3); valine (–1 5); leucine (–1.8), isoleucine (–1.8); tyrosine (–2.3); phenylalanine (–2.5); tryptophan (–3.4).

It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within ±2 is preferred, those that are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions which take the various foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

Peptides, polypeptides, and proteins biologically functionally equivalent to TIC901, TIC1201, TIC407, TIC417, TIC431 and related proteins and the like include amino acid sequences containing conservative amino acid changes in the fundamental sequence shown in SEQ ID NO. 4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO·10, and SEQ ID NO:33. In particular, with reference to SEQ ID NO:4 from about amino acid 44 through about amino acid 367, with reference to SEQ ID NO.6 from about amino acid 44 through about amino acid 364, with reference to SEQ ID NO:8 from about amino acid 44 through about amino acid 368, with reference to SEQ ID NO 10 from about amino acid 44 through about amino acid 364, and with reference to SEQ ID NO·33 from about amino acid 44 through about amino acid 364, for such amino acid sequences, one or more amino acids in the fundamental sequence can be substituted with another amino acid(s), the charge and polarity of which is similar to that of the native amino acid, *i.e.* a conservative amino acid substitution, resulting in a silent change.

Substitutes for an amino acid within the fundamental polypeptide sequence can be selected from other members of the class to which the naturally occurring amino acid belongs. Amino acids can be divided into the following four groups: (1) acidic amino acids; (2) basic amino acids, (3) neutral polar amino acids; and (4) neutral non-polar amino acids Representative amino acids within these various groups include, but are not limited to. (1) acidic (negatively charged) amino acids such as aspartic acid and glutamic acid; (2) basic (positively charged) amino acids such as arginine, histidine, and lysine; (3) neutral polar amino acids such as glycine, serine, threonine, cyteine, cystine, tyrosine, asparagine, and glutamine; (4) neutral nonpolar (hydrophobic) amino acids such as alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine.

Conservative amino acid changes within the fundamental polypeptide sequences of the present invention can be made by substituting one amino acid within one of these groups with another amino acid within the same group. Biologically functional equivalents of TIC901, TIC1201, TIC407, TIC417, TIC431 and the like and related sequences can have 10 or fewer conservative amino acid changes, more preferably seven or fewer conservative amino acid changes, and most preferably five or

fewer conservative amino acid changes. The encoding nucleotide sequence (gene, plasmid DNA, cDNA, or synthetic DNA) will thus have corresponding base substitutions, permitting it to encode biologically functional equivalent forms of the TIC901, TIC1201, TIC407, TIC417, and TIC431 proteins and related sequences

Amino acid sequence variants of the proteins of the present invention and related sequences can be made by procedures well known in the art. For example, a TIC901 amino acid sequence variant protein that is not secreted into the extracellular milieu can be obtained through ethylmethane sulfonate (EMS) mutagenesis of a nucleotide sequence encoding TIC901. The mutants can also be constructed using ultraviolet light and nitrosoguanidine by procedures well known in the art, or by constructing a coding sequence that lacks all or a part of the coding sequence encoding a signal peptide amino acid sequence.

Site-specific mutagenesis is another technique useful in the preparation of individual peptides, or biologically functional equivalent proteins or peptides, through specific mutagenesis of the underlying DNA. The technique further provides a ready ability to prepare and test sequence variants, for example, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the DNA. Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the sequence targeted for modification. Typically, a primer of from about 17 to about 25 nucleotides in length is preferred, with at least from about 5 to about 10 residues of identity being available on both sides of the target sequence being modified.

The preparation of sequence variants of the selected peptide- encoding DNA segments using site-directed mutagenesis is provided as a means of producing potentially useful species and is not meant to be limiting as there are other ways in which sequence variants of peptides and the DNA sequences encoding them may be obtained. For example, recombinant vectors encoding the desired peptide sequence may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants.

The *tic*901 and related nucleotide coding sequences isolated from *B thuringiensis* strains as set forth herein in SEQ ID NO:3, SEQ ID NO.5, SEQ ID NO:7, SEQ ID NO:9, and SEQ ID NO:32 and the like may be used as hybridization probes to identify and isolate naturally occurring variants of these and related nucleotide coding sequences from other strains of *B. thuringiensis* or from other microorganisms such as from microbial species such as *Clostridium*, *Bacillus*, *Xenorhabdus*, and *Photorhabdus*. The present invention encompasses nucleotide sequences from microorganisms, where the nucleotide sequences are isolatable by hybridization with all, or part, of the *Bacillus* nucleotide sequence of the invention. Proteins encoded by such nucleotide sequences can be tested for insecticidal activity. The invention also encompasses the proteins encoded by the nucleotide sequences. For example, an alignment of these four nucleotide sequences provides regions of identity that would be preferred for use as probes and primers. Also, an alignment of the amino acid sequences encoded by these and related nucleotide coding sequences provides information regarding regions of

identity and/or substantial similarity between the proteins, and provides a basis for constructing probes and or primers that can be used to identify these and other related nucleotide sequences encoding such insecticidal proteins For example, SEQ ID NO:23-SEQ IDNO:29 are representative of such probes and primers and are exemplified herein in Example 10.

Antibodies raised in response to immune challenge by TIC901, TIC1201, TIC407, TIC417, and TIC431 and the like or related proteins of the present invention may be produced using standard immunological techniques for production of polyclonal antisera and, if desired, immortalizing the antibody-producing cells of the immunized host for sources of monoclonal antibody production. Techniques for producing antibodies to any substance of interest are well known, e.g., as in Harlow and Lane (1988) and as in Goding (1986). For example, antibodies that bind to epitopes on or within TIC901 may be used as probes to identify *B. thuringiensis* strains or other microorganisms that produce variants of TIC901 or related proteins that are encoded by variations of a *tic*901 or related gene The present invention encompasses insecticidal proteins that cross-react with antibodies raised against one or more of the insecticidal proteins of the present invention.

The antibodies produced in the present invention are also useful in immunoassays for determining the amount or presence of a TIC901, TIC1201, TIC407, TIC417, and TIC431 or related protein in a biological sample. Such assays are also useful in quality-controlled production of compositions containing one or more of the proteins of the present invention or related proteins. In addition, the antibodies can be used to assess the efficacy of recombinant production of one or more of the proteins of the present invention or a related protein, as well as for screening expression libraries for the presence of a nucleotide sequence encoding one or more of the proteins of the present invention or related protein coding sequences Antibodies are useful also as affinity ligands for purifying and/or isolating any one or more of the proteins of the present invention and related proteins The proteins of the present invention and proteins containing related antigenic epitopes may be obtained by over expressing full or partial lengths of a sequence encoding all or part of a protein of the present invention or a related protein in a preferred host cell.

The peptides of the present invention are primarily, though not exclusively, intended for use in plants, and in certain preferred embodiments, nucleotide sequences modified for encoding the proteins of the present invention in plants are contained within one or more plasmid vectors. Such vectors may contain a variety of regulatory and other elements intended to allow for optimal expression of the proteins of the present invention in plant cells. These additional elements may include promoters, terminators, and introns as outlined above. Any vector containing the DNA construct and any regulatory or other elements may be selected from the group consisting of a yeast artificial chromosome, bacterial artificial chromosome, a plasmid, or a cosmid, and the like. Further, the expression vectors themselves may be of a variety of forms. These forms may differ for various reasons, and will likely be comprised of varying components depending upon whether they are intended to be used to transform a monocotyledonous plant or a dicotyledonous plant.

Vectors further envisioned to be within the scope of the present invention include those vectors capable of containing a *tic*901, *tic*1201, *tic*407, *tic*417 or related nucleic acid compositions

disclosed above, as well as any other DNA constructs which further comprise plant-expressible coding regions for other insecticidal proteins derived from *Bacillus* species.

The nucleotide sequence encoding a TIC901 (SEQ ID NO: 3 encoding SEQ ID NO:4) or encoding a related peptide sequence such as TIC1201 (SEQ ID NO:5 encoding SEQ ID NO:6), TIC407 (SEQ ID NO:7 encoding SEQ ID NO:8), TIC417 (SEQ ID NO:9 encoding SEQ ID NO:10), and TIC431 (SEQ ID NO:32 encoding SEQ ID NO:33) may be introduced into a variety of microorganism hosts without undue experimentation, using procedures well known to those skilled in the art of transforming suitable hosts, and under conditions which allow for stable maintenance and expression of the introduced nucleotide sequence (Sambrook et al., 1989, Molecular Cloning. A Laboratory Manual 2nd Ed., Cold Spring Harbor Press, New York). Suitable hosts that allow for expression of the proteins of the present invention and related sequences include *B. thuringiensis* and other *Bacillus* species such as *Bacillus subtilis*, *Bacillus sphaericus*, *Bacillus laterosporous*, *Bacillus megaterium*, or *Bacillus anthracis*. Genetically altered or engineered microorganisms containing a gene encoding one or more of the proteins of the present invention, including TIC901, TIC1201, TIC407, TIC417, and TIC431 and the like can also contain nucleotide sequences encoding other toxin proteins present in the same microorganism; these coding sequences could concurrently produce insecticidal proteins different from the proteins of the present invention or related proteins In particular, it would be preferable to produce two or more different insecticidal proteins in a host cell, wherein each protein is toxic to the same insect species and each protein exhibits a mode of action different from the other(s).

Plant-colonizing or root-colonizing microorganisms may also be employed as host cells for the production of one or more of the proteins of the present invention or related protein. Exemplary microorganism hosts for *B thuringiensis* toxin genes include the plant-colonizing microbe *Clavibacter xyli* as described by Turner et al. (1993, Endophytes: an alternative genome for crop improvement.; International Crop Science Congress, Ames, Iowa, USA, 14-22 July 1992., pp.555-560) and root-colonizing pseudomonad strains, as described by Obukowicz et al. (US Patent No. 5,229,112).

The toxin-encoding nucleotide sequences obtainable from the isolates of the present invention can be introduced into a wide variety of microbial or plant hosts. Expression of the toxin gene results, directly or indirectly, in the intracellular production and maintenance of the pesticide. With suitable microbial hosts, e.g., *Pseudomonas*, the microbes can be applied to the situs of the pest, where they will proliferate and be ingested by the pest. The result is a control of the pest exhibited by reduced plant damage, increased plant yield, decreased prevalence of the plant pest in the general local environment of the transgenic organism expressing the toxin protein(s), and the death or stunted growth of the plant pest, generally without any additional impact on the microbial flora surrounding the plant or transgenic organism expressing the toxin protein(s), and without any additional impact on the environment in general. Alternatively, the microbe hosting the toxin gene can be treated under conditions that prolong the activity of the toxin and stabilize the cell. The treated cell, which retains the toxic activity, then can be applied to the environment of the target pest.

Where the toxin gene of the present invention or a related nucleotide coding sequence is introduced by means of a suitable vector into a microbial host, and the host is applied to the environment in a living state, it is advantageous to use certain host microbes. For example,

microorganism hosts can be selected which are known to occupy the pest's habitat. Microorganism hosts may also live symbiotically with a specific species of pest. These microorganisms are selected so as to be capable of successfully competing in the particular environment with the wild-type microorganisms, provide for stable maintenance and expression of the gene expressing the polypeptide pesticide, and, desirably, provide for improved protection of the pesticide from environmental degradation and inactivation.

A large number of microorganisms are known to inhabit the habitat of pests. These microorganisms include bacteria, algae, and fungi. Of particular interest are microorganisms, such as bacteria, e.g., genera *Bacillus*, *Escherichia*, *Pseudomonas*, *Erwinia*, *Serratia*, *Klebsiella*, *Salmonella*, *Pasteurella*, *Xanthomonas*, *Streptomyces*, *Rhizobium*, *Rhodopseudomonas*, *Methylophilius*, *Agrobacterium*, *Acetobacter*, *Lactobacillus*, *Arthrobacter*, *Azotobacter*, *Leuconostoc*, and *Alcaligenes*; fungi, e.g., genera *Metarhizium*, *Bavaria*, *Saccharomyces*, *Cryptococcus*, *Kluyveromyces*, *Sporobolomyces*, *Rhodotorula*, and *Aureobasidium*.

A wide variety of means are available for introducing a toxin gene encoding a toxin into a microorganism host under conditions that allow for stable maintenance and expression of the gene. These methods are well known to those skilled in the art and are described, for example, in U.S. Patent No. 5,135,867.

As mentioned above, *B. thuringiensis* or recombinant cells expressing a protein of the present invention or a related toxin protein can be treated to prolong the toxin activity and stabilize the cell. The pesticide microcapsule that is formed comprises one or more of the proteins of the present invention or one or more related toxins within a structure that has been stabilized and which functions to protect the toxin or toxins when the microcapsule is applied to the environment of the target pest. Suitable host cells may include either prokaryotes or eukaryotes, normally being limited to those cells that do not produce substances toxic to higher organisms, such as mammals. However, organisms which produce substances toxic to higher organisms could be used, where the toxic substances are unstable or the level of application sufficiently low as to avoid any possibility of toxicity to a mammalian host. Of particular interest as hosts will be prokaryotes as well as lower eukaryotes such as fungi. The cells of these organisms will usually be intact and be substantially in the proliferative form when treated, rather than in a spore form, although in some instances spores may be employed. Such microcapsules can also contain one or more of the proteins of the present invention or one or more related proteins along with one or more unrelated insecticidal protein compositions including but not limited to delta endotoxins such as Cry1, Cry2, Cry3, Cry9, Cry22, ET70, TIC851, and/or binary toxins such as ET80/76, ET33/34, PS149B1, and ET100/101 and the like, and insecticidal proteins or insecticidal protein complexes from such diverse organisms such as *Xenorhabdus* and/or *Photorhabdus*, or VIP, WAR, and/or MIS protein toxins and related proteins.

Treatment of the microbial cell, e.g., a microbe containing a nucleotide sequence or a nucleotide segment of the present invention or a related coding sequence, can be by chemical or physical means, or by a combination of chemical and/or physical means, so long as the technique does not deleteriously affect the properties of the toxin, nor diminish the cellular capability of protecting the toxin. Examples of chemical reagents are halogenating agents, particularly halogens of atomic no. 17-

80. More particularly, iodine can be used under mild conditions and for sufficient time to achieve the desired results. Other suitable techniques include treatment with aldehydes, such as glutaraldehyde; anti-infectives, such as zephiran chloride and cetylpyridinium chloride; alcohols, such as isopropyl and ethanol; various histologic fixatives, such as Lugol iodine, Bouin's fixative, various acids, and Helly's fixative (See: Humason, 1967); or a combination of physical (heat) and chemical agents that preserve and prolong the activity of the toxin produced in the cell when the cell is administered to the host animal. Examples of physical means are short wavelength radiation such as gamma- radiation and X-radiation, freezing, UV irradiation, lyophilization, and the like. Methods for treatment of microbial cells are disclosed in U.S. Patent Nos. 4,695,455 and 4,695,462.

The cells generally will have enhanced structural stability that will enhance resistance to environmental conditions. Where the pesticide is in a proform or precursor form, the method of cell treatment should be selected so as not to inhibit processing of the proform to the mature form of the pesticide by the target pest pathogen. For example, formaldehyde will crosslink proteins and could inhibit processing of the proform of a polypeptide pesticide. The method of cell treatment retains at least a substantial portion of the bio-availability or bioactivity of the toxin.

Characteristics of particular interest in selecting a host cell for purposes of production include ease of introducing one or more of the coding sequences of the present invention or one or more related coding sequences into the host, availability of expression systems, efficiency of expression, stability of the pesticide in the host, and the presence of auxiliary genetic capabilities. Characteristics of interest for use as a pesticide microcapsule include protective qualities for the pesticide, such as thick cell walls, pigmentation, and intracellular packaging or formation of inclusion bodies; survival in aqueous environments, lack of mammalian toxicity; attractiveness to pests for ingestion; ease of killing and fixing without damage to the toxin; and the like. Other considerations include ease of formulation and handling, economics, storage stability, and the like.

The cellular host containing a nucleotide sequence encoding a protein of the present invention or a related protein may be grown in any convenient nutrient medium, where the DNA construct provides a selective advantage, providing for a selective medium so that all or substantially all of the cells retain the nucleotide sequence encoding the protein of the present invention or related coding sequence. These cells may then be harvested in accordance with conventional ways. Alternatively, the cells can be treated prior to harvesting.

The coding sequences of the present invention, including those as set forth in SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, and SEQ ID NO 32 and the like can be used as the basis for constructing modified nucleotide sequences for incorporation into plant cells. Even more preferable is the synthesis of a non-naturally occurring nucleotide sequence that encodes one or more of the proteins of the present invention or a related insecticidal protein or its equivalent for expression in a plant cell, the synthesis of the non-naturally occurring nucleotide sequence being based on the amino acid sequence of the native protein without reference to the native nucleotide sequence from which the native amino acid sequence was deduced. Expression of such sequences in plant cells would render a plant comprised of such cells more resistant to insect attack by coleopteran species and the like. Genetic engineering of plants with modified sequences encoding one or more of the proteins of the

present invention or a related protein or a related insecticidal amino acid sequence may be accomplished by introducing the desired plantized (the word 'plantized' being synonymous with the words 'modified' or 'synthetic') DNA containing the coding sequence into plant tissues or cells, using DNA molecules of a variety of forms and origins that are well known to those skilled in plant genetic engineering. Examples of techniques for introducing DNA into plant tissue are disclosed at least by Perlak et al. (1991).

DNA containing a modified gene encoding a protein of the present invention such as a TIC901, a TIC1201, a TIC407, a TIC417, or a TIC431 and the like, or a related insecticidal protein, operatively linked to a plant functional promoter, may be delivered into the plant cells or tissues directly by a number of means including but not limited to *Agrobacterium* mediated transformation, plant viruses, electroporation, microinjection, vacuum infiltration, liposome fusion means, and ballistic methods and the like. The plant promoter may be a constitutive promoter; or the promoter may be a temporally, spatially, chemically, photosynthetically, thermally, or artificially regulated promoter; a tissue-specific promoter; or even a chimeric or hybrid promoter assembled from parts of other plant functional promoters For example, the promoter may be a cauliflower mosaic virus (CaMV) 35S promoter or a plant functional derivative thereof.

Native bacterial genes and coding sequences are often poorly expressed in transgenic plant cells. Plant codon usage more closely resembles that of other higher organisms than unicellular organisms, such as bacteria. Several reports have disclosed methods for improving expression of recombinant genes in plants (Murray *et al.*, 1989, Nucleic Acids Research, Vol.17:477-498, Diehn *et al*, 1998, Plant Physiology, 117:1433-1443; Rocher et al., 1998, Plant Phys. 117:1445-1461). These reports disclose various methods for engineering coding sequences to represent sequences which are more efficiently translated based on plant codon frequency tables, improvements in codon third base position bias, using recombinant sequences which avoid suspect polyadenylation or A/T rich domains or intron splicing consensus sequences. While these methods for synthetic gene construction are notable, synthetic genes of the present invention for expression in particular plants are prepared substantially according to the method of Brown et al. (U. S. Patent No. 5,689,052).

The work described herein takes advantage of methods of potentiating *in planta* expression of one or more of the proteins of the present invention and related insecticidal proteins, which confer resistance to coleopteran or even lepidopteran plant insect or even nematode pathogens, by incorporation or localization of coding sequences into the nuclear, plastid, or chloroplast genome of susceptible plants. U. S. Patent No. 5,500,365 and related patents describe methods for synthesizing plant genes to achieve optimum expression levels of the protein for which the synthesized, non-naturally occurring, synthetic, or artificial gene encodes. These methods relate to the modification of native Bt structural gene sequences to produce a coding sequence that is more "plant-like" and therefore more likely to be translated and expressed by the plant, monocot or dicot. However, the method as disclosed in Brown et al. (U. S. Patent No. 5,689,052) provides for enhanced expression of transgenes, preferably in monocotyledonous plants.

Thus, the amount of a gene or nucleotide sequence or nucleotide segment coding for a polypeptide of interest, *e.g* all or an insecticidal part of a TIC901, TIC1201, TIC407, TIC417, TIC431

or related polypeptide, can be increased in plants by transforming those plants using transformation methods mentioned above. In particular, transformation of chloroplast or plastid organelles can result in desired coding sequences being present in up to about 10,000 copies per cell in tissues containing these subcellular organelle structures (McBride et al., WO 95/24492).

DNA encoding the peptides of the present invention and related proteins can also be introduced into plants by utilizing a direct DNA transfer method into pollen as described (Zhou *et al.*, 1983, Mol. Cell Biol., 10:4529-4537; Hess, 1987, Intern Rev. Cytol., 107:367.). Expression of polypeptide coding sequences, i.e., *tic*901 and the like, can be obtained by injection of the DNA into reproductive organs of a plant as described (Pena *et al.*, 1987, Nature, 325:274). The DNA can also be injected directly into the cells of immature embryos and into rehydrated desicated embryos as described (Neuhaus *et al.*, 1987, Theor. Appl. Genet., 75:30).

After effecting delivery of exogenous nucleotide sequences encoding the insecticidal proteins of the present invention or related proteins to recipient cells, the next step to obtain a transgenic plant generally concerns identifying the transformed cells for further culturing and plant regeneration, i.e., selection of the transformed cells. As mentioned herein, in order to improve the ability to identify transformants, one may desire to employ a selectable or screenable marker gene as, or in addition to, the expressible gene of interest In this case, one would then generally assay the potentially transformed cell population by exposing the cells to a selective agent or agents, or one would screen the cells for the desired marker gene trait

An exemplary embodiment of methods for identifying transformed cells involves exposing the transformed cultures to a selective agent, such as a metabolic inhibitor, an antibiotic, herbicide or the like. Cells that have been transformed and have stably integrated a marker gene conferring resistance to the selective agent used, will grow and divide in culture. Sensitive cells will not be amenable to further culturing. One example of a preferred marker gene confers resistance to the herbicide glyphosate. When this gene is used as a selectable marker, the putatively transformed cell culture is treated with glyphosate. Upon exposure to glyphosate, transgenic cells containing a recombinant GOX enzyme or a recombinant glyphosate insensitive EPSPS enzyme will be available for further culturing while sensitive, or non-transformed cells, will not. (U. S. Patent No. 5,569,834). Another example of a preferred selectable marker system is the neomycin phosphotransferase (*npt*II) resistance system by which resistance to the antibiotic kanamycin is conferred, as described in U S. Patent No 5,569,834. Again, after transformation with this system, transformed cells will be available for further culturing upon treatment with kanamycin, while non-transformed cells will not. Yet another preferred selectable marker system involves the use of a gene construct conferring resistance to paromomycin Use of this type of a selectable marker system is described in U. S. Patent No. 5,424,412. Other selectable markers are well known in the art, including but not limited to antibiotic resistance markers such at *npt*II, *tet*, *aad*, and the like, *phn*O and other various acetylases (US Patent No 6,448,476), various esterases (6,107,549), barnase (Hartley, 1988), J. Mol. Biol. 202: 913), bacterial enzymes conferring glyphosate oxidase activity upon the transformed cell (*gox*) (Barry et al, 1992, Inhibitors of amino acid biosynthesis: Strategies for imparting glyphosate tolerance to crop plants. In: Biosynthesis and

Molecular Regulation of Amino Acids in Plants. pp. 139-145. Singh, Flores, and Shannon Eds., American Society of Plant Physiologists, Rockville, Md.) and the like.

Transplastonomic selection (selection of plastid or chloroplast transformation events) is simplified by taking advantage of the sensitivity of chloroplasts or plastids to spectinomycin, an inhibitor of plastid or chloroplast protein synthesis, but not of protein synthesis by the nuclear genome encoded cytoplasmic ribosomes. Spectinomycin prevents the accumulation of chloroplast proteins required for photosynthesis so spectinomycin resistant transformed plant cells may be distinguished on the basis of their difference in color: the resistant, transformed cells are green, whereas the sensitive cells are white, due to inhibition of plastid-protein synthesis. Transformation of chloroplasts or plastids with a suitable bacterial *aad* gene, or with a gene encoding a spectinomycin resistant plastid or chloroplast functional ribosomal RNA provides a means for selection and maintenance of transplastonomic events (Maliga, 1993, Trends in Biotechnology 11:101-106).

It is further contemplated that combinations of screenable and selectable markers will be useful for identification of transformed cells. In some cell or tissue types a selection agent, such as glyphosate or kanamycin, may either not provide enough killing activity to clearly recognize transformed cells or may cause substantial nonselective inhibition of transformants and nontransformants alike, thus causing the selection technique to not be effective. It is proposed that selection with a growth inhibiting compound, such as glyphosate or AMPA (amino-methyl phosponic acid) at concentrations below those that cause 100% inhibition, followed by screening of growing tissue for expression of a screenable marker gene such as kanamycin would allow one to recover transformants from cell or tissue types that are not amenable to selection alone. It is proposed that combinations of selection and screening may enable one to identify transformants in a wider variety of cell and tissue types.

The development or regeneration of plants from either single plant protoplasts or various explants is well known in the art (Weissbach and Weissbach, 1988). This regeneration and growth process typically includes the steps of selection of transformed cells, culturing those individualized cells through the usual stages of embryonic development through the rooted plantlet stage. Transgenic embryos and seeds are similarly regenerated The resulting transgenic rooted shoots are thereafter planted in an appropriate plant growth medium such as soil.

The development or regeneration of plants containing a foreign, exogenous gene that encodes all or an insecticidal part of a TIC901, TIC1201, TIC407, TIC417, TIC431 or a related polypeptide introduced into the plant genome by *Agrobacterium* transformation of leaf explants can be achieved by methods well known in the art (Horsch *et al.*, 1985). In this procedure, transformants are cultured in the presence of a selection agent and in a medium that induces the regeneration of shoots in the plant strain being transformed as described (Fraley *et al.*, 1983). In particular, U S. Patent No. 5,349,124 details the creation of genetically transformed lettuce cells and plants resulting therefrom which express hybrid crystal proteins conferring insecticidal activity against Coleopteran larvae to such plants.

This procedure typically produces shoots within two to four months and those shoots are then transferred to an appropriate root-inducing medium containing the selective agent and an antibiotic to

prevent bacterial growth. Shoots that rooted in the presence of the selective agent to form plantlets are then transplanted to soil or other media to allow the production of roots. These procedures vary depending upon the particular plant strain employed, such variations being well known in the art.

Preferably, the regenerated plants are self-pollinated to provide homozygous transgenic plants, or pollen obtained from the regenerated plants is crossed to seed-grown plants of agronomically important, preferably inbred lines. Conversely, pollen from plants of those important lines is used to pollinate regenerated plants. A transgenic plant of the present invention containing a nucleotide sequence encoding a desired insecticidal protein of the present invention or a related polypeptide is cultivated using methods well known to one skilled in the art.

A transgenic plant of the present invention contains at least a coding region encoding one or more polypeptides of the present invention or a related polypeptide, such as an insecticidal fragment of a TIC901, a TIC1201, a TIC407, a TIC417, or a TIC431 protein or a chimera of these proteins. A preferred transgenic plant is an independent segregant and can transmit that gene and its activity to its progeny. A more preferred transgenic plant is homozygous for that gene, and transmits that gene to all of its offspring on sexual mating Seed from a transgenic plant may be grown in the field or greenhouse, and resulting sexually mature transgenic plants are self-pollinated to generate true breeding plants. The progeny from these plants become true breeding lines that are evaluated for increased expression of the *B. thuringiensis* transgene.

Transgenic plants expressing more than one insecticidal agent are preferred, each agent being toxic to a target insect pest species, and each insecticidal agent exhibiting a separate mode of action or exhibiting a different means for introducing pores into the midgut epithelium of the target insect pest either by binding to separate and independent receptors or by forming pores that are measurably different than the pores formed by the other one or more insecticidal agents present in the same transgenic plant. Such plants conceivably may be transformed to express at least two or more of the proteins of the present invention, or will be transformed to express at least one of the proteins of the present invention along with at least one or more unrelated insecticidal protein including but not limited to delta-endotoxin proteins such as one or more of a Cry1, a Cry2, a Cry3, a Cry9, a Cry22, an ET70, a TIC851, and/or any one or more of the binary toxins ET80/76, ET33/34, PS149B1, and ET100/101 and the like and fusions, chimeras, and variants thereof, insecticidal proteins or insecticidal proteins complexes from such diverse organisms such as *Xenorhabdus* and/or *Photorhabdus*, or VIP, WAR, and/or MIS protein toxins and related proteins, and or one or more transgenic double stranded RNA's for which expression in the plant cell results in suppression of one or more genes in one or more target insect pests.

To identify a transgenic plant expressing high levels of a polypeptide of the present invention or a related protein from a preferred nucleotide sequence, it may be necessary to screen the selected transgenic event, (R_0 generation) for insecticidal activity and/or expression of the gene. This can be accomplished by various methods well known to those skilled in the art, including but not limited to: 1) obtaining small tissue samples from the transgenic R_0 plant and directly assaying the tissue for activity against susceptible insects in parallel with tissue derived from a non-expressing, negative control plant. [For example, R_0 transgenic corn plants expressing an insecticidal fragment of a TIC901

or a related protein can be identified by assaying leaf tissue derived from such plants for activity against Colorado Potato Beetle (CPB, *Leptinotarsa decemlineata*) and Southern Corn Rootworm (SCR, *Diabrotica undecimpunctata howardi*)]; 2) analysis of protein extracts by enzyme linked immunoassays (ELISAs) specific for the insecticidal protein fragment or related protein; or 3) reverse transcriptase thermal amplification (also known in the art as RTPCR) to identify events expressing the sequence encoding the insecticidal protein fragment or related protein.

The pesticidal agents of the present invention can be applied alone or in combination with other pesticidal agents to a seed as a component of a seed coating, or the agents of the present invention can be produced within a transgenic seed and combined with other pesticidal agents in the form of a seed coating. Seed coating methods and compositions that are known in the art are useful when they are modified by the addition of one of the embodiments of the combination of pesticides of the present invention. Such coating methods and apparatus for their application are disclosed in, for example, U.S. Patent Nos. 5,918,413, 5,891,246, 5,554,445, 5,389,399, 5,107,787, 5,080,925, 4,759,945 and 4,465,017. Seed coating compositions are disclosed, for example, in U.S. Patent Nos 5,939,356, 5,882,713, 5,876,739, 5,849,320, 5,834,447, 5,791,084, 5,661,103, 5,622,003, 5,580,544, 5,328,942, 5,300,127, 4,735,015, 4,634,587, 4,383,391, 4,372,080, 4,339,456, 4,272,417 and 4,245,432, among others.

As used herein, the term "biological sample", or "sample", is intended to include nucleic acids, polynucleotides, DNA, RNA, tRNA, cDNA, and the like in a composition or fixed to a substrate which enables the sample to be subjected to molecular probe analysis or thermal amplification using oligonucleotide probes and/or primers. A plant or plant product or the fruit or seed from a plant is considered to be a biological sample, and any extract from such plant or plant product, plant part, fruit, or seed, is also considered to be a biological sample. As such, biological samples can be derived from agronomically and commercially important products and/or compositions of matter including but not limited to animal feed, commodities, and corn products and by-products that are intended for use as food for human consumption or for use in compositions that are intended for human consumption including but not limited to corn flour, corn meal, corn syrup, corn oil, corn starch, popcorn, corn cakes, cereals containing corn and corn by-products, and the like are intended to be within the scope of the present invention if these products and compositions of matter contain detectable amounts of the nucleotide sequences or the pesticidal proteins encoded therefrom as set forth herein.

The following examples further illustrate the characteristics of the nucleotide sequences disclosed herein and the insecticidal activity of the proteins encoded by the disclosed nucleotide sequences. In addition, methods and procedures for practicing the invention are disclosed. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

EXAMPLES

Example 1

This example illustrates the identification of an insecticidal protein secreted into the media by *B. thuringiensis* strain EG2158.

EG2158 is a wild-type *B thuringiensis* strain originally isolated from soybean grain dust (Donovan et al, 1988). During sporulation cells of EG2158 produce rhomboid-shaped crystals composed of a 73 kDa protein identified to be Cry3C. Spore crystal mixtures of EG2158 are toxic to Colorado potato beetle larvae. *Bacillus megaterium* transformed with the cloned *cry*3C gene exhibit toxicity to CPB larvae similar to strain EG2158 suggesting that the coleopteran-specific toxicity of EG2158 is due to the Cry3C crystal protein (Donovan et al, 1988). Therefore, it was unexpected that spent media used for fermentation of strain EG2158, which does not contain spores or Cry3C crystal protein, would be toxic to CPB because the processed spent media does not contain sufficient spores or Cry3 crystal protein to exhibit measurable insecticidal toxicity.

B. thuringiensis strains were grown in 60 mL of PYG culture medium with shaking overnight at 30° C. PYG medium consisted of 11.8 grams peptone, 23.6 grams yeast extract, 4 milliliters glycerol, 19.4 grams anhydrous K_2HPO_4, and 2 2 grams anhydrous KH_2PO_4 per liter of deionized water *B. thuringiensis* cultures fermented overnight were centrifuged at 11,000x g for 30 minutes and the cell free supernatants were transferred to clean flasks. Supernatants are also referred to herein as spent medium and/or clarified spent medium or clarified supernatant

The supernatants were chilled to 4° C. 34 grams of ammonium sulfate plus 1 milliliter of 1 M NaOH were slowly added to 60 milliliters of supernatant while stirring. The ammonium sulfate saturated supernatant mixtures were centrifuged to collect the precipitable proteins into a pellet. The pellets were dissolved in 2 milliliters of 20 mM Tris.HCl pH 7 5. The Tris suspensions were transferred to dialysis tubing (6000 MWCO) and dialyzed at 4°C against 20mM Tris-HCl pH 7 5 The dialyzed suspensions are referred to herein as am.sulf-dialysates (ASD)

The am.sulf.-dialysates (ASD) were tested for toxicity to Colorado potato beetle larvae (CPB, also formally known as *Leptinotarsa decemlineata*) 50 µl of each ASD was applied topically to 2 mL of insect diet in a diet cup. A total of sixteen diet cups were treated for each dialysate. One first-instar CPB larva was placed in each diet cup and insect mortality was scored after 3 days. Numerous repetitions of this procedure were completed with a number of different *B. thuringiensis* strains. The dialysate from *B. thuringiensis* strain EG2158 exhibited toxicity to CPB larvae.

Example 2

This example illustrates a means for purifying an insecticidal protein, designated herein as TIC901, from the spent culture media produced by fermentation of *Bacillus thuringiensis* strain EG2158.

Proteins present in the EG2158 ASD were evaluated by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE). 30 µl of the dialysate was mixed with 15 µl of Laemmli protein solubilization buffer (Mol. Biol. 80:575-599; 1973) and heated to 100°C for 5 minutes. Approximately 25 µl of the mixture was loaded onto a 10% SDS polyacrylamide gel. The

proteins were visualized after electrophoresis after staining with Coomassie Brilliant Blue. The results indicated that the EG2158 ASD contained about thirty proteins ranging in size from approximately 20 kDa to about 120 kDa.

2 milliliters of the EG2158 ASD was applied to a diethylaminoethyl (DEAE) column equilibrated with 20 mM Tris-HCl pH 7.5. Proteins were eluted from the column with a 20 milliliter gradient of NaCl (0 to 500 mM) in 20 mM Tris-HCl, pH 7.5. 1 milliliter fractions of the eluate were collected and each fraction was dialyzed against 20 mM Tris-HCl, pH 7.5. Individual fractions were tested after dialysis for toxicity to CPB using a bioassay similar to that described in Example 1. Fractions exhibiting toxicity to CPB larvae were pooled and dialyzed. The pooled fractions were referred to as the EG2158 DEAE-toxic fraction. Proteins in the EG2158 DEAE-toxic fraction were analyzed by SDS-PAGE and Coomassie staining. The results indicated that the EG2158 DEAE-toxic fraction contained five primary proteins as well as several minor proteins of various sizes.

2 milliliters of the EG2158 DEAE-toxic fraction was applied to a quaternary ammonium (QA) column equilibrated with 20 mM Tris-HCl pH 7 5 Protein was eluted from the column with 20 milliliters of a linear NaCl gradient (0 to 500 mM) in 20 mM Tris-HCl, pH 7.5. 1 milliliter fractions were collected and dialyzed separately, again in 20 mM Tris-HCl, pH 7.5. Dialyzed fractions were tested for toxicity to CPB larvae Fractions exhibiting the highest CPB toxicity were pooled. The pooled fraction was referred to as the EG2158 QA-toxic fraction. Proteins in the EG2158 QA-toxic fraction were analyzed by SDS-PAGE and Coomassie staining. The results indicated that the EG2158 QA-toxic fraction contained one major protein of approximately 38 kDa, referred to as TIC901, and several other minor protein species of various sizes. The results suggested that the purified 38 kDa protein was responsible for CPB toxicity exhibited by the samples.

Example 3

This example illustrates the isolation and identification of a nucleotide sequence from EG2158 that encodes the TIC901 protein.

Proteins in the EG2158 QA-toxic fraction were size-separated by SDS-PAGE without Coomassie staining Separated proteins were transferred from the SDS-PAGE to a polyvinylidene difluoride (PVDF) membrane by electrotransfer. The PVDF membrane was stained briefly with Coomassie dye and the portion of the membrane containing the purified TIC901 protein was excised with a clean razor blade and subjected to automated Edmund degradation sequencing. The results indicated that the TIC901 protein contained an amino terminal amino acid sequence corresponding to $^{NH}{}_3$-V I G P Y A E S Y I D R V Q D-$^{COO-}$ as set forth in SEQ ID NO:1.

It is widely recognized that most proteins produced *in vivo* in bacterial systems exhibit a methionine (M) residue at their N-terminus. The fact that the N-terminus of the TIC901 protein did not contain an amino terminal methionine residue, along with the fact that the protein was found to be localized to the spent medium, suggested that the TIC901 protein might be formed by proteolytic digestion of a secretory signal peptide from the N-terminal region of a larger precursor protein.

Based on the partial amino acid sequence of the gel purified TIC901 protein as determined by Edmund degradation, and based on the codon usage preference exhibited by native *Bacillus*

thuringiensis genes encoding δ-endotoxins, a degenerate oligonucleotide probe was designed for use as a probe to detect nucleotide sequences from *Bacillus thuringiensis* strain EG2158 that might encode the gel purified TIC901 protein. The probe was identified as WD444 and comprised the sequences as set forth in SEQ ID NO:2 (5'-GTA ATT GGA CCA TAT GCA GAA TCA TAT ATT GAT XGA GTA CAA GA-3', where X is either A or C).

DNA was purified from *B. thuringiensis* strain EG2158 cells by standard procedures (e.g., Sambrook et al., 1989). A sample of the DNA extract was digested with *Hind*III restriction endonuclease, subjected to 0.8% agarose gel electrophoresis in TAE buffer, blotted to a nylon filter, and analyzed by Southern blot by probing with an alkaline phosphatase conjugated WD444 probe mixture for approximately 16 hours at 40°C in hybridization buffer under low to moderate stringency, and washed at 40°C in wash buffer, then exposed to chemi-luminescence reagents, and exposed to film (hybridization and wash buffers were supplied along with AMERSHAM/PHARMACIA BIOTECH Kit, Catalog number RPN3690). The results indicated that the probe appeared to specifically hybridize with a single approximately 8 kilobase pair *Hind*III fragment.

An *E. coli* library containing EG2158 *Hind*III restriction fragments of approximately 8 kilobase pairs was constructed in plasmid pUC18. Recombinant colonies were blotted to nylon filters and denatured with NaOH by standard procedures (e g., Sambrook et al , 1989). The filters were incubated in hybridization solution with labeled WD444 probe mixture. The membranes were washed after incubation with the probe under conditions similar to those described above, exposed to chemiluminescence reagents, and exposed to film. Several colonies were identified that appeared to specifically hybridize to the WD444 probe mixture. Plasmid DNA was extracted from several of these colonies and analyzed by Southern blot using the WD444 oligonucleotide mixture as a probe. The results indicated that the plasmids consisted of the pUC18 vector plus a *Hind*III insert fragment of approximately 8 kb. One typical plasmid was selected for further characterization and was designated as pEG1379 (and later as pMON74007). The colony purified *E. coli* strain containing pMON1379 was designated EG12447 and was deposited with the NRRL on February 6, 2002 and provided with the deposit accession number NRRL B 30549.

Example 4

This example illustrates the expression of an insecticidal protein from an approximately 8 kilobase pair *Hind*III fragment in plasmid pEG1381 in an acrystalliferous strain of *Bacillus thuringiensis*.

Native *Bacillus thuringiensis* genes are often poorly expressed in non-*Bacillus* host cells such as *E. coli*. pEG518 was constructed as an *Bacillus-E. coli* shuttle vector and is a chimera of a *Bacillus* replicon pMM101 (Norton et al., Plasmid 13:211-214; 1985) and pUC18 (Yanisch-Perron et al, Gene 33:103-119; 1985) and therefore is capable of replication either in *E. coli* and closely related gram negative bacteria or in *Bacillus* and closely related gram positive bacteria. pEG518 confers β-lactam antibiotic resistance to *E. coli* and related bacteria and tetracycline resistance to *Bacillus* and related bacteria transformed to contain derivatives of this particular plasmid. The 8 kilobase pair *Hind*III fragment present in plasmid pMON1379 was excised and inserted into the *Hind*III site in plasmid

pEG518. The resulting plasmid, pEG1381, was transformed by electroporation into the acrystalliferous and non-insecticidal *Bacillus thuringiensis* strain EG10650 (NRRL Accession Number NRRL B-30217).

(*Bacillus thuringiensis* strain EG10650. *B. thuringiensis* EG10650 is a derivative of strain EG10368 (U.S. Patent No. 5,759,538, June 2, 1998) that is deficient in neutral protease and alkaline protease activities and contains only one known extra-chromosomal plasmid element of 7.5 kb. A deletion in the alkaline protease gene was introduced in strain EG10368, resulting in strain EG EG10654 (NRRL Accession Number NRRL B-21344). A deletion in the neutral protease gene was introduced in strain EG10368, resulting in strain EG10624 (NRRL Accession Number NRRL B-21347). These deletions were then combined into one strain to produce strain EG10650, lacking any mega-Dalton plasmids and therefore lacking the potential for expressing any known insecticidal proteins as well as being deficient in the production of both the alkaline and neutral proteases.)

A single colony representing a single transformant and designated as EG12450 was selected on media containing tetracycline for further analysis.

Plasmid pEG1381 was isolated from a culture of strain EG12450 grown overnight in the presence of tetracycline. pEG1381 DNA was digested with *Hind*III and compared by agarose gel electrophoresis and ethidium bromide staining to the *Hind*III fragment present in pEG1379. Both plasmids contained an apparently identical about 8 kb *Hind*III fragment.

Strains EG12450, EG2158, and EG10650 were grown overnight in PYG medium An ammonium sulfate dialysate (ASD) was prepared from each of the culture supernatants as described in Example 1. ASD samples were applied to a DEAE column, proteins were eluted from the column, and fractions were collected as described in Example 2. Proteins in the DEAE fractions were analyzed by SDS-PAGE and stained with Coomassie Brilliant Blue. The results demonstrated that DEAE fractions of the control acrystalliferous strain EG10650 did not contain significant amounts of any protein corresponding in size to the TIC901 protein, although some minor proteins of approximately the size of the TIC901 protein were present. In contrast, the DEAE fractions obtained from spent culture supernatants in which strain EG2158 was grown contained a prevalent amount of a protein corresponding to the size of the TIC901 protein. DEAE fractions obtained from spent culture supernatants in which the recombinant strain EG12450 was grown contained significant amounts of a protein corresponding to the size of the TIC901 protein produced by strain EG2158. This result indicated that the 8 kb *Hind*III fragment derived from strain EG2158 present in the recombinant plasmid pEG1381 contained a native *Bacillus thuringiensis* gene a protein exhibiting a mass about equal to that of the the TIC901 protein isolated from strain EG2158 spent fermentation media

Example 5

This example illustrates the identification of the DNA sequence of the native nucleotide sequence encoding the TIC901 protein, and the deduced TIC901 amino acid sequence.

The approximately 8 kilobase pair *Hind*III fragment present in plasmid pMON1378 was subjected to dideoxy-nucleotide sequence analysis. Part of the nucleotide sequence derived from this 8 kb fragment is set forth in SEQ ID NO:5. An open reading frame consisting of 1101 nucleotides was

identified within the approximately 8 kilobase pair *Hind*III fragment, beginning with an ATG methionine codon at nucleotide position 153-155 and terminating at a leucine codon TTG at nucleotide position 1251-1253, immediately upstream of a TAG termination codon as set forth in SEQ ID NO:5 The open reading frame identified within this nucleotide sequence is comprised of 429 adenosine residues (about 39% A), 334 thymidine residues (about 30% T), 196 guanidine residues (about 18% G), and 145 cytosine residues (about 13% C), and exhibits approximately 69% AT and about 31% GC. Three methionine codons are represented within the first eighty four (84) nucleotides of the ORF. Other than the ATG codon at nucleotide positions 153-155 as set forth in SEQ ID NO:3, a second ATG codon in-frame with that at position 153-155 is positioned at nucleotides 182-184 and still a third ATG codon in-frame with these previously mentioned two ATG codons is positioned at nucleotides 201-203. No consensus ribosome binding sequence was observed upstream of either the second or third ATG codons, however a consensus ribosome binding sequence was identified upstream of the first ATG codon, centered at about 10 nucleotides upstream of the first ATG. The consensus ribosome binding sequence consists of nucleotides from about 141 to about 147 as set forth in SEQ ID NO:3.

The amino acid sequence deduced from this open reading frame consists of 367 amino acids as set forth in SEQ ID NO:4 The open reading frame is predicted to encode a precursor protein that has a calculated molecular weight of about 41,492Daltons. 40 amino acids are strongly basic (lysine and arginine), 43 amino acids are strongly acidic (aspartate and glutamate), 113 amino acids are characteristically hydrophobic in nature (alanine, isoleucine, leucine, phenylalanine, tryptophan, and valine), and 122 amino acids are characteristically polar in nature (asparagine, cysteine, glutamine, serine, threonine, and tyrosine). The precursor protein exhibits a calculated isoelectric point of about 6.368 and exhibits a calculated net charge of about –2.102 at pH 7.0. The first thirty (30) amino acid residues encoded by the open reading frame are predicted to comprise the amino terminal signal peptide that is processed, likely by a type II signal peptidase, to release the insecticidal protein into the extracellular space. These first thirty (30) amino acid residues are predicted to comprise an amino terminal signal peptide based on the algorithm of Nielsen et al. (1997, Protein Engineering 10 1-6) The predicted signal peptide in addition to the amino acid residues from position thirty-one (31) through forty-three (43) as set forth in SEQ ID NO:4 are not found to be linked to or associated with the mature protein localized into the spent media and are believed to be proteolytically removed during membrane translocation and release of the mature insecticidal protein into the extracellular space.

Nucleotides 130 through 174 as set forth in SEQ ID NO:3 correspond most closely to the probe sequence as set forth in SEQ ID NO:2. The SEQ ID NO 2 nucleotides, when read in frame with the corresponding coding sequence of the TIC901 amino acid sequence, correspond to the codons encoding the amino terminal amino acids of the insecticidal protein purified from the spent media of cultures of *Bacillus thuringiensis* strain EG2158. The amino acids encoded by the TIC901 reading frame throughout this nucleotide sequence region were expected to correspond to the amino acid sequence of the amino acids deduced by Edmund Degradation of the insecticidal protein purified from the spent media of cultures of strain EG2158 and as set forth in SEQ ID NO:1. The amino acid sequence as set forth in SEQ ID NO:1 is identical to the amino acid sequence deduced by this region of the open reading frame for TIC901, except that amino acid position 12 as set forth in SEQ ID NO:1

corresponds to an arginine residue, whereas the corresponding amino acid residue deduced by the sequence of the open reading frame in SEQ ID NO:3 corresponds to an isoleucine residue at amino acid position 55 as set forth in SEQ ID NO:4. This result suggested either that an error was encountered as a result of Edmund sequence analysis of the protein purified from EG2158 spent culture media and believed to be TIC901 protein, or that more than one TIC901-like protein was present in the concentrated and purified sample submitted for Edmund sequence analysis. It is well known that mistakes in Edmund sequencing are not uncommon. However, this result alone suggested that more than one gene encoding an extracellular secreted insecticidal protein may be present in *Bacillus thuringiensis* strain EG2158. A full analysis of this hypothesis is set forth herein from about Example 9.

Most bacterial proteins released into the extracellular spaces tend to be synthesized as larger precursors that are processed by a secretion apparatus that cleaves a portion of the precursor protein away from the mature protein that is released into the extracellular milieu. It is possible that the insecticidal protein found in the spent media from cultures of strain EG2158 is synthesized in strain EG2158 as a precursor protein that is targeted for secretion by all or part of the forty three (43) amino acids deduced as a part of the open reading frame as set forth in SEQ ID NO:3 Secreted proteins produced by gram positive and gram negative bacteria that are targeted to a type II secretion apparatus typically exhibit a secretory or signal peptide that is between from about eighteen (18) to about twenty eight (28) to thirty (30) or more amino acids in length. As indicated above, there are two methionine codons corresponding to amino acid residues eighteen (18) and twenty seven (27) respectively, as deduced from the open reading frame as set forth in SEQ ID NO:3, upstream of the valine residue (codon at nucleotides 282-284 as set forth in SEQ ID NO:3) deduced by Edmund Degradation to be the amino terminal amino acid of the TIC901 protein. The valine residue as set forth in SEQ ID NO 4 corresponding to the Edmund degradation predicted amino terminal residue of the secreted protein found in the spent media from cultures of EG2158 is positioned at residue number forty four (44). It is conceivable that the two proximal methionine residues within the amino acid sequence amino terminal to the valine residue at position 44 comprise alternative translation initiation sites for synthesis of the proposed precursor TIC901 protein produced in EG2158. However, very poor consensus Shine-Delgarno sequences are available upstream of these two alternative translation initiation codons. In addition, because the protein is found in the spent media, an analysis of the predicted amino acid sequence encoded by the nucleotides following the MET codon at position 153-155 using signal peptide prediction algorithms well known in the art indicates that the first thirty amino acids fit well within the parameters established for amino acid sequences that function as signal peptides. Therefore, it is believed that the first ATG at nucleotide positions 153-155 as set forth in SEQ ID NO:3 is the transcription initiation codon actually utilized by *Bacillus thuringiensis* strain EG2158 for expression of the precursor TIC901 protein that is subsequently processed to the mature form of the protein secreted from the cell.

The predicted signal peptide, as indicated above, is comprised of amino acid residues one (1) through thirty (30) as set forth in SEQ ID NO:4 and exhibits the hallmarks of a classic type II signal peptide, including a basic charge at the amino terminus, a neutral charged predicted alpha-helical core

region from about amino acid residue six (6) through about amino acid residue twenty-two (22) and a consensus signal peptidase II recognition sequence Ala-Xaa-Ala at residues twenty-eight (28) through thirty (30) which would be recognized as a proteolytic cleavage substrate for cleavage of the peptide bond between amino acid residues thirty (30) and thirty-one (31). The carboxyl terminus of the signal peptide has been identified as being crucial as a recognition sequence for the signal peptidase responsible for proteolytic cleavage of the signal peptide from the precursor protein, releasing the mature protein into the extracytoplasmic spaces, i.e., periplasm or extracellular space for gram negative bacteria, or into the extracellular space for gram positive bacteria. The stereotypical consensus signal peptidase II recognition sequence is Ala-Xaa-Ala The rule for the Xaa residue is that it typically cannot be a large or charged amino acid residue. In contrast, the three residues adjacent to the amino terminal amino acid residue (+1) in the mature TIC901 protein do not conform to the consensus type II signal peptidase recognition sequence. Instead, the TIC901 signal peptide amino acid sequence that may be recognized by an alternative or presently unknown and/or uncharacterized signal peptidase is Ser-Gln-Gln (S-Q-Q), which is remarkably uncharacteristic of any signal peptidase recognition sequence previously observed Alternatively, there could be some as yet unidentified peptidase that recognizes the overall structure of amino acids thirty-one (31) through forty-three (43) as set forth in SEQ ID NO:4 and removes these amino acids to release the active insecticidal protein into the media or other surrounding extracellular spaces. These last thirteen amino acids from thirty-one (31) through forty-three (43) as set forth in SEQ ID NO 4 may serve to buffer toxicity of the protein until it is released from the cell.

Example 6

This example illustrates a comparison of the insecticidal biological activity of the protein purified from the native *Bacillus thuringiensis* strain EG2158 to the insecticidal protein isolated from the recombinant *Bacillus thuringiensis* strain EG12450.

Spent media samples from cultures of strains EG12450, EG2158, and EG10650 were obtained and soluble proteins present in each of the samples were concentrated and subjected to DEAE fractionation according to the method as set forth in Example 3 herein. Fractions containing the protein corresponding to TIC901 for each strain were pooled. Pooled fractions were compared side by side in bioassay against Colorado potato beetle The DEAE fractions from EG12450 and EG2158 spent media contained 87 µg/µl and 22 µg/µl TIC901 protein respectively as measured by gel densitometry using known amounts of BSA as a standard. Toxicity to CPB larvae was measured by applying various amounts of the fractions to the surface of insect diet in cups, as described in Example 1. The results are presented in Table 2.

Table 2. Percent mortality of CPB larvae.

	Amount of DEAE Fraction Applied				
Strain	50µl	25µl	12.5µl	6.3µl	3.1µl
EG2158	97%	78%	47%	25%	19%
EG10650	3	0	6	3	9
EG12450	91	94	72	38	16

The results as shown in Table 2 demonstrate that DEAE fractions containing the TIC901 protein from *B. thuringiensis* strains EG2158 and EG12450 are toxic to CPB larvae. Similar DEAE fractions from strain EG10650 were used as a negative control and were not toxic to CPB larvae. Mean standard errors were not calculated.

Example 7

This example illustrates the isolation of a 1 96 kb *NsiI-HincII* restriction fragment containing the entire TIC901 coding region from pEG1379, and an analysis of the corn rootworm inhibitory biological activity of the TIC901 protein isolated from spent cultures of a recombinant strain SIC8098 expressing the TIC901 protein from a plasmid containing the *NisI-HincII* fragment.

A 1 96 kb *NsiI-HincII* restriction fragment containing the entire TIC901 coding region from pEG1379 was subcloned into compatible restriction sites (*PstI* and *SmaI*) in the *B. thuringiensis- E. coli* shuttle vector pEG597 (Baum et al , 1990) The only open reading frame within this restriction fragment that is capable of encoding a protein the size of TIC901 is the coding sequence as set forth in SEQ ID NO:3. The resulting recombinant plasmid, designated pIC17048, was introduced into the acrystalliferous *B. thuringiensis* host strain EG10650 by electroporation. The recombinant *B. thuringiensis* strain containing pIC17048 was designated as strain SIC8098.

B. thuringiensis strains EG10650 and SIC8098 were grown in 300 ml Terrific broth (in 1 L flasks) for 40 hours at 30°C with vigorous shaking. Spent cultures were centrifuged at 8,000 rpm and 4°C for 30 minutes in a Sorvall GS3 rotor Aliquots of the culture supernatants were analyzed by SDS-polyacrylamide gel electrophoresis. Strain SIC8098 accumulates high levels of the TIC901 protein in the culture supernatant. Evaluation of growth conditions for EG2158 and for recombinant *B. thuringiensis* strains encoding TIC901 indicated that a substantially greater amount of TIC901 protein accumulated in the spent media when the strains were fermented at 25°C for forty eight (48) hours instead of under standard conditions of 30°C for eighteen (18) hours. This suggests that the TIC901 protein may not be optimally produced until late log or early stationary phases of growth have been attained.

EG10650 and SIC8098 culture supernatants were brought to 85% saturation with ammonium sulfate at 4°C. Precipitated proteins were collected by centrifugation The protein precipitates were solubilized in 10 mM Tris-HCl, 0.1 mM EDTA, 0.005% Triton X-100 (pH 6.8) and dialyzed against a 200-fold excess volume of the same buffer. Insoluble material was pelleted by centrifugation and the clarified dialysate containing the substantially purified TIC901 protein was used directly for insect bioassays. The amount of purified TIC901 protein in the dialysate was determined by SDS-

polyacrylamide gel electrophoresis and densitometry, using bovine serum albumin as a protein standard. Bioassays were conducted in 96-well plates, each well containing an artificial rootworm diet. Samples containing toxin or control samples were deposited onto the diet providing a toxin impregnated surface overlay. An average of one rootworm egg was deposited per well. An average of twenty-four (24) wells were utilized per treatment. Plates were sealed and incubated for one week before larvae were characterized for mortality and mass.

Bioassays were performed against both southern and western corn rootworm larvae, using the rootworm toxin Cry3Bb (amino acid sequence variant 11231, English et. al., US Patent No. 6,063,597) as a positive control Assay results are shown in Tables 3 and 4.

Table 3. TIC901 Bioassay vs Western corn rootworm

Source Strain	Toxin mg/ml	Mean larval mass	95% CI[1]	Significance[2]
EG10650	0	0.44	0.35-0.53	
Cry3Bb	0.25	0 18	0 13-0.22	*
Cry3Bb	0.50	0.15	0.10-0.20	*
Cry3Bb	1.0	0.13	0.10-0.16	*
TIC901	0.25	0.36	0 25-0.48	
TIC901	0 50	0.48	0.36-0.61	
TIC901	1.0	0.29	0.26-0.31	*
TIC901	2.0	0.27	0.23-0 32	*

[1] 95% confidence interval
[2] An asterisk (*) indicates that the results were significantly different from the EG10650 negative control, Dunnett's test, alpha = 0.05

Table 4. TIC901 Bioassay vs Southern corn rootworm

Source Strain	Toxin mg/ml	Mean larval mass	95% CI[1]	Significance[2]
EG10650	0	0.64	0.55-0.72	
Cry3Bb	0.25	0.45	0.40-0.51	*
Cry3Bb	0.50	0.42	0.39-0 44	*
Cry3Bb	1 0	0.31	0.19-0 42	*
TIC901	0.25	0.38	0.31-0.45	*
TIC901	0 50	0.33	0.20-0.45	*
TIC901	1.0	0.26	0 14-0 38	*

[1] 95% confidence interval
[2] An asterisk (*) indicates that the results were significantly different from the EG10650 negative control, Dunnett's test, alpha = 0.05

The results demonstrate that TIC901 is active against both rootworm species, but exhibits greater activity against the southern corn rootworm species.

Example 8

This example illustrates the construction of a nucleotide sequence encoding a TIC901 protein for enhanced expression in plants.

The amino acid sequence of the TIC901 protein as set forth in SEQ ID NO:4 was used to construct a nucleotide sequence essentially according to Fischhoff and Perlak, US Patent Serial No. 5,500,365, and essentially according to Brown et al., US Patent Serial No. 5,689,052. Briefly, a codon frequency table was derived from more than 53,000 sequence segments of corn, rice, and wheat genomic DNA believed to encode protein sequence in those plants. More than ten million nine hundred thousand codons (10,900,000) were derived from these sequence segments. Codons were then selected for use in constructing the non-naturally occurring or synthetic sequence encoding TIC901 for use in plants. The resulting coding sequence is composed of a GC composition which resembles the sequences present in plants, and in particular, because monocot plants were used to derive the codon usage table, the coding sequence resembles a monocot gene in architecture more so than a dicot gene and is therefore preferable for use in expressing TIC901 protein or protein variants in monocot plant species. One such sequence for a non-naturally occurring coding sequence preferable for use in monocot species is set forth in SEQ ID NO:13, and the amino acid sequence of the encoded protein is set forth in SEQ ID NO:14. The skilled artisan would recognize that SEQ ID NO:13 is but one of a myriad number of coding sequences that would function to express the insecticidal protein in a monocot plant species. Furthermore, the skilled artisan would also recognize that a synthetic coding sequence such as SEQ ID NO.13 encoding all or an insecticidal portion of the TIC901 protein or an insecticidal amino acid sequence variant thereof consisting of other than the sequence as set forth in SEQ ID NO:4, but which hybridized to SEQ ID NO:13 under stringent hybridization conditions would also be within the scope of the present invention

Example 9

This example illustrates the identification and characterization of homologues of *tic*901 and peptides encoded by these nucleotide sequence homologues that are substantially similar to and thus related to the TIC901 protein.

As set forth in Table 1, a number of strains that produced extracellular protein profiles exhibiting coleopteran insecticidal biological activity were identified. The TIC901 protein coding sequence was identified as exemplified in the Examples herein-above from a strain EG2158 plasmid library by probing with SEQ ID NO:2, corresponding to sequences encoding SEQ ID NO:1 and biased toward the codon usage preference exhibited by native *Bacillus thuringiensis* genes encoding δ-endotoxins. The *tic*901 coding sequence was then used as a probe to identify homologues of *tic*901 in genomic plasmid library clones prepared from genomes of other *Bacillus thuringiensis* strains exhibiting secreted insecticidal protein activity an RFLP profile different from the RFLP profile exhibited by strain EG2158.

As set forth in Table 1, *Bacillus thuringiensis* strain 86833 contained a *tic*901 homologous sequence different from that of strain EG2158. The secreted protein profile from strain 86833 contains a major band at 38 kDa which is similar to that generated by strain EG2158. These facts together

suggested that strain 86833 contained a *tic*901 homolog. The N-terminal sequence of the strain 86833 protein was determined by Edmund degradation, and a fifteen (15) amino acid sequence was identified (amino acids seventy six (76) through ninety one (91) as set forth in SEQ ID NO:6, which also correspond to residues 2-16 of the predicted mature TIC1201 protein). The amino acid sequence did not correspond to any previously known amino acid sequence, but exhibited an 87% identity with the corresponding amino terminal amino acid sequence of the predicted mature TIC901 protein described herein above. A plasmid clone, pJP263, derived from a Lambda ZAP Express DNA library prepared using *Sau*IIIA partially digested, size selected, strain 86833 total DNA containing an insert of about 5 kilobase pairs was identified by probing essentially as described in Example 2 with *tic*901 DNA labeled with digoxygenin under thermal amplification conditions. Plasmid pJP263 was subjected to thermal amplification conditions to excise the entire strain 86833 insert which was subsequently inserted into the *E. coli* / *Bacillus* shuttle vector pEG597. The resulting plasmid pJP274-1 was used to transform the acrystalliferous strain EG10650, and an isolate containing this plasmid was selected and designated as the recombinant *B. thuringiensis* strain SIC4018. Strain SIC4018was fermented as described in Example 1 and the proteins present in the spent culture media were concentrated and tested for insecticidal activity in bioassay, and analyzed by SDS-PAGE. The spent culture media tested positive for coleopteran insecticidal bioactivity, and a protein of about 38 kDa was identified by SDS PAGE analysis as being present in overabundance in the spent culture media. The insert in pJP274-1 was sequenced by the Sanger dideoxy method and a single open reading frame as set forth in SEQ ID NO:5 from nucleotide four hundred thirty seven (437) through sixteen-hundred-twenty-one (1621)was identified. The amino acid sequence of the precursor insecticidal protein deduced from the open reading frame set forth in SEQ ID NO:5 was designated as TIC1201 (SEQ ID NO:6). The nucleotide sequence of a clone corresponding to the "B" type RFLP profile from another strain, EG3618, was also characterized, and determined to contain an ORF identical in sequence to that as set forth in SEQ ID NO:5.

Strain EG6618 exhibited a "C" type RFLP profile as exemplified in Table 1. A plasmid clone, pJP264, derived from a Lambda ZAP Express DNA library prepared using *Sau*IIIA partially digested, size selected, total DNA from strain EG6618 containing an about 5 kilobase pair fragment was identified by probing essentially as described in Example 2 with *tic*901 DNA labeled with digoxygenin under thermal amplification conditions. . DNA sequence analysis of the inserted DNA in pJP264 indicated that there was an incomplete, or partial, open reading frame within this clone, that possibly encoded the C-terminal portion of a new and previously uncharacterized protein. Alignment of this partial ORF with the ORF encoding TIC901 as set forth in SEQ ID NO:3 revealed that the nucleotide sequence of the inserted DNA in pJP264 (nucleotides 481 through 1302 as set forth in SEQ ID NO:7) encoded a peptide, the deduced amino acid sequence of which was substantially similar to the amino acid sequence corresponding to TIC901 amino acid residues 96 through 367. The new amino acid sequence encoded by the nucleotide sequence in pJP264 exhibited an approximately 78 percent identity with the analogous TIC901 peptide sequence. The nucleotide sequences encoding these peptides exhibited an approximately 81 percent identity. The partial coding sequence within pJP264 corresponds to nucleotide position 481 through 1302 as set forth in SEQ ID NO:7 and is analogous to

the TIC901 coding sequence as set forth in SEQ ID NO:3 from nucleotide position 438 through nucleotide 1253. The substantial identity between the carboxyl terminus of each of these two proteins suggested that the peptide encoded by the sequence in pJP264 represented a new member of the genus of insecticidal proteins related to the TIC901 protein. Therefore, this protein was designated as TIC407, and the nucleotide sequence encoding the TIC407 protein was designated as *tic*407.

It was believed that a full length ORF existed for TIC407 within the genome of EG6618. In order to identify the nucleotide sequence of the balance of the 5' portion of the TIC407 coding sequence, a sample of *Bacillus thuringiensis* strain EG6618 total DNA was digested with the restriction enzyme *Nde*I, the products of which were then diluted and circularized by ligation to generate a population of templates for use in an inverse thermal amplification reaction The enzyme *Nde*I was selected because (1) it did not cut the insert sequence in pJP264, (2) it cut often within the *Bacillus thuringiensis* genome because of the prevalence of A and T in the restriction enzyme recognition sequence, and (3) Southern blot analysis of EG6618 DNA indicated that *Nde*I cut, within reason, near to the endpoints of the DNA inserted into plasmid pJP264 but sufficiently far enough away from the endpoints to incorporate a full length ORF encoding *TIC407*.

Two divergent inverse thermal amplification primers were prepared based on the partial *tic*407 nucleotide sequence present in pJP264. Primer prJWP151 (SEQ ID NO:15 CCTTTGGCAGAAACTTTAACTCC) corresponds to the reverse complement of nucleotides 766-788 as set forth in SEQ ID NO:7. Primer prJWP152 (SEQ ID NO 16 GTGTATTCTGGTACGCATGAC) corresponds to nucleotides 955-975 as set forth in SEQ ID NO:7. These primers were included along with *Nde*I circularlized template molecules described above and the necessary reagents, buffers and *Pfu* polymerase in a thermal amplification reaction consisting of a five minute denaturation step at 95°C for five minutes, followed by thirty cycles consisting of a one minute annealing step at 37°C, a two minute extension step at 55°C, and a one minute denaturation step at 95°C, terminating with a five minute extension step at 55°C and a 4°C soak until the samples could be further processed. A single thermal amplification product was identified by gel electrophoresis, which was excised and cloned into a T-vector to construct plasmid pJP300-2. The nucleotide sequence of the inserted DNA sequence in pJP300-2 was obtained and provided sufficient nucleotide sequence upstream and downstream of the previous sequence present in pJP264 to identify the complete open reading frame encoding the TIC407 insecticidal protein as well as upstream and downstream flanking sequences. The nucleotide sequence information obtained from pJP300-2 was assembled by computer using sequence analysis software because the sequence present in pJP300-2 was constructed using inverse thermal amplification of a restriction fragment that was circularized from the genome of strain EG6618. Therefore, to confirm that the sequence that had been assembled was accurate and in fact encoded the TIC407 amino acid sequence, first from the partial open reading frame identified in pJP264, and then from the partial sequences identified in pJP300-2, two new thermal amplification primers flanking the open reading frame encoding TIC407 were designed (prJWP186 and prJWP183) for use in the amplification of a completeTIC407 coding sequence from EG6618 genomic DNA.

Construction of primer prJWP186 (SEQ ID NO.17 gccggatccCTAGCTGAATATGCAGTAGATAATG), was based in part on the sequence identified as

being upstream of the TIC407 open reading frame in pJP300-2. The first nine (9) nucleotides of this primer correspond to nucleotide sequence that is not present in the sequence identified as being present in either pJP264 or in pJP300-2, and also contains a sequence that incorporates a *Bam*HI restriction site into the sequence upstream of the amplicon produced using these primers. The terminal twenty-five (25) nucleotides in primer prJWP186 correspond to the sequence as set forth at nucleotides twenty-eight (28) to fifty-two (52) in SEQ ID NO:7 which lie upstream of the proposed ATG initiation sequence in the TIC407 ORF, and enable primer extension in toward the TIC407 ORF.

Construction of primer prJWP183 (SEQ ID NO.18 GTGGCACGTTTATAGGCCATTGTTC) was based entirely on the sequence deduced as being within sequence downstream of the TIC407 ORF within pJP300-2, and is the reverse complement of the sequence from nucleotide position seventeen-hundred-thirty-five (1,735) to seventeen-hundred-fifty-nine (1,759) as set forth in SEQ ID NO.7. While there are no restriction site sequences incorporated into this primer, a naturally occurring *Hind*III recognition sequence is present downstream of the TIC407 ORF as set forth in SEQ ID NO·7 from nucleotide fifteen-hundred to fifteen-hundred-five (1,500-1,505).

Primers prJWP186 and prJWP183 were included in a thermal amplification reaction along with a sample of EG6618 total DNA and reagents necessary to carry out the reaction. Amplification conditions were similar to those described above for amplification of the *Nde*I circularized sample. An amplicon comprising 1,732 nucleotides corresponding to the TIC407 ORF, including flanking nucleotide sequence up and down stream of the ORF, was isolated. This amplicon was cloned into a T-vector to construct pJP306-4, and the complete nucleotide sequence of the amplicon was obtained. As expected, the nucleotide sequence of the amplicon in pJP306-4 corresponded to nucleotides twenty-eight (28) through seventeen-hundred-fifty-nine (1,759) as set forth in SEQ ID NO 7, and contained an open reading frame as set forth in SEQ ID NO:7 from nucleotide position one hundred ninety six (196) through twelve-hundred-ninety-nine (1,299) encoding a protein consisting of 368 amino acids. Alignment of the nucleotide sequence encoding TIC407 to the sequence encoding TIC901 demonstrated that the two ORF's are about eighty percent (80%) identical, however, the amino acid sequences deduced from the ORF's exhibit only about 74% identity.

The amino acid sequences of TIC901, TIC1201, and TIC407 were aligned to form a consensus amino acid sequence. Even though the primary amino acid sequences were significantly different, regions of substantial identity were readily identifiable. Two regions, corresponding to amino acid sequence ASN-ASN-ASN-HIS-GLN-THR-ASN-ARG from amino acid sequence position 96-103 as set forth in SEQ ID NO·4 and the amino acid sequence GLN-LYS-PHE-ILE-TYR-PRO-ASN from amino acid sequence position 276-282 as set forth in SEQ ID NO:4 were 100% conserved in primary sequence and position within each of the three insecticidal protein sequences. The nucleotide sequence encoding these protein sequences in each of the corresponding open reading frames, i e , as set forth in each of the sequences in SEQ ID NO:3, SEQ ID NO:5, and SEQ ID NO.7, (TIC901, TIC1201, and TIC407, respectively) was also substantially conserved, providing a means for identifying other homologous sequences encoding such secreted insecticidal proteins present in other *Bacillus* or *Bacillus thuringiensis* strains.

Two thermal amplification primer sequences were prepared for use in a thermal amplification reaction intended for use in identifying nucleotide sequences encoding other homologs of the three insecticidal proteins TIC901, TIC1201, and TIC407. The amplification primer sequences are set forth herein as (1) SEQ ID NO:11, a forward amplification twenty-one-mer primer sequence designated as prJWP139, corresponding to the coding sequence as set forth in SEQ ID NO.3 from nucleotide position 438 through 458 except that the nucleotide at position eighteen (18) within the oligonucleotide is redundant in that either an adenosine or a thymidine nucleotide is incorporated therein, resulting in two possible primer sequences, and (2) SEQ ID NO:12, a reverse amplification twenty-mer primer sequence designated as prJWP143, corresponding to the reverse complement of the sequence as set forth in SEQ ID NO:3 from nucleotide position 998 through 978, except that the nucleotide within the oligonucleotide at position four (4) and/or at position ten (10) is/are redundant in that either an adenosine or a thymidine nucleotide is incorporated at either position, resulting in four possible primer sequences. SEQ ID NO.11 and SEQ ID NO:12 are comprised of a nucleotide sequence that is biased toward codons preferred for use in gene sequences derived from *Bacillus thuringiensis* or other *Bacillus* species strains, in which the codons contain A and/or T in the third position for each codon represented within the sequence.

As indicated above, the amino acid sequence identified by Edmund Degradation of the about 38 kDa protein isolated and purified from spent culture media from fermentations using strain EG2158 was different by one amino acid from the corresponding sequence ultimately deduced from the open reading frame for TIC901 as set forth in SEQ ID NO 4 It was believed that the difference in amino acid sequences observed by Edmund degradation and the resulting TIC901 amino acid sequence deduced from the *tic*901 ORF could be the result of more than one gene present in strain EG2158, each gene encoding a different secreted insecticidal protein, and each secreted protein being about 38 kDa in size. In fact, if two or more genes were present in EG2158, each would have to be present on virtually identical *Hind*III fragments, or both could be present on the same *Hind*III fragment, or one may be sufficiently different from the *tic*901 coding sequence that it would not hybridize to a nucleotide sequence comprising SEQ ID NO:2 under stringent hybridization conditions. If each gene indeed encoded a homologous insecticidal protein, or at least a protein that was sufficiently similar to that protein encoded by the other such that the conserved sequences identified above were present within each coding sequence, then redundant primers such as those set forth in SEQ ID NO:11 and SEQ ID NO.12 may allow for the identification of the genes by amplification of the related coding sequences from a common sample in a thermal amplification reaction For example, primers comprising the sequences as set forth in SEQ ID NO:11 and SEQ ID NO:12 would be expected to produce an amplicon of about 560 base pairs when using the TIC901, TIC1201, or TIC407 coding sequences as templates in a sample. In order to test this hypothesis, EG2158 total DNA was used as the template in a thermal amplification reaction using the amplification primers set forth in SEQ ID NO:11 and SEQ ID NO:12 under standard reaction conditions comprising 10-100 nanograms of bacterial genomic DNA template, 50 picomoles of each primer, 1X ROCHE amplification reaction buffer containing a final 1.5 mM divalent Mg^{2+} cation concentration, 0.2 mM each deoxynucleotide triphosphate (dNTP), and 2.5 units of ROCHE TAQ Polymerase per 50 µl reaction volume, an initial denaturation cycle at 94°C for

two minutes, thirty (30) amplification cycles comprising a 94°C denaturation phase for one minute, a 45°C anneal phase for two (2) minutes, and a 72°C primer extension phase for one (1) minute, followed by a final 72°C primer extension and finishing phase for five (5) minutes, and a 4°C soak. In addition, template DNA obtained from the following strains was also used in parallel thermal amplification reactions along with the amplification primers as set forth in SEQ ID NO:11 and SEQ ID NO:12: strain EG10650, strain EG4332, strain EG5552, strain EG5858, strain EG6489, strain EG6561, strain EG3618, strain EG6555, strain EG6618, and strain 86833. Samples of each thermal amplification reaction were analyzed by gel electrophoresis. The results are presented in Table 5.

Table 5. Amplicons Produced From B. thuringiensis Strains

Template DNA	*Hind*III RFLP Type	Presence of Amplicon[1]
EG10650	-	-
EG4332	-	-
EG5552	-	-
EG5858	-	-
EG2158	A	+
EG6489	A	+
EG6561	A	+
EG3618	B	+
EG6555	B	+
86833	B	+
EG6618	C	+

1- amplicon produced in thermal amplification reaction using template DNA from indicated strain, amplification primers comprising the sequence as set forth in SEQ ID NO:11 and SEQ ID NO:12 under standard conditions as described above.

No amplicon was detected when DNA from strains EG10650, EG4332, EG5552, or EG5858 was used as template along with this particular primer set. This result suggests that these strains do not contain nucleotide sequences encoding a homolog of TIC901, TIC1201, and TIC407. However, each of the other strains tested appeared to produce a single amplicon of about 560 base pairs in length within the limits of resolution of the gel system used. Amplicons from each of the samples producing a positive reaction under these amplification conditions were cloned into a T-vector, and the nucleotide sequence was determined for randomly selected cloned amplicons from each thermal amplification reaction. Strains EG2158, EG6489, and EG6561 each produced an amplicon exhibiting a nucleotide sequence identical to the corresponding TIC901 coding sequence. Strains EG3618 and 86833 produced an amplicon exhibiting a nucleotide sequence identical to the corresponding TIC1201 coding sequence. Strain EG6555, which is a strain exhibiting a "B" type *Hind*III RFLP profile identical to that exhibited by strains EG3618 and 86833, also produced a 560 base pair amplicon. Strain EG6555 was

not previously fully characterized, however, it was previously shown to generate a "B" type RFLP as set forth in Table 1, which corresponds to the RFLP phenotype exhibited by strain 86833 from which TIC1201 was isolated. Therefore it was anticipated that the homolog present in strain EG6555 encoded a protein substantially similar if not identical to TIC1201. Sequence analysis of an amplicon obtained from EG6555 was identical to the corresponding sequence obtained from strain 86833. As expected, strain EG6618 produced an amplicon exhibiting a nucleotide sequence identical to the corresponding TIC407 coding sequence.

Surprisingly, in addition to amplicons comprising the corresponding *tic*901 sequence, amplicons were derived from total DNA isolated from EG2158 that contained a second coding sequence that was different from the corresponding TIC901 nucleotide sequence identified using the redundant probe as set forth in SEQ ID NO 2. As anticipated, one amplicon obtained from EG2158 DNA was also comprised of about 560 nucleotides, but DNA sequence analysis of the clone resulted in the identification of an ORF (nucleotides 377 through 937 as set forth in SEQ ID NO:9) encoding a peptide sequence (amino acid sequence 96 through 282 as set forth in SEQ ID NO.10) that was substantially different from the corresponding peptide sequence for TIC901. This result suggested that there was yet at least one other nucleotide sequence present in strain EG2158 that may encode a protein homolog of TIC901 that could be considered as another member within the genus of proteins related to TIC901, TIC1201, and TIC407.

In order to fully characterize this second sequence identified from strain EG2158 DNA, inverse thermal amplification primers were designed based upon the second sequence that would specifically amplify the flanking sequences up and downstream of the second sequence. Inverse thermal amplification of EG2158 DNA was used to identify the complete open reading frame containing this second sequence. First, enzymes were selected for use in digesting EG2158 DNA that would (1) not cut within the second sequence identified above, and (2) cut frequently within *Bacillus* DNA as a result of having an AT rich sequence comprising the enzymes' recognition sequence. Considering these parameters, *Xba*I and *Spe*I were selected, but neither enzyme when used alone digested EG2158 DNA into fragments that could be easily circularized for use as inverse thermal amplification templates. However, since both enzymes produce compatible 5' overhanging ends, double-digestion of EG2158 DNA using both enzymes produced products of sufficiently small size that, when circularized, could be used as templates for inverse thermal amplification. Inverse amplification primers were designed based on the second sequence identified above, particularly sequences in which the corresponding *tic*901 sequence (SEQ ID NO:3) and the second sequence exhibited significant differences upon alignment.

Primer prJWP155 (SEQ ID NO:19) corresponds to and is the reverse complement of the second sequence as set forth in SEQ ID NO:9 from nucleotide 454-476). Primer prJWP156 (SEQ ID NO:20) corresponds to the second sequence as set forth in SEQ ID NO:9 from nucleotide 692-717. These two primers were used in an amplification reaction with double-digested, circularized EG2158 template DNA as indicated above, under conditions similar to those set forth above for inverse thermal amplification of EG6618 DNA. A single amplicon was obtained that was inserted into a T-vector to construct plasmid pJP290-1, and the nucleotide sequence of the inserted amplicon was determined.

The inverse thermal amplification sequence derived using this method was assembled using DNA sequence analysis software and aligned with the second sequence identified above to construct a sequence as set forth in SEQ ID NO:9 that included a complete open reading frame encoding a peptide sequence designated as TIC417 as well as upstream and downstream flanking sequence.

To confirm the sequence, assembled in part from the second amplicon identified from EG2158 DNA as set forth above, and in part from the inverse thermal amplification reaction using primers prJWP155 and 156, two additional thermal amplification primers were designed that were complementary to the sequences flanking the open reading frame encoding TIC417 (prJWP168 and prJWP170). These primers would allow for the amplification of the complete TIC417 open reading frame along with a short length of DNA upstream and downstream of the TC417 ORF. Primer prJWP168 (SEQ ID NO.21) was constructed based in part on the sequence identified as being downstream of or 3' to the TIC417 ORF and functions to extend nucleotide sequence polymerization into the TIC417 ORF. The first five nucleotides of this primer correspond to nucleotides that are not present in the sequence identified as being present in plasmid pJP290-1and also serve to incorporate a *Hind*III restriction site when coupled with bases 7-8 of SEQ ID NO:21. The terminal 39 nucleotides of prJWP168 correspond to the reverse complement of nucleotides 1148 through 1186 as set forth in SEQ ID NO.9. Primer prJWP170 (SEQ ID NO:22) was constructed based in part of the sequence identified as being upstream of or 5' to the TIC417 ORF and functions to enable primer extension in toward the 5' end of the TIC417 ORF. The first 9 nucleotides of this primer correspond to nucleotides that are not present in the sequence identified as being present in pJP290-1 and also serve to incorporate a *Bam*HI restriction site recognition sequence upstream of the predicted TIC417 ATG initiation codon. The terminal 30 nucleotides of this primer correspond to nucleotides The protein encoded by this different coding sequence was designated as TIC417. The amino acid sequence encoded by the cloned portion of the TIC417 amplicon is set forth in SEQ ID NO:10, and the cloned amplicon coding sequence is set forth in SEQ ID NO:9. The amino acid sequence of TIC417 is about 85% identical to the corresponding sequence from TIC901 and TIC1201, and is about 76% identical to the corresponding amino acid sequence of TIC407. It is believed that the protein fragment corresponding to the TIC417 coding sequence revealed in the TIC417 amplicon derived from EG2158 DNA is an insecticidal protein similar to TIC901, TIC1201 and TIC407, corresponding to a protein that is expressed as a precursor protein exhibiting a signal peptide of about 43 amino acid residues in length that terminates in a consensus signal peptide cleavage sequence corresponding to SER-GLN-GLN, which is processed by a signal peptidase to release an about 38 kDa, plus or minus about 2 kDa, mature protein into the extracellular space. It is also anticipated that the TIC417 mature amino terminal amino acid sequence may correspond to the amino acid sequence as set forth in SEQ ID NO:1, and that the TIC417 protein exhibits coleopteran species inhibitory insecticidal biological activity against corn rootworms and Colorado potato beetles.

Example 10

This example illustrates the design and use of degenerate probes and primers for use in identifying naturally occurring nucleotide sequences encoding all or part of insecticidal proteins derived from the proteins of the present invention or homologues thereof.

Based on an amino acid sequence alignment of the TIC901, TIC1201, TIC407 and TIC417 precursor proteins as set forth in FIGURE 1, a number of regions of amino acid sequence identity were observed, and degenerate oligonucleotide sequence primers were designed based on the corresponding native nucleotide coding sequence for each protein, taking into consideration the profile for known *Bacillus thuringiensis* insecticidal proteins. While there are many conserved amino acid sequence segments to choose from within the alignment as set forth in FIGURE 1 for designing degenerate primers, three segments were selected based on the position of the primer sequences within the coding region of the known TIC proteins and the unique size of the amplicons that would be expected to be generated in a thermal amplification reaction with these primers.

A first region or segment of conserved amino acid sequence corresponding to amino acid seventy-five (75) through amino acid eighty-three (83) as set forth in SEQ ID NO.4 was identified, and the corresponding nucleotide sequence encoding this amino acid segment from each of the sequences as set forth in SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, and SEQ ID NO.9 were then aligned. The consensus nucleotide sequence segment encoding this particular amino acid sequence segment was utilized to determine the scope of nucleotide sequences comprising the forward amplification primer sequence as set forth in SEQ ID NO:23 (prJWP200), as well as two other degenerate primers consisting of subsets of the SEQ ID NO:23 degenerate primer (SEQ ID NO:24-25, corresponding to prJWP201 and prJWP202 respectively).

A second region or segment of conserved amino acid sequence between the proteins of the present invention is exemplified by the amino acid sequence from about amino acid one-hundred-forty-seven (147) through about one-hundred-fifty-three (153) as set forth in SEQ ID NO:4. The nucleotide sequence encoding this amino acid segment from each of the sequences as set forth in SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, and SEQ ID NO:9 were then aligned. The consensus nucleotide sequence segment encoding this particular amino acid sequence segment was utilized to determine the scope of nucleotide sequences comprising the forward amplification primer sequence as set forth in SEQ ID NO:26 (prJWP203)

A third region or segment of conserved amino acid sequence between the proteins of the present invention is exemplified by the amino acid sequence from about amino acid two-hundred-seventy-five (275) through about amino acid two-hundred-eighty-three (283) as set forth in SEQ ID NO.4. The nucleotide sequence encoding this amino acid segment from each of the sequences as set forth in SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, and SEQ ID NO:9 were then aligned. The consensus nucleotide sequence segment encoding this particular amino acid sequence segment was utilized to determine the scope of nucleotide sequences comprising the reverse amplification primer sequence as set forth in SEQ ID NO:27 (prJWP204), as well as two other degenerate primers consisting of subsets of the SEQ ID NO:27 degenerate primer (SEQ ID NO.28-29, corresponding to prJWP205 and prJWP206 respectively).

--

Any combination of primers prJWP200, prJWP201, prJWP202, or prJWP203 with prJWP204, prJWP205, or prJWP206 can be used together in a thermal amplification reaction with a template nucleotide sequence encoding a TIC901, TIC1201, TIC407, TIC417 or a related secreted insecticidal protein or homologue thereof to produce an amplicon corresponding to a sequence segment that hybridizes to the corresponding sequences between the positions within the TIC coding sequences used for designing the primer sequences prJWP200-206. If any of the primers prJWP200-202 are used along with any of the primers prJWP204-206 in a thermal amplification reaction with an appropriate template in a sample, the production of an amplicon of from about 590 to about 650 base pairs in size is characteristically diagnostic for the presence of one or more coding sequences in the sample that are related to the sequences of the present invention. Generally, the diagnostic amplicon using primers within these ranges would consist of from about 617 to about 626 base pairs.

The use of any of the primers comprising SEQ ID NO:26 (prJWP203) along with any of the primers comprising SEQ ID NO:27-29 (prJWP204-206) in a thermal amplification reaction with an appropriate template in a sample, the production of an amplicon of from about 390 to about 415 base pairs in size is characteristically diagnostic for the presence of one or more coding sequences in the sample that are related to the sequences of the present invention Generally, the diagnostic amplicon using primers within these ranges would consist of from about 400 to about 410 base pairs.

To exemplify the utility of these primers, primer prJWP200 (SEQ ID NO.23) was combined in a thermal amplification reaction with primer prJWP204 (SEQ ID NO:27), each at a concentration of at least about 1 pico-mole per micro-liter along with 1X TAQ amplification buffer, 0 2 molar each deoxy-nucleotide tri-phosphate (dATP, dTTP, dCTP, and dGTP), 2 millimolar $MgCl_2$, 2 units TAQ polymerase, and from about ten (10) to about one hundred (100) nano-grams of a sample of genomic DNA from strain EG2158 known to contain the coding sequences for TIC901 and TIC417. Thermal amplification cycle conditions consisted of an initial denaturation of about 2 minutes at 94°C followed by 35 cycles of a denaturation step of 30 seconds at 94°C, an annealing step of 30 seconds at 50°C, and an extension step of 45 seconds at 72°C followed by a final extension step of 7 minutes at 72°C. The temperature of the annealing step was decreased by 0.3°C for each successive cycle so that the final annealing temperature was about 39.8°C.

A ten microliter sample was analyzed on a 1.2% TAE agarose gel and compared to a co-migrating 100 base pair ladder, and stained with ethidium bromide. The results indicated the production of an amplicon segment corresponding to about 620 base pairs. The about 620 base pair band was excised and the DNA recovered for cloning. Several independent clones representing this segment were isolated and the nucleotide sequence of each was obtained. As expected, a first sequence identical to the corresponding nucleotide sequence encoding TIC901 was identified (not including the primer sequences at either end of the clone, from about nucleotide position four-hundred-two (402) through about nucleotide position nine-hundred-seventy-four (974) as set forth in SEQ ID NO.3), and a second sequence identical to the corresponding nucleotide sequence encoding TIC417 was identified (not including the primer sequences at either end of the clone, from about nucleotide position four-hundred-sixty-four (464) through about nucleotide position one-thousand-thirty-six (1,036) as set forth in SEQ ID NO:9).

Surprisingly, a third sequence (SEQ ID NO:30) was also identified that did not correspond identically to any of the sequences set forth herein . However, the third sequence was substantially similar in nucleotide sequence to each of the sequences encoding a secreted insecticidal protein set forth herein, including tic901, *tic*1201, *tic*407, and *tic*417. As indicated in the examples above, it was quite unexpected to find the *tic*417 coding sequence in the strain EG2158 genomic DNA. It was even more surprising to identify yet a third nucleotide sequence that likely corresponds to a nucleotide segment that encodes yet a third secreted insecticidal protein from strain EG2158 that is different from those that have been set forth herein above, but which is sufficiently similar in sequence to be classified as one of the species within the genus of secreted insecticidal proteins encoded by a nucleotide sequence that hybridizes to one or more of the sequences set forth herein, and is exemplary of the novelty and utility of the degenerate oligonucleotide probes and primers exemplified herein for use in identifying sequences that encode secreted insecticidal proteins and that hybridize under stringent conditions to the related *tic*901, *tic*1201 *tic*07, and *tic*417 coding sequences as set forth herein.

The amino acid sequence encoded by the uninterrupted open reading frame as set forth in SEQ ID NO:30, which has had the twenty-six-mer degenerate oligonucleotide sequences deleted from both the 5' and 3' ends, is set forth in SEQ ID NO:31. The amino acid sequence set forth in SEQ ID NO:31 is substantially similar to the amino acid sequence of TIC901 from about amino acid position eighty-five (85) through about amino acid position two-hundred-seventy-four (274) as set forth in SEQ ID NO:4, containing only two (2) amino acids that are different from the analogous sequence in SEQ ID NO:4, corresponding to an about 98.9% identity between SEQ ID NO:31 and the corresponding sequence in SEQ ID NO:4. The amino acid sequence set forth in SEQ ID NO:31 is substantially similar to the amino acid sequence of TIC1201 from about amino acid position eighty-five (85) through about amino acid position two-hundred-seventy-four (274) as set forth in SEQ ID NO:6, containing only thirteen (13) amino acids that are different from the analogous sequence in SEQ ID NO.6, corresponding to an about 93.2% identity between SEQ ID NO:31 and the corresponding sequence in SEQ ID NO 6. The amino acid sequence set forth in SEQ ID NO:31 is substantially similar to the amino acid sequence of TIC417 from about amino acid position eighty-five (85) through about amino acid position two-hundred-seventy-four (274) as set forth in SEQ ID NO.10, containing only thirty (30) amino acids that are different from the analogous sequence in SEQ ID NO:10, corresponding to an about 83.7% identity between SEQ ID NO:31 and the corresponding sequence in SEQ ID NO.10. The amino acid sequence set forth in SEQ ID NO.31 is also substantially similar to the amino acid sequence of TIC401 from about amino acid position eighty-five (85) through about amino acid position two-hundred-seventy-four (274) as set forth in SEQ ID NO:8, containing forty-one (41) amino acids that are different from the analogous sequence in SEQ ID NO:8, corresponding to an about 78.4% identity between SEQ ID NO:31 and the corresponding sequence in SEQ ID NO:8.

In contrast to the results obtained when using DNA template from strains EG5858, EG4332, and EG5552 along with the primer set comprising SEQ ID NO:11 and SEQ ID NO:12 (as set forth in Table 5, above), DNA from these strains each produce an amplicon, the sequence of which appears to correspond to the TIC407 coding sequence, when using a primer set corresponding to primer prJWP200 and prJWP204. The corresponding coding sequence for a TIC407 homolog may be present

on a large *HindIII* fragment in these strains and therefore may not have been readily available for identification when carrying out the initial screen as described above with a Southern blot using *tic*901 DNA as the probe, the results of which were set forth in Table 1.

Example 11

This example illustrates the identification of unique TIC901 related RFLP patterns and insecticidal protein coding sequences from *Bacillus* strains that produce in culture supernatant immunologically related TIC901extracellular protein as determined using ELISA detection methods but whose DNA fails to result in the production of detectable thermal amplification products using standard *tic*901 thermal amplification detection methods.

Initial analyses as set forth using the examples above relied on the detection of insecticidal activity in culture supernatants grown overnight at 30C in PYG medium followed by the detection of *tic*901 related sequences in *B. thuringiensis* strains when probed by the method of Southern using a TIC901 specific sequence. The total DNA from several hundreds of *B. thuringiensis* strains was analyzed. RFLP profiles were generated and compared Thermal amplification from these DNA's using primers specific for *tic*901, *tic*1201, *tic*407, and *tic*417 determined that about twenty to twenty five percent of Bt strains contain such related sequences. It was believed that some TIC901 related sequences may not be detected using the RFLP and thermal amplification detection methods. Based on the extracellular protein profiles of Bt strains allowed to continue fermentation well into late log and through stationary phase and sporulation, a more direct approach using immunological methods would be useful for detecting Bt strains that produced TIC901 related proteins.

A Bt culture collection was arrayed into a 96-well format and stored as glycerol stocks at –80C, and these were used as inoculum for micro-scale fermentations by transferring samples to deep well blocks (DWB's, QUIAGEN INC). Each sample was fermented in 1 milliliter of terrific broth (TB) per well. DWB's were sealed with AIRPORE tape sheets (QUIAGEN) and incubated at 30C for from between 20 and 45 hours with shaking in a HiGro incubarot (GENOMIC SOLUTIONS/GENOMIC INSTRUMENTATION SERVICES) at 300 RPM. At the end of the fermentation time, twenty five units per milliliter of benzonase (SIGMA ALDRICH) were added to the culture broth to reduce viscosity and incubated at 30C for an additional one hour. Solids, including cells and spores, were removed by centrifugation at 1900 x g in a swinging bucket rotor ar 4C for thirty minutes and clarified supernatants were transferred in 200 microliter aliquots to shallow 96 well plates for storage at –80C and further processing by ELISA analysis.

An ELISA method was developed for identification of supernatant broths containing TIC901 related proteins. 96 well NUNC Immuno MaxSorb plates were coated overnight at 4C with 100 microliters per well of a 1:1000 diluted reverse-affinity purified rabbit polyc lonal anti-TIC901 IgG antibody in coating buffer containing 15 millimolar sodium carbonate, 35 millimolar sodium bicarbonate, 150 millimolar NaCl, pH 9.6. Wells were washed three times with PBST (50 mM K/Na phosphate, 150 mM NaCl, 0.05% Tween 20, pH 7 4) and excess binding sites were blocked with 250 microliters of a 1% BSA solution in PBST for one hour at room temperature. Blocking buffer was discarded and each well was loaded with 200 microliters of a Bt culture supernatant diluted 1:75

(volume to volume) in PBST containing 0.2% BSA, followed by incubation at 4C overnight. Each well was washed three times with PBST followed by the addition of 200 microliters per well of a 1:2000 diluted biotinylated polyclonal anti-TIC901 IgG in PBST containing 0.2% BSA. The biotinylated antibody solution was allowed to incubated for two hours at room temperature and washed three times as described above. The biotinylated antibody-TIC901 related protein complexes were detected by incubating 200 microliters per well of a 1:3000 diluted HRP conjugated streptavidin for two hours at room temperature, followed by three washes in PBST and the addition o f 100 microliters per well of a TMP peroxidase substrate. Reactions were terminated by the addition of 100 microliters per well of a 3M phosphate solution. Absorbance of the individual wells was measured at a wavelength of 450 nanometers, and Bt strain supernatants determined to contain TIC901 related protein were determined by the average absorbance of negative controls plus three standard deviations.

Several strains were identified using this ELISA method that exhibited a positive ELISA result when compared to supernatants obtained from negative control cultures, but which also failed to result in the production of a thermal amplification product when using TIC901, TIC1201, TIC407, or TIC417 specific amplification primer pairs. Genomic DNA was obtained from one such ELISA positive/PCR negative strain previously designated as *Bacillus thuringiensis* strain EG4653 and a genomic plasmid expression library was constructed in the acrystalliferous strain EG10650. The EG10650 plasmid library derived from DNA of strain EG4653 was blotted to membranes and probed with the rabbit polyclonal anti-TIC901 IgG antibody described above. Positive clones were further purified and amplified and the DNA insert present in one such plasmid was sequenced. A complete TIC901 homologous protein was deduced from a single open reading frame. The homologous protein was designated as TIC431 and exhibits the amino acid sequence as set forth in SEQ ID NO:33. SEQ ID NO:32 consists of the open reading frame from which the TIC431 amino acid sequence was deduced. An alignment of the TIC431 amino acid sequence with the amino acid sequences of TIC901 and homologous proteins TIC1201, 407, and 417 is set forth in FIGURE 1. A comparison of the deduced mature TIC431m to TIC901m, TIC1201m, TIC407m, and TIC417m shows that 431 exhibits about 75% identity to 407, about 79% identity to 901, about 80% identity to 1201, and about 95% identity to 417. Therefore, it is anticipated that other TIC431 homologous proteins would exhibit from about 75 to about 95% identity to TIC431 amino acid sequences The deduced precursor protein consists of a predicted about 30 amino terminal amino acid sequence consistent with a signal peptide that exhibits a consensus type II signal peptidase cleavage sequence. In addition, the same or a substantially similar amino acid sequence exists between the signal peptide sequence and the deduced amino terminal amino acid of the mature peptide beginning at amino acid position 44 as set forth in SEQ ID NO:33. Thermal amplification primers as set forth in SEQ ID NO:23-26 and SEQ ID NO:27-29, consistent with the consensus amino acid sequences encoded by the nucleotide sequences as set forth in SEQ ID NO.32 from about 223 to about 249, or from about 439 to about 456, or from about 823 to about 846 are expected to produce amplicons that are diagnostic for the presence of SEQ ID NO:32, as well as other *tic*901 homologous insecticidal protein coding sequences.

The ELISA method coupled with the thermal amplification method should provide the skilled artisan with the means for identifying any nucleotide sequence encoding an insecticidal protein derived

from a *Bacillus* species that exhibits from about 67% to about 99% or greater amino acid sequence identity to a TIC901 or a TIC901 homologous amino acid sequence as set forth herein.

One skilled in the art will recognize that other embodiments are possible using the probes and primers of the present invention to identify sequences encoding proteins of the present invention and to identify proteins of the present invention genes and proteins related thereto.

In summary, the above specification describes preferred embodiments of the present invention. It will be understood by those skilled in the art that, without departing from the scope and spirit of the present invention and without undue experimentation, the present invention can be performed within a wide range of equivalent parameters. While the present invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications. The present invention is intended to cover any uses, variations, or adaptations of the invention following the principles of the invention in general. Various permutations and combination of the elements provided in all the claims that follow are possible and fall within the scope of this invention. All publications and patents mentioned in this specification are herein incorporated by reference as if each individual publication or patent was specially and individually stated to be incorporated by reference.

REFERENCES

Barry et al., 1992, Inhibitors of amino acid biosynthesis: Strategies for imparting glyphosate tolerance to crop plants. In: Biosynthesis and Molecular Regulation of Amino Acids in Plants. pp. 139-145. Singh, Flores, and Shannon Eds., American Society of Plant Physiologists, Rockville, Md.

Baum, J. A.; Coyle, D. M.; Gilbert, M. P.; Jany, C. S.; Gawron-Burke, C., 1990 Novel cloning vectors for *Bacillus thuringiensis*; 1983, Applied and Environmental Microbiology 56 (11): 3420-3428.

Bevan , Nature 304:184-187.

Callis *et al.*, 1987, Genes & Develop. 1·1183-1200.

Crickmore et al., 1998, *Microbiol. Molecular Biol. Rev.* 62, pp. 807-813.

DeBarjac and Frachon, 1990, *Entomophaga* 35, pp. 233-240.

Diehn *et al.*, 1996, *In: Genetic Engineering*, Ed. J K. Setlow, Plenum Press, New York, NY, 18:83-99.

Diehn *et al.*, 1998, Plant Physiology, 117:1433-1443

Donovan et al., 1988, *Mol. Gen. Genet.* 214:365-372.

Ellis *et al.*, 1987, EMBO Journal 6:11-16.

Estruch *et al.*, 1996, *Proc. Natl Acad. Sci.* U.S.A , 93:5389-5394.

Fraley et al., 1983, *Proc. Natl. Acad. Sci.* U S.A., 80:4803.

Goding, Monoclonal Antibodies: Principles and Practice, 2nd eds, Academic Press, NY, 1986.

Harlow and Lane, 1988, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Hartley, J., 1988, Mol. Biol. 202 913.

Herrera-Estrella , 1983, Nature 303:209-213.

Hess, 1987, *Intern Rev. Cytol.*, 107: 367.

Höfte et al., 1989, *Microbiol. Rev.* 53: 242-255,

Horsch *et al.*, 1985, *Science*, 227:1229-1231.

Humason, Gretchen L., 1967, Animal Tissue Techniques, W. H. Freeman and Company.

Joshi, 1987, Nucl Acids Res. 15:9627-9640.

Klee, 1985, Bio/Technol. 3:637-642.

Kostichka et al , 1996, J. Bacteriol 178:2141-2144.

Kyte and Doolittle, 1982, J. Mol. Biol., 157:105-132.

Laemmli , 1973, Mol. Biol. 80:575-599

Lu et al., 2000, J. Plant Phys., 156:277-283.

Ma *et al.*, 1988, Nature 334:631-633.

Maliga, 1993, Trends in Biotechnology 11:101-106.

McElroy *et al.*, 1990, Plant Cell 2:163-171.

Murray et al., 1989, Nucl. Acids. Res , 17:477-498.

Neuhaus *et al.*, 1987, Theor. Appl. Genet., 75:30.

Nielsen et al., 1997, Protein Engineering 10:1-6.

Norton et al., 1985, Plasmid 13:211-214

Odell *et al.*, 1985, Nature 313:810-812.

Pena *et al.*, 1987, Nature, 325:274.

Perlak et al., 1991, Proc. Natl. Acad. Sci. USA, 88:3324-3328,.

Sambrook et al., 1989, Molecular Cloning - A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Schnepf et al., 1998, Microbiol. Molec. Biol. Rev. 62 :775-806.

Turner et al., 1993, Endophytes: an alternative genome for crop improvement; International Crop Science Congress, Ames, Iowa, USA, 14-22 July 1992., pp.555-560.

Vasil *et al* , 1989, Plant Phys. 91:1575-1579.

Yanisch-Perron et al, 1985, Gene 33:103-119.

Zhou *et al.*, 1983, Mol. Cell Biol., 10:4529-4537.

U.S. Patent No. 6,489,542

U.S. Patent No. 6,448,476

U.S. Patent No. 6,279,369

U.S. Patent No. 6,242,669

U.S. Patent No. 6,204,435

U.S. Patent No. 6,107,549

U.S. Patent No. 6,063,597

U.S. Patent No. 5,888,801

U.S. Patent No. 5,872,212

U.S. Patent No. 5,877,012

U.S. Patent No. 5,866,326

U.S. Patent No. 5,849,870

U.S. Patent No. 5,837,848

U.S. Patent No. 5,759,538

U.S. Patent No. 5,689,052

U.S. Patent No. 5,569,834

U.S. Patent No. 5,500,365

U.S. Patent No. 5,424,412

U.S. Patent No. 5,349,124

U.S. Patent No. 5,229,112

U.S. Patent No. 5,135,867

U.S. Patent No. 5,080,897

U.S. Patent No. 4,935,353

U.S. Patent No. 4,695,462

U.S. Patent No. 4,695,455

U.S. Patent No. 4,554,101

WO96/10083.

WO94/21795.

WO95/24492.

EPO 0 120 516.

What is Claimed is:

1. An isolated polynucleotide which encodes a *Bacillus thuringiensis* insecticidal toxin or insecticidal fragment thereof, wherein said polynucleotide hybridizes under conditions selected from the group consisting of stringent hybridization conditions and specific hybridization conditions with one or more of the nucleotide sequences selected from the group consisting of SEQ ID NO:2, SEQ ID NO:3 (tic901), SEQ ID NO:5(tic1201), SEQ ID NO:7 (tic407), SEQ ID NO.9 (tic417), SEQ ID NO:11, SEQ ID NO.12, SEQ ID NO:15, SEQ ID NO·16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO.22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO.26, SEQ ID NO·27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, and SEQ ID NO:32 or with the complement thereof.

2. The isolated polynucleotide of claim 1 wherein said toxin is active against a coleopteran insect pest.

3. The isolated polynucleotide according to claim 2 wherein said coleopteran insect pest is selected from the group consisting of a corn rootworm and a Colorado potato beetle.

4. The polynucleotide according to claim 3 wherein said corn rootworm is selected from the group consisting of a western corn rootworm, a southern corn rootworm, or a northern corn rootworm.

5. The polynucleotide according to claim 1 wherein said nucleotide sequence is SEQ ID NO·3, SEQ ID NO·5, SEQ ID NO.7, SEQ ID NO 9, and SEQ ID NO·32.

6. A polynucleotide comprising a nucleotide sequence which encodes an approximately 34 to about 39 kDa toxin active against a coleopteran pest, wherein said nucleotide sequence has been optimized for expression in plants, and wherein said toxin comprises the amino acid sequence selected from the group consisting of SEQ ID NO·2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO.8, SEQ ID NO:10, SEQ ID NO:31, and SEQ ID NO·33

7. A host cell transformed to contain a polynucleotide encoding an insecticidal protein or insecticidal fragment thereof wherein said polynucleotide comprises a nucleotide sequence as set forth in a sequence selected from the group consisting of SEQ ID NO.3, SEQ ID NO:5, SEQ ID NO.7, SEQ ID NO:9, SEQ ID NO:13, SEQ ID NO 30, and SEQ ID NO:32.

8. The host cell of claim 7 wherein said host cell is a plant cell.

9. A recombinant bacterium comprising an isolated polynucleotide that encodes an insecticidal protein wherein said protein comprises an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO.4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:31, and SEQ ID NO:33.

10. A method for controlling a coleopteran insect pest comprising contacting said pest with a pesticidal amount of an approximately 34 to about 39 kDa *Bacillus thuringiensis* toxin protein or insecticidal fragment thereof, wherein said toxin protein is selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:31, and SEQ ID NO:33.

11. A method for controlling a coleopteran insect pest comprising contacting said pest with a pesticidal amount of an approximately 34 to about 39 kDa *Bacillus thuringiensis* toxin protein or insecticidal fragment thereof, wherein said toxin protein is encoded by a nucleotide sequence that is or that hybridizes under stringent conditions to a nucleotide sequence comprising at least 18 consecutive nucleotides selected from the group consisting of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO.7, SEQ ID NO 9, SEQ ID NO 30, and SEQ ID NO:32.

12. An isolated insecticidal protein derived from *Bacillus thuringiensis* that is from about 34 kDa to about 39 kDa and is encoded by a nucleotide sequence that that is or that hybridizes under stringent conditions to a nucleotide sequence comprising at least 18 consecutive nucleotides selected from the group consisting of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO 7, SEQ ID NO:9, SEQ ID NO:30, and SEQ ID NO:32.

13. An isolated insecticidal protein derived from *Bacillus thuringiensis* that is from about 34 kDa to about 39 kDa and is encoded by a nucleotide sequence that is or hybridizes under stringent conditions to at least 18 consecutive nucleotides selected from the group consisting of SEQ ID NO 1, SEQ ID NO.3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, and SEQ ID NO:32, and wherein said *Bacillus thuringiensis* is selected from the group consisting of EG2158, EG6489, EG6561, EG3618, EG6555, EG6618, 86833, and EG4653.

14. A method for detecting a first nucleotide sequence encoding an insecticidal protein wherein said first nucleotide sequence hybridizes to a second nucleotide sequence that is selected from the group consisting of SEQ ID NO.2, SEQ ID NO.3 (tic901), SEQ ID NO:5(tic1201), SEQ ID NO:7 (tic407), SEQ ID NO:9 (tic417), SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO.18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO 30, and SEQ ID NO 32 (tic431) or with the complement thereof under stringent hybridization conditions.

15. A nucleotide sequence that encodes an insecticidal protein or insecticidal fragment or homolog thereof, wherein said insecticidal protein or insecticidal fragment or homolog thereof is selected from the group of sequences consisting of SEQ ID NO:1, SEQ ID NO:4, SEQ ID NO 6, SEQ ID NO.8, SEQ ID NO 10, SEQ ID NO:31, and SEQ ID NO.33.

16. A method for detecting a first nucleotide sequence encoding an insecticidal protein, comprising the steps of (a) producing an amplicon from a template comprising said first nucleotide sequence in a sample using amplification primers comprising the sequences as set forth in SEQ ID NO:13 and SEQ ID NO:14; (b) isolating and purifying the amplicon; (c) isolating and purifying the complete first nucleotide sequence encoding said insecticidal protein based on the isolated and purified amplicon sequence; and (d) expressing said insecticidal protein in a host cell from said first nucleotide sequence, wherein said first nucleotide sequence hybridizes to a second nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO 3, SEQ ID NO:5, SEQ ID NO 7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:30, and SEQ ID NO:32 or complements thereof under stringent hybridization conditions.

17. The method according to claim 16 wherein said host cell is selected from the group of cells comprising a plant cell, a bacterial cell, a fungal cell, an insect cell, and a mammalian cell

18. The method according to claim 17 wherein said host cell is a plant cell selected from the group of plant cells comprising a monocot plant cell and a dicot plant cell.

19. The method according to claim 18 wherein said monocot plant cell is selected from the group of plant cells comprising a corn plant cell, a wheat plant cell, a rice plant cell, an oat plant cell, an onion plant cell, and a grass plant cell.

20. The method according to claim 18 wherein said dicot plant cell is selected fro the group of plant cells comprising a cotton plant cell, a canola plant cell, a soybean plant cell, a fruit tree plant cell, an okra plant cell, a pepper plant cell, an ornamental plant cell, a sunflower plant cell, a cucurbit plant cell, and a melon plant cell.

21. An isolated nucleic acid molecule comprising a polynucleotide sequence encoding a toxin protein wherein said toxin protein comprises a sequence that exhibits at least about 70% sequence identity to a nucleotide sequence selected from the group of amino acid sequences consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:31, and SEQ ID NO 33, or a coleopteran-active variant or portion thereof.

22 An isolated nucleic acid molecule comprising a polynucleotide sequence which encodes a protein toxic to coleopteran insect pests, wherein said polynucleotide sequence hybridizes to a nucleic acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:3 (tic901), SEQ ID NO.5(tic1201), SEQ ID NO:7 (tic407), SEQ ID NO·9 (tic417), SEQ ID NO.11, SEQ ID NO.12, SEQ ID NO:15, SEQ ID NO·16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO·19, SEQ ID NO:20, SEQ ID NO 21, SEQ ID NO:22, SEQ ID NO 23, SEQ ID NO·24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO 27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO.30, and SEQ ID NO.32 (tic431), or with the complement thereof under stringent hybridization conditions.

23. The isolated nucleic acid of claim 1 or 5 wherein said polynucleotide sequence is isolated from a bacterium.

24. The isolated nucleic acid of claim 6 wherein said polynucleotide sequence is isolated from *Bacillus thuringiensis.*

25. A substantially purified toxin protein wherein said toxin protein comprises a sequence that has at least about 70% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO.2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO 8, SEQ ID NO.10, SEQ ID NO:31, and SEQ ID NO:33.

26. The substantially purified toxin protein of claim 25, wherein said toxin protein comprises SEQ ID NO· 4.

27. A recombinant DNA construct, comprising a polynucleotide sequence encoding an insecticidal protein, wherein said polynucleotide sequence is at least about 70% identical to SEQ ID NO 13, said insecticidal protein being selected from the group consisting of all or an

insecticidal fragment of a protein as set forth in SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:6, SEQ IDNO:10, SEQ ID NO:31, and SEQ ID NO.33.

28. The recombinant DNA construct of claim 27, wherein said polynucleotide sequence is as set forth in SEQ ID NO.3.

29. The recombinant DNA construct of claim 28 further comprising a promoter or partial promoter region operably linked to said polynucleotide sequence.

30 The recombinant DNA construct of claim 29 wherein said promoter is a constitutive promoter, an inducible promoter, or a tissue-specific promoter.

31. The recombinant DNA construct of claim 30, wherein the promoter is a cauliflower mosaic virus (CaMV) 35S promoter.

32. The recombinant DNA construct of claim 31 further comprising an intron.

33. The recombinant DNA construct of claim 32 wherein said intron is an hsp70 intron.

34. The recombinant DNA construct of claim 33 further comprising an operably linked 3' non-translated region.

35 A recombinant host cell transformed with a polynucleotide sequence of claim 1 or claim 27 encoding an insecticidal protein, wherein said recombinant host cell expresses said protein that is at least about 70% similar in amino acid sequence to an insecticidal protein selected from the group consisting of SEQ ID NO.4, SEQ ID NO:6, SEQ ID NO·8, SEQ ID NO·10, SEQ ID NO 31, and SEQ ID NO·33.

36. The recombinant host cell of claim 35, wherein said recombinant host cell is a plant cell.

37. A method for expressing an insecticidal protein in a plant cell, comprising the steps of

(a) inserting into the genome of a plant cell a nucleic acid sequence comprising:

(i) a promoter that functions in said plant cell operably linked to,

(ii) a structural DNA encoding a protein selected from the group consisting of SEQ ID NO.4, SEQ ID NO:6, SEQ ID NO.8, SEQ ID NO.10, SEQ ID NO·31, SEQ ID NO.33, and insecticidal proteins that are at least about 70% similar to said group of proteins, and

(iii) a 3' non-translated DNA sequence that functions in the plant cell to cause transcription termination and polyadenylation;

(b) obtaining said plant cell transformed to contain the nucleic acid sequence of step (a), and

(c) regenerating from the transformed plant cell a genetically transformed plant that expresses an insecticidal amount of said insecticidal protein.

38. A seed produced from the transformed plant of claim 37.

39. A plant grown from the seed of claim 38.

40. A biological sample derived from a plant, tissue, or seed, wherein said sample comprises a nucleotide sequence encoding an insecticidal protein selected from the group consisting of SEQ ID NO 4, SEQ ID NO:6, SEQ ID NO.8, SEQ ID NO:10, SEQ ID NO:31, and SEQ ID NO.33.

41. The biological sample of claim 40 comprising plant, tissue, or seed of a transgenic plant encoding said protein.

42. The biological sample of claim 41 wherein said sample is selected from the group consisting of an extract obtainable from the transgenic plant, and wherein said extract comprises a nucleotide sequence encoding an insecticidal fragment of a protein selected from the group consisting of SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:31, and SEQ ID NO:33.

43. The biological sample of claim 42 wherein said sample is selected from the group consisting of corn flour, corn meal, corn syrup, corn oil, corn starch, and cereals manufactured in whole or in part to contain corn by-products

44 An extract derived from a transgenic corn plant, tissue, or seed comprising a nucleotide sequence encoding an insecticidal fragment of a protein selected from the group consisting of SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:31, and SEQ ID NO.33.

45 The extract of claim 44 wherein said sequence is detectable in said extract using a nucleic acid amplification or nucleic acid hybridization method.

46 The extract of claim 45 comprising plant, tissue, or seed of a transgenic corn plant expressing said protein.

47 The extract of claim 46 wherein said sample is selected from the group consisting of corn flour, corn meal, corn syrup, corn oil, corn starch, and cereals manufactured in whole or in part to contain corn by-products.

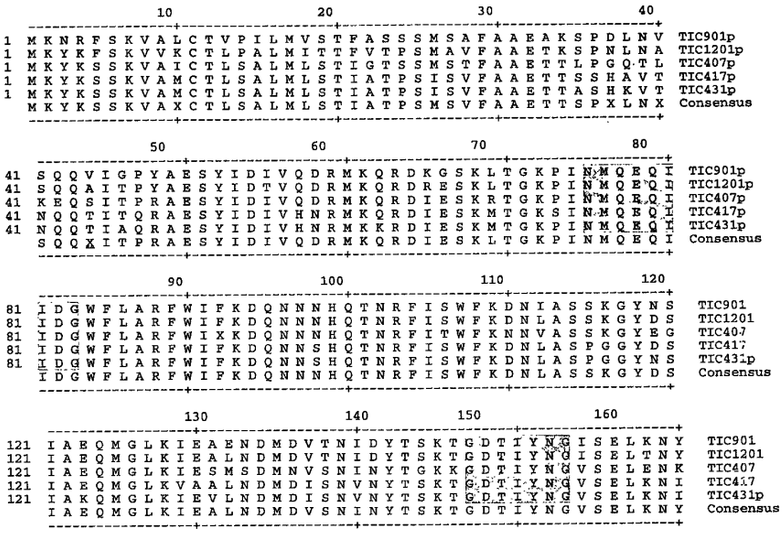

10 20 30 40
1 MKNRFSKVALCTVPILMVSTFASSSMSAFAAEAKSPDLNV TIC901p
1 MKYKFSKVVKCTLPALMITTFVTPSMAVFAAETKSPNLNA TIC1201p
1 MKYKSSKVAICTLSALMLSTIGTSSMSTFAAETTLPGQTL TIC407p
1 MKYKSSKVAMCTLSALMLSTIATPSISVFAAETTSSHAVT TIC417p
1 MKYKSSKVAMCTLSALMLSTIATPSISVFAAETTASHKVT TIC431p
MKYKSSKVAXCTLSALMLSTIATPSMSVFAAETTSPXLNX Consensus
50 60 70 80
41 SQQVIGPYAESYIDIVQDRMKQRDKGSKLTGKPINMQEQI TIC901p
41 SQQAITPYAESYIDTVQDRMKQRDRESKLTGKPINMQEQI TIC1201p
41 KEQSITPRAESYIDIVQDRMKQRDIESKRTGKPINMQEQI TIC407p
41 NQQTITQRAESYIDIVHNRMKQRDIESKMTGKSINMQEQI TIC417p
41 NQQTIAQRAESYIDIVHNRMKKRDIESKMTGKPINMQEQI TIC431p
SQQXITPRAESYIDIVQDRMKQRDIESKLTGKPINMQEQI Consensus
90 100 110 120
81 IDGWFLARFWIFKDQNNNNHQTNRFISWFKDNIASSKGYNS TIC901
81 IDGWFLARFWIFKDQNNNNHQTNRFISWFKDNLASSKGYDS TIC1201
81 IDGWFLARFWIXKDQNNNNHQTNRFITWFKNNVASSKGYEG TIC407
81 IDGWFLARFWIFKDQNNNSHQTNRFISWFKDNLASPGGYDS TIC417
81 IDGWFLARFWIFKDQNNNSHQTNRFISWFKDNLASPGGYNS TIC431p
IDGWFLARFWIFKDQNNNNHQTNRFISWFKDNLASSKGYDS Consensus
130 140 150 160
121 IAEQMGLKIEAENDMDVTNIDYTSKTGDTIYNGISELKNY TIC901
121 IAEQMGLKIEALNDMDVTNIDYTSKTGDTIYNGISELTNY TIC1201
121 IAEQMGLKIESMSDMNVSNINYTGKKGDTIYNGVSELENK TIC407
121 IAEQMGLKVAALNDMDISNVNYTSKTGDTIYNGVSELKNI TIC417
121 IAKQMGLKIEVLNDMDISNVNYTSKTGDTIYNGVSELKNI TIC431p
IAEQMGLKIEALNDMDVSNINYTSKTGDTIYNGVSELKNY Consensus

Figure 1A

```
                  170                 180                 190                 200
161 T G S T Q K M K T D S F Q R D Y T K S E S T S V T N G L Q L G F K V A A K G V V  TIC901
161 T G T T Q K M K T D S F Q R D Y T K S E S T S V T N G L Q L G F K V A A K G V V  TIC1201
161 M G T P Q K M K S D S F Q R D Y T K S Q S T S V T N G L Q L G V K V S A K G T V  TIC407
161 T G T T Q K M K T D S F Q R D Y T K S Q S T S I T D G L Q L G F K V S A K G I V  TIC417
161 T G T T Q K M K T D S F Q R D Y T K S Q S T S I T N G L Q L G F K V S A K G V I  TIC431p
    T G T T Q K M K T D S F Q R D Y T K S Q S T S V T N G L Q L G F K V S A K G V V  Consensus

                  210                 220                 230                 240
201 A L A G A D F E T S V T Y N L S T T T T E T N T I S D K F T V P S Q E V T L P P  TIC901
201 A L A G A D F E T S V T Y N L S S T T T E T N T I S D K F T V P S Q E V T L S P  TIC1201
201 V L G E A S L E T S V T Y N L S S T A T E T D T T S D K F T V P S Q E V T L P P  TIC407
201 A L A G A D F E A S V N Y N L S T T A T E T N T I S D K F T V P S Q E V T L A P  TIC417
201 A L A G A D F E A S V N Y N L S T T A T E T N I I S D K F T V P S Q E V T L A P  TIC431p
    A L A G A D F E T S V T Y N L S T T A T E T N T I S D K F T V P S Q E V T L P P  Consensus

                  250                 260         270         280
241 G H K A I V K H D L R K M V Y S G T H D L K G D L I V S F N D K E I V Q K F I Y  TIC901
241 G H K A V V K H D L R K M V Y F G T H D L K G D L K V G F N D K E I V Q K F I Y  TIC1201
241 G H K A V I K H D L R K M V Y S G T H D L K G D L K V A F N D K A I M Q K F I Y  TIC407
241 G H K A I V K H S L K K M V Y S G T H D L K G D L T I T F N D K D L V Q K F I Y  TIC417
241 G H K A I V K H S L K K M V Y S G T H D L K G D L T I T F N D K D L V Q K F I Y  TIC431p
    G H K A I V K H D L R K M V Y S G T H D L K G D L K V X F N D K E I V Q K F I Y  Consensus

                  290                 300                 310                 320
281 P N Y R E I N L S D I R E T M I E I D E W N H V N P V N F Y E L V G V K N H I K  TIC901
281 P N Y R S I D L S D I R K T M I E I D K W N H V N T I D F Y Q L V G V K N H I K  TIC1201
281 P N Y R S I N L S D I R K T M K E I D E W N H V K P I D F Y Q L V G I K N H I K  TIC407
281 P N Y K A I D L S N I R K A M T E I D E W N H V K P T D F Y Q L V G N K N Y I K  TIC417
281 P N Y K A I D L S N I R K A L T E I D E W N H V K P T D F Y Q L V G N K N Y I K  TIC431p
    P N Y R S I D L S D I R K T M I E I D E W N H V N P I D F Y Q L V G V K N H I K  Consensus
```

Figure 1B

```
                          330                 340                 350
      ------------------+-------------------+-------------------+------------------+
321   N G E T L Y I D T P A K F M F N G A N P Y Y R A T F T E Y D G N N N P V Q T - K   TIC901
321   N G D T L Y I D T P A E F T F N G A N P Y Y R A T F T E Y D E N G N P V Q T - K   TIC1201
321   N G D T L Y I E T P A K F I F N G A N V Y Y R A T F T E Y D K D G K P V Q F N K   TIC407
321   N G D T L Y I E T P A K F T L N G G N P Y Y T A T F T E Y D E N G N Q V K T - K   TIC417
321   N G D T L Y I E T P A K F T L N G G N P Y Y T A T F T E Y D E S G N Q V K T - K   TIC431p
      N G D T L Y I D T P A K F T F N G A N P Y Y R A T F T E Y D E N G N P V Q T - K   Consensus
      ------------------+-------------------+-------------------+------------------+

      ---------------
360   V L S E N F K L 367                                          TIC901 p
360   I L S G N     364                                            TIC1201p
361   F L S E N Y K L 368                                          TIC407p
360   R L - N N - K   364                                          TIC417p
360   H L S V - - K   364                                          TIC431p
      X L S E N - K L                                              Consensus
      ---------------
```

Figure 1C

SEQUENCE LISTING

<110> Monsanto Technology LLC

Baum, James A.
Donovan, Judith C.

Donovan, William P.

Engleman, James T.

Krasomil-Osterfeld, Karina

Pitkin, John W.

Roberts, James K.

<120> Insecticidal Proteins Secreted From Bacillus Species and Uses Therefor

<130> 38-21(52806)PCT

<150> US 60/485,483

<151> 2003-07-07

<160> 33

<170> PatentIn version 3.1

<210> 1

<211> 15

<212> PRT

<213> Bacillus thuringiensis

<400> 1

Val Ile Gly Pro Tyr Ala Glu Ser Tyr Ile Asp Arg Val Gln Asp
1 5 10 15

<210> 2

<211> 45

<212> DNA

<213> Artificial Sequence

<220>

<223> tic gene probe encoding SEQ ID NO 1

<400> 2
gtaattggac catatgcaga atcatatatt gatacgagta caaga 45

<210> 3

<211> 1253

<212> DNA

<213> Bacillus thuringiensis

<220>

<221> CDS

<222> (153)..(1253)

<223>

<400> 3
aattatgatt ttaatattct tatgttattc ctataatata caataaaagc ataattatcc 60

ttcatattat gtttataaat ttaataaaat acataaaaat agagtgttat aatatttttg 120

aaagcgttat caagagtgat ggagggataa tt atg aaa aat aga ttt tca aaa 173
Met Lys Asn Arg Phe Ser Lys
1 5

gtg gca tta tgc acc gta ccg att tta atg gtt tct aca ttc gcc agt 221
Val Ala Leu Cys Thr Val Pro Ile Leu Met Val Ser Thr Phe Ala Ser
10 15 20

tca agc atg tca gct ttt gct gca gaa gcc aaa tca cca gat tta aat 269
Ser Ser Met Ser Ala Phe Ala Ala Glu Ala Lys Ser Pro Asp Leu Asn
25 30 35

gta tct caa caa gta ata ggt ccc tat gcc gaa tct tat att gat att 317
Val Ser Gln Gln Val Ile Gly Pro Tyr Ala Glu Ser Tyr Ile Asp Ile
40 45 50 55

gtg cag gat aga atg aaa caa agg gat aag gga tca aaa tta act ggt 365
Val Gln Asp Arg Met Lys Gln Arg Asp Lys Gly Ser Lys Leu Thr Gly
60 65 70

aaa cca ata aat atg caa gaa caa ata ata gat ggg tgg ttt cta gct 413
Lys Pro Ile Asn Met Gln Glu Gln Ile Ile Asp Gly Trp Phe Leu Ala
75 80 85

```
aga ttt tgg ata ttt aag gat caa aac aat aac cat cag aca aat aga      461
Arg Phe Trp Ile Phe Lys Asp Gln Asn Asn Asn His Gln Thr Asn Arg
        90                  95                  100

ttt ata tcc tgg ttt aaa gat aat att gct agt tca aaa ggg tat aat      509
Phe Ile Ser Trp Phe Lys Asp Asn Ile Ala Ser Ser Lys Gly Tyr Asn
    105                 110                 115

agt att gcg gag caa atg ggt tta aaa ata gaa gca gaa aac gat atg      557
Ser Ile Ala Glu Gln Met Gly Leu Lys Ile Glu Ala Glu Asn Asp Met
120                 125                 130                 135

gat gta aca aat ata gat tat aca tct aag aca ggc gat acc att tat      605
Asp Val Thr Asn Ile Asp Tyr Thr Ser Lys Thr Gly Asp Thr Ile Tyr
                140                 145                 150

aat ggt att tca gaa ttg aaa aat tat aca gga tca act caa aag atg      653
Asn Gly Ile Ser Glu Leu Lys Asn Tyr Thr Gly Ser Thr Gln Lys Met
            155                 160                 165

aaa aca gat agt ttt caa aga gat tat aca aaa tca gaa tct act tca      701
Lys Thr Asp Ser Phe Gln Arg Asp Tyr Thr Lys Ser Glu Ser Thr Ser
        170                 175                 180

gta act aat gga tta caa tta gga ttt aaa gtt gct gct aaa gga gta      749
Val Thr Asn Gly Leu Gln Leu Gly Phe Lys Val Ala Ala Lys Gly Val
    185                 190                 195

gtt gct ttg gct ggg gca gac ttt gaa acc agt gtt act tat aat cta      797
Val Ala Leu Ala Gly Ala Asp Phe Glu Thr Ser Val Thr Tyr Asn Leu
200                 205                 210                 215

tca act act aca act gaa aca aat aca ata tca gac aag ttt act gtc      845
Ser Thr Thr Thr Thr Glu Thr Asn Thr Ile Ser Asp Lys Phe Thr Val
                220                 225                 230

cca tct caa gaa gtt aca ttg cct cca gga cat aaa gcg ata gtg aaa      893
Pro Ser Gln Glu Val Thr Leu Pro Pro Gly His Lys Ala Ile Val Lys
            235                 240                 245

cat gat tta aga aaa atg gtt tat tct ggt act cat gat cta aag ggt      941
His Asp Leu Arg Lys Met Val Tyr Ser Gly Thr His Asp Leu Lys Gly
        250                 255                 260

gat tta att gtg agt ttt aat gat aaa gag att gta caa aaa ttt att      989
Asp Leu Ile Val Ser Phe Asn Asp Lys Glu Ile Val Gln Lys Phe Ile
    265                 270                 275

tat cca aat tat aga gaa att aat tta tct gat atc cgt gaa act atg     1037
Tyr Pro Asn Tyr Arg Glu Ile Asn Leu Ser Asp Ile Arg Glu Thr Met
280                 285                 290                 295

att gaa att gat gaa tgg aat cat gta aac cct gtg aat ttt tat gaa     1085
Ile Glu Ile Asp Glu Trp Asn His Val Asn Pro Val Asn Phe Tyr Glu
                300                 305                 310

tta gtt ggg gtc aaa aat cat ata aaa aat ggt gaa act ttg tat ata     1133
Leu Val Gly Val Lys Asn His Ile Lys Asn Gly Glu Thr Leu Tyr Ile
            315                 320                 325

gat act cca gct aaa ttt atg ttt aat ggt gct aat cca tat tat aga     1181
Asp Thr Pro Ala Lys Phe Met Phe Asn Gly Ala Asn Pro Tyr Tyr Arg
        330                 335                 340
```

gca aca ttt aca gaa tac gac ggg aat aat aat cct gtt caa aca aag 1229
Ala Thr Phe Thr Glu Tyr Asp Gly Asn Asn Asn Pro Val Gln Thr Lys
345 350 355

gta tta agt gaa aac ttt aaa ttg 1253
Val Leu Ser Glu Asn Phe Lys Leu
360 365

<210> 4

<211> 367

<212> PRT

<213> Bacillus thuringiensis

<400> 4

Met Lys Asn Arg Phe Ser Lys Val Ala Leu Cys Thr Val Pro Ile Leu
1 5 10 15

Met Val Ser Thr Phe Ala Ser Ser Ser Met Ser Ala Phe Ala Ala Glu
20 25 30

Ala Lys Ser Pro Asp Leu Asn Val Ser Gln Gln Val Ile Gly Pro Tyr
35 40 45

Ala Glu Ser Tyr Ile Asp Ile Val Gln Asp Arg Met Lys Gln Arg Asp
50 55 60

Lys Gly Ser Lys Leu Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile
65 70 75 80

Ile Asp Gly Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn
85 90 95

Asn Asn His Gln Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Ile
100 105 110

Ala Ser Ser Lys Gly Tyr Asn Ser Ile Ala Glu Gln Met Gly Leu Lys
115 120 125

Ile Glu Ala Glu Asn Asp Met Asp Val Thr Asn Ile Asp Tyr Thr Ser
130 135 140

Lys Thr Gly Asp Thr Ile Tyr Asn Gly Ile Ser Glu Leu Lys Asn Tyr
145 150 155 160

Thr Gly Ser Thr Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr

165 170 175

Thr Lys Ser Glu Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Phe
180 185 190

Lys Val Ala Ala Lys Gly Val Val Ala Leu Ala Gly Ala Asp Phe Glu
195 200 205

Thr Ser Val Thr Tyr Asn Leu Ser Thr Thr Thr Thr Glu Thr Asn Thr
210 215 220

Ile Ser Asp Lys Phe Thr Val Pro Ser Gln Glu Val Thr Leu Pro Pro
225 230 235 240

Gly His Lys Ala Ile Val Lys His Asp Leu Arg Lys Met Val Tyr Ser
245 250 255

Gly Thr His Asp Leu Lys Gly Asp Leu Ile Val Ser Phe Asn Asp Lys
260 265 270

Glu Ile Val Gln Lys Phe Ile Tyr Pro Asn Tyr Arg Glu Ile Asn Leu
275 280 285

Ser Asp Ile Arg Glu Thr Met Ile Glu Ile Asp Glu Trp Asn His Val
290 295 300

Asn Pro Val Asn Phe Tyr Glu Leu Val Gly Val Lys Asn His Ile Lys
305 310 315 320

Asn Gly Glu Thr Leu Tyr Ile Asp Thr Pro Ala Lys Phe Met Phe Asn
325 330 335

Gly Ala Asn Pro Tyr Tyr Arg Ala Thr Phe Thr Glu Tyr Asp Gly Asn
340 345 350

Asn Asn Pro Val Gln Thr Lys Val Leu Ser Glu Asn Phe Lys Leu
355 360 365

<210> 5

<211> 1621

<212> DNA

<213> Bacillus thuringiensis

<220>

38-21(52806) Sequence Listing_PCT_2.5725.txt

<221> CDS

<222> (530)..(1621)

<223>

<400> 5

```
actagttttt tcataatggc ataagcggga tgatgatctt gtttttttacg atgttcaata     60

tccaatgtgt gcctttctat atcaatcgca cgatataaat aacactattt ccctttgaat    120

tttatatagg tttcatctaa ttttcaagac atgtggttgt tttgcgtttt cttcttccaa    180

atttgataaa tcaagctccc atattcatga atccagcgca taatgattgt gggatgaact    240

gaaacatcac gatagcttaa agcaaaacga caatagtagc ggacggctac cataataata    300

tcttgtttga actgtttccc tttaaaatat cacatttgtg attctcctcg atgctttttt    360

tagagtgtag cttcatctag aacactttgc aatagaacca ttcctttgat atacaattaa    420

accacattta tccttcatgg aatgtttata tattaaagaa tataaaaaaa catacgatgt    480

tataattaat ttgaaagcgt taacaaaaat gaatgatgga gggataatt atg aaa tac    538
                                                Met Lys Tyr
                                                1

aag ttt tca aaa gtc gtt aag tgt act tta cca gct tta atg att act      586
Lys Phe Ser Lys Val Val Lys Cys Thr Leu Pro Ala Leu Met Ile Thr
    5                  10                  15

aca ttc gtt act cca agt atg gca gtt ttt gcc gca gaa acc aag tcg      634
Thr Phe Val Thr Pro Ser Met Ala Val Phe Ala Ala Glu Thr Lys Ser
20                  25                  30                  35

cca aat cta aat gca tct caa caa gca ata act cca tat gct gaa tct      682
Pro Asn Leu Asn Ala Ser Gln Gln Ala Ile Thr Pro Tyr Ala Glu Ser
                40                  45                  50

tat att gat act gtt caa gat aga atg aaa caa aga gat agg gaa tca      730
Tyr Ile Asp Thr Val Gln Asp Arg Met Lys Gln Arg Asp Arg Glu Ser
            55                  60                  65

aaa cta act ggt aaa cca ata aat atg caa gaa caa ata ata gat gga      778
Lys Leu Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile Ile Asp Gly
        70                  75                  80

tgg ttt tta gct aga ttc tgg ata ttt aaa gat caa aat aac aat cat      826
Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn Asn Asn His
    85                  90                  95

caa aca aat aga ttt ata tcc tgg ttt aaa gat aat ctt gct agt tcg      874
Gln Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Leu Ala Ser Ser
100                 105                 110                 115

aag ggg tat gac agt ata gca gaa caa atg ggc tta aaa ata gaa gca      922
Lys Gly Tyr Asp Ser Ile Ala Glu Gln Met Gly Leu Lys Ile Glu Ala
                120                 125                 130

tta aat gat atg gat gta aca aat att gat tat aca tct aaa aca ggt      970
Leu Asn Asp Met Asp Val Thr Asn Ile Asp Tyr Thr Ser Lys Thr Gly
```

```
                135                 140                 145

gat acc ata tat aat gga att tct gaa ctc aca aat tat aca gga aca      1018
Asp Thr Ile Tyr Asn Gly Ile Ser Glu Leu Thr Asn Tyr Thr Gly Thr
        150                 155                 160

acc caa aaa atg aaa acc gat agt ttt caa aga gat tat aca aaa tct      1066
Thr Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr Thr Lys Ser
    165                 170                 175

gaa tcc act tca gta aca aat ggg tta caa tta gga ttt aaa gtt gct      1114
Glu Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Phe Lys Val Ala
180                 185                 190                 195

gct aag gga gta gtt gca tta gca ggt gca gat ttt gaa aca agt gtt      1162
Ala Lys Gly Val Val Ala Leu Ala Gly Ala Asp Phe Glu Thr Ser Val
                200                 205                 210

acc tat aat tta tca tct act aca act gaa aca aat aca ata tcg gat      1210
Thr Tyr Asn Leu Ser Ser Thr Thr Thr Glu Thr Asn Thr Ile Ser Asp
            215                 220                 225

aag ttt act gtt cca tct caa gaa gtt aca tta tcc cca gga cat aaa      1258
Lys Phe Thr Val Pro Ser Gln Glu Val Thr Leu Ser Pro Gly His Lys
        230                 235                 240

gca gtg gtg aaa cat gat ttg aga aaa atg gtg tat ttt ggg act cat      1306
Ala Val Val Lys His Asp Leu Arg Lys Met Val Tyr Phe Gly Thr His
    245                 250                 255

gat tta aag ggt gat tta aaa gta ggt ttt aat gat aaa gag att gta      1354
Asp Leu Lys Gly Asp Leu Lys Val Gly Phe Asn Asp Lys Glu Ile Val
260                 265                 270                 275

caa aaa ttt att tat cca aat tat aga tca att gat tta tct gat att      1402
Gln Lys Phe Ile Tyr Pro Asn Tyr Arg Ser Ile Asp Leu Ser Asp Ile
                280                 285                 290

cgt aaa aca atg att gaa att gat aaa tgg aat cat gta aat acc att      1450
Arg Lys Thr Met Ile Glu Ile Asp Lys Trp Asn His Val Asn Thr Ile
            295                 300                 305

gac ttt tat caa tta gtt gga gtt aaa aat cat ata aaa aat ggt gat      1498
Asp Phe Tyr Gln Leu Val Gly Val Lys Asn His Ile Lys Asn Gly Asp
        310                 315                 320

act tta tat ata gat acc ccg gcc gaa ttt aca ttt aat gga gct aat      1546
Thr Leu Tyr Ile Asp Thr Pro Ala Glu Phe Thr Phe Asn Gly Ala Asn
    325                 330                 335

cca tat tat aga gca aca ttt aca gaa tac gac gag aac gga aat cct      1594
Pro Tyr Tyr Arg Ala Thr Phe Thr Glu Tyr Asp Glu Asn Gly Asn Pro
340                 345                 350                 355

gtt caa aca aag att tta agt gga aat                                 1621
Val Gln Thr Lys Ile Leu Ser Gly Asn
                360
```

<210> 6

<211> 364

<212> PRT

<213> Bacillus thuringiensis

<400> 6

Met Lys Tyr Lys Phe Ser Lys Val Val Lys Cys Thr Leu Pro Ala Leu
1 5 10 15

Met Ile Thr Thr Phe Val Thr Pro Ser Met Ala Val Phe Ala Ala Glu
20 25 30

Thr Lys Ser Pro Asn Leu Asn Ala Ser Gln Gln Ala Ile Thr Pro Tyr
35 40 45

Ala Glu Ser Tyr Ile Asp Thr Val Gln Asp Arg Met Lys Gln Arg Asp
50 55 60

Arg Glu Ser Lys Leu Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile
65 70 75 80

Ile Asp Gly Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn
85 90 95

Asn Asn His Gln Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Leu
100 105 110

Ala Ser Ser Lys Gly Tyr Asp Ser Ile Ala Glu Gln Met Gly Leu Lys
115 120 125

Ile Glu Ala Leu Asn Asp Met Asp Val Thr Asn Ile Asp Tyr Thr Ser
130 135 140

Lys Thr Gly Asp Thr Ile Tyr Asn Gly Ile Ser Glu Leu Thr Asn Tyr
145 150 155 160

Thr Gly Thr Thr Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr
165 170 175

Thr Lys Ser Glu Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Phe
180 185 190

Lys Val Ala Ala Lys Gly Val Val Ala Leu Ala Gly Ala Asp Phe Glu
195 200 205

Thr Ser Val Thr Tyr Asn Leu Ser Ser Thr Thr Thr Glu Thr Asn Thr
210 215 220

Ile Ser Asp Lys Phe Thr Val Pro Ser Gln Glu Val Thr Leu Ser Pro
225 230 235 240

Gly His Lys Ala Val Val Lys His Asp Leu Arg Lys Met Val Tyr Phe
245 250 255

Gly Thr His Asp Leu Lys Gly Asp Leu Lys Val Gly Phe Asn Asp Lys
260 265 270

Glu Ile Val Gln Lys Phe Ile Tyr Pro Asn Tyr Arg Ser Ile Asp Leu
275 280 285

Ser Asp Ile Arg Lys Thr Met Ile Glu Ile Asp Lys Trp Asn His Val
290 295 300

Asn Thr Ile Asp Phe Tyr Gln Leu Val Gly Val Lys Asn His Ile Lys
305 310 315 320

Asn Gly Asp Thr Leu Tyr Ile Asp Thr Pro Ala Glu Phe Thr Phe Asn
325 330 335

Gly Ala Asn Pro Tyr Tyr Arg Ala Thr Phe Thr Glu Tyr Asp Glu Asn
340 345 350

Gly Asn Pro Val Gln Thr Lys Ile Leu Ser Gly Asn
355 360

<210> 7

<211> 1552

<212> DNA

<213> Bacillus thuringiensis

<220>

<221> CDS

<222> (169)..(1272)

<223>

<400> 7

ctagctgaat atgcagtaga taatgatttt gactatacta atgagatttt ttatgtaata 60

aaccatattt atccttaaat aaatgttcac atatttaaaa agaataaact accatgcggt 120

gctagaatat aaatgagagc gctaacaaaa aataatggag ggataatc atg aaa tac 177

Met Lys Tyr
1

```
aaa tca tca aaa gta gca ata tgt act tta tca gct tta atg ctt tca      225
Lys Ser Ser Lys Val Ala Ile Cys Thr Leu Ser Ala Leu Met Leu Ser
  5                  10                  15

aca att ggt act tcg agt atg tcc act ttt gct gca gaa aca aca tta      273
Thr Ile Gly Thr Ser Ser Met Ser Thr Phe Ala Ala Glu Thr Thr Leu
20                  25                  30                  35

cca ggt caa act ctt aag gaa caa tca ata acc cca cgt gca gaa tct      321
Pro Gly Gln Thr Leu Lys Glu Gln Ser Ile Thr Pro Arg Ala Glu Ser
                40                  45                  50

tat att gat att gta caa gat aga atg aaa caa agg gat ata gaa tcg      369
Tyr Ile Asp Ile Val Gln Asp Arg Met Lys Gln Arg Asp Ile Glu Ser
            55                  60                  65

aaa cgt act ggt aaa ccg att aat atg caa gaa caa ata ata gat gga      417
Lys Arg Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile Ile Asp Gly
        70                  75                  80

tgg ttt tta gca aga ttc tgg ata ttt aaa gat caa aat aat aac cat      465
Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn Asn Asn His
    85                  90                  95

caa aca aat aga ttc ata aca tgg ttt aaa aat aat gtt gcc agc tca      513
Gln Thr Asn Arg Phe Ile Thr Trp Phe Lys Asn Asn Val Ala Ser Ser
100                 105                 110                 115

aaa ggt tat gag ggt att gca gaa caa atg ggt ttg aaa ata gaa tcg      561
Lys Gly Tyr Glu Gly Ile Ala Glu Gln Met Gly Leu Lys Ile Glu Ser
                120                 125                 130

atg agt gat atg aat gta tcg aat ata aat tat aca ggt aaa aag ggt      609
Met Ser Asp Met Asn Val Ser Asn Ile Asn Tyr Thr Gly Lys Lys Gly
            135                 140                 145

gat act ata tat aat ggc gtt tcg gaa tta gaa aat aaa atg gga aca      657
Asp Thr Ile Tyr Asn Gly Val Ser Glu Leu Glu Asn Lys Met Gly Thr
        150                 155                 160

cct caa aaa atg aaa tca gat agt ttt caa aga gat tat acc aaa tct      705
Pro Gln Lys Met Lys Ser Asp Ser Phe Gln Arg Asp Tyr Thr Lys Ser
    165                 170                 175

caa tca acc tca gta aca aat ggg tta caa tta gga gtt aaa gtt tct      753
Gln Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Val Lys Val Ser
180                 185                 190                 195

gcc aaa ggt acg gtt gtc tta gga gag gca agc ctt gaa aca agc gtt      801
Ala Lys Gly Thr Val Val Leu Gly Glu Ala Ser Leu Glu Thr Ser Val
                200                 205                 210

act tat aat tta tcg tct act gca act gaa aca gat aca aca tcg gac      849
Thr Tyr Asn Leu Ser Ser Thr Ala Thr Glu Thr Asp Thr Thr Ser Asp
            215                 220                 225

aag ttt act gtc cca tcc caa gaa gtt aca tta cca cca gga cat aaa      897
Lys Phe Thr Val Pro Ser Gln Glu Val Thr Leu Pro Pro Gly His Lys
        230                 235                 240
```

38-21(52806) Sequence Listing_PCT_2.ST25.txt

```
gca gta att aag cat gat tta aga aaa atg gtg tat tct ggt acg cat      945
Ala Val Ile Lys His Asp Leu Arg Lys Met Val Tyr Ser Gly Thr His
    245                 250                 255

gac tta aag ggg gat tta aaa gta gct ttt aac gat aaa gca att gta      993
Asp Leu Lys Gly Asp Leu Lys Val Ala Phe Asn Asp Lys Ala Ile Val
260                 265                 270                 275

caa aaa ttt att tat cca aat tat aga tct ata aat tta tct gat att     1041
Gln Lys Phe Ile Tyr Pro Asn Tyr Arg Ser Ile Asn Leu Ser Asp Ile
                280                 285                 290

cgt aaa aca atg aaa gaa att gat gaa tgg aat cat gta aaa ccc att     1089
Arg Lys Thr Met Lys Glu Ile Asp Glu Trp Asn His Val Lys Pro Ile
            295                 300                 305

gat ttt tat caa ctg gtt gga ata aaa aat cat ata aaa aat ggg gat     1137
Asp Phe Tyr Gln Leu Val Gly Ile Lys Asn His Ile Lys Asn Gly Asp
        310                 315                 320

acc tta tat ata gag act cca gct aaa ttt att ttt aat gga gct aat     1185
Thr Leu Tyr Ile Glu Thr Pro Ala Lys Phe Ile Phe Asn Gly Ala Asn
    325                 330                 335

gta tat tat aga gca act ttt aca gaa tat gat aag gat gga aaa cct     1233
Val Tyr Tyr Arg Ala Thr Phe Thr Glu Tyr Asp Lys Asp Gly Lys Pro
340                 345                 350                 355

gtt caa ttc aac aaa ttt tta agt gaa aat tac aag tta tagaggaagt      1282
Val Gln Phe Asn Lys Phe Leu Ser Glu Asn Tyr Lys Leu
                360                 365

aaagatgccg tagtgagatc gtttcacagc tactgagtat tcaaataata cacgggaaaa   1342

ttcaccttcc tggaaggacg gatttacttt ttttacggag gaacttgttt tatacatcaa   1402

aatgtttttt tatgaggttt gtgtattctt atttgagcct ggaacggaac cattttgagt   1462

aagcttaatt tgacttggaa atgtattttt attaccttat tacgtgaaca atggcctata   1522

aacgtgccac acaggaatgg gaggacgagt                                    1552
```

<210> 8

<211> 368

<212> PRT

<213> Bacillus thuringiensis

<400> 8

```
Met Lys Tyr Lys Ser Ser Lys Val Ala Ile Cys Thr Leu Ser Ala Leu
1               5                   10                  15

Met Leu Ser Thr Ile Gly Thr Ser Ser Met Ser Thr Phe Ala Ala Glu
            20                  25                  30
```

```
Thr Thr Leu Pro Gly Gln Thr Leu Lys Glu Gln Ser Ile Thr Pro Arg
        35                  40                  45

Ala Glu Ser Tyr Ile Asp Ile Val Gln Asp Arg Met Lys Gln Arg Asp
    50                  55                  60

Ile Glu Ser Lys Arg Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile
65                  70                  75                  80

Ile Asp Gly Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn
                85                  90                  95

Asn Asn His Gln Thr Asn Arg Phe Ile Thr Trp Phe Lys Asn Asn Val
            100                 105                 110

Ala Ser Ser Lys Gly Tyr Glu Gly Ile Ala Glu Gln Met Gly Leu Lys
        115                 120                 125

Ile Glu Ser Met Ser Asp Met Asn Val Ser Asn Ile Asn Tyr Thr Gly
    130                 135                 140

Lys Lys Gly Asp Thr Ile Tyr Asn Gly Val Ser Glu Leu Glu Asn Lys
145                 150                 155                 160

Met Gly Thr Pro Gln Lys Met Lys Ser Asp Ser Phe Gln Arg Asp Tyr
                165                 170                 175

Thr Lys Ser Gln Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Val
            180                 185                 190

Lys Val Ser Ala Lys Gly Thr Val Val Leu Gly Glu Ala Ser Leu Glu
        195                 200                 205

Thr Ser Val Thr Tyr Asn Leu Ser Ser Thr Ala Thr Glu Thr Asp Thr
    210                 215                 220

Thr Ser Asp Lys Phe Thr Val Pro Ser Gln Glu Val Thr Leu Pro Pro
225                 230                 235                 240

Gly His Lys Ala Val Ile Lys His Asp Leu Arg Lys Met Val Tyr Ser
                245                 250                 255

Gly Thr His Asp Leu Lys Gly Asp Leu Lys Val Ala Phe Asn Asp Lys
            260                 265                 270

Ala Ile Val Gln Lys Phe Ile Tyr Pro Asn Tyr Arg Ser Ile Asn Leu
        275                 280                 285
```

Ser Asp Ile Arg Lys Thr Met Lys Glu Ile Asp Glu Thr Asn His Val
290 295 300

Lys Pro Ile Asp Phe Tyr Gln Leu Val Gly Ile Lys Asn His Ile Lys
305 310 315 320

Asn Gly Asp Thr Leu Tyr Ile Glu Thr Pro Ala Lys Phe Ile Phe Asn
325 330 335

Gly Ala Asn Val Tyr Tyr Arg Ala Thr Phe Thr Glu Tyr Asp Lys Asp
340 345 350

Gly Lys Pro Val Gln Phe Asn Lys Phe Leu Ser Glu Asn Tyr Lys Leu
355 360 365

<210> 9

<211> 1378

<212> DNA

<213> Bacillus thuringiensis

<220>

<221> CDS

<222> (215)..(1306)

<223>

<400> 9

cagtggatag gaatttgttt tcgtgctagg tatcaattta atttgttcta taagataagt 60

gaagtacgat caaaatgaat acttttgtgt attagatcaa taggtaaaat aataataaat 120

tttatatttg aaccttaaaa aattatttaa tcaaatcttt ttcactttaa aaacaaaata 180

tccagaaaaa acaatagtta acggagggat aata atg aaa tac aag tca tca aaa 235
Met Lys Tyr Lys Ser Ser Lys
1 5

gta gca atg tgt aca tta tca gct tta atg ctt tcg aca atc gcc act 283
Val Ala Met Cys Thr Leu Ser Ala Leu Met Leu Ser Thr Ile Ala Thr
10 15 20

cca agt ata tct gtt ttc gct gca gaa aca act tcg tca cat gcg gtt 331
Pro Ser Ile Ser Val Phe Ala Ala Glu Thr Thr Ser Ser His Ala Val
25 30 35

act aat cag caa aca att acg cag cgt gca gaa tct tat att gat att 379
Thr Asn Gln Gln Thr Ile Thr Gln Arg Ala Glu Ser Tyr Ile Asp Ile
40 45 50 55

```
gtg cac aat aga atg aaa caa aga gat atc gaa tca aaa atg aca ggt      427
Val His Asn Arg Met Lys Gln Arg Asp Ile Glu Ser Lys Met Thr Gly
                60                  65                  70

aaa tcc att aat atg caa gaa caa ata att gat gga tgg ttt tta gct      475
Lys Ser Ile Asn Met Gln Glu Gln Ile Ile Asp Gly Trp Phe Leu Ala
            75                  80                  85

aga ttt tgg ata ttt aag gat caa aat aat agt cac caa aca aat aga      523
Arg Phe Trp Ile Phe Lys Asp Gln Asn Asn Ser His Gln Thr Asn Arg
        90                  95                  100

ttt att tca tgg ttt aag gat aat ttg gct agc cca gga ggg tat gat      571
Phe Ile Ser Trp Phe Lys Asp Asn Leu Ala Ser Pro Gly Gly Tyr Asp
    105                 110                 115

agt atc gct gaa cag atg ggc cta aaa gta gca gca tta aat gat atg      619
Ser Ile Ala Glu Gln Met Gly Leu Lys Val Ala Ala Leu Asn Asp Met
120                 125                 130                 135

gat ata tca aat gta aat tat act tct aag aca ggg gat act ata tat      667
Asp Ile Ser Asn Val Asn Tyr Thr Ser Lys Thr Gly Asp Thr Ile Tyr
                140                 145                 150

aat ggt gtt tca gaa tta aaa aat atc aca gga aca act caa aaa atg      715
Asn Gly Val Ser Glu Leu Lys Asn Ile Thr Gly Thr Thr Gln Lys Met
            155                 160                 165

aaa aca gat agt ttt caa aga gat tat aca aaa tcc cag tca act tca      763
Lys Thr Asp Ser Phe Gln Arg Asp Tyr Thr Lys Ser Gln Ser Thr Ser
        170                 175                 180

atc acc aat gga tta caa tta gga ttt aaa gtt tca gct aaa gga ata      811
Ile Thr Asn Gly Leu Gln Leu Gly Phe Lys Val Ser Ala Lys Gly Ile
    185                 190                 195

gtg gcc tta gcc ggt gcg gat ttt gaa gca agt gta aac tat aat tta      859
Val Ala Leu Ala Gly Ala Asp Phe Glu Ala Ser Val Asn Tyr Asn Leu
200                 205                 210                 215

tcc act acc gca act gaa acc aat aca ata tct gat aaa ttt acc gtt      907
Ser Thr Thr Ala Thr Glu Thr Asn Thr Ile Ser Asp Lys Phe Thr Val
                220                 225                 230

cct tca caa gaa gtc aca tta gcg cca gga cat aag gcg atc gta aaa      955
Pro Ser Gln Glu Val Thr Leu Ala Pro Gly His Lys Ala Ile Val Lys
            235                 240                 245

cat agt ttg aag aaa atg gta tac tct gga acg cat gat tta aaa gga     1003
His Ser Leu Lys Lys Met Val Tyr Ser Gly Thr His Asp Leu Lys Gly
        250                 255                 260

gat tta aca att act ttt aat gat aag gat tta gtt caa aaa ttt att     1051
Asp Leu Thr Ile Thr Phe Asn Asp Lys Asp Leu Val Gln Lys Phe Ile
    265                 270                 275

tat cca aat tat aaa gct att gat tta tct aat att cgt aaa gca atg     1099
Tyr Pro Asn Tyr Lys Ala Ile Asp Leu Ser Asn Ile Arg Lys Ala Met
280                 285                 290                 295

aca gaa att gat gaa tgg aat cat gta aaa cct acc gat ttc tat caa     1147
Thr Glu Ile Asp Glu Trp Asn His Val Lys Pro Thr Asp Phe Tyr Gln
```

300 305 310

tta gtt ggg aat aaa aat tat ata aaa aac ggg gac act tta tac atc 1195
Leu Val Gly Asn Lys Asn Tyr Ile Lys Asn Gly Asp Thr Leu Tyr Ile
315 320 325

gaa aca cct gct aaa ttc act ttg aat gga ggc aac cct tat tat aca 1243
Glu Thr Pro Ala Lys Phe Thr Leu Asn Gly Gly Asn Pro Tyr Tyr Thr
330 335 340

gca acc ttt acg gaa tat gat gaa aat gga aat caa gtc aaa aca aag 1291
Ala Thr Phe Thr Glu Tyr Asp Glu Asn Gly Asn Gln Val Lys Thr Lys
345 350 355

cgt tta aat aac aaa taagttactt aaaggtaatt cattaacaat gtatccatta 1346
Arg Leu Asn Asn Lys
360

tataattaat ttataaaaat aatgttttaa aa 1378

<210> 10

<211> 364

<212> PRT

<213> Bacillus thuringiensis

<400> 10

Met Lys Tyr Lys Ser Ser Lys Val Ala Met Cys Thr Leu Ser Ala Leu
1 5 10 15

Met Leu Ser Thr Ile Ala Thr Pro Ser Ile Ser Val Phe Ala Ala Glu
20 25 30

Thr Thr Ser Ser His Ala Val Thr Asn Gln Gln Thr Ile Thr Gln Arg
35 40 45

Ala Glu Ser Tyr Ile Asp Ile Val His Asn Arg Met Lys Gln Arg Asp
50 55 60

Ile Glu Ser Lys Met Thr Gly Lys Ser Ile Asn Met Gln Glu Gln Ile
65 70 75 80

Ile Asp Gly Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn
85 90 95

Asn Ser His Gln Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Leu
100 105 110

Ala Ser Pro Gly Gly Tyr Asp Ser Ile Ala Glu Gln Met Gly Leu Lys
115 120 125

Val Ala Ala Leu Asn Asp Met Asp Ile Ser Asn Val Asn Tyr Thr Ser
130 135 140

Lys Thr Gly Asp Thr Ile Tyr Asn Gly Val Ser Glu Leu Lys Asn Ile
145 150 155 160

Thr Gly Thr Thr Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr
165 170 175

Thr Lys Ser Gln Ser Thr Ser Ile Thr Asn Gly Leu Gln Leu Gly Phe
180 185 190

Lys Val Ser Ala Lys Gly Ile Val Ala Leu Ala Gly Ala Asp Phe Glu
195 200 205

Ala Ser Val Asn Tyr Asn Leu Ser Thr Thr Ala Thr Glu Thr Asn Thr
210 215 220

Ile Ser Asp Lys Phe Thr Val Pro Ser Gln Glu Val Thr Leu Ala Pro
225 230 235 240

Gly His Lys Ala Ile Val Lys His Ser Leu Lys Lys Met Val Tyr Ser
245 250 255

Gly Thr His Asp Leu Lys Gly Asp Leu Thr Ile Thr Phe Asn Asp Lys
260 265 270

Asp Leu Val Gln Lys Phe Ile Tyr Pro Asn Tyr Lys Ala Ile Asp Leu
275 280 285

Ser Asn Ile Arg Lys Ala Met Thr Glu Ile Asp Glu Trp Asn His Val
290 295 300

Lys Pro Thr Asp Phe Tyr Gln Leu Val Gly Asn Lys Asn Tyr Ile Lys
305 310 315 320

Asn Gly Asp Thr Leu Tyr Ile Glu Thr Pro Ala Lys Phe Thr Leu Asn
325 330 335

Gly Gly Asn Pro Tyr Tyr Thr Ala Thr Phe Thr Glu Tyr Asp Glu Asn
340 345 350

Gly Asn Gln Val Lys Thr Lys Arg Leu Asn Asn Lys
355 360

<210> 11

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> probe sequence, or amplification primer sequence for use with primer as set forth in SEQ ID NO 12, corresponding to CDS as set forth in SEQ ID NO 3 from 438-458, biased toward codons preferred in Bacillus species genes containing A or T in 3rd position

<400> 11
aataataatc atcaaacwaa t 21

<210> 12

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> probe sequence, or amplification primer sequence for use with SEQ ID NO 11 corresponding to SEQ ID NO 3 from nucleotide position 978- 998, biased toward codons preferred in Bacillus species genes in which A or T is in 3rd position

<400> 12
attwggataw ataaattttt g 21

<210> 13

<211> 1101

<212> DNA

<213> Artificial Sequence

<220>

<223> coding sequence preferred for use in monocot species encoding a Bt TIC901 amino acid sequence variant

<220>

<221> CDS

<222> (1)..(1101)

<223>

<400> 13

```
atg aag aac cgc ttc agc aag gtc gcc ctc tgc acg gtg cct atc ctc      48
Met Lys Asn Arg Phe Ser Lys Val Ala Leu Cys Thr Val Pro Ile Leu
1               5                  10                  15

atg gtt tct acg ttc gcg tcc agc tcg atg tcc gcg ttc gca gcg gag      96
Met Val Ser Thr Phe Ala Ser Ser Ser Met Ser Ala Phe Ala Ala Glu
            20                  25                  30

gcc aaa agt cct gac ttg aac gtg tcc caa cag gtc ata ggc cct tac     144
Ala Lys Ser Pro Asp Leu Asn Val Ser Gln Gln Val Ile Gly Pro Tyr
        35                  40                  45

gca gaa tct tac atc gac atc gtc cag gac aga atg aag cag aga gac     192
Ala Glu Ser Tyr Ile Asp Ile Val Gln Asp Arg Met Lys Gln Arg Asp
    50                  55                  60

aaa gga tcc aaa ctc act ggc aaa ccc atc aac atg caa gag cag atc     240
Lys Gly Ser Lys Leu Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile
65                  70                  75                  80

atc gat ggg tgg ttt ctc gca cga ttc tgg att ttc aag gat cag aac     288
Ile Asp Gly Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn
                85                  90                  95

aat aac cac cag aca aac agg ttc atc tca tgg ttt aag gat aac atc     336
Asn Asn His Gln Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Ile
            100                 105                 110

gcc tca tct aag gga tac aac tca ata gcc gaa cag atg ggc ctc aaa     384
Ala Ser Ser Lys Gly Tyr Asn Ser Ile Ala Glu Gln Met Gly Leu Lys
        115                 120                 125

atc gaa gca gag aat gat atg gac gtg aca aat atc gac tac act agt     432
Ile Glu Ala Glu Asn Asp Met Asp Val Thr Asn Ile Asp Tyr Thr Ser
    130                 135                 140

aag acc gga gac aca atc tac aac ggc att tcg gaa ctt aaa aac tat     480
Lys Thr Gly Asp Thr Ile Tyr Asn Gly Ile Ser Glu Leu Lys Asn Tyr
145                 150                 155                 160

acg ggc agc acc cag aaa atg aag acc gat agc ttt caa agg gac tac     528
Thr Gly Ser Thr Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr
                165                 170                 175

aca aaa tcc gag tcg acc tcc gtg acc aat ggc ctc cag ctg ggc ttc     576
Thr Lys Ser Glu Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Phe
            180                 185                 190

aag gtg gca gca aag ggc gtc gtc gct tta gcc ggc gca gac ttc gag     624
Lys Val Ala Ala Lys Gly Val Val Ala Leu Ala Gly Ala Asp Phe Glu
        195                 200                 205

act tcg gtg acc tac aat ctg tct aca act acg act gag acg aac aca     672
Thr Ser Val Thr Tyr Asn Leu Ser Thr Thr Thr Thr Glu Thr Asn Thr
    210                 215                 220

att tcc gac aag ttt acg gtt ccg tct cag gag gtt acg ttc cct cca     720
Ile Ser Asp Lys Phe Thr Val Pro Ser Gln Glu Val Thr Phe Pro Pro
225                 230                 235                 240
```

```
ggc cac aag gca atc gtc aag cac gac ctg agg aaa atg gtc tac agc      768
Gly His Lys Ala Ile Val Lys His Asp Leu Arg Lys Met Val Tyr Ser
                245                 250                 255

ggc acc cat gat ctc aaa ggc gac ctc atc gtg tcg ttc aac gac aag      816
Gly Thr His Asp Leu Lys Gly Asp Leu Ile Val Ser Phe Asn Asp Lys
            260                 265                 270

gag ata gtc cag aag ttc atc tac cca aat tac cgc gac atc aac ctc      864
Glu Ile Val Gln Lys Phe Ile Tyr Pro Asn Tyr Arg Asp Ile Asn Leu
        275                 280                 285

agt gac atc cga gag acc atg atc gag atc gac gag tgg aac cac gtg      912
Ser Asp Ile Arg Glu Thr Met Ile Glu Ile Asp Glu Trp Asn His Val
    290                 295                 300

aac cct gtc aat ttc tac gaa ctc gta gga gtt aag aac cac atc aag      960
Asn Pro Val Asn Phe Tyr Glu Leu Val Gly Val Lys Asn His Ile Lys
305                 310                 315                 320

aac ggt gaa aca ttg tac atc gac acg ccg gct aag ttc atg ttc aac     1008
Asn Gly Glu Thr Leu Tyr Ile Asp Thr Pro Ala Lys Phe Met Phe Asn
                325                 330                 335

gga gcg aat cct tac tat cga gct acc ttc acg gag tac gat ggc aac     1056
Gly Ala Asn Pro Tyr Tyr Arg Ala Thr Phe Thr Glu Tyr Asp Gly Asn
            340                 345                 350

aac aat cct gtt cag acc aag gtg ttg agt gag aat ttc aag ctg         1101
Asn Asn Pro Val Gln Thr Lys Val Leu Ser Glu Asn Phe Lys Leu
        355                 360                 365
```

<210> 14

<211> 367

<212> PRT

<213> Artificial Sequence

<220>

<223> coding sequence preferred for use in monocot species encoding a B t TIC901 amino acid sequence variant

<400> 14

```
Met Lys Asn Arg Phe Ser Lys Val Ala Leu Cys Thr Val Pro Ile Leu
1               5                   10                  15

Met Val Ser Thr Phe Ala Ser Ser Ser Met Ser Ala Phe Ala Ala Glu
            20                  25                  30

Ala Lys Ser Pro Asp Leu Asn Val Ser Gln Gln Val Ile Gly Pro Tyr
        35                  40                  45
```

Ala Glu Ser Tyr Ile Asp Ile Val Gln Asp Arg Met Lys Gln Arg Asp
50 55 60

Lys Gly Ser Lys Leu Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile
65 70 75 80

Ile Asp Gly Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn
85 90 95

Asn Asn His Gln Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Ile
100 105 110

Ala Ser Ser Lys Gly Tyr Asn Ser Ile Ala Glu Gln Met Gly Leu Lys
115 120 125

Ile Glu Ala Glu Asn Asp Met Asp Val Thr Asn Ile Asp Tyr Thr Ser
130 135 140

Lys Thr Gly Asp Thr Ile Tyr Asn Gly Ile Ser Glu Leu Lys Asn Tyr
145 150 155 160

Thr Gly Ser Thr Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr
165 170 175

Thr Lys Ser Glu Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Phe
180 185 190

Lys Val Ala Ala Lys Gly Val Val Ala Leu Ala Gly Ala Asp Phe Glu
195 200 205

Thr Ser Val Thr Tyr Asn Leu Ser Thr Thr Thr Thr Glu Thr Asn Thr
210 215 220

Ile Ser Asp Lys Phe Thr Val Pro Ser Gln Glu Val Thr Phe Pro Pro
225 230 235 240

Gly His Lys Ala Ile Val Lys His Asp Leu Arg Lys Met Val Tyr Ser
245 250 255

Gly Thr His Asp Leu Lys Gly Asp Leu Ile Val Ser Phe Asn Asp Lys
260 265 270

Glu Ile Val Gln Lys Phe Ile Tyr Pro Asn Tyr Arg Asp Ile Asn Leu
275 280 285

Ser Asp Ile Arg Glu Thr Met Ile Glu Ile Asp Glu Trp Asn His Val
290 295 300

Asn Pro Val Asn Phe Tyr Glu Leu Val Gly Val Lys Asn His Ile Lys
305 310 315 320

Asn Gly Glu Thr Leu Tyr Ile Asp Thr Pro Ala Lys Phe Met Phe Asn
325 330 335

Gly Ala Asn Pro Tyr Tyr Arg Ala Thr Phe Thr Glu Tyr Asp Gly Asn
340 345 350

Asn Asn Pro Val Gln Thr Lys Val Leu Ser Glu Asn Phe Lys Leu
355 360 365

<210> 15

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer coupled with prJPW152 SEQ ID NO 16 in thermal amplification reactions

<400> 15
cctttggcag aaactttaac tcc 23

<210> 16

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer coupled with prJPW151 SEQ ID NO 15 in thermal amplification reactions

<400> 16
gtgtattctg gtacgcatga c 21

<210> 17

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer coupled with prJPW183 SEQ ID NO 18 in thermal amplification reactions

<400> 17
gccggatccc tagctgaata tgcagtagat aatg 34

<210> 18

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer coupled with prJPW186 SEQ ID NO 17 in thermal amplification reactions

<400> 18
gtggcacgtt tataggccat tgttc 25

<210> 19

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer coupled with prJPW156 SEQ ID NO 20 in thermal amplification reactions

<400> 19
cttttaggcc catctgttca gcg 23

<210> 20

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer coupled with prJPW155 SEQ ID NO 19 in thermal amplification reactions

<400> 20
gccttagccg gtgcggattt tgaagc 26

<210> 21

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer coupled with prJPW170 SEQ ID NO 22 in thermal amplification reactions

<400> 21
ggagcttatt tgttatttaa acgctttgtt ttgacttgat ttcc 44

<210> 22

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer coupled with prJPW168 SEQ ID NO 21 in thermal amplification reactions

<400> 22
gccggatccc agtggatagg aatttgtttt cgtgctagg 39

<210> 23

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> universal forward amplification primer similar to SEQ ID NO:24 and SEQ ID NO:25 that, when used in a thermal amplification reaction with any of SEQ ID NO:SEQ ID NO:27-29 and template DNA homologous to tic901, 1201, 407, 417, or 431 and the like result in amplicon of from about 600 to about 650 base pairs

<400> 23
aayatgcarg arcarathat hgaygg 26

<210> 24

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> universal forward amplification primer similar to SEQ ID NO:23 and SEQ ID NO:25 that, when used in a thermal amplification reaction with any of SEQ ID NO:27-29 and template DNA homologous to tic901, 1201, 407, 417, or 431 and the like result in amplicon of from about 600 to about 650 base pairs

<400> 24
aayatgcarg arcarathat hga 23

<210> 25

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> universal forward amplification primer similar to SEQ ID NO:23 and SEQ ID NO:24 that, when used in a thermal amplification reaction with any of SEQ ID NO:27-29 and template DNA homologous to tic901, 1201, 407, 417, or 431 and the like result in amplicon of from about 600 to about 650 base pairs

<400> 25
aayatgcarg arcarathat 20

<210> 26

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> universal forward amplification primer that, when used in a thermal amplification reaction with any of SEQ ID NO:27-29 and template DNA homologous to tic901, 1201, 407, 417, 431 and the like result in amplicon of from about 395 to about 425 base pairs

<220>

<221> misc_feature

<222> (3)..(3)

<223> inosine

<220>

<221> misc_feature

<222> (9)..(9)

<223> inosine

<400> 26
ggngayacna thtayaaygg 20

<210> 27

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> universal reverse amplification primer similar to SEQ ID NO:28 and SEQ ID NO:29

<220>

<221> misc_feature

<222> (6)..(6)

<223> inosine

<220>

<221> misc_feature

<222> (24)..(24)

<223> inosine

<400> 27
tarttnggrt adatraaytt ytgnac 26

<210> 28

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> universal reverse amplification primer similar to SEQ ID NO:27 and SEQ ID NO:29

<220>

<221> misc_feature

<222> (6)..(6)

<223> inosine

<400> 28
tarttnggrt adatraaytt ytg 23

<210> 29

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> universal reverse amplification primer similar to SEQ ID NO:27 and SEQ ID NO:28

<220>

<221> misc_feature

<222> (18)..(18)

<223> inosine

<400> 29
ggrtadatra ayttytgnac 20

<210> 30

<211> 570

<212> DNA

<213> Bacillus thuringiensis

<220>

<221> CDS

<222> (1)..(570)

<223>

<400> 30

```
ttt tta gct aga ttt tgg ata ttt gag gat caa aat aat agt cac caa        48
Phe Leu Ala Arg Phe Trp Ile Phe Glu Asp Gln Asn Asn Ser His Gln
1               5                  10                  15

aca aat aga ttt att tca tgg ttt aag gat aat att gct agt tca aaa        96
Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Ile Ala Ser Ser Lys
            20                  25                  30

ggg tat aat agt att gcg gag caa atg ggt tta aaa ata gaa gca gaa       144
Gly Tyr Asn Ser Ile Ala Glu Gln Met Gly Leu Lys Ile Glu Ala Glu
        35                  40                  45

aac gat atg gat gta aca aat ata gat tat aca tct aag aca ggc gat       192
Asn Asp Met Asp Val Thr Asn Ile Asp Tyr Thr Ser Lys Thr Gly Asp
    50                  55                  60

acc att tat aat ggt att tca gaa ttg aaa aat tat aca gga tca act       240
Thr Ile Tyr Asn Gly Ile Ser Glu Leu Lys Asn Tyr Thr Gly Ser Thr
65                  70                  75                  80

caa aag atg aaa aca gat agt ttt caa aga gat tat aca aaa tca gaa       288
Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr Thr Lys Ser Glu
                85                  90                  95

tct act tca gta act aat gga tta caa tta gga ttt aaa gtt gct gct       336
Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Phe Lys Val Ala Ala
            100                 105                 110

aaa gga gta gtt gct ttg gct ggg gca gac ttt gaa acc agt gtt act       384
Lys Gly Val Val Ala Leu Ala Gly Ala Asp Phe Glu Thr Ser Val Thr
        115                 120                 125

tat aat cta tca act act aca act gaa aca aat aca ata tca gac aag       432
Tyr Asn Leu Ser Thr Thr Thr Thr Glu Thr Asn Thr Ile Ser Asp Lys
    130                 135                 140

ttt act gtc cca tct caa gaa gtt aca ttg cct cca gga cat aaa gcg       480
Phe Thr Val Pro Ser Gln Glu Val Thr Leu Pro Pro Gly His Lys Ala
145                 150                 155                 160

ata gtg aaa cat gat tta aga aaa atg gtt tat tct ggt act cat gat       528
Ile Val Lys His Asp Leu Arg Lys Met Val Tyr Ser Gly Thr His Asp
                165                 170                 175

cta aag ggt gat tta att gtg agt ttt aat gat aaa gag att               570
Leu Lys Gly Asp Leu Ile Val Ser Phe Asn Asp Lys Glu Ile
            180                 185                 190
```

<210> 31

<211> 190

<212> PRT

<213> Bacillus thuringiensis

<400> 31

Phe Leu Ala Arg Phe Trp Ile Phe Glu Asp Gln Asn Asn Ser His Gln
1 5 10 15

Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Ile Ala Ser Ser Lys
20 25 30

Gly Tyr Asn Ser Ile Ala Glu Gln Met Gly Leu Lys Ile Glu Ala Glu
35 40 45

Asn Asp Met Asp Val Thr Asn Ile Asp Tyr Thr Ser Lys Thr Gly Asp
50 55 60

Thr Ile Tyr Asn Gly Ile Ser Glu Leu Lys Asn Tyr Thr Gly Ser Thr
65 70 75 80

Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr Thr Lys Ser Glu
85 90 95

Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Phe Lys Val Ala Ala
100 105 110

Lys Gly Val Val Ala Leu Ala Gly Ala Asp Phe Glu Thr Ser Val Thr
115 120 125

Tyr Asn Leu Ser Thr Thr Thr Thr Glu Thr Asn Thr Ile Ser Asp Lys
130 135 140

Phe Thr Val Pro Ser Gln Glu Val Thr Leu Pro Pro Gly His Lys Ala
145 150 155 160

Ile Val Lys His Asp Leu Arg Lys Met Val Tyr Ser Gly Thr His Asp
165 170 175

Leu Lys Gly Asp Leu Ile Val Ser Phe Asn Asp Lys Glu Ile
180 185 190

<210> 32

<211> 1095

<212> DNA

<213> Bacillus thuringiensis

<220>

<221> CDS

<222> (1)..(1092)

<223> sequence encoding TIC431 precursor amino acid sequence

<400> 32

```
atg aaa tac aag tct tca aaa gta gca atg tgt aca tta tcg gct tta       48
Met Lys Tyr Lys Ser Ser Lys Val Ala Met Cys Thr Leu Ser Ala Leu
1               5                  10                  15

atg ctt tcg aca atc gcc act cca agt ata tct gtt ttc gct gct gaa       96
Met Leu Ser Thr Ile Ala Thr Pro Ser Ile Ser Val Phe Ala Ala Glu
            20                  25                  30

aca act gca tca cat aag gtt act aat cag caa aca att gca cag cgt      144
Thr Thr Ala Ser His Lys Val Thr Asn Gln Gln Thr Ile Ala Gln Arg
        35                  40                  45

gca gaa tct tat atc gat att gtg cat aat aga atg aaa aaa cga gat      192
Ala Glu Ser Tyr Ile Asp Ile Val His Asn Arg Met Lys Lys Arg Asp
    50                  55                  60

att gaa tca aaa atg aca ggt aaa cct att aat atg caa gaa caa ata      240
Ile Glu Ser Lys Met Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile
65                  70                  75                  80

att gat gga tgg ttt tta gct aga ttt tgg ata ttc aag gac caa aat      288
Ile Asp Gly Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn
                85                  90                  95

aat agt cac caa aca aat aga ttt att tca tgg ttt aaa gat aat tta      336
Asn Ser His Gln Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Leu
            100                 105                 110

gct agt cca gga ggg tat aat agt atc gct aaa caa atg ggg tta aaa      384
Ala Ser Pro Gly Gly Tyr Asn Ser Ile Ala Lys Gln Met Gly Leu Lys
        115                 120                 125

ata gaa gta tta aat gat atg gat ata tca aat gta aat tat act tct      432
Ile Glu Val Leu Asn Asp Met Asp Ile Ser Asn Val Asn Tyr Thr Ser
    130                 135                 140

aag aca ggg gat act ata tat aat ggt gtt tcc gaa tta aaa aat atc      480
Lys Thr Gly Asp Thr Ile Tyr Asn Gly Val Ser Glu Leu Lys Asn Ile
145                 150                 155                 160

aca ggt aca act caa aaa atg aaa aca gat agt ttt caa aga gat tat      528
Thr Gly Thr Thr Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr
                165                 170                 175
```

```
aca aaa tca cag tca act tca atc acc aat gga tta caa tta gga ttt      576
Thr Lys Ser Gln Ser Thr Ser Ile Thr Asn Gly Leu Gln Leu Gly Phe
            180                 185                 190

aaa gtt tct gcc aaa ggg gtg ata gct tta gca gga gca gac ttc gaa      624
Lys Val Ser Ala Lys Gly Val Ile Ala Leu Ala Gly Ala Asp Phe Glu
        195                 200                 205

gca agt gtc aac tat aat tta tcc act acc gca act gaa acc aat ata      672
Ala Ser Val Asn Tyr Asn Leu Ser Thr Thr Ala Thr Glu Thr Asn Ile
    210                 215                 220

ata tct gat aaa ttt acc gtt cct tca caa gaa gtt aca tta gcg cca      720
Ile Ser Asp Lys Phe Thr Val Pro Ser Gln Glu Val Thr Leu Ala Pro
225                 230                 235                 240

gga cat aag gcg atc gta aaa cat agt tta aag aaa atg gta tac tcc      768
Gly His Lys Ala Ile Val Lys His Ser Leu Lys Lys Met Val Tyr Ser
                245                 250                 255

gga acg cat gat tta aaa gga gat tta aca att act ttt aat gat aag      816
Gly Thr His Asp Leu Lys Gly Asp Leu Thr Ile Thr Phe Asn Asp Lys
            260                 265                 270

gat tta gtt caa aaa ttt att tat cca aat tat aaa gct att gat tta      864
Asp Leu Val Gln Lys Phe Ile Tyr Pro Asn Tyr Lys Ala Ile Asp Leu
        275                 280                 285

tct aat att cgt aaa gca ctg act gaa att gat gaa tgg aat cat gta      912
Ser Asn Ile Arg Lys Ala Leu Thr Glu Ile Asp Glu Trp Asn His Val
    290                 295                 300

aaa cct acc gat ttc tat caa tta gtt ggg aac aaa aat tat ata aaa      960
Lys Pro Thr Asp Phe Tyr Gln Leu Val Gly Asn Lys Asn Tyr Ile Lys
305                 310                 315                 320

aac ggg gac act tta tac atc gaa aca cct gct aaa ttc act ttg aat     1008
Asn Gly Asp Thr Leu Tyr Ile Glu Thr Pro Ala Lys Phe Thr Leu Asn
                325                 330                 335

gga gga aac cct tat tat aca gca acc ttt acg gaa tat gat gaa agt     1056
Gly Gly Asn Pro Tyr Tyr Thr Ala Thr Phe Thr Glu Tyr Asp Glu Ser
            340                 345                 350

gga aat caa gtc aaa aca aag cat tta agt gtc aaa taa                 1095
Gly Asn Gln Val Lys Thr Lys His Leu Ser Val Lys
        355                 360
```

<210> 33

<211> 364

<212> PRT

<213> Bacillus thuringiensis

<400> 33

Met Lys Tyr Lys Ser Ser Lys Val Ala Met Cys Thr Leu Ser Ala Leu

1 5 10 15

Met Leu Ser Thr Ile Ala Thr Pro Ser Ile Ser Val Phe Ala Ala Glu
20 25 30

Thr Thr Ala Ser His Lys Val Thr Asn Gln Gln Thr Ile Ala Gln Arg
35 40 45

Ala Glu Ser Tyr Ile Asp Ile Val His Asn Arg Met Lys Lys Arg Asp
50 55 60

Ile Glu Ser Lys Met Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile
65 70 75 80

Ile Asp Gly Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn
85 90 95

Asn Ser His Gln Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Leu
100 105 110

Ala Ser Pro Gly Gly Tyr Asn Ser Ile Ala Lys Gln Met Gly Leu Lys
115 120 125

Ile Glu Val Leu Asn Asp Met Asp Ile Ser Asn Val Asn Tyr Thr Ser
130 135 140

Lys Thr Gly Asp Thr Ile Tyr Asn Gly Val Ser Glu Leu Lys Asn Ile
145 150 155 160

Thr Gly Thr Thr Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr
165 170 175

Thr Lys Ser Gln Ser Thr Ser Ile Thr Asn Gly Leu Gln Leu Gly Phe
180 185 190

Lys Val Ser Ala Lys Gly Val Ile Ala Leu Ala Gly Ala Asp Phe Glu
195 200 205

Ala Ser Val Asn Tyr Asn Leu Ser Thr Thr Ala Thr Glu Thr Asn Ile
210 215 220

Ile Ser Asp Lys Phe Thr Val Pro Ser Gln Glu Val Thr Leu Ala Pro
225 230 235 240

Gly His Lys Ala Ile Val Lys His Ser Leu Lys Lys Met Val Tyr Ser
245 250 255

Gly Thr His Asp Leu Lys Gly Asp Leu Thr Ile Thr Phe Asn Asp Lys
260 265 270

Asp Leu Val Gln Lys Phe Ile Tyr Pro Asn Tyr Lys Ala Ile Asp Leu
275 280 285

Ser Asn Ile Arg Lys Ala Leu Thr Glu Ile Asp Glu Trp Asn His Val
290 295 300

Lys Pro Thr Asp Phe Tyr Gln Leu Val Gly Asn Lys Asn Tyr Ile Lys
305 310 315 320

Asn Gly Asp Thr Leu Tyr Ile Glu Thr Pro Ala Lys Phe Thr Leu Asn
325 330 335

Gly Gly Asn Pro Tyr Tyr Thr Ala Thr Phe Thr Glu Tyr Asp Glu Ser
340 345 350

Gly Asn Gln Val Lys Thr Lys His Leu Ser Val Lys
355 360

PROTEÍNAS INSECTICIDAS SECRETADAS DE *BACILLUS THURINGIENSIS* Y USOS DE LAS MISMAS

CAMPO DE LA INVENCIÓN

La presente invención se relaciona con una nueva familia de secuencias de nucleótidos que codifican proteínas insecticidas y fragmentos insecticidas de las mismas. En particular, la presente invención está dirigida a proteínas ilustrativas denominadas TIC901, TIC1201, TIC407, TIC417 y TIC431 en la presente, y fragmentos insecticidas de las mismas, cada una codificada por las secuencias de nucleótidos codificantes ilustrativas denominadas, respectivamente, *tic*901, *tic*1201, *tic*407, *tic*417 y *tic*431 en la presente, así como homólogos de las secuencias de nucleótidos que (1) codifican proteínas insecticidas y (2) se hibridizan con las secuencias de codificación *tic*901, *tic*1201, *tic*407, *tic*417 y *tic*431 bajo condiciones de hibridación seleccionadas del grupo formado por condiciones de hibridación severas y condiciones de hibridación específicas. La presente invención también se relaciona con células huésped transformadas con las secuencias de nucleótidos de la presente invención o transformadas con variantes de las secuencias de nucleótidos basadas en el gen *tic*901, con genes relacionados y/u homólogos de las mismas, en particular aquellas secuencias que fueron modificadas para mejorar la expresión en plantas. En una realización preferida, las células huésped transformadas son células vegetales.

ANTECEDENTES DE LA INVENCIÓN

Bacillus thuringiensis es una bacteria Gram-positiva que produce inclusiones cristalinas proteináceas durante la esporulación. Estas proteínas cristalinas de *B. thuringiensis* a menudo son muy tóxicas para algunos insectos específicos. Se han identificado actividades insecticidas para las proteínas cristalinas de diversas cepas de *B. thuringiensis* contra larvas de insectos de las órdenes de insecto Lepidóptera (orugas), Díptera (mosquitos, moscas) y Coleóptera (escarabajos).

Las proteínas cristalinas individuales de *B. thuringiensis*, también conocidas como delta-endotoxinas o cristales paraesporales o proteínas de toxinas, pueden

diferir ampliamente en cuanto a su estructura y actividad insecticida Estas proteínas insecticidas están codificadas por genes que típicamente están ubicados sobre plásmidos grandes, con un tamaño de más de 30 mega Daltons (MDa), presentes en las cepas de *B thuringiensis*. Se han clonado numerosos genes de toxinas de *B thuringiensis* y se ha caracterizado asimismo los productos de proteínas cristalinas insecticidas por sus propiedades insecticidas específicas. Hay numerosas revisiones sobre genes y proteínas cristalinas de toxinas de *B thuringiensis* (por ejemplo, Höfte et al., 1989, Schnepf et al., 1998).

Las propiedades insecticidas de *B thuringiensis* se conocen desde hace mucho tiempo y se han incorporado cepas de *B. thuringiensis* en productos biológicos insecticidas comerciales por más de cuarenta años. Las formulaciones insecticidas comerciales de *B thuringiensis* contienen típicamente cultivos por fermentación esporulados secos de *B. thuringiensis* cuyas proteínas cristalinas son tóxicas para diversas especies de insectos.

Los productos bioinsecticidas comerciales tradicionales de *B. thuringiensis* derivan de cepas de *B. thuringiensis*" de tipo salvaje", es decir, cultivos purificados de cepas de *B thuringiensis* aislados de fuentes naturales Los productos bioinsecticidas comerciales de *B. thuringiensis* más nuevos se basan en cepas de *B. thuringiensis* genéticamente alteradas, tal como las cepas transconjugadas de *B. thuringiensis* que se describen en las Patentes de EE.UU. Nº. 5.080.897 y 4.935.353.

Se han clasificado numerosas cepas de *B. thuringiensis* sobre la base de las reacciones de los flagelos de *B. thuringiensis* con anticuerpos. Una cepa de *B thuringiensis* cuyos flagelos reaccionan con un único anticuerpo se clasifica como una serovariedad única, y se han descrito más de treinta serovariedades o subespecies diferentes de *B. thuringiensis* (DeBarjac y Frachon, 1990).

Cada subespecie de *B thuringiensis* a menudo produce tipos únicos de proteínas cristalinas insecticidas Por ejemplo, la subespecie *kurstaki* de *B. thuringiensis* produce proteínas cristalinas de aproximadamente 130 kilo Daltons (kD)

y 70 kD que son tóxicas para orugas, en tanto la subespecie *tenebrionis* de *B. thuringiensis* produce una proteína cristalina de aproximadamente 72 kD que es tóxica para escarabajos.

Una característica de las proteínas cristalinas es su capacidad para coalescer formando cristales dentro de la célula madre de *B. thuringiensis*. Con la lisis de la célula madre las proteínas son liberadas como cristales hacia el entorno externo Además, *B thuringiensis* también produce proteínas no cristalinas que, a diferencia de las proteínas cristalinas, son secretadas por las células de *B. thuringiensis* como proteínas solubles hacia el medio de cultivo. Las proteínas no cristalinas secretadas de *B. thuringiensis* incluyen fosfolipasas, proteasas y β-lactamasas que tienen poca actividad insecticida, si alguna Sin embargo, se han descrito tres proteínas no cristalinas secretadas de *B. thuringiensis* denominadas Vip1, Vip2 y Vip3 que son tóxicas para insectos coleópteros o lepidópteros (Estruch et al., 1996; Patente de EE.UU. Nº.5.866 326; WO 94/21795; WO 96/10083). Se ha descrito que una proteína no cristalina de *B thuringiensis* denominada CryV es tóxica para insectos lepidópteros (Kostichka et al , 1996). Se han identificado previamente numerosas formas aisladas de *Bacillus thuringiensis* que producen toxinas de proteínas secretadas extracelulares insecticidas (Patente de EE.UU. Nº:5.840 868, Patente de EE UU. Nº.5.849.870, Patente de EE.UU. Nº.5.866.326, Patente de EE UU Nº·5 872 212; Patente de EE UU Nº·5.877 012; Patente de EE.UU. Nº.5 888 801, Patente de EE UU. Nº.6.204 435, Patente de EE.UU. Nº 6.242.669; Patente de EE.UU Nº·6 279.369). Se ha demostrado que todas dichas cepas producen una o más de estas proteínas de toxinas VIP o CryV u homólogos estrechamente relacionados. Sorpresivamente, los inventores describen en la presente una nueva clase de toxinas de proteínas secretadas extracelulares insecticidas que no exhiben homología con las clases de proteínas conocidas, VIP o CryV.

Las comparaciones de las secuencias de aminoácidos indican que las clases de proteínas Vip1, Vip2, Vip3, WAR, MIS y CryV no están relacionadas con las

proteínas de la presente invención. Otra comparación muestra que ninguna de las ciento treinta y siete proteínas tóxicas más o menos conocidas para insectos de *B. thunngiensis* (Crickmore et al , 1998) está relacionada con las proteínas de la presente invención De hecho, no se encontró una homología significativa entre las secuencias de las proteínas de la presente invención y ninguna de las miles de secuencias proteicas contenidas en National Center for Genome Resources (GenBank), Santa Fe, NM Una búsqueda con BLAST solamente permitió identificar dos proteínas en la base de datos GenBank que sugería una posible homología con TIC901 La proteína insecticida Mtx2 de *Bacillus sphaericus* solamente exhibió un 21% de identidad de secuencia de aminoácidos con una secuencia de 135 aminoácidos contiguos de longitud cuando se alineaba con TIC901 Una secuencia de aminoácidos putativa que se puede expresar a partir del genoma del virus Fowlpox solamente presentaba una identidad del 27% para la secuencia de aminoácidos con una secuencia contigua de 147 aminoácidos de longitud cuando era alineada con TIC901.

RESUMEN DE LA INVENCIÓN

En una realización, la presente invención se relaciona con una proteína insecticida aislada y purificada, cuya secuencia de aminoácidos es sustancialmente como se muestra en las SEQ ID Nº: 4, SEQ ID Nº 6, SEQ ID Nº.8, SEQ ID Nº 10 y SEQ ID Nº 33, ya sea como secuencias de aminoácidos precursoras o como formas maduras y/o procesadas y secretadas de estas secuencias de aminoácidos o secuencias de aminoácidos relacionadas y homólogos de las mismas. Se ha demostrado la actividad insecticida de TIC901 y de las proteínas relacionadas en bioensayos con el escarabajo de papa de Colorado (CPB) y con los gusanos de la raíz del maíz del oeste y del sur. En particular, las proteínas son tóxicas para insectos coleópteros incluyendo el escarabajo de papa de Colorado (*Lymantria dispar*) y los gusanos de la raíz del maíz (CRW), como se muestra en la presente

En otra realización, la presente invención también se relaciona con una

secuencia de nucleótidos aislada y purificada, es decir una secuencia de codificación, que comprende la secuencia de nucleótidos que se muestra en las SEQ ID Nº 3, SEQ ID Nº 5, SEQ ID Nº:7, SEQ ID Nº·9 y/o SEQ ID Nº:32 y secuencias relacionadas u homólogos de las mismas La secuencia de codificación nativa o de tipo salvaje *tic*901 que se muestra en la SEQ ID Nº:3 codifica un precursor, una preproteína o una proteína pretoxina de la TIC901 nativa que presenta la secuencia de aminoácidos que se muestra en la SEQ ID Nº:4. Los organismos productores de la proteína TIC901 exhiben actividad insecticida y/o propiedades de resistencia a insectos. La secuencia de aminoácidos insecticida correspondiente a la proteína localizada en el espacio extracelular que rodea a la célula de *Bacillus* que expresa la proteína de la SEQ ID Nº:3 corresponde a una proteína que comprende desde la posición de aminoácido 44 aproximadamente hasta la posición de aminoácido 367 aproximadamente como se muestra en la SEQ ID Nº 4 La secuencia de codificación nativa o de tipo salvaje *tic*1201 que se muestra en la SEQ ID Nº.5 codifica la proteína precursora TIC1201 que presenta la secuencia de aminoácidos que se muestra en la SEQ ID Nº.6 La secuencia de aminoácidos insecticida correspondiente a la proteína localizada en el espacio extracelular que rodea a la célula de *Bacillus* que expresa la proteína de la SEQ ID Nº 5 corresponde a la proteína madura que comprende desde la posición de aminoácido 44 aproximadamente hasta la posición de aminoácido 364 aproximadamente como se muestra en la SEQ ID Nº·6 La secuencia de codificación nativa o de tipo salvaje *tic*407 que se muestra en la SEQ ID Nº:7 codifica el precursor, la preproteína o la proteína pretoxina TIC407 que presenta la secuencia de aminoácidos que se muestra en la SEQ ID Nº·8. La secuencia de aminoácidos insecticida correspondiente a la proteína madura localizada en el espacio extracelular que rodea a la célula de *Bacillus* que expresa la proteína de la SEQ ID Nº·7 corresponde a una proteína que comprende desde la posición de aminoácido 44 aproximadamente hasta la posición de aminoácido 367 aproximadamente como se muestra en la SEQ ID Nº:8 La secuencia de codificación nativa o de tipo salvaje

*tic*417 que se muestra en la SEQ ID Nº:9 codifica al precursor, a la preproteína o a la proteína pretoxina TIC417 que presenta la secuencia de aminoácidos que se muestra en las SEQ ID Nº.10 La secuencia de aminoácidos insecticida correspondiente a la proteína madura localizada en el espacio extracelular que rodea a la célula de *Bacillus* que expresa la proteína de la SEQ ID Nº.9 corresponde a una proteína que comprende desde la posición de aminoácido 44 aproximadamente hasta la posición de aminoácido 364 aproximadamente como se muestra en la SEQ ID Nº.10. La secuencia de codificación nativa o de tipo salvaje *tic*431 que se muestra en la SEQ ID Nº 32 codifica al precursor, a la preproteína o a la proteína pretoxina TIC431 que presenta la secuencia de aminoácidos que se muestra en las SEQ ID Nº:33. La secuencia de aminoácidos insecticida correspondiente a la proteína madura localizada en el espacio extracelular que rodea a la célula de *Bacillus* que expresa la proteína de la SEQ ID Nº 33 corresponde a una proteína que comprende desde el aminoácido 44 aproximadamente hasta el aminoácido 364 aproximadamente como se muestra en la SEQ ID Nº:33 Los homólogos de las secuencias de nucleótidos, es decir, las proteínas insecticidas codificadas por secuencias de nucleótidos que se hibridizan con cada una o cualquiera de las secuencias descritas en la presente bajo condiciones de hibridación severas, están específicamente incluidos en el alcance de la presente invención.

En una realización adicional, la presente invención se relaciona con un cultivo biológicamente puro de una bacteria *Bacillus thuringiensis* transformada con un vector plásmido que contiene la secuencia de nucleótidos que se muestra en las SEQ ID Nº.3, SEQ ID Nº.5, SEQ ID Nº:7, SEQ ID Nº.9 y/o SEQ ID Nº.32, y/o secuencias relacionadas u homólogos que producen una proteína insecticida y secreta la proteína hacia el espacio extracelular que rodea a la cepa bacteriana durante la fermentación. Se ha depositado un ejemplar de la cepa EG12450 en una colección de cultivo permanente en cumplimiento con el Tratado de Budapest y se le ha asignado el Nº de Acceso NRRL B-30357

En una realización adicional, la invención también se relaciona con un cultivo biológicamente puro de una bacteria *B. thuringiensis* designada como cepa EG2158 que presenta actividad insecticida contra insectos coleópteros La cepa EG2158 de *B thuringiensis* representa la cepa de *B. thuringiensis* de tipo salvaje a partir de la cual se aisló la secuencia de codificación *tic*901 y fue depositada en una colección de cultivo permanente en cumplimiento con el Tratado de Budapest y ha recibido el Nº de Acceso NRRL B-18213. En la presente se muestra que la cepa EG2158 produce al menos dos proteínas insecticidas que comprenden las secuencias de aminoácidos seleccionadas del grupo formado por las SEQ ID Nº.4 y SEQ ID Nº.10.

En una realización adicional, la presente invención provee un vector que comprende la secuencia de nucleótidos que se muestra en la SEQ ID Nº:3 que codifica la secuencia de aminoácidos TIC901 que se muestra en la SEQ ID Nº 4. Se ha depositado una cepa de *Escherichia coli* que contiene un vector que comprende la SEQ ID Nº:3, el 6 de febrero, 2002, en las instalaciones de *Northern Regional Research Lab of the Agricultural Research Service Center Collection* (NRRL), USDA, bajo las disposiciones del "Tratado de Budapest sobre el reconocimiento internacional del depósito de microorganismos a los efectos del procedimiento de patentes" y recibió el Nº de Acceso NRRL B-30549. El plásmido que contiene dicha secuencia de nucleótidos se muestra aquí como el pEG1381.

En una realización adicional, la presente invención provee la secuencia de nucleótidos que se muestra en las SEQ ID Nº.3 que codifica la secuencia de aminoácidos TIC901 y una porción de oligonucleótidos que se puede marcar y usar como sonda de hibridación para identificar otros genes relacionados que codifican proteínas insecticidas relacionadas u homólogos de las mismas. Otras secuencias de nucleótidos relacionadas ilustradas específicamente en la presente comprenden las secuencias que se muestran en las SEQ ID Nº:5, SEQ ID Nº:7, SEQ ID Nº:9 y SEQ ID Nº:32, cada una de las cuales codifica las toxinas proteicas insecticidas que se muestran en las SEQ ID Nº: 4, SEQ ID Nº:6, SEQ ID Nº:8, SEQ ID Nº:10 y SEQ ID

Nº.33, respectivamente

En aún otra realización adicional, la invención provee células vegetales y plantas que fueron transformadas con la secuencia de nucleótidos que codifica la proteína TIC901 que se muestra en la SEQ ID Nº:4 o un fragmento insecticida de la misma. La secuencia de nucleótidos puede ser traducida y expresada por células vegetales y en tejidos vegetales a niveles suficientes para inhibir o matar plagas por insectos coleópteros. Tanto las plantas monocotiledóneas como dicotiledóneas se encuentran dentro del alcance de la invención Puede ser necesario modificar la secuencia con el fin de lograr el nivel máximo de expresión y para aumentar la capacidad de la planta que contiene la secuencia para producir niveles insecticidas de la proteína TIC901.

En aún otra realización adicional, la presente invención también provee un método para producir una planta transgénica que presenta mayores niveles de expresión de la secuencia de nucleótidos que codifica TIC901, y después mayores niveles de la proteína insecticida TIC901 Las plantas así transformadas con las secuencias de nucleótidos modificadas con respecto a las descritas en la presente, exhiben niveles mayores y mejores de resistencia a plagas de coleópteros en comparación con una planta que no contiene la secuencia de nucleótidos que codifica una TIC901 o una proteína relacionada

Para poder lograr lo mencionado, se provee un método para expresar una secuencia de nucleótidos que codifica una proteína TIC901 en una planta que comprende los pasos de insertar en el genoma de una célula vegetal una secuencia de ácido nucleico que comprende en la dirección 5' a 3', un promotor funcional en plantas ligado operativamente a una secuencia de ADN estructural optimizada para la expresión en plantas que causa la producción de una secuencia de ARN que codifica la secuencia de polipéptidos TIC901 que se muestra en la SEQ ID Nº: 4, o una secuencia que presenta desde al menos un 80% aproximadamente o desde al menos un 85% aproximadamente o desde al menos un 90% aproximadamente o

desde al menos un 95% aproximadamente o desde al menos un 99% de identidad de secuencia con la secuencia de aminoácidos que se muestra en la SEQ ID Nº: 4, y una secuencia de ADN no traducida 3' cuya función en las células de la planta causa la terminación de la transcripción y poliadenilación; obtener células vegetales transformadas que contienen a dicha secuencia de ácido nucleico; y regenerar a partir de las células vegetales transformadas plantas transformadas genéticamente que expresan la secuencia de nucleótidos que codifica la proteína TIC901, donde las plantas transformadas son morfológicamente normales y exhiben niveles elevados o mejorados de resistencia a una plaga de coleópteros en comparación con una planta que no fue transformada para expresar dicha proteína.

Otra realización de la presente invención comprende la provisión de anticuerpos que se unen específicamente a los epitopes presentados solamente por la proteína TIC901 u homólogos de la misma Los anticuerpos se pueden usar para identificar la presencia de una proteína TIC901 o de un homólogo, para purificar dicha proteína u homólogo, para identificar la secuencia de nucleótidos a partir de la cual se expresa una proteína TIC901 u homólogo y para su uso en conjuntos de elementos diseñados para poder detectar una proteína TIC901 u homólogo o para detectar la secuencia de nucleótidos que expresa dicha proteína u homólogo.

Una ventaja particular de la presente invención comprende la mejora en el manejo de la resistencia a insectos. La capacidad para combinar dos o más agentes insecticidas, cada uno tóxico para la misma especie de plaga de insecto, en una sola composición y el hecho que cada agente exhiba un modo de acción diferente de los demás agentes insecticidas con los cuales es combinado, presenta un medio para controlar con mayor eficacia una especie de plaga de insecto particular porque reduce sustancialmente la probabilidad de desarrollo de resistencia a la composición insecticida en la población La proteína TIC901 de la presente invención se puede combinar con numerosos de agentes insecticidas conocidos para lograr el alcanzar un nivel de manejo de la resistencia en una composición particular, preferentemente

por expresión de la combinación de agentes insecticidas en plantas. En particular, las composiciones con la proteína TIC901 se pueden combinar con variantes de la secuencia de aminoácidos Cry3 o Cry3 para controlar diversas especies de coleópteros que son plagas vegetales, o con otras proteínas Cry apropiadas, tal como PS149B1, CryET33/34, CryET80/76, CryET70, Cry22, CryET39, CryET76, Cry5Ba, Cry6a y Cry12a, y semejantes, y con proteínas VIP, WAR o MIS y semejantes, y con diversas composiciones insecticidas derivadas de especies de las bacterias *Xenorhabdus* y *Photorhabdus* que han demostrado una bioactividad insecticida dirigida a especies de coleópteros que son plagas vegetales. Preferentemente, el uso *in planta* de estas composiciones estará dirigido al mejoramiento de la expresión de las proteínas en las partes de la planta que presentan mayor vulnerabilidad a la depredación por insectos coleópteros. Para la protección de papa contra CPB, será preferible lograr los mayores niveles de expresión en las hojas y los tallos de la planta Para las especies de maíz susceptibles a gusanos de alambre y a gusanos de la raíz, será preferible alcanzar los mayores niveles de expresión en las partes subterráneas de la planta, es decir, en los sistemas radiculares de la planta.

Otra realización comprende un polinucleótido aislado que codifica una toxina insecticida de *Bacillus thuringiensis* o un fragmento insecticida de la misma, que sea activo contra una plaga de insecto, donde dicha toxina o fragmento insecticida tiene un peso molecular entre aproximadamente 36.000 Daltons y aproximadamente 42.500 Daltons. Además, la secuencia de nucleótidos que codifica la toxina, o el complemento de la misma, se hibridiza bajo condiciones severas con la SEQ ID Nº:3. La toxina exhibe preferentemente una actividad biológica que permite controlar o matar una plaga de insectos coleópteros, preferentemente el escarabajo de papa de Colorado y/o gusanos de la raíz del maíz En una realización la secuencia de nucleótidos que codifica la toxina es optimizada para su expresión en plantas, aún cuando codifica sustancialmente la toxina o un fragmento insecticida de la misma, es

decir, codifica a misma, o sustancialmente la misma, secuencia de aminoácidos que está presente en la secuencia de aminoácidos nativa.

Otra realización de la presente invención provee células huésped transformadas para contener un polinucleótido que codifica una proteína insecticida de la presente invención o un fragmento insecticida de la misma Preferentemente las secuencias de nucleótidos de la presente invención son modificadas para mejorar la expresión de las proteínas de la presente invención en una célula huésped preferida La célula huésped de la presente invención se selecciona del grupo formado por una célula bacteriana, una célula fúngica y una célula vegetal. La expresión en una célula vegetal puede comprender la expresión para poder acumular la proteína insecticida en el citoplasma o puede dar como resultado la acumulación de la proteína insecticida en un organela subcelular, tal como un plástido, cloroplasto o mitocondrias. Como alternativa, la proteína insecticida de la presente invención, o fragmentos insecticidas de la misma, se puede localizar en la maquinaria de secreción de proteínas de dicha célula huésped particular y dar como resultado la acumulación del producto proteico fuera de la célula y en los espacios extracelulares que rodean a dicha célula.

Una realización adicional de la presente invención provee un método para controlar la infestación de una planta por una especie de insecto coleóptero Preferentemente, se provee una cantidad plaguicida de una proteína insecticida de la presente invención, o un fragmento insecticida de la misma, para ser consumida por dicha plaga de insecto con la dieta del insecto. La dieta puede consistir de una parte de la planta de la cual el insecto se alimenta normalmente, tal como un tejido vegetal o una célula vegetal. La proteína insecticida, o un fragmento insecticida de la misma, se puede proveer en una composición que es aplicada a la superficie del tejido vegetal, a una parte de la planta o a células vegetales o, más preferentemente, puede ser producida por la maquinaria de síntesis de proteínas de la célula y, como se describió previamente, acumulada dentro de la célula vegetal o secretada fuera de

la célula vegetal, en tanto la cantidad de la proteína toxina provista sea una cantidad insecticida suficiente como para inhibir que la plaga de insectos continúe alimentándose, o como para inhibir el crecimiento y desarrollo adicional de la plaga de insecto o como para causar la muerte de dicha plaga de insecto La dieta provista al insecto también puede ser una dieta artificial que contiene a la proteína toxina uniformemente distribuida en el interior o aplicada tópicamente a una o más superficies expuestas del sustrato de dieta o incluida como un gradiente de concentraciones en el interior o aplicado tópicamente a una o más superficies del sustrato de dieta. La toxina insecticida, o un fragmento de la misma, deriva de una secuencia de nucleótidos codificada en *Bacillus thuringiensis* por una secuencia de nucleótidos que se hibridiza bajo condiciones severas con una secuencia de nucleótidos sustancialmente complementaria de la SEQ ID Nº.3.

La presente invención también provee un método para detectar una primera secuencia de nucleótidos que se hibridiza con una segunda secuencia de nucleótidos que se muestra en la SEQ ID Nº.3, donde dicha primera secuencia de nucleótidos codifica una proteína insecticida, o un fragmento insecticida de la misma, y se hibridiza bajo condiciones de hibridación severas con la segunda secuencia de nucleótidos Los ejemplos de dichas secuencias son las SEQ ID Nº:2 y SEQ ID Nº:3.

La presente invención también provee secuencias de nucleótidos no naturales o sintéticas que codifica una proteína insecticida TIC901, o un fragmento insecticida de la misma o un homólogo de la misma, donde dicha proteína TIC901, o un fragmento insecticida de la misma o un homólogo de la misma, se selecciona del grupo de secuencias formado por las SEQ ID Nº.5 y SEQ ID Nº:7 Preferentemente, la secuencia de nucleótidos no natural, o las secuencias provistas en la presente, que codifica una proteína insecticida o un fragmento insecticida de la misma se proveen para expresar una proteína TIC901 o una proteína relacionada en células vegetales Por ello, se proveen en la presente las células vegetales transformadas con dichas secuencias Las plantas cultivadas a partir de las células vegetales

transformadas se proveen asimismo con la presente invención También se proveen las semillas de las plantas transformadas de la presente invención en tanto las semillas contengan al menos las secuencias que codifican las proteínas insecticidas o fragmentos de dichas proteínas insecticidas.

Los ejemplos de secuencias de la presente invención incluyen, además de las que están relacionadas con las SEQ ID Nº.3 y SEQ ID Nº:4, al menos (1) la secuencia de nucleótidos que se muestra en la SEQ ID Nº:5 y la secuencia de aminoácidos codificada por la SEQ ID Nº·5 que se muestra en la SEQ ID Nº·6, también denominada proteína insecticida TIC1201 en la presente, (2) la secuencia de nucleótidos que se muestra en la SEQ ID Nº:7 y la secuencia de aminoácidos codificada por la SEQ ID Nº:7 que se muestra en la SEQ ID Nº 8, también denominada proteína insecticida TIC407 en la presente, (3) la secuencia de nucleótidos que se muestra en la SEQ ID Nº.9 y la secuencia de aminoácidos codificada por la SEQ ID Nº·9 que se muestra en la SEQ ID Nº 10, también denominada proteína insecticida TIC417 en la presente y (4) la secuencia de nucleótidos que se muestra en la SEQ ID Nº.32 y la secuencia de aminoácidos codificada por la SEQ ID Nº:32 que se muestra en la SEQ ID Nº:33, también denominada proteína insecticida TIC431 en la presente. Cada una de estas proteínas y las secuencias de nucleótidos nativas de *B. t.* que codifican estas proteínas están relacionadas con TIC901 según se define en la presente. Por ejemplo, y respectivamente, la SEQ ID Nº·5 es una secuencia de nucleótidos que codifica la proteína insecticida TIC1201 que se muestra en la SEQ ID Nº:6 La SEQ ID Nº:5 que se muestra en la presente se puede identificar por hibridación con la SEQ ID Nº·3 bajo condiciones severas. La SEQ ID Nº·5 codifica una proteína que exhibe actividad biológica tóxica para coleópteros, exhibe toxicidad para los gusanos de la raíz del maíz y los escarabajos de papa de Colorado. Las SEQ ID Nº.5, SEQ ID Nº 3, SEQ ID Nº:7, SEQ ID Nº·9 y SEQ ID Nº:32 tienen la capacidad, cada una, de hibridizarse entre sí bajo condiciones de hibridación seleccionadas del grupo formado por

condiciones de hibridación severas y condiciones de hibridación específicas. Cada secuencia también se puede identificar por hibridación con la SEQ ID Nº.2 bajo condiciones seleccionadas del grupo formado por condiciones de hibridación severas y condiciones de hibridación específicas. Cada secuencia también se puede identificar por amplificación usando, por ejemplo, el par de cebadores de oligonucleótidos que se muestra en las SEQ ID Nº.11 y SEQ ID Nº:12 y el par de cebadores de oligonucleótidos que se muestra en las SEQ ID Nº:23 y SEQ ID Nº:27. El par de cebadores que se muestra en las SEQ ID Nº:11 y SEQ ID Nº.12 y el par de cebadores que se muestra en las SEQ ID Nº.23 y SEQ ID Nº.27 son ilustrativos y elementos de diagnóstico para identificar la presencia de la secuencia de nucleótidos que codifica una proteína TIC901 o una proteína insecticida relacionada en una muestra Estos pares de oligonucleótidos, cuando se usan solos o juntos bajo condiciones de amplificación definidas y en presencia de una secuencia de nucleótidos adecuada como sustrato, produce un amplicón que comprende entre aproximadamente 540 y aproximadamente 640 pares de bases. Las reacciones de amplificación térmica que emplean estos juegos de cebadores son de utilidad para detectar la presencia de un gen *B.t* que codifica una proteína insecticida correspondiente a la proteína TIC901 o una proteína relacionada en una muestra, y simplifica enormemente la búsqueda e identificación de dichas secuencias relacionadas. También se contemplan otros amplicones derivados del uso de otros pares de cebadores basados en la alineación de secuencias de nucleótidos de, por ejemplo, las SEQ ID Nº:3, SEQ ID Nº:5, SEQ ID Nº:7, SEQ ID Nº 9 y SEQ ID Nº·32. Las regiones de identidad sustancial de secuencia de aminoácidos de las proteínas codificadas por estas secuencias de nucleótidos corresponden a secuencias de nucleótidos que se pueden emplear para preparar secuencias complementarias o sustancialmente complementarias para su uso como sondas o cebadores en reacciones de amplificación térmica que permitan detectar las secuencias relacionadas con las proteínas TIC901, TIC1201, TIC407, TIC417 y TIC431

Se proveen además las sondas y los cebadores de oligonucleótidos degenerados que se muestran en las SEQ ID Nº:23 a SEQ ID Nº.29 como un medio para identificar cualquier secuencia de nucleótidos que codifique una proteína insecticida secretada de al menos una especie de *Bacillus thuringiensis* en la cual la secuencia de nucleótidos identificada con las sondas de oligonucleótidos degeneradas se hibridiza bajo condiciones severas con una o más de las secuencias seleccionadas del grupo formado por SEQ ID Nº:2, SEQ ID Nº.3, SEQ ID Nº.5, SEQ ID Nº·7, SEQ ID Nº 9, SEQ ID Nº 11, SEQ ID Nº·12, SEQ ID Nº:15, SEQ ID Nº.16, SEQ ID Nº:17, SEQ ID Nº.18, SEQ ID Nº 19, SEQ ID Nº 20, SEQ ID Nº:21, SEQ ID Nº:22, SEQ ID Nº.23, SEQ ID Nº:24, SEQ ID Nº:25, SEQ ID Nº.26, SEQ ID Nº:27, SEQ ID Nº:28, SEQ ID Nº.29 y SEQ ID Nº·32. Los ejemplos de secuencias que se pueden identificar usando dichos oligonucleótidos incluyen las secuencias seleccionadas del grupo formado por las SEQ ID Nº:3, SEQ ID Nº:5, SEQ ID Nº·7, SEQ ID Nº.9, SEQ ID Nº 30 y SEQ ID Nº 32, cada una de las cuales codifica, respectivamente, los péptidos que se muestran en las SEQ ID Nº·4, SEQ ID Nº·6, SEQ ID Nº 8, SEQ ID Nº.10, SEQ ID Nº:31 y SEQ ID Nº 33

Otra realización comprende un método de detección de *tic*901 y secuencias de codificación de proteínas relacionadas en cepas de *Bacillus,* que comprende los pasos de cultivar una especies de *Bacillus* por un lapso entre 16 y 45 horas aproximadamente en un caldo rico bajo condiciones aeróbicas, detectar una proteína en el sobrenadante del cultivo que posea identidad antigénica relacionada con TIC901 y/o reactividad cruzada con un anticuerpo que se una específicamente a una o más péptidos o antígenos de la proteína TIC901, identificar y purificar la secuencia de nucleótidos que codifica la proteína detectada, expresar la proteína a partir de la secuencia de nucleótidos, y demostrar la actividad insecticida usando la proteína expresada.

También se contemplan conjuntos de elementos para detectar la presencia de las secuencias de nucleótidos de la presente invención, así como sondas,

cebadores, análogos y derivados de las mismas. Dichos conjuntos de elementos contienen una o más secuencias de nucleótidos, cada una de las cuales se puede usar como una sonda para detectar la presencia de una secuencia de nucleótidos que codifica una proteína insecticida de la presente invención, o un fragmento de la misma, o secuencias de nucleótidos relacionadas, o se puede usar en combinación con una o más sondas o cebadores distintos incluidos en dicho conjunto de elementos para amplificar una o más secuencias de la presente invención o una secuencia de nucleótidos relacionada. Dichos conjuntos de elementos también podrían, o como alternativa podrían, contener un anticuerpo específico que se unirá con uno o más péptidos o proteínas de la presente invención, así como reactivos para usar con la sonda o el anticuerpo, y dichos conjuntos de elementos también podrían contener muestras control cuyo uso permite asegurar que los nucleótidos o péptidos identificados con la sonda y/o el anticuerpo y los reactivos están funcionando de acuerdo con las instrucciones del proveedor. Todos los reactivos necesarios para llevar a cabo los métodos de identificación de cualquiera de las secuencias de nucleótidos o péptidos estarán envasados juntos en el conjunto de elementos junto con las instrucciones para su uso El conjunto de elementos ilustrativo va a contener una secuencia de nucleótidos derivada de la TIC901, TIC1201, TIC407, TIC417, y/o TIC431 secuencia de codificación junto con una muestra de cebadores de amplificación de la secuencia de nucleótidos, por ejemplo, la que se muestra en la SEQ ID Nº:11 y la SEQ ID Nº:12 o diversas combinaciones de las SEQ ID Nº.23-26 y de las SEQ ID Nº:27-29 junto con los reactivos necesarios para una reacción de amplificación, todos envasados juntos en dicho conjunto de elementos.

Por ello se contempla que las composiciones y los métodos descritos en la presente invención proporcionarán muchas ventajas con respecto al arte anterior, incluyendo los que específicamente se definieron precedentemente Además, la presente invención provee una clase completamente nueva de proteínas insecticidas

y las secuencias de nucleótidos que codifican dichas proteínas que no eran conocidas previamente en el arte.

DESCRIPCIÓN BREVE DE LAS FIGURAS

La Figura 1 muestra la alineación de secuencias de aminoácidos de las proteínas precursoras TIC901p, TIC407p, TIC417p, TIC1201p y TIC431p; cada secuencia de aminoácidos contiene una secuencia predicha de treinta (30) aminoácidos amino-terminal característica de un péptido señal de tipo II seguida de trece aminoácidos adicionales desde la posición del aminoácido treinta y uno (31) hasta la posición de aminoácido cuarenta y tres (43) de las respectivas secuencias, que no están presentes en la proteína madura aislada del medio de fermentación utilizado El aminoácido subrayado en la posición 44 de la secuencia consenso representa el aminoácido amino terminal de la proteína madura. La secuencia de nucleótidos nativa que codifica los correspondientes aminoácidos sombreados en las posiciones 75-83, 147-153 y 275-283 se usó como base para construir las sondas y los cebadores de nucleótidos redundantes empleados para identificar las secuencias que codifican estas y otras proteínas insecticidas relacionadas a partir de secuencias de nucleótidos de distintas especies de *Bacillus*.

DESCRIPCIÓN BREVE DE LAS SECUENCIAS

La SEQ ID Nº. 1 representa la secuencia de aminoácidos deducida a partir de los resultados de una degradación de Edmund de una proteína insecticida purificada sobre gel de aproximadamente 38 kDa secretada hacia el medio por células de la cepa EG2158 de *B. thuringiensis* y corresponde sustancialmente a la secuencia de aminoácidos que se muestra en la SEQ ID Nº:4 entre las posiciones de aminoácidos 44 a 58.

La SEQ ID Nº: 2 representa una secuencia de nucleótidos sintética de utilidad como sonda para detectar la secuencia *tic*901 o una secuencia de nucleótidos relacionada, o de utilidad como uno de un par de cebadores de amplificación térmica para amplificar todo o una parte de una secuencia *tic*901 o una secuencia de

nucleótidos relacionada, y corresponde a tripletes de codones preferidos para su uso con *B. thuringiensis* y otras especies de *Bacillus*, en particular, los que presentan una utilización de codones que tienden a contener A o T en la posición del tercer par de bases en cada codón

La SEQ ID Nº· 3 representa una secuencia de nucleótidos nativa (también denominada de tipo salvaje en la presente) de *Bacillus thuringiensis* que codifica una proteína TIC901. Hay una caja Pribnow o secuencia de Shine & Dalgarno predicha ubicada aproximadamente en los nucleótidos 141-147 El ORF predicho que codifica a la proteína precursora TIC901 corresponde a los nucleótidos desde la posición 153 hasta la posición 1.253. Los nucleótidos de la posición 282-325 corresponden sustancialmente a la secuencia de la sonda de oligonucleótidos que se muestra en las SEQ ID Nº:2, que se hibridiza con el complemento de los nucleótidos 282-325 que se muestra en la SEQ ID Nº.3. El codón GTA de valina en la posiciones de nucleótidos 282-284 corresponde al aminoácido amino-terminal en la forma secretada de TIC901.

La SEQ ID Nº. 4 representa una secuencia de aminoácidos de TIC901 de 367 residuos deducida del marco de lectura abierto que se muestra en la SEQ ID Nº.3 desde la posición de nucleótido 153 hasta la posición de nucleótido 1.253. La secuencia de aminoácidos de longitud completa de 367 residuos corresponde a la secuencia de aminoácidos predicha para la proteína precursora expresada a partir de la secuencia de codificación nativa/de tipo salvaje en *B. thuringiensis* La secuencia de aminoácidos entre los números de residuos 1 a 30 que se muestra en la SEQ ID Nº·4 corresponde al péptido señal amino-terminal o al péptido señal secretor predicho que es producido en *B. thuringiensis* a partir de la expresión de la secuencia de nucleótidos que se muestra en la SEQ ID Nº·3, seguida de trece (13) aminoácidos que no están presentes en la forma madura/secretada de la secuencia de la proteína madura insecticida de 324 residuos de aminoácidos cuando se expresa en *B. thuringiensis*. La secuencia de la proteína madura insecticida de 324 residuos de

aminoácidos corresponde a la secuencia de la proteína madura y secretada eficaz como insecticida TIC901. Se prevé que la secuencia de aminoácidos amino-terminal de 43 residuos sea clivada proteolíticamente de la proteína precursora, ya sea en parte durante la translocación a través de la membrana citoplasmática bacteriana o en parte por una señal peptidasa aún no definida u otra proteasa que reconozca la secuencia consenso que comprende la secuencia de aminoácidos con los residuos XAA1-XAA2-GLN inmediatamente antes del punto de ruptura sésil, liberando la proteína madura insecticida hacia el medio extracelular, donde XAA1 corresponde a serina (SER), lisina (LYS) o asparagina (ASN) y XAA2 corresponde a glutamato (GLU) o glutamina (GLN).

La SEQ ID Nº:5 representa una secuencia de nucleótidos nativa de *B thuringiensis* que codifica una proteína insecticida denominada TIC1201 en la presente. La secuencia incluye 529 nucleótidos hacia el extremo 5' de la secuencia del codón de inicio ATG predicho ubicado en las posiciones de nucleótido 530-532. Hay una caja consenso Pribnow o una secuencia de Shine & Dalgarno predicha ubicada 5' con respecto al predicho codón de inicio ATG desde la posición de nucleótido 518 hasta la posición 524. El marco de lectura abierto que codifica la proteína precursora TIC1201 predicha, como TIC901, comprende una secuencia de aminoácidos amino-terminal correspondiente al péptido señal o péptido de direccionamiento de la secreción predicho Se prevé que la secuencia que codifica al péptido señal TIC1201 codifica treinta (30) aminoácidos, seguido de trece (13) aminoácidos adicionales que no están presentes en la forma madura/secretada de la proteína insecticida. Estos trece aminoácidos adicionales terminan en una secuencia que codifica la secuencia de reconocimiento de peptidasa SER-GLN-GLN idéntica a la secuencia presente en la secuencia de la proteína precursora TIC901. Se predijo a partir de la secuencia de codificación que la secuencia de aminoácidos de la proteína insecticida TIC1201 liberada hacia el medio de las cepas de *B thuringiensis* que expresan esta secuencia comprende 321 residuos de aminoácidos, que son

codificados por la secuencia de nucleótidos en las posiciones 659 a 1621, como se muestra en la SEQ ID Nº.5. Aunque el ORF predicho que codifica la proteína precursora TIC1201 se identifica en la presente entre los nucleótidos 530-1621, el ORF se podría extender posiblemente entre los nucleótidos 437 y 1621 Se considera que esto es poco probable porque la similitud del péptido señal con el de la TIC901, y la falta de cualquier caja Pribnow o secuencia de Shine & Dalgarno consenso en una proximidad razonable a cualquier codón de inicio ATG ubicado 5' con respecto al ubicado en los nucleótidos 518 a 524

La SEQ ID Nº:6 representa la secuencia de 395 aminoácidos deducida de la proteína precursora TIC1201 codificada por la secuencia de nucleótidos entre los nucleótidos 530 y 1621 que se muestra en la SEQ ID Nº 5

La SEQ ID Nº.7 representa la secuencia de nucleótidos que codifica la proteína insecticida denominada TIC407 en la presente

La SEQ ID Nº:8 representa la secuencia de 368 aminoácidos deducida para TIC407 codificada por la secuencia de nucleótidos entre las posiciones de nucleótidos 196 a 1299 que se muestra en la SEQ ID Nº·7.

La SEQ ID Nº.9 representa la secuencia de nucleótidos que codifica la proteína insecticida denominada TIC417 en la presente.

La SEQ ID Nº.10 representa la secuencia de 364 aminoácidos deducida para TIC417 codificada por la secuencia de nucleótidos entre las posiciones de nucleótidos 92 y 1173 que se muestra en la SEQ ID Nº.9.

La SEQ ID Nº 11 representa una secuencia de un cebador directo para una amplificación térmica, o una secuencia de sonda, denominada prJWP139 en la presente, correspondiente a la secuencia de codificación que se muestra en la SEQ ID Nº:3 entre las posiciones de nucleótidos 438 y 458, y correspondiente además a los codones que codifican la secuencia de aminoácidos ASN-ASN-ASN-HIS-GLN-THR-ASN-ARG de la secuencia de aminoácidos de las posiciones 96-103 que se muestra en la SEQ ID Nº.4, que tienden a los codones preferidos para su uso en las

secuencias genéticas derivadas de *Bacillus thuringiensis* u otras cepas de especies de *Bacillus*, en lo cuales los codones contienen A y o T en la tercera posición.

La SEQ ID Nº:12 representa la secuencia de un cebador inverso de amplificación térmica, o la secuencia de una sonda, denominada prJWP143 en la presente, correspondiente al complemento inverso de la secuencia de codificación que se muestra en la SEQ ID Nº:3 entre las posiciones de nucleótidos 978 y 998, y correspondiente además a los codones que codifican la secuencia de aminoácidos GLN-LYS-PHE-ILE-TYR-PRO-ASN de la secuencia de aminoácidos de las posiciones 276-282 que se muestra en la SEQ ID Nº 4, que tienden a los codones preferidos para su uso en las secuencias genéticas derivadas de *Bacillus thuringiensis* o de otras cepas de especies de *Bacillus*, en los cuales los codones contienen A y o T en la tercera posición.

La SEQ ID Nº.13 representa la secuencia de nucleótidos sintética, artificial o no natural que codifica una variante de la secuencia de aminoácidos TIC901.

La SEQ ID Nº.14 representa la secuencia de aminoácidos deducida de la secuencia de codificación que se muestra en la SEQ ID Nº:13 entre la posición de nucleótidos 1 y la posición de nucleótidos 1104.

La SEQ ID Nº 15 representa una secuencia de nucleótidos artificial para su uso como sonda o cebador, descrita en la presente como prJWP151

La SEQ ID Nº·16 representa una secuencia de nucleótidos artificial para su uso como sonda o cebador, descrita en la presente como prJWP152.

La SEQ ID Nº.17 representa una secuencia de nucleótidos artificial para su uso como sonda o cebador, descrita en la presente como prJWP186.

La SEQ ID Nº:18 representa una secuencia de nucleótidos artificial para su uso como sonda o cebador, descrita en la presente como prJWP183.

La SEQ ID Nº.19 representa una secuencia de nucleótidos artificial para su uso como sonda o cebador, descrita en la presente como prJWP155.

La SEQ ID Nº:19 representa una secuencia de nucleótidos artificial para su

uso como sonda o cebador, descrita en la presente como prJWP155.

La SEQ ID Nº 20 representa una secuencia de nucleótidos artificial para su uso como sonda o cebador, descrita en la presente como prJWP156.

La SEQ ID Nº 21 representa una secuencia de nucleótidos artificial para su uso como sonda o cebador, descrita en la presente como prJWP168

La SEQ ID Nº:22 representa una secuencia de nucleótidos artificial para su uso como sonda o cebador, descrita en la presente como prJWP170.

La SEQ ID Nº.23 representa una secuencia de oligonucleótidos artificial degenerada para su uso como sonda o cebador, descrita en la presente como prJWP200

La SEQ ID Nº·24 representa una secuencia de oligonucleótidos artificial degenerada para su uso como sonda o cebador, descrita en la presente como prJWP201.

La SEQ ID Nº.25 representa una secuencia de oligonucleótidos artificial degenerada para su uso como sonda o cebador, descrita en la presente como prJWP202.

La SEQ ID Nº·26 representa una secuencia de oligonucleótidos artificial degenerada para su uso como sonda o cebador, descrita en la presente como prJWP203.

La SEQ ID Nº.27 representa una secuencia de oligonucleótidos artificial degenerada para su uso como sonda o cebador, descrita en la presente como prJWP204.

La SEQ ID Nº:28 representa una secuencia de oligonucleótidos artificial degenerada para su uso como sonda o cebador, descrita en la presente como prJWP205.

La SEQ ID Nº·29 representa una secuencia de oligonucleótidos artificial degenerada para su uso como sonda o cebador, descrita en la presente como prJWP206.

La SEQ ID Nº:30 representa un fragmento de una secuencia de nucleótidos de codificación derivado de la amplificación térmica del genoma de EG2158 con oligonucleótidos prJWP200 y prJWP204.

La SEQ ID Nº·31 representa la secuencia de aminoácidos del marco de lectura abierto primario que se muestra en la SEQ ID Nº·30.

La SEQ ID Nº.32 representa la secuencia de nucleótidos que codifica la proteína insecticida denominada TIC431 en la presente

La SEQ ID Nº 33 representa la secuencia de 364 aminoácidos deducida para TIC431 codificada por la secuencia de nucleótidos entre las posiciones de nucleótidos 1 y 1092 que se muestra en la SEQ ID Nº·32

DESCRIPCIÓN DETALLADA DE LA INVENCIÓN

De acuerdo con la presente invención, se ha descubierto un nuevo grupo de secuencias de nucleótidos que codifica un nuevo grupo de proteínas insecticidas derivadas de *Bacillus thuringiensis* y cepas relacionadas de *Bacillus.* Tal como se define en otra parte en la presente, estas secuencias de codificación de nucleótidos se hibridizan entre sí bajo las condiciones de hibridación apropiadas y las proteínas codificadas por estas secuencias de nucleótidos reacciones de forma cruzada con el antisuero generado contra cualquiera de las demás proteínas. La alineación de las secuencias de nucleótidos que codifican las formas maduras/secretadas de las proteínas TIC1201, TIC901, TIC407, TIC417 y TIC431 revela que la secuencia de codificación que codifica un fragmento de la proteína madura TIC901 (forma secretada de TIC901 desde la posición de aminoácido 44 hasta la posición de aminoácido 367 que se muestra en la SEQ ID Nº:4) es entre un 79 y un 91 por ciento idéntica aproximadamente a cada una de las secuencias que codifican los otros tres fragmentos de proteína madura descritos en la presente; la secuencia que codifica un fragmento de la proteína madura/secretada TIC417 predicha (posición de aminoácido 44 a la posición de aminoácido 364 que se muestra en la SEQ ID Nº:10) es entre un 75 y un 95 por ciento idéntica aproximadamente a cada una de las

secuencias que codifican los otros tres fragmentos de proteína madura descritas en la presente; la secuencia que codifica un fragmento de la proteína madura/secretada TIC407 predicha (posición de aminoácido 44 a la posición de aminoácido 368 que se muestra en la SEQ ID Nº:8) es entre un 75 y un 82 por ciento idéntica aproximadamente a cada una de las secuencias que codifican los otros tres fragmentos de la proteína madura descrita en la presente; la secuencia que codifica el fragmento maduro/secretado TIC1201 (posición de aminoácido 44 a la posición de aminoácido 395 que se muestra en la SEQ ID Nº:6) es entre un 80 y un 91 por ciento idéntica aproximadamente a cada una de las secuencias que codifican los otros tres fragmentos de la proteína madura descrita en la presente y la secuencia que codifica un fragmento de la proteína madura/secretada TIC431 predicha (posición de aminoácido 44 a la posición de aminoácido 364 que se muestra en la SEQ ID Nº:8) es entre un 75 y un 82 por ciento idéntica aproximadamente a cada una de las secuencias que codifican los otros tres fragmentos de la proteína madura descrita en la presente. Las proteínas codificadas por cada una de estas secuencias de nucleótidos codificadoras exhiben actividad biológica inhibidora de especies de coleópteros, exhiben en parte una identidad sustancial de secuencias de aminoácidos y en parte una similitud sustancial de secuencias de aminoácidos, y por ello se considera que son proteínas insecticidas relacionadas. La secuencia de aminoácidos de la forma madura/secretada predicha para la proteína insecticida TIC417 (TIC417m) es aproximadamente un 78,9 por ciento idéntica a la correspondiente secuencia de aminoácidos de la forma madura/secretada para TIC901 (TIC901m). La secuencia de aminoácidos de la forma madura/secretada predicha para la proteína insecticida TIC1201 (TIC1201m) es aproximadamente un 90,1 por ciento idéntica a la correspondiente secuencia de aminoácidos para TIC901m y es aproximadamente un 80,7 por ciento idéntica a la correspondiente secuencia de aminoácidos para TIC417m y TIC431m La secuencia de aminoácidos de la forma madura/secretada predicha para la proteína insecticida TIC407

(TIC407m) es aproximadamente un 80% idéntica a la correspondiente a TIC901m, aproximadamente un 75% idéntica a la correspondiente a TIC417m y aproximadamente un 82% idéntica a la correspondiente a TIC1201m. La secuencia de aminoácidos de la forma madura/secretada predicha para la proteína insecticida TIC431 (TIC431m) es un 75% idéntica aproximadamente a la TIC407 madura, un 79% idéntica aproximadamente a la TIC901 madura, un 80% idéntica aproximadamente a la secuencia de aminoácidos de la proteína madura TIC1201 y un 95% idéntica a la secuencia de aminoácidos de la proteína madura TIC417. Cada una de las proteínas codificadas por las secuencias de nucleótidos descritas en la presente se pueden expresar solas en las plantas o como diversas combinaciones entre sí o con otros agentes insecticidas inhibidores coleópteros, tal como proteínas, proteínas cristalinas, δ-endotoxinas, lectinas, patatinas y otras toxinas y semejantes para alcanzar un medio de resistencia al manejo de insectos en el campo que no ha sido factible antes con el mero uso de las proteínas insecticidas para coleópteros conocidas derivadas de las cepas de *Bacillus thuringiensis*, tal como proteínas Cry3, proteínas VIP y/o WAR y/o MIS y diversas proteínas insecticidas inhibidoras de coleópteros derivadas de especies de *Bacillus laterosporous*, especies de *Bacillus sphaericus* y especies de las bacterias *Xenorhabdus* y *Photorhabdus* Las proteínas de la presente invención también se pueden usar en plantas en combinación con otros tipos de agentes insecticidas y/o toxinas insecticidas para obtener plantas transformadas que contengan al menos un medio para controlar una o más de cada una de las plagas de plantas comunes seleccionadas de los grupos formados por plagas de insectos coleópteros, plagas de insectos lepidópteros, plagas de insectos perforadores y chupadores y semejantes. Otras proteínas y/o agentes para el control de insectos que se pueden expresar en una planta en combinación con las proteínas de la presente invención incluyen, a modo de ejemplo, proteínas insecticidas para lepidópteros de especies de *Bacillus*, tales como las proteínas Cry derivadas de las especies *Bacillus thuringiensis*, *Bacillus laterosporous* y *Bacillus sphaericus*, y las

proteínas WAR, MIS y/o VIP que se pueden aislar de diversas especies de *Bacillus*, proteínas insecticidas derivadas de las especies bacterianas *Xenorhabdus* y *Photorhabdus* y composiciones tales como ARNcd dirigidos expresamente a la supresión de uno o más genes de una o más plagas de insectos. Tal como se utiliza en la presente, un "polipéptido insecticida" o una "proteína insecticida" o un "fragmento insecticida de la misma" se refiere a un polipéptido que presenta propiedades insecticidas, por ejemplo, un polipéptido que inhibe el crecimiento, desarrollo, viabilidad o fecundidad de las plagas de insectos de interés y un agente insecticida que incluye a todas ellas así como a los ARN de cadena doble dirigidos a la supresión de uno o más genes de una o más plagas blanco.

Sorpresivamente, las proteínas de la presente invención no parecen estar relacionadas con ninguna de las proteínas insecticidas de *Bacillus thuringiensis* descubiertas hasta el presente en el arte. Se demuestra aquí que las proteínas de la presente invención son excretadas hacia el espacio extracelular que rodea a las especies de *Bacillus* de las cuales derivan. Se demuestra aquí que estas proteínas son significativamente menores que las proteínas Cry, VIP, WAR y MIS conocidas previamente en el arte y que se pueden expresar durante la etapa de crecimiento vegetativo de cultivos aislados y purificados de células bacterianas Esto se diferencia de la expresión de las proteínas Cry que en general se expresan en la fase de crecimiento de esporulación y que forma diversos cuerpos cristalinos dentro de la endospora de la célula

Como será evidente para los especialistas en el arte, los inventores describen aquí el aislamiento y la purificación de una secuencia de nucleótidos, *tic*901, que codifica una proteína precursora TIC901 (TIC901p) que posteriormente es procesada proteolíticamente para liberar una proteína madura TIC901 (TIC901m) que presenta actividad biológica inhibidora de especies de coleópteros. Los inventores de la presente describen el uso de la secuencia *tic*901 como un medio para identificar una pluralidad de otras secuencias relacionadas, cada una de las cuales también codifica

proteínas insecticidas relacionadas con TIC901, TIC901p y TIC901m, y describen el uso de anticuerpos generados contra TIC901m en un método con ELISA para detectar cepas de Bt que producen proteínas relacionadas con TIC901 expresadas a partir de las secuencias de codificación relacionadas con la que codifica TIC901.

Las secuencias de nucleótidos descritas en la presente y que codifican TIC901 fueron derivadas de cepas de *Bacillus thuringiensis,* incluyendo las cepas EG2158, EG6489, EG6561, EG12450 y EG4653 Estas cepas se han depositado bajo las disposiciones del Tratado de Budapest con la colección permanente de *Agricultural Research Service Culture Collection, Northern Regional Research Laboratory* (NRRL), Departamento de Agricultura de los EE.UU. (USDA), 1815 North University Street, Peoria, IL 61604. Las cepas relevantes se depositaron en NRRL entre el 29 de abril, 1987, y el 6 de febrero, 2002. La cepa EG2158 de *B. thuringiensis* recibió el Nº de Acceso NRRL B-18213, y la cepa EG12450 de *B. thuringiensis* recibió el Nº de Acceso NRRL B-30357. Las secuencias de nucleótidos relacionadas con *tic*901, y las secuencias de aminoácidos relacionadas con TIC901 (incluyendo las especies precursoras y maduras de TIC901) que se describen en la presente incluyen, sin limitaciones, *tic*1201 y la proteína insecticida codificada por la misma TIC1201 aislada y producida al menos por las cepas EG3618 y 86833 de *B.t.*, *tic*407 y la proteína insecticida codificada por la misma TIC407 aislada y producida al menos por la cepa EG6618 de *B.t.* y *tic*417 y la proteína insecticida codificada por la misma TIC417 aislada y producida al menos por las cepas EG2158, EG6489 y EG6561, y TIC431 codificada por al menos la cepa EG4653 de *B.t.*.

Se evaluó la actividad insecticida de un caldo de cultivo derivado de la cepa EG2158 de *B.t* purificada y se determinó que presentaba actividad biológica inhibidora de insectos coleópteros dirigida contra el escarabajo de papa de Colorado (CPB) Se purificó una proteína que presentaba una masa de aproximadamente 38 kDa según un análisis por SDS-PAGE a partir del caldo de cultivo y se observó que contenía la toxicidad indicada para coleópteros. Esta proteína fue sometida a

degradación de Edmund automatizada, y como resultado de la misma se obtuvo una secuencia de aminoácidos que se consideró que era la secuencia de 15 aminoácidos amino-terminales (SEQ ID Nº·1) de la proteína de aproximadamente 38 kDa, es decir, la proteína TIC901m. Se construyó una secuencia de oligonucleótidos sintética semi-redundante (WD444; SEQ ID Nº.2) correspondiente a los codones naturales preferidos para su uso en las secuencias de codificación aisladas de *Bacillus thuringiensis* o de especies bacterianas relacionadas con *Bacillus*, es decir, que presentaban una preferencia por A o T en la tercera posición de cada codón, con el fin de utilizarla como sonda para detectar secuencias homólogas que posiblemente codificaran la proteína insecticida de aproximadamente 38 kDa en *Bacillus thuringiensis*. Se construyó una biblioteca de secuencias de nucleótidos a partir del ADN purificado a partir de la cepa EG2158 de *Bt*. El ADN purificado fue digerido por completo con *Hind*III y los fragmentos se insertaron en el plásmido vector pUC18 digerido con *Hind*III, tratado con fosfatasa de intestino bovino, para construir una biblioteca genómica con la cepa EG2158 de *Bt*. La biblioteca de ADN se transformó en una cepa de *E. coli* y la mezcla de transformación se plaqueó sobre un medio de selección sólido. Las colonias que aparecieron fueron sometidas a pruebas con sondas utilizando una muestra de la sonda con la secuencia de nucleótidos sintética conjugada con fosfatasa alcalina WD444 (SEQ ID Nº:2) Una cepa recombinante de *E. coli* denominada EG12447, que contiene al plásmido pEG1379 (también denominado pMON74007), un derivado del pUC18 que contenía un fragmento *Hind*III de 8 kb aislado de la cepa EG2158 de *B. thuringiensis*, se hibridizó con la sonda de oligonucleótidos conjugada con fosfatasa alcalina Se determinó por análisis de la secuencia de nucleótidos que el fragmento *Hind*III de 8 kb en el plásmido pEG1379 contenía toda la secuencia de nucleótidos que codifica la proteína TIC901p. Se efectuó un depósito de la cepa EG12447 en NRRL y la muestra depositada recibió el Nº de Acceso NRRL B-30549 el 6 de febrero, 2002

Se determinó la secuencia de nucleótidos del marco de lectura abierto que

codifica la proteína insecticida de aproximadamente 38 kDa. El marco de lectura abierto que codifica la proteína TIC901p se denominó *tic*901 El marco de lectura abierto consiste de una secuencia de nucleótidos de 1101 nucleótidos (los nucleótidos 153-1.253 que se muestran en la SEQ ID Nº:3) y se prevé que codifica una proteína precursora formada por 367 aminoácidos (SEQ ID Nº·4). El peso molecular predicho de la secuencia de aminoácidos deducida a partir del marco de lectura abierto es de 41.492 Daltons, un valor que se encuentra dentro de las expectativas razonables para la masa de la proteína secretada estimada mediante análisis por SDS-PAGE de aproximadamente 38 kDa teniendo en cuenta la pérdida de masa de un péptido señal que es de tres a cuatro kDa Se ha predicho que la proteína precursora (TIC901p) tiene un punto isoeléctrico de 6,368 y una carga neta de –2,102 a pH 7,0. La composición total de la secuencia de nucleótidos que codifica la proteína precursora comprende un 69% de AT, lo cual es coherente con otras secuencias de codificación identificadas en *B thuringiensis* y en otras cepas de *Bacillus* La composición de nucleótidos de la secuencia de codificación también es coherente con otros genes caracterizados a partir de especies de *Bacillus*, que contiene aproximadamente un 39% de adenosina, aproximadamente un 18% de guanosina, aproximadamente un 30% de timidina y aproximadamente un 13% de citosina.

La secuencia de codificación nativa de *tic*901 en el pEG1379 no produjo una cantidad mensurable de la proteína TIC901m a partir de cultivos recombinantes de *E. coli* que contenían a este plásmido Esto no era inesperado dada la falta de funcionalidad conocida de los promotores de *Bacillus* en *E. coli*, y las diferencias en la preferencia de codones entre los dos organismos. No se observó inhibición por CPB con los sobrenadantes de cultivo con la cepa EG12447 o con células que contenían a este plásmido. Se recortó el inserto de 8 kB y se introdujo en un vector ambivalente de *E. coli / B. thuringiensis* para formar el plásmido pEG12450. El pEG12450 se transformó en una cepa acristalífera de *Bacillus thuringiensis*,

EG10650 (Patente de EE UU Nº:6.468.523), para producir la cepa EG12450. Dicha cepa EG10650 derivaba de la cepa EG10368 acristalífera de *B. thunngiensis* (identificada en la Patente de EE.UU. Nº:5.322.687) por sustitución de *npr* y *apr* (genes de proteasas neutras y proteasas ácidas, respectivamente) con los alelos mutantes de supresión de estos dos genes de proteasas, *npr*3 y *apr*1 respectivamente (Patente de EE UU Nº.5 759 538). Los sobrenadantes de cultivos derivados de EG12450 resultaron positivos para la actividad inhibidora del escarabajo de papa de Colorado (CPB) La proteína purificada a partir de los sobrenadantes de cultivo de EG12450 también resultó positiva para la actividad inhibidora de CPB. No se observaron estructuras cristalinas en los cultivos esporulados, pero la biomasa de cultivo de células en pelets/esporuladas también resultó positiva para la bioactividad inhibidora de CPB lo cual sugiere que alguna porción de la proteína TIC901 aún permanecía asociada con las esporas/cultivo o que las esporas consumidas por las especies de insectos de prueba germinaron dentro del insecto y produjeron suficiente proteína insecticida TIC901 como para causar un efecto inhibidor observable

Los sobrenadantes de cultivo y la proteína purificada de la cepa EG12450 también fueron evaluados por su actividad biológica contra gusanos de la raíz del maíz Se observó bioactividad inhibidora para los gusanos de la raíz del maíz del sur y del oeste (*Diabrotica undecimpunctata howardii* y *Diabrotica virgifera virgifera*, respectivamente)

Se examinó una colección variada de cepas de *Bacillus* como se describen los inventores en la presente con el fin de determinar si estas cepas de *B. thuringiensis* y/o *B. sphaericus* también producían proteínas extracelulares relacionadas con TIC901. En particular, se procesó la pasta de células y el medio usado de 279 cepas y se usaron en un bioensayo con larvas de gusanos de la raíz del maíz del sur. Aproximadamente un tercio de las cepas produjo proteínas secretadas hacia el medio de crecimiento que eran positivas para la actividad inhibidora de los gusanos

de la raíz. Se priorizaron treinta y seis (36) cepas que produjeron las proteínas extracelulares con mayor inhibición eficaz para los gusanos de la raíz. Estas cepas fueron examinadas adicionalmente para determinar si la secuencia de codificación TIC901 se hibridizaba con las secuencias presentes en el genoma de cada una de las cepas y comparar los resultados para estas cepas con los resultados obtenidos en las pruebas con sonda con la cepa EG2158 que contenía la secuencia de codificación nativa TIC901. Los resultados se muestran en la Tabla 1.

Tabla 1. Isotipos RFLP de las cepas de *Bacillus* productoras de proteínas insecticidas secretadas contra CRW

Cepa	RFLP tic901[1]	ID del homólogo de la proteína TIC
EG2158[2]	A	901/417
EG2211	B	*1201*
EG2874	B	*1201*
EG2904	B	*1201*
EG3109	B	*1201*
EG 3111	B	*1201*
EG 3116	B	*1201*
EG 3117	B	*1201*
EG 3119	B	*1201*
EG 3120	B	*1201*
EG 3171	B	*1201*
EG 3173	B	*1201*
EG 3177	B	*1201*
EG 3354	B	*1201*
EG 3458	B	*1201*
EG 3461	B	*1201*

EG 3618	B	1201[3]
EG 3619	B	*1201*
EG 3620	B	*1201*
EG 3753	B	*1201*
EG 3787	B	*1201*
EG 4189	B	*1201*
EG 4191	B	*1201*
EG 4193	B	*1201*
EG 4332[4]	-	
EG 4834	B	*1201*
EG 5194	B	*1201*
EG 5552	-	
EG 5858[4]	-	
EG 6489	A	901/417
EG 6555	B	*1201*[3]
EG 6561	A	901/417
EG 6564	B	*1201*
EG 6618	C	407[3]
EG 6890	B	*1201*
EG 10650	-	
86833	B	1201

1- El ADN total aislado de cada cepa fue digerido por completo con *Hind*III y analizado por transferencia Southern Las letras A, B, C o D corresponden al tamaño de un fragmento de restricción particular (polimorfismo) obtenido con tic901 o la sonda de la SEQ ID Nº:2.

La cepa EG2158 también contiene una tercera secuencia de nucleótidos homóloga relacionada con *tic*901, *tic*1201, *tic*407 y *tic*417 que se muestra en la SEQ

ID Nº:30.

3- Solamente indica la identificación de la secuencia de nucleótidos parcial obtenida, pero toda o una porción de la secuencia obtenida codifica una secuencia de proteína que es un 100% idéntica a la proteína TIC indicada

4- Indica que se ha demostrado que estas cepas producen sustancias extracelulares que presentan actividad biológica inhibidora de Lygus

Cuatro (4) de las treinta y seis (36) cepas no produjeron un fragmento de nucleótidos que se hibridizaba con las sondas VIP1 o VIP2 (cepas EG4332, EG5858, EG6618 y 86833) bajo condiciones de gran severidad, es decir, lavados con SSC 0,5 X y a 65 °C. Todas las demás cepas produjeron un fragmento de nucleótidos que se hibridiza con la sonda TIC901, sin embargo, se observaron variaciones interesantes en la intensidad de las señal de hibridación, específicamente con referencia a uno o más fragmentos de restricción *Hind*III obtenidos con la sonda TIC901 Esencialmente se trataba de tres fragmentos de restricción de distinto tamaño (polimorfismos) identificados por hibridación con la sonda *tic*901 Primero se creyó que solamente uno de los tres fragmentos diferentes que se hibridizaban bajo estas condiciones con la sonda *tic*901 estaba presente en cualquiera de las cepas Dos polimorfismos de fragmento de restricción presentaron una intensidad de señal diferente cuando se hibridizaba con la sonda *tic*901 en comparación con la señal producida por hibridación de esta sonda con el fragmento del gen *tic*901 de la cepa EG2158. Este resultado sugiere que existen al menos tres alelos de la secuencia de codificación *tic*901 en estas treinta y seis (36) cepas. El análisis de secuencia, descrito en la presente, de cada una de estos tres polimorfismos de longitud de fragmentos de restricción ha permitido identificar cada uno de estos ORF (marcos de lectura abiertos) relacionados que codifican una proteína TIC901 relacionada contenida dentro de cada secuencia. Se ha llevado a cabo una comparación de estas secuencias de nucleótidos, descritas en la presente, por alineación de las secuencias con la secuencia de codificación nativa TIC901 para determinar el grado de

identidad. Además, se evaluaron las propiedades insecticidas de las proteínas codificadas por los ORF TIC901 y TIC1201 identificados en estos fragmentos de restricción y cada una presenta toxicidad para las plagas de coleópteros. Por ello se cree que las proteínas TIC407 y TIC417 también exhibirán actividad biológica insecticida para coleópteros basado en su alto grado de relación con las proteínas TIC901 y TIC1201. Como consecuencia de la relación significativa de identidad de secuencias de aminoácidos entre las proteínas TIC407, TIC417 y TIC1201 en comparación con TIC901, las proteínas se describen en la presente como variantes de dichas secuencias de aminoácidos. Por ejemplo, TIC1201 contiene tres aminoácidos menos que TIC901 y contiene 31 variaciones en la secuencia de aminoácidos en comparación con TIC901. Por ello, cuando se compara con TIC901, las otras secuencias de aminoácidos TIC contienen variaciones de aminoácido que pueden contribuir en las diferencias en cuanto a espectro insecticida y/o virulencia y potencia. El posterior análisis de estas y otras cepas usando métodos de amplificación térmica con pares de cebadores que han incorporado una degeneración en sus secuencias basada en las alineaciones de secuencias de nucleótidos de las secuencias de codificación para las proteínas TIC901, TIC407, TIC1201 y TIC417, dio como resultado amplicones que podrían corresponder a la secuencia de codificación TIC407 presente en el genoma de las cepas EG5858, EG5552 y EG4332 Posiblemente, estas secuencias no aparecieron con una transferencia específica de TIC901 porque se determinó que la secuencia TIC407 se encontraba sobre un fragmento *Hind*III grande, de aproximadamente 18-19 kb, que no había sido transferido eficazmente a la membrana de transferencia.

Se pretende que las proteínas de la presente invención se utilicen para fines agronómicos, es decir, para proteger a las plantas de una infestación por plagas de insectos y, más particularmente, para proteger a las plantas de una infestación con plagas de insectos coleópteros. Tal como se muestra aquí a modo de ejemplo, las proteínas de la presente invención son de utilidad para proteger a las plantas por lo

menos contra una infestación con el escarabajo de papa de Colorado y por lo menos contra una infestación con gusanos de la raíz del maíz La protección de las plantas se puede lograr con la aplicación tópica a una planta o a partes de la planta, tal como por aplicación a la superficie de la planta, es decir, las hojas, flores, tallos, ramas y raíces, de una composición en la forma de un polvillo, aerosol, polvo o emulsión u otro excipiente aceptable para uso agronómico que contenga una cantidad eficaz como insecticida de una o más de las proteínas de la presente invención. Como alternativa, dicho excipiente agronómico puede contener, además de una o más de las proteínas de la presente invención, una o más proteínas insecticidas adicionales eficaces para inhibir el mismo espectro de plagas de insectos que serían controladas por las proteínas de la presente invención, tal como proteínas Cry3, CryET33/34, CryET80/76, PS149B1 y otras proteínas de toxinas binarias inhibidores de coleópteros, proteínas VIP/MIS/WAR y semejantes, y/o proteínas que son eficaces para controlar un espectro completamente diferente de plagas de insectos en plantas, tal como Cry1's, Cry2's, Cry9's y semejantes. También se considera dentro del alcance de la presente invención que el excipiente agronómico descrito previamente contenga otros tipos de composiciones plaguicidas, tal como fungicidas y/o acaricidas y semejantes. Como alternativa, y preferentemente, la propia planta será transformada para contener una o más secuencias de nucleótidos modificadas para mejorar la expresión de una o más de las proteínas de la presente invención *in planta* o la expresión de una porción insecticida de las mismas, ya sea solas o en combinación con otros agentes insecticidas, tales como los que se pueden producir *in planta* usando métodos de biología molecular, incluyendo métodos mediados por ARN de cadena doble para la supresión de genes en las células de la plaga de interés.

Las proteínas TIC901, TIC1201, TIC407, TIC417 y TIC431 son compuestos insecticidas que son activos contra insectos coleópteros, tal como CPB y gusanos de la raíz Estas proteínas que se muestran en las SEQ ID Nº 4, SEQ ID Nº 6, SEQ ID

Nº:8, SEQ ID Nº.10 y SEQ ID Nº·33, respectivamente, y fragmentos insecticidas de las mismas, y las proteínas insecticidas relacionadas se pueden utilizar como ingrediente activo en las formulaciones insecticidas útiles para controlar insectos coleópteros. Tal como se utiliza en la presente y con referencia a las proteínas insecticidas que están relacionadas con estas proteínas, se considera que las proteínas insecticidas relacionadas son aquellas que se identifican como homólogos de estas proteínas o aquellas que se identifican como codificadas por una secuencia de nucleótidos que se hibridiza ya sea bajo condiciones de hibridación severas o bajo condiciones de hibridación específicas a toda o una parte de la secuencia nativa de *Bacillus thuringiensis* que codifica la proteína TIC901, la proteína TIC1201, la proteína TIC417, la proteína TIC407, la proteína TIC431 o una porción insecticida de las mismas. Las condiciones de severidad, definidas en la presente, comprenden al menos una hibridación a 42 °C seguido de dos lavados de cinco minutos cada uno a temperatura ambiente con SSC 2X, SDS 0,1%, seguido de dos lavados de treinta minutos cada uno a 65 °C en SSC 0,5X, SDS 0,1%. Por supuesto, el especialista en el arte comprenderá que, debido a la redundancia del código genético, hay muchas otras secuencias que pueden codificar dichas proteínas relacionadas, y se considera que dichas secuencias, hasta el punto que funcionen en la expresión de proteínas insecticidas ya sea en cepas de *Bacillus* o en células vegetales, están comprendidas por la presente invención, teniendo en cuenta, por supuesto, que muchas de dichas secuencias de codificación redundantes no se hibridizarán bajo estas condiciones con las secuencias nativas que codifican TIC901, TIC1201, TIC407, TIC417 y/o TIC431 Se debe tener en cuenta que cuando se hace referencia a una TIC901 o una proteína insecticida relacionada o un fragmento insecticida de la misma, o cuando se hace referencia a una secuencia de nucleótidos que codifica una TIC901 o una proteína insecticida relacionada o un fragmento insecticida de la misma, que la referencia a TIC901 es utiliza indistintamente de la referencia a TIC407, TIC417, TIC431 y TIC1201 y semejantes, incluyendo la secuencia de aminoácidos de la SEQ

ID Nº:31 codificada por la secuencia de nucleótidos que se muestra en la SEQ ID Nº 30, que incluye la proteína insecticida de longitud completa codificada por la secuencia de codificación de longitud completa de la proteína, una parte de la cual se muestra a modo de ejemplo en la secuencia de aminoácidos de la SEQ ID Nº:31.

Algunas secuencias de nucleótidos que se pueden relacionar con las secuencias de nucleótidos de la presente invención no se hibridizan bajo condiciones severas, pero de hecho se pueden hibridizar con una secuencia *tic*901, *tic*1201, *tic*407, *tic*431 y/o *tic*417, o una secuencia relacionada usando condiciones de hibridación específicas Dichas secuencias pueden codificar una proteína que tiene al menos un 30% de identidad de secuencia de aminoácidos aproximadamente con las proteínas de la presente invención Las proteínas que presentan al menos un 30% aproximadamente de identidad de secuencia también pueden presentar estructuras terciarias muy similares y asimismo también pueden presentar una actividad biológica similar o relacionada. Con referencia a la presente invención, dicha similitud en cuanto a la estructura terciaria incluirá la actividad biológica insecticida. Las condiciones de hibridación específicas que permiten identificar secuencias de nucleótidos menos relacionadas incluyen una primera hibridación a una temperatura baja, típicamente de 40 °C aproximadamente o semejante, seguido de los lavados indicados previamente a temperatura ambiente para eliminar la sonda que no se unió específicamente, seguido de exposición a una película (en los casos en que el investigador utilice medios de marcación isotópica) o exposición a reactivos inmunológicos y a reactivos de revelado químico para identificar los fragmentos de nucleótidos que se hibridizan con la sonda de un gen específico Una indicación que la hibridación no es específica es cuando se observan muchos fragmentos de hibridación. En los casos en que se observa una cantidad de fragmentos de hibridación, el transferido se lava una o más veces, cada vez a una temperatura ligeramente superior a la del lavado previo (por ejemplo, cada lavado se podría llevar a cabo a una temperatura de aproximadamente 5 °C más que el lavado previo) hasta

que solamente se observen uno o unos pocos (dos o tres) fragmentos de hibridación, donde dicho uno o unos pocos fragmentos presentan una hibridación específica y pueden entonces ser clonados y secuenciados para determinar el grado de homología y/o identidad con la secuencia de la sonda original. Dichas secuencias estarán relacionadas específicamente por cuanto codifican proteínas que tienen una función relacionada, por ejemplo, una actividad insecticida, con la proteína codificada por la secuencia original derivada de la sonda de nucleótidos.

Se consideran secuencias de codificación las que codifican todo o una porción insecticida de la TIC901 o de una proteína relacionada que no se hibridiza bajo condiciones severas Sin embargo, dichas secuencias son derivadas de la secuencia de nucleótidos nativa de modo tal que la secuencia de nucleótidos nativa puede ser modificada para presentar una secuencia no nativa que aún codifique la misma o sustancialmente la misma secuencia de aminoácidos nativa, o que la secuencia de aminoácidos nativa se pueda usar junto con una tabla de codones para volver a traducir la secuencia de aminoácidos, permitiendo al especialista obtener una secuencia de nucleótidos que codifica todo o una porción insecticida de una TIC901 o una proteína relacionada. Todas dichas secuencias se consideran dentro del alcance de la presente invención.

Se pueden producir composiciones insecticidas a partir de cepas bacterianas que expresan las proteínas de la presente invención. Se puede cultivar una cepa de *B thuringiensis* que contiene una o más secuencias de nucleótidos codificadoras de una o más proteínas TIC901 o proteínas relacionadas y/o equivalentes sustanciales de la misma, usando medios y técnicas de fermentación estándar conocidos Dado que las proteínas de la presente invención son secretadas preferentemente hacia el medio extracelular, una vez completado el ciclo de fermentación, las bacterias que expresan TIC901 o un homólogo de la misma, pueden ser cosechadas separando primero la *B. thuringiensis* junto con todas las esporas y cristales producidas por la misma, del caldo de fermentación usado empleando medios bien conocidos en el

arte. Las esporas y cristales recuperadas de *B. thuringiensis* pueden ser formuladas en un polvo humectable, un concentrado líquido, gránulos, solución, emulsión, aerosol, suspensión, polvo, espuma, pasta, aerosol, cápsula u otro material grueso o finamente dividido o como un impregnante para materiales naturales o sintéticos, u otra formulación, mezclado con vehículos, diluyentes, adyuvantes, conservantes, dispersantes, solventes, agentes emulsionantes, vehículos inertes y otros componentes adecuados para una asociación física o química con plantas o el locus de las mismas, para la captación oral por los patógenos vegetales de interés y para facilitar el manejo y la aplicación de plagas particulares de interés. Los procedimientos de formulación y aplicación son bien conocidos en el arte.

Se pueden formular los gránulos formulados como cebo al entorno de la plaga, donde dichos gránulos contienen una sustancia de atracción y esporas y cristales de las formas aisladas de *B thuringiensis* o medio de fermentación usado concentrado o proteínas insecticidas purificadas a partir de las esporas o del medio de fermentación usado, o microbios recombinantes que comprenden las secuencias de nucleótidos que codifican TIC901 o proteínas insecticidas relacionadas que se pueden obtener de las formas aisladas de *B thuringiensis* descritas en la presente El cebo se puede aplicar generosamente dado que la toxina no afecta a animales o seres humanos. El producto también se puede formular como un aerosol o polvo. Las plagas recogen el producto con las patas o abdomen y lo llevan al nido donde otros individuos serán expuestos a la toxina. La forma aislada de *B. thuringiensis* o el huésped recombinante que expresa la secuencia de nucleótidos o el gen que codifica la TIC901 o una proteína relacionada de la presente invención, también puede ser incorporada en un cebo o fuente de alimento para la plaga.

Los especialistas en el arte comprenderán que la concentración plaguicida varía ampliamente dependiendo de la naturaleza de la formulación particular, en especial si se trata de un concentrado o si se utilizará directamente El plaguicida comprenderá al menos un 1% en peso y puede comprender hasta un 100% en peso.

Las formulaciones secas comprenderán aproximadamente entre un 1-95% en peso de plaguicida, en tanto las formulaciones líquidas generalmente comprenderán aproximadamente un 1-60% en peso de sólidos suspendidos o que se puedan suspender en la fase líquida. Las formulaciones comprenderán en general entre aproximadamente 10^2 y aproximadamente 10^4 células/mg o entre aproximadamente 5 y aproximadamente 100 partes por millón de componente activo de proteína insecticida, es decir, la proteína TIC901, una variante de la secuencia de aminoácidos de la misma, una porción insecticida o un fragmento de la misma, o un homólogo de la misma tal como TIC1201, TIC417, TIC417 y TIC407. Estas formulaciones serán administradas a una dosis que comprende entre aproximadamente 50 mg (líquido o seco) y aproximadamente 1 kg o más por hectárea. Las formulaciones pueden ser aplicadas al entorno de las plagas de coleópteros, por ejemplo, plantas, tierra o agua por rociado, pulverización, aspersión o semejantes, y también se pueden aplicar a las superficies de semillas como un tratamiento para semillas o un recubrimiento de semillas y pueden ser permeadas en el recubrimiento y/o en los cotiledones de las semillas. El especialistas en el arte también tendrán en cuenta que las combinaciones de las proteínas de la presente invención, cuando se combinan juntos en una composición o formulación, pueden tener efectos particularmente benéficos y útiles, por ejemplo, porque proveen un rango de huéspedes más amplio para controlar la infestación por insectos, o porque incrementan la virulencia y potencia de una composición de uso pretendido como agente insecticida.

Es propio de la especialización del arte construir una variante o secuencia de nucleótidos modificada que codifica la proteína insecticida de la presente invención, o un fragmento insecticida de la misma, o una variante de la secuencia de aminoácidos insecticida de la misma con una mayor actividad insecticida en comparación con la secuencia de aminoácidos nativa, e introducir dicha secuencia de nucleótidos en un casete de expresión cuya función en las plantas causa la transcripción de la

secuencia de codificación en un ARN mensajero que posteriormente es traducido en las células de la planta, de modo tal que se produce una cantidad eficaz como insecticida de la proteína insecticida dentro de los tejidos vegetales. También es propio del arte transformar una célula vegetal, preferentemente una célula vegetal de maíz, algodón, soja, canola, arroz, trigo, avena, mijo, pasto, plantas forrajeras, árboles frutales, flores ornamentales, tomate, papa, zanahoria, brócoli y tabaco y semejantes, con una secuencia de nucleótidos incluida en un casete de expresión funcional en plantas, para seleccionar las células que contienen la secuencia y que expresan cantidades eficaces como insecticidas de una proteína TIC901 (y/o una variante de la secuencia de aminoácidos de la misma, una porción insecticida o un fragmento de la misma o un homólogo de la misma, tal como TIC431, TIC1201, TIC417 y TIC407) y/o producir plantas a partir de dichas células transformadas El especialista en el arte está familiarizado con el uso de los métodos de electroporación, infusión, balísticos o los métodos mediados por *Agrobacterium tumefaciens* y semejantes, para introducir las secuencias de nucleótidos de la presente invención, o modificaciones de la misma, en una célula vegetal. Dichos métodos son bien conocidos en el arte.

El término "variante o modificada" con respecto a secuencias de nucleótidos hace referencia a secuencias de nucleótidos que codifican las mismas toxinas o que codifican toxinas equivalentes de actividad insecticida similar, donde el término "toxina equivalente" se refiere a una toxina que presenta la misma, esencialmente la misma o una mayor actividad biológica contra las plagas de interés que la toxina nativa o de referencia reivindicada. Una variante o secuencia de nucleótidos modificada destinada para su uso en plantas dicotiledóneas codificará sustancialmente la misma secuencia de aminoácidos que la secuencia de codificación nativa, es decir, la secuencia de codificación presente en la naturaleza, pero comprenderá una composición total combinada de aproximadamente 49 a aproximadamente 58 por ciento de GC y evitará el uso de los codones menos

preferidos por las plantas dicotiledóneas, determinado compilando dichas frecuencias de preferencia y utilización a partir de un conjunto de secuencias de codificación derivadas de una o más especies vegetales dicotiledóneas individuales que se pretenden transformar con la variante o secuencia de nucleótidos modificada. La variante o secuencia de nucleótidos modificada que se utilizará en una planta monocotiledóneas también codificará sustancialmente la misma secuencia de aminoácidos que la secuencia de codificación nativa, y comprenderá una composición total combinada entre aproximadamente un 52 y aproximadamente un 64 por ciento o más de GC y también evitará el uso de los codones menos preferidos codificadores de cualquiera de los aminoácidos, determinado compilando dichas frecuencias de preferencia y utilización a partir de un conjunto de secuencias de codificación derivadas de una o más especies vegetales monocotiledóneas individuales que se pretenden transformar con la variante o secuencia de nucleótidos modificada La frecuencia de utilización de codones se refiere a la cantidad de veces, en promedio, que se usa un codón particular en una secuencia de codificación Para una especie vegetal particular, el codón que causará la incorporación de un aminoácido particular en una secuencia de aminoácidos naciente se utilizará en promedio con una cierta frecuencia fija relativa. Para los aminoácidos que solamente utilizan dos codones, esta frecuencia es en general de aproximadamente cincuenta-cincuenta, es decir, cada codón es utilizado aproximadamente la mitad de las veces, a menos que uno de los codones emplea una cantidad sustancialmente mayor de purinas o pirimidinas que no son típicamente representativos del contenido de GC de la especie vegetal particular. Para las especies de *Bacillus*, por ejemplo, las secuencias de codificación comprenden, en general, entre aproximadamente 60 y aproximadamente casi el 70 por ciento de AT. La utilización de codones en las especies de *Bacillus* tienden al uso de codones enriquecidos en la presencia de A o T en un codón particular y más particularmente en A o T en la posición de la tercera base de cualquier codón particular. Por ello, se emplean los codones que utilizan

primariamente G o C en una secuencia de codificación nativa y/o natural de *Bacillus* con una frecuencia mucho menor que los codones que contienen A o T. Por ello, cuando se produce una variante o secuencia de nucleótidos modificada para su uso en una planta particular, monocotiledónea o dicotiledónea, es importante asegurar que se preste la atención apropiada al hecho de no usar los codones menos preferidos con gran frecuencia para dicha planta particular. De hecho, para monocotiledóneas, se prefiere que la secuencia de codificación imite la distribución de GC que se observa en la mayoría de las plantas monocotiledóneas nativas o naturales; siendo preferido aproximadamente un 65% de GC para aproximadamente el primer 10% de la secuencia de codificación, disminuyendo hasta aproximadamente un 60% de GC para los segundos 10-15% de la secuencia de codificación y nivelándose en aproximadamente un 50-55% de GC para la mitad o más de la secuencia de codificación, y luego aumentando gradualmente el % de GC de la secuencia de codificación hasta alcanzar aproximadamente un 60-64% en los últimos 15-20% de la secuencia de codificación. Esta distribución de % GC para imitar tan estrechamente como fuera posible la distribución de % GC de un gen natural de monocotiledóneas y por ello se cree que un gen o secuencia de codificación sintético o producido artificialmente también debería imitar esta arquitectura.

Tal como se utiliza en la presente, las "secuencias de codificación sintéticas" o las "secuencias de codificación no naturales" que codifican las proteínas TIC901 de *B thuringiensis* u homólogos o derivados de las mismas como toxinas insecticidas de la presente invención son aquellas que se preparan de una manera que comprende cualquier tipo de aislamiento o manipulación genética que da como resultado la preparación de una secuencia de codificación de una proteína insecticida TIC901 o una secuencia de aminoácidos relacionada o un homólogo o variante o equivalente sustancial de la misma, incluyendo secuencias de codificación que codifican al menos la porción insecticida de una proteína TIC901, de una proteína TIC431, de

una proteína TIC1201, de una proteína TIC407 o de una proteína TIC417. Esto incluye el aislamiento de la secuencia de codificación de su estado natural, la manipulación de la secuencia de codificación con modificación de sus codones (como se describe en la presente), la síntesis química, tal como la química de fosforamidita y semejantes, o mutagénesis específica del sitio (como se describe en la presente), truncamiento de la secuencia de codificación o cualquier otro método de manipulación o aislamiento.

Tal como se utiliza en la presente, la frase "porcentaje de identidad de secuencia" se determina por comparación de dos secuencias alineadas de manera óptima sobre una ventana de comparación, donde la porción de la secuencia en dicha ventana de comparación puede comprender adiciones o supresiones (es decir, huecos o faltas de coincidencia) en comparación con la secuencia de referencia (que no comprende adiciones o supresiones) para lograr una alineación óptima de las dos secuencias El porcentaje se calcula determinando la cantidad de posiciones en las cuales aparecen las bases de ácido nucleico o los residuos de aminoácidos idénticos en ambas secuencias para obtener la cantidad de posiciones coincidentes, dividiendo la cantidad de posiciones coincidentes por la cantidad total de posiciones en la ventana de comparación y multiplicando el resultado por 100 para obtener el porcentaje de identidad de secuencia. Se dice que una secuencia que es idéntica en cada una de las posiciones en comparación con la secuencia de referencia es idéntica a la secuencia de referencia y viceversa Se dice que una primera secuencia de nucleótidos, observada en la dirección 5' a 3', es el "complemento" de una segunda secuencia de nucleótidos o una secuencia de referencia, observada en la dirección 3' a 5', si dicha primera secuencia de nucleótidos presenta una complementariedad completa con la segunda secuencia o secuencia de referencia Tal como se utiliza en la presente, se dice que las moléculas de ácido nucleico de la secuencia presentan "complementariedad completa" cuando cada uno de los nucleótidos de una de las secuencias leída de 5' a 3' es complementario de cada uno

de los nucleótidos de la otra secuencia leída de 3' a 5'. Se dice que una secuencia de nucleótidos que es idéntica en cada posición cuando es leída 5' a 3' en comparación con una secuencia de referencia de nucleótidos leída de 5' a 3' es idéntica a la secuencia de referencia y viceversa Una secuencia de nucleótidos que es complementaria a una secuencia referencia de nucleótidos de referencia presentará una secuencia idéntica a la secuencia del complemento inverso de la secuencia de nucleótidos de referencia. Estos términos y descripciones están bien definidos en el arte y son comprendidos fácilmente por los especialistas en el arte.

Tal como se utiliza en la presente, una "homología sustancial", con referencia secuencias de ácido nucleico, se refiere a secuencias de nucleótidos que se hibridizan bajo condiciones severas a las secuencias de codificación que se muestran en las SEQ ID Nº 3, SEQ ID Nº:5, SEQ ID Nº:7, SEQ ID Nº:9 o SEQ ID Nº:32 o los complementos de las mismas. Las secuencias que se hibridizan bajo condiciones severas con la SEQ ID Nº·3, SEQ ID Nº·5, SEQ ID Nº·7, SEQ ID Nº:9 o SEQ ID Nº:32 o los complementos de las mismas, en particular con los nucleótidos entre la posición de nucleótido 153 aproximadamente y la posición 1.253 aproximadamente de la SEQ ID Nº·3, y más particularmente entre la posición de nucleótido 282 aproximadamente y la posición de nucleótido 1.253 aproximadamente de la SEQ ID Nº:3; o entre la posición de nucleótido 437 aproximadamente y la posición de nucleótido 1621 aproximadamente de la SEQ ID Nº·5, más particularmente entre la posición de nucleótido 530 aproximadamente y la posición de nucleótido 1621 aproximadamente de la SEQ ID Nº:5, y más particularmente aún entre la posición de nucleótido 659 aproximadamente y la posición de nucleótido 1621 aproximadamente de la SEQ ID Nº·5; o entre la posición de nucleótido 196 aproximadamente y la posición de nucleótido 1299 aproximadamente de la SEQ ID Nº:7, o más particularmente entre la posición de nucleótido 325 aproximadamente y la posición de nucleótido 1299 aproximadamente de la SEQ ID Nº 7, o entre la posición de nucleótido 215 aproximadamente y la posición de nucleótido 1306

aproximadamente de la SEQ ID Nº.9, o más particularmente entre la posición de nucleótido 344 aproximadamente y la posición de nucleótido 1306 aproximadamente de la SEQ ID Nº 9, o entre la posición de nucleótido 1 aproximadamente y hasta la posición de nucleótido 1092 aproximadamente como se muestra en SEQ ID Nº.32, o más particularmente entre la posición de nucleótido 130 aproximadamente y la posición de nucleótido 1092 aproximadamente como se muestra en SEQ ID Nº.32, contiene una o más secuencias de nucleótidos contiguas que son suficientemente idénticas de una o más secuencias de nucleótidos contiguas de la SEQ ID Nº.3, SEQ ID Nº.5, SEQ ID Nº:7, SEQ ID Nº:32 o SEQ ID Nº:9, de modo tal que se puede efectuar una alineación entre las dos secuencias y dichas dos secuencias pueden entonces, bajo condiciones severas, formar uniones de hidrógeno con las bases correspondientes en la cadena opuesta formando así una molécula doble que sea suficientemente estable bajo condiciones severas por un período de tiempo suficientemente prologando como para ser detectable con los métodos bien conocidos en la técnica. Dichas secuencias sustancialmente homólogas son preferentemente entre un 67% aproximadamente y un 70% aproximadamente idénticas o más preferentemente entre un 80% aproximadamente y un 85% aproximadamente idénticas o con mayor preferencia entre un 90% aproximadamente y un 95% aproximadamente idénticas, hasta un 99% aproximadamente idénticas o más a las secuencias de nucleótidos de referencia que se muestran en la SEQ ID Nº 32, SEQ ID Nº 3, SEQ ID Nº 5, SEQ ID Nº 7 o la SEQ ID Nº 9, o los complementos de las mismas. Además, se considera también las secuencias de nucleótidos que codifican proteínas insecticidas que se pueden aislar de *Bacillus thuringiensis* u otras cepas de la especie *Bacillus* y semejantes; que se hibridizan bajo condiciones severas con la SEQ ID Nº.2, presentan una homología sustancial con las secuencias de nucleótidos de referencia enumeradas previamente que se hibridizan bajo condiciones severas con las secuencias de codificación *TIC901, TIC1201, TIC407, TIC417 y TIC431* que se muestran en la SEQ ID Nº:3 o SEQ ID

Nº·5, SEQ ID Nº.7, SEQ ID Nº.9 y SEQ ID Nº:32, respectivamente, o los complementos de las mismas. En la presente dichas secuencias de nucleótidos se denominan homólogos de la SEQ ID Nº.3 y semejantes y comprenden las SEQ ID Nº·5, SEQ ID Nº:7, SEQ ID Nº 9 y SEQ ID Nº:32 y secuencias relacionadas y homólogos de las mismas.

Con referencia a las secuencias de polipéptidos, los términos "identidad sustancial" o "similitud sustancial" se refieren a polipéptidos que presentan una identidad sustancial de secuencia de aminoácidos o una similitud sustancial secuencia de aminoácidos con respecto a una secuencia de aminoácidos de referencia, en particular en vista del hecho, descrito en la presente, que determinados aminoácidos pueden ser sustituidos por otros aminoácidos basado en los índices de hidropaticidad o hidrofilicidad y aún dar como resultado una proteína con una actividad biológica similar, es decir, aún obtener una proteína biológica y funcionalmente equivalente. Por ello, una secuencia de aminoácidos que presenta una identidad sustancial o una similitud sustancial a un polipéptido o secuencia de aminoácidos de referencia presentará entre aproximadamente un 30% y aproximadamente un 50% de identidad de secuencia de aminoácidos con la secuencia de referencia y más preferentemente presentará entre aproximadamente un 70% y aproximadamente un 80% de identidad de secuencia de aminoácidos con la secuencia de referencia, más preferentemente entre aproximadamente un 86% y aproximadamente un 90% identidad de secuencia de aminoácidos con la secuencia de referencia, y más preferentemente aún entre aproximadamente un 95% y aproximadamente un 99% identidad de secuencia de aminoácidos con la secuencia de referencia de polipéptidos. Más específicamente, los inventores consideran que los péptidos que presentan una actividad insecticida que está relacionada con los péptidos de la presente invención presentan una identidad de péptidos sustancial o una similitud de péptidos sustancial con los péptidos de la presente invención y presentan al menos entre aproximadamente un 67, 68, 69, 70, 71, 72, 73, 74, 75, 76,

77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98 y 99 por ciento de identidad o similitud con las secuencias de polipéptidos de referencia como se describe en la presente y seleccionadas del grupo formado por las SEQ ID Nº:1, SEQ ID Nº·4, SEQ ID Nº·6, SEQ ID Nº:8, SEQ ID Nº:10 y SEQ ID Nº.33.

Con referencia a las proteínas de la presente solicitud, los términos "variante de la secuencia de aminoácidos" o "secuencia de aminoácidos variante" o "secuencia de aminoácidos variante modificada" hacen referencia a las secuencias de aminoácidos que son sustancialmente equivalentes a las secuencias de aminoácidos de la presente invención. Por ejemplo, una proteína producida por introducción de un sitio de restricción para una mayor conveniencia con las manipulaciones moleculares en la secuencia de codificación de la presente invención que da como resultado la adición o eliminación de uno o más codones sin (1) interrumpir la secuencia de codificación nativa, (2) interrumpir el marco de lectura abierto nativo y (3) interrumpir la actividad biológica insecticida de la proteína, constituye (a) una variante de la secuencia de aminoácidos en comparación con la toxina insecticida nativa, (b) una secuencia de aminoácidos variante en comparación con la toxina insecticida nativa o (c) una secuencia de aminoácidos variante modificada en comparación con la toxina insecticida nativa. El especialista en el arte comprenderá que es posible efectuar otros tipos de modificaciones a la secuencia de aminoácidos de la presente invención sin interrumpir la actividad biológica de la proteína El uso del término "interrumpir", con referencia a la actividad biológica, hace referencia a las modificaciones de la secuencia de aminoácidos nativa obtenidas por inserción o supresión de uno o más aminoácidos, o intercambio o sustitución de un aminoácido por otro, que no disminuye la actividad biológica insecticida de la proteína Las inserciones, supresiones y sustituciones se encuentran dentro del alcance de la presente descripción hasta el grado que la secuencia de aminoácidos variante resultante no presente una actividad insecticida menor que la de la proteína insecticida nativa. Las

quimeras de las proteínas descritas en la presente, las fusiones de las proteínas o partes de las proteínas descritas en la presente, y las permuteínas de las proteínas descritas en la presente están específicamente contempladas.

Las proteínas que son sustancialmente equivalentes de las proteínas de la presente solicitud son consideradas como biológica y funcionalmente equivalentes Tal como se utiliza en la presente, la frase "equivalentes biológicos y funcionales", con respecto a las proteínas insecticidas de la presente invención, son péptidos, polipéptidos y proteínas que contienen una secuencia o porción que presenta similitud de secuencia con los nuevos péptidos de la presente invención, tal como una proteína TIC901 o un fragmento insecticida de la misma, o una proteína TIC1201, TIC410, TIC407 y TIC431 o fragmentos insecticidas de las mismas, y que presentan las mismas propiedades funcionales o propiedades similares a las de los polipéptidos descritos en la presente, incluyendo la actividad insecticida. Los equivalentes biológicos también incluyen péptidos, polipéptidos y proteínas que reaccionan, es decir, que se unen específicamente, con anticuerpos generados contra epitopes presentes sobre o dentro de la TIC901 y de las proteínas relacionadas tal como TIC1201, TIC417, TIC407 y TIC431 y fragmentos insecticidas de las mismas, y que presentan la misma actividad de unión o reactiva o actividad similar, incluyendo anticuerpos tanto monoclonales como policlonales.

También se contempla que las proteínas de la presente invención puedan ser de utilidad para proteger a plantas dicotiledóneas de una infestación por insectos Dichas infestaciones podrían ser el resultado de infestaciones por coleópteros, dípteros, lepidópteros o aún ácaros, gusanos de la harina, gorgojos o una amplia variedad de insectos que dañan a las plantas porque perforan los tejidos vegetales y extraen los nutrientes pretendidas para el crecimiento y desarrollo de la planta. Las modificaciones a la secuencia de aminoácidos primaria de las proteínas de la presente invención podrían dar como resultado una proteína con un rango de huéspedes diferente del rango de la proteína nativa.

Las proteínas de la presente invención, localizadas en el espacio extracelular cuando son expresadas por cepas de *Bacillus*, pueden ser de utilidad para direccionar otras proteínas hacia el espacio extracelular. Por ejemplo, el especialista conoce la manera de ligar una primera proteína que no es secretada normalmente hacia el espacio extracelular a una segunda proteína que normalmente es secretada hacia el espacio extracelular con el fin de localizar dicha primera proteína en el espacio extracelular. Las proteínas de la presente invención se podrían fusionar, utilizando cualquiera de los numerosos medios bien conocidos en el arte, con una o más toxinas insecticidas tal como delta-endotoxinas cristalinas para formar una proteína quimérica cuya secreción será direccionada hacia el espacio extracelular que rodea a una célula huésped particular. Se considera aún que el propio evento de secreción podría llevar a la separación de las dos partes proteicas de modo tal que se van a liberar dos proteínas insecticidas separadas y distintas en el espacio extracelular que rodea a la célula huésped particular. Las dos proteínas podrían ser (1) tóxicas ambas para la misma especie de insecto pero efectuar su actividad insecticida usando diferentes modos de acción o (2) cada una tóxica para una especie de insecto diferente. Se considera que se podrían ligar cualquier número de proteínas insecticidas extremo-contra-extremo a las proteínas de la presente invención para formar quimeras multiméricas que serán dirigidas hacia el espacio extracelular que rodea a la célula huésped particular. Dichas "otras" proteínas podrían ser la proteína fluorescente verde y las proteínas relacionadas y variantes de la misma, quinasas y fosfatasas para modular los procesos de señalamiento de las células, nucleasas, lipasas, proteínas de tolerancia a herbicidas expresadas por genes tales como *gox*, diversos homólogos de *epsps*, *bar* y homólogos y semejantes, PhnO, NptII, Aad y semejantes Todas estas proteínas también se podrían usar como marcadores seleccionables, en particular cuando se ligan a un gen que codifica una o más de las proteínas de la presente invención, con el fin de rastrear la presencia de los genes que codifican una o más de las proteínas de la presente invención en una

planta u otra célula huésped.

Las proteínas de la presente invención podrían ser direccionadas para su importación en un organela subcelular. Por ejemplo, se podría fusionar o ligar operativamente una primera secuencia de nucleótidos codificadora de una secuencia de direccionamiento a cloroplastos o plástidos a una segunda secuencia de nucleótidos codificadora de una proteína insecticida de la presente invención para producir una proteína quimérica precursora buscada para su inserción en el cloroplasto o plástido dentro de una célula vegetal. La expresión de dichas proteínas quiméricas daría como resultado la importación de las proteínas de la presente invención en los cloroplastos o plástidos de la planta, dando como resultado la localización de la toxina insecticida, o un fragmento insecticida de la misma, en los cloroplastos o plástidos Además, la secuencia de nucleótidos, que codifica una o más proteínas de la presente invención, se podría localizar en los cloroplastos o plástidos para su expresión. La localización de las secuencias de nucleótidos en los plástidos o cloroplastos podría dar como resultado la incorporación de las secuencias de nucleótidos en el genoma de los cloroplastos o plástidos o podría dar como resultado la presencia de una secuencia de ácido nucleico de replicación autónoma codificadora de la proteína de la presente invención En cualquier caso, las proteínas de la presente invención se podrían localizar en los cloroplastos o plástidos Por ello, tal como se utiliza en la presente, la frase "localizado en cloroplastos o plástidos" se refiere a una molécula biológica, ya sea un polinucleótido o un polipéptido, localizado dentro de cloroplastos o plástidos de modo tal que la molécula queda aislada del entorno del citoplasma celular y funciona dentro del citoplasma de cloroplastos o plástidos proporcionando los efectos reivindicados en la presente invención. La localización de la molécula biológica en cloroplastos o plástidos puede tener lugar, con referencia a los polinucleótidos, utilizando medios mecánicos artificiales, tal como electroporación, microinyección mecánica, o por bombardeo con microproyectiles recubiertos con polinucleótidos o, con referencia a polipéptidos,

utilizando medios de secreción o importación, en los cuales se usa una secuencia peptídica de direccionamiento a plástidos o cloroplastos natural, sintética o heteróloga cuya función consiste en dirigir, insertar, asistir o localizar el polipéptido ligado en cloroplastos o plástidos

Tal como se utiliza en la presente, la frase "unido operativamente" o "ligado operativamente" se refiere a segmentos de ácido nucleico unidos en marco de modo tal que las propiedades de uno afecta la expresión del otro. Estas frases y grupos de palabras también se pueden usar con referencia a secuencias de aminoácidos que presentan alguna función cuando se ligan a otra secuencia de aminoácidos, por ejemplo, cuando un péptido señal está ligado a una proteína de interés se dice que está ligado operativamente a dicha la proteína de interés a los efectos de dirigir dicha proteína de interés hacia el aparato secretor de la célula huésped en la cual es producida la proteína o hacia un compartimiento subcelular, tal como el retículo endoplasmático, un cloroplasto o un plástido, una mitocondria, una vacuola, el núcleo o nucleolo u otro compartimiento subcelular y semejantes.

A los efectos de la presente invención, la palabra "gen" se refiere a una secuencia de nucleótidos que contiene un marco de lectura abierto que codifica una proteína TIC901, una proteína TIC1201, una proteína TIC417, una proteína TIC407, una proteína TIC431 o un fragmento insecticida de las mismas, o una secuencia de aminoácidos variante de las mismas, o una proteína relacionada homóloga o un fragmento insecticida de la misma o una secuencia de aminoácidos variante de la misma que está ligado operativamente a por lo menos una secuencia promotora y una secuencia de terminación de la transcripción, donde dichas secuencias promotora y de terminación de la transcripción son funcionales en la célula huésped en la cual es producida la proteína. Tal como se utiliza en la presente, un "gen estructural" se refiere a un gen que se expresa produciendo un polipéptido. Un gen estructural de la presente invención puede contener, además de las secuencias promotora y de terminación de la transcripción, secuencias no traducidas 5',

secuencias intrónicas y elementos potenciadores que sean funcionales en determinadas plantas, y preferentemente los que derivan de maíz o de otras plantas monocotiledóneas que, cuando se ligan juntos en la secuencia apropiada con una o más secuencias de codificación de la presente invención, dan como resultado mayores niveles de expresión en tejidos vegetales particulares, y preferentemente dan como resultado una mayor expresión en los tejidos radiculares de plantas de maíz.

La información de secuencia de nucleótidos provista por la presente invención permite preparar de secuencias de ADN relativamente cortas, denominadas sondas o cebadores en la presente, que tienen la capacidad para hibridizarse específicamente con las secuencias de polinucleótidos seleccionadas descritas en la presente. Dichas sondas de ácido nucleico de longitud apropiada se preparan sobre la base de consideraciones obre las secuencias de polipéptidos seleccionadas que codifican los polipéptidos insecticidas de la presente invención, por ejemplo, una secuencia tal como la que se muestra en la SEQ ID Nº:2, toda o una parte específica de la sonda de la SEQ ID Nº 3 aproximadamente entre el nucleótido 153 y aproximadamente el nucleótido 1.253; toda o una parte específica de la sonda de la SEQ ID Nº:5 aproximadamente entre el nucleótido 530 y aproximadamente el nucleótido 1621, toda o una parte específica de la sonda de la SEQ ID Nº:7, toda o una parte específica de la sonda de la SEQ ID Nº.9, toda o una parte específica de la sonda de la SEQ ID Nº·32 o toda o una parte específica de la sonda o específica del cebador de las secuencias seleccionadas del grupo formado por las SEQ ID Nº.11, SEQ ID Nº·12, SEQ ID Nº:13, SEQ ID Nº:15, SEQ ID Nº.16, SEQ ID Nº.17, SEQ ID Nº:18, SEQ ID Nº 19, SEQ ID Nº 20, SEQ ID Nº.21 y SEQ ID Nº·22, y complementos de las mismas y semejantes. La frase "toda o una parte específica de la sonda" hace referencia a una sonda que comprende una secuencia de al menos entre 15 y 50, más o menos, nucleótidos contiguos seleccionados del grupo de nucleótidos que se muestra en la secuencia de referencia particular, tal como las SEQ ID Nº:2, SEQ ID

Nº·3, SEQ ID Nº:5, SEQ ID Nº:7, SEQ ID Nº.9, SEQ ID Nº·11, SEQ ID Nº 12, SEQ ID Nº.13, SEQ ID Nº·15, SEQ ID Nº.16, SEQ ID Nº:17, SEQ ID Nº·18, SEQ ID Nº.19, SEQ ID Nº.20, SEQ ID Nº 21, SEQ ID Nº:22 y SEQ ID Nº 32 y complementos de las mismas y semejantes. La capacidad de dichas sondas de ácido nucleico para hibridizarse específicamente con una secuencia de nucleótidos que codifica una secuencia de polipéptidos insecticida los vuelve de particular utilidad en diversas realizaciones. Más importante aún, las sondas se pueden utilizar en diversos ensayos para detectar la presencia de secuencias complementarias en una muestra dada.

En determinadas realizaciones, resulta ventajoso usar cebadores de oligonucleótidos. La secuencia de dichos cebadores es diseñada usando un polinucleótido de la presente invención con el fin de emplearla para detectar, amplificar o modificar un segmento definido de la secuencia de codificación de una proteína insecticida de *B. thuringiensis* o de *Bacillus sphaericus* y semejantes usando para ello tecnología de amplificación térmica Los segmentos de las secuencias de nucleótidos relacionadas con los polinucleótidos que codifican los polipéptidos insecticidas de la presente invención también se pueden aislar y caracterizar usando tecnología de amplificación térmica y los mencionados cebadores.

Para proveer algunas de las ventajas acordes con la presente invención, la secuencia de ácido nucleico preferida empleada en estudios o ensayos de hibridación o como un cebador incluye secuencias que son complementarias de una extensión de por lo menos 14 a 30 o más nucleótidos contiguos de una secuencia de polinucleótidos que codifica todo o una parte de una proteína insecticida de la presente invención, tal como las que se muestran en las SEQ ID Nº.2, SEQ ID Nº.3, SEQ ID Nº:5, SEQ ID Nº:7, SEQ ID Nº.9, SEQ ID Nº·11, SEQ ID Nº·12, SEQ ID Nº:13, SEQ ID Nº 15, SEQ ID Nº·16, SEQ ID Nº.17, SEQ ID Nº:18, SEQ ID Nº 19, SEQ ID Nº.20, SEQ ID Nº 21, SEQ ID Nº:22 y SEQ ID Nº.32 y complementos de las mismas y semejantes.

Un tamaño de cebador o sonda de al menos 14 nucleótidos de longitud permite asegurar que el fragmento será de una longitud suficiente como para formar una molécula doble que es tanto estable como selectiva. En general, se prefieren las moléculas que presentan secuencias complementarias sobre segmentos de más de 14 bases de longitud. Con el fin de incrementar la estabilidad y selectividad del híbrido, y de esa manera mejorar la calidad y el grado de moléculas híbridas específicas obtenidas, en general se prefiere diseñar moléculas de ácido nucleico que tienen secuencias complementarias de *tic*901 y semejantes de 14 a 20 nucleótidos, o aún más cuando se deseara, o que tienen secuencias complementarias de *tic*1201 y semejantes de 14 a 20 nucleótidos, o aún más cuando se deseara; secuencias complementarias de *tic*417y semejantes de 14 a 20 nucleótidos, o aún más cuando se deseara; secuencias complementarias de *tic*407 y semejantes de 14 a 20 nucleótidos, o aún más cuando se deseara y secuencias complementarias de *tic*431 y semejantes de 14 a 20 nucleótidos, o aún más cuando se deseara Dichos fragmentos se pueden preparar fácilmente, por ejemplo, por síntesis directa del fragmento por medios químicos, por aplicación de tecnología de reproducción de ácido nucleico o por recorte de fragmentos seleccionados de ADN de secuencias recombinantes localizadas en plásmidos o en otros vectores que contiene los insertos apropiados y los sitios de restricción adecuados.

Los inventores también han diseñado sondas y cebadores universales degenerados para su uso en la identificación de secuencias de nucleótidos naturales que codifican secuencias de aminoácidos derivadas de proteínas insecticidas que son homólogas de las proteínas de la presente invención. Las secuencias de nucleótidos identificadas usando los ejemplos de sondas y cebadores que se muestran en las SEQ ID Nº:23-SEQ ID Nº.29 se hibridizan bajo condiciones severas con las secuencias de nucleótidos que codifican las proteínas insecticidas secretadas descritas en la presente.

La alineación de secuencias de aminoácidos que se muestra en la Figura 1

proporcionó la base para identificar secuencias de aminoácidos que están altamente conservadas entre las cuatro proteínas precursoras insecticidas alineadas Por ejemplo, la secuencia de aminoácidos que se muestra en la SEQ ID Nº:4 entre aproximadamente el aminoácido setenta y cinco (75) y aproximadamente el aminoácido ochenta y tres (83) es una secuencia que está conservada tanto en cuanto a secuencia como a posición en las secuencias primarias de las proteínas TIC901, TIC1201, TIC407, TIC417 y TIC431 Esto se describe en la presente como una 'primera secuencia de aminoácidos conservada' La secuencia de aminoácidos que comprende entre aproximadamente el aminoácido ciento cuarenta y siete (147) y aproximadamente el aminoácido ciento cincuenta y tres (153) que se muestra en la SEQ ID Nº 4 también está conservada en cuanto a secuencia y posición en las secuencias primarias de las proteínas TIC901, TIC1201, TIC407, TIC431 y TIC417. Esto se describe en la presente como una 'segunda secuencia de aminoácidos conservada' De manera similar, la secuencia de aminoácidos que comprende entre aproximadamente el aminoácido doscientos setenta y cinco (275) y aproximadamente el aminoácido doscientos ochenta y tres (283) que se muestra en la SEQ ID Nº:4 también está conservada en cuanto a secuencia y posición en las secuencias primarias de las proteínas TIC901, TIC1201, TIC407, TIC431 y TIC417. Esto se describe en la presente como una 'tercera secuencia de aminoácidos conservada' Estas secuencias corresponden cada una a secuencias de nucleótidos sustancialmente conservadas pero ligeramente degeneradas en la respectiva secuencia de codificación para cada proteína que se pueda usar ya sea como sonda o como cebador en la identificación de segmentos de nucleótidos homólogos a la secuencia que comprende al menos una secuencia de catorce bases seleccionadas del grupo formado por las SEQ ID Nº·3, SEQ ID Nº 5, SEQ ID Nº.7, SEQ ID Nº:32 y SEQ ID Nº:9 o el complemento de las mismas. Por ejemplo, una reacción de amplificación térmica que emplea una secuencia cebadora degenerada, cuya síntesis se basa en secuencias compiladas seleccionadas entre las SEQ ID Nº.3, SEQ ID

Nº:5, SEQ ID Nº 7, SEQ ID Nº:32 y SEQ ID Nº:9, y los complementos de las mismas cuando se contempla una amplificación térmica inversa, correspondiente a la secuencia que codifica la 'primera secuencia de aminoácidos conservada' como se describió previamente, comprende un oligonucleótido de veintiséis mer (26-mer) correspondiente a la primera secuencia de nucleótidos conservada, por ejemplo la que se muestra en la SEQ ID Nº·3 entre la posición de nucleótido trescientos setenta y cinco (375) aproximadamente y la posición de nucleótido cuatrocientos uno (401) aproximadamente, que sería una de las secuencias degeneradas que se podrían usar en pruebas con sondas con las muestras para determinar la presencia de una secuencia de nucleótidos homóloga correspondiente a la secuencia que se hibridiza con la correspondiente secuencia de las SEQ ID Nº.3, SEQ ID Nº.5, SEQ ID Nº·7, SEQ ID Nº:32 y SEQ ID Nº.9. Como alternativa, la secuencia de oligonucleótidos se podría usar como uno de un par de cebadores de oligonucleótidos en una reacción de amplificación térmica para producir una secuencia de una amplicón que se hibridizará con las correspondientes secuencias contenidas en una o más de las secuencias definidas en la presente, por ejemplo, contenida en una secuencia que comprende las SEQ ID Nº.3, SEQ ID Nº.5, SEQ ID Nº:7, SEQ ID Nº:32 y SEQ ID Nº 9, bajo condiciones de hibridación severas

Por ello, los inventores de la presente han construido juegos de cebadores y sondas que se pueden usar solos o en combinación para identificar secuencias que están relacionadas con las proteínas de la presente invención, incluyendo TIC901, TIC1201, TIC407, TIC431 y TIC417, que codifican proteínas insecticidas secretadas, y que se hibridizan con una o más de las secuencias descritas en la presente bajo condiciones de hibridación severas. Un juego de cebadores consiste de secuencias de oligonucleótidos degeneradas que corresponden a las secuencias que se muestran en las SEQ ID Nº 23 a SEQ ID Nº:25 La SEQ ID Nº:23 corresponde a un juego de secuencias de oligonucleótidos degeneradas correspondiente a una extensión que comprende entre la posición de nucleótido trescientos setenta y cinco

(375) aproximadamente y la posición de nucleótido cuatrocientos uno (401) aproximadamente de la SEQ ID Nº 3, entre la posición de nucleótido setecientos cincuenta y dos (752) aproximadamente y la posición de nucleótido setecientos setenta y ocho (778) aproximadamente de la SEQ ID Nº:5, entre la posición de nucleótido trescientos noventa y uno (391) aproximadamente y la posición de nucleótido cuatrocientos diecisiete (417) aproximadamente de la SEQ ID Nº 7, entre la posición de nucleótido cuatrocientos treinta y siete (437) aproximadamente y la posición de nucleótido cuatrocientos sesenta y tres (463) aproximadamente de la SEQ ID Nº:9 y entre la posición de nucleótido doscientos veintitrés (223) aproximadamente y la posición de nucleótido doscientos cuarenta y nueve (249) aproximadamente de la SEQ ID Nº·32. Todas las combinaciones posibles de secuencias de nucleótidos que codifican la correspondiente 'primera secuencia de aminoácidos conservada' descrita previamente en la presente que razonablemente podrían estar presentes en una especie de *Bacillus* están contempladas aquí, tal como la que se muestra en la SEQ ID Nº:23. Las SEQ ID Nº.24 y SEQ ID Nº.25 son secuencias de oligonucleótidos degeneradas que comprenden subconjuntos de las secuencias correspondientes a la SEQ ID Nº:23 y que, al igual que la SEQ ID Nº.23, contienen codones que tienden a la utilización de codones preferida de las secuencias de codificación de *Bacillus*.

Se contempla un segundo juego de cebadores y sondas que se puede utilizar para identificar las secuencias descritas en la presente relacionadas con las SEQ ID Nº:3, SEQ ID Nº.5, SEQ ID Nº:7, SEQ ID Nº.9 y SEQ ID Nº:32 y que codifican las proteínas insecticidas secretadas con las secuencias degeneradas que se muestran en la SEQ ID Nº·26 Estas secuencias de oligonucleótidos corresponden al rango de secuencias de nucleótidos anticipadas que serían de preferencia para una especie de *Bacillus* que codifica la secuencia de aminoácidos descrita precedentemente como la 'segunda secuencia de aminoácidos conservada', y que además corresponde entre la posición de nucleótido quinientos noventa y uno (591)

aproximadamente y la posición de nucleótido seiscientos once (611) aproximadamente que se muestra en la SEQ ID Nº.3, entre la posición de nucleótido novecientos sesenta y ocho (968) aproximadamente y la posición de nucleótido novecientos ochenta y ocho (988) aproximadamente que se muestra en la SEQ ID Nº.5, entre la posición de nucleótido seiscientos siete (607) aproximadamente y la posición de nucleótido seiscientos veintisiete (627) aproximadamente que se muestra en la SEQ ID Nº·7, entre la posición de nucleótido seiscientos cincuenta y tres (653) aproximadamente y la posición de nucleótido seiscientos setenta y tres (673) aproximadamente que se muestra en la SEQ ID Nº:9 y entre la posición de nucleótido cuatrocientos treinta y nueve (439) aproximadamente y la posición de nucleótido cuatrocientos cincuenta y seis (456) aproximadamente que se muestra en SEQ ID Nº·32. Todas las combinaciones posibles de secuencias de nucleótidos que codifica la correspondiente 'segunda secuencia de aminoácidos conservada' descrita precedentemente que razonablemente podrían estar presentes en una especie de *Bacillus* están contempladas aquí, tal como la que se muestra en la SEQ ID Nº·26 y los codones seleccionados para su incorporación en la secuencia de oligonucleótidos degenerada tienden a la utilización de codones preferida de las secuencias de codificación de *Bacillus*

Aún más, un tercer juego de cebadores y sondas que se puede usar para identificar las secuencias descritas en la presente que están relacionadas con las SEQ ID Nº·3, SEQ ID Nº 5, SEQ ID Nº:7, SEQ ID Nº:9 y SEQ ID Nº·32 y que codifican las proteínas insecticidas secretadas están contempladas por las secuencias degeneradas como se muestra en las SEQ ID Nº.27 - SEQ ID Nº:29 Estas secuencias de oligonucleótidos corresponden al rango de secuencias de nucleótidos anticipadas como preferidas por una especie de *Bacillus* que codifica la secuencia de aminoácidos descrita precedentemente como la 'tercera secuencia de aminoácidos conservada', y además corresponde al complemento inverso de la secuencia de nucleótidos entre la posición de nucleótido novecientos setenta y cinco

(975) aproximadamente y la posición de nucleótido mil uno (1.001) aproximadamente como se muestra en la SEQ ID Nº·3, el complemento inverso de la secuencia de nucleótidos entre la posición de nucleótido mil trescientos cincuenta y dos (1 352) aproximadamente y la posición de nucleótido mil trescientos setenta y ocho (1.378) aproximadamente como se muestra en la SEQ ID Nº·5, el complemento inverso de la secuencia de nucleótidos entre la posición de nucleótido novecientos noventa y uno (991) aproximadamente y la posición de nucleótido mil diecisiete (1.017) aproximadamente como se muestra en la SEQ ID Nº.7, el complemento inverso de la secuencia de nucleótidos entre la posición de nucleótido mil treinta y siete (1 037) aproximadamente y la posición de nucleótido mil sesenta y tres (1.063) aproximadamente como se muestra en SEQ ID Nº 9, y el complemento inverso de la secuencia de nucleótidos entre la posición de nucleótido ochocientos veintitrés (823) aproximadamente y la posición de nucleótido ochocientos cuarenta y seis (846) aproximadamente que se muestra en la SEQ ID Nº 32 Todas las combinaciones posibles de secuencias de oligonucleótidos que codifica la correspondiente 'tercera secuencia de aminoácidos conservada' descritas precedentemente que razonablemente podrían estar presentes en una especie de *Bacillus* están contemplada aquí, tal como la que se muestra en la SEQ ID Nº·27 Las SEQ ID Nº 28 y SEQ ID Nº.29 también son secuencias de nucleótidos degeneradas que comprenden subconjuntos de las secuencias correspondientes a la SEQ ID Nº:27 y que, al igual que la SEQ ID Nº 27, contienen codones que tienden a la utilización de codones preferida de las secuencias de codificación de *Bacillus*.

Cualquiera de las secuencias contempladas por las secuencias que se muestran en las SEQ ID Nº:23 a SEQ ID Nº 29 se puede usar sola como una sonda para identificar la presencia en una muestra de una secuencia de nucleótidos que codifica una proteína insecticida secretada relacionada con cualquiera de las proteínas que se ofrecen como ejemplos en la presente, en particular una secuencia de nucleótidos que se hibridiza con una o más de las secuencias de nucleótidos de la

presente invención, incluyendo, sin limitaciones, las SEQ ID Nº·2, SEQ ID Nº:3, SEQ ID Nº.5, SEQ ID Nº 7, SEQ ID Nº·9, SEQ ID Nº32 y cualquiera de las secuencias de nucleótidos que se muestran aquí como sondas y/o cebadores seleccionadas del grupo formado por las SEQ ID Nº:11, SEQ ID Nº 12, SEQ ID Nº:15, SEQ ID Nº:16, SEQ ID Nº·17, SEQ ID Nº.18, SEQ ID Nº·19, SEQ ID Nº:20, SEQ ID Nº·21, SEQ ID Nº:22, SEQ ID Nº:23, SEQ ID Nº:24, SEQ ID Nº.25, SEQ ID Nº 26, SEQ ID Nº·27, SEQ ID Nº:28 y SEQ ID Nº:29.

Como alternativa, y preferentemente, se pueden usar varias combinaciones de las sondas y cebadores descritas aquí juntas en una reacción de amplificación térmica para producir uno o más amplicones que constituyen elementos de diagnóstico de la presencia en una muestra de una secuencia de nucleótidos que codifica una proteína insecticida secretada relacionada con las proteínas de la presente invención, donde dicha secuencia de nucleótidos que codifica la proteína insecticida se hibridiza con una o más de las secuencias de nucleótidos que se ofrecen a modo de ejemplo en la presente que comprenden las SEQ ID Nº·2, SEQ ID Nº.3, SEQ ID Nº·5, SEQ ID Nº·7, SEQ ID Nº.9, SEQ ID Nº:11, SEQ ID Nº·12, SEQ ID Nº 15, SEQ ID Nº:16, SEQ ID Nº:17, SEQ ID Nº 18, SEQ ID Nº 19, SEQ ID Nº:20, SEQ ID Nº·21, SEQ ID Nº:22, SEQ ID Nº.23, SEQ ID Nº:24, SEQ ID Nº 25, SEQ ID Nº:26, SEQ ID Nº:27, SEQ ID Nº.28, SEQ ID Nº.29 y SEQ ID Nº:32 o los complementos de las mismas, bajo condiciones de hibridación severas. Por ejemplo, la combinación de uno o más de los nucleótidos que se muestran en las SEQ ID Nº.23-25, tal como el cebador prJWP200 (SEQ ID Nº:23), con uno o más de los nucleótidos que se muestran en las SEQ ID Nº.27-29, tal como el cebador prJWP204 (SEQ ID Nº·27), cada uno a una concentración de al menos 1 picomol por microlitro aproximadamente en una reacción de amplificación térmica que contiene solución amortiguadora de amplificación TAQ 1X, 0,2 molar de cada uno de los desoxinucleótidos trifosfato (dATP, dTTP, dCTP y dGTP), $MgCl_2$ 2 milimolar, 2 unidades de TAQ polimerasa y entre aproximadamente diez (10) y aproximadamente

cien (100) nanogramos de una muestra que contiene un templado de ADN, el cual contiene una o más de las secuencias de la presente invención codificadoras de una proteína insecticida secretada o un fragmento de la misma, da como resultado la síntesis de un fragmento de ADN de cadena doble que es un amplicón que comprende entre aproximadamente 600 y aproximadamente 650 pares de bases, más preferentemente entre aproximadamente 615 y aproximadamente 630 pares de bases y más preferentemente aún entre aproximadamente 623 y aproximadamente 626 pares de bases. Un amplicón de este tamaño constituye un elemento de diagnóstico de la presencia en una muestra de una secuencia de nucleótidos que codifican todo o parte de una proteína insecticida secretada que está relacionada con una o más de las proteínas de la presente invención, y es de un tamaño que razonablemente puede ser anticipado, en particular si las condiciones de los ciclos de amplificación térmica consisten de una desnaturalización inicial de aproximadamente 2 minutos a 94 °C seguido de 35 ciclos de un paso de desnaturalización de 30 segundos a 94 °C, un paso de alineación a menor temperatura de 30 segundos a 50 °C y un paso de extensión de 45 segundos a 72 °C seguido de un paso final de extensión de 7 minutos a 72 °C La temperatura del paso de alineación a menor temperatura se debe disminuir en 0,3 °C en cada ciclo sucesivo de modo tal que la temperatura de alineación final sea de aproximadamente 39,8 °C Además, la combinación de uno o más de los nucleótidos que se muestran en la SEQ ID Nº:26, tal como el prJWP203, con uno o más de los cebadores que se muestran en las SEQ ID Nº:27-29 en una reacción de amplificación térmica similar, daría como resultado la síntesis de un amplicón que comprende entre aproximadamente 400 y aproximadamente 415 pares de bases, o un amplicón que comprende entre aproximadamente 400 y aproximadamente 415 pares de bases, o que comprende entre aproximadamente 405 y aproximadamente 410 pares de bases, que constituye un elemento de diagnóstico de la presencia en una muestra de una secuencia de nucleótidos que codifica una proteína insecticida secretada

relacionada las proteínas de la presente invención.

Con estas condiciones, y usando los cebadores prJWP200 (SEQ ID Nº 23) y prJWP204 (SEQ ID Nº:27) junto con ADN genómico de la cepa EG2158, se obtuvo como resultado un segmento de amplicón correspondiente a 620 pares de bases aproximadamente. Se aislaron aleatoriamente numerosos clones individuales que representaban a este segmento de la reacción de amplificación y se obtuvo la secuencia de nucleótidos de cada clon individual. Tal como se esperaba, se identificó una primera secuencia idéntica a la secuencia de nucleótidos correspondiente a TIC901 en la población de clones (sin incluir las secuencias de cebadores de cualquiera de los extremos del clon, entre la posición de nucleótido cuatrocientos dos (402) aproximadamente y la posición de nucleótido novecientos setenta y cuatro (974) aproximadamente de en la SEQ ID Nº:3). Además, y como se esperaba, se identificó una segunda secuencia idéntica a la secuencia de nucleótidos correspondiente a TIC417 en la población de clones (sin incluir las secuencias de cebadores de cualquiera de los extremos del, entre la posición de nucleótido cuatrocientos sesenta y cuatro (464) aproximadamente y la posición de nucleótido mil treinta y seis (1,036) aproximadamente de la SEQ ID Nº:9)

Inesperadamente, también se identificó una tercera secuencia (SEQ ID Nº·30) en la población de clones que no correspondía de forma idéntica a ninguna de las secuencias definidas aquí pero que era sustancialmente similar en cuanto a secuencia de nucleótidos a cada una de las secuencias de la presente invención, incluyendo la secuencia de codificación correspondiente a las secuencias de codificación *TIC901, TIC1201, TIC407, TIC417 y TIC431*. Tal como se ilustra en la presente, resultó bastante inesperado encontrar la secuencia de codificación *tic*417 en el ADN genómico de la cepa EG2158. Sin embargo, fue aún más sorpresivo identificar una tercera secuencia de nucleótidos en el genoma de EG2158 que probablemente correspondía a un segmento de nucleótidos que codificaba aún una tercera proteína insecticida secretada diferente de TIC901, TIC1201, TIC407, TIC417

y TIC431, pero que era suficientemente similar en cuanto a secuencia a las proteínas de la presente invención como para poder clasificarla como una de las especies pertenecientes al grupo de proteínas insecticidas secretadas codificadas por una secuencia de nucleótidos que se hibridiza con una o más de las secuencias definidas aquí, e ilustra la novedad y utilidad de las sondas y cebadores de oligonucleótidos degenerados ilustrados aquí para su uso en la identificación de secuencias que codifican proteínas insecticidas secretadas y que se hibridizan bajo condiciones severas con las secuencias de codificación *tic*901, *tic*1201 *tic*07, *tic*417 y *tic431* relacionadas descritas en la presente.

La secuencia de aminoácidos codificada por el marco de lectura abierto no interrumpido que se muestra en la SEQ ID Nº·30, de la cual se han suprimido las secuencias de oligonucleótidos degeneradas de veintiséis mer de ambos extremos 5' y 3', se muestra en la SEQ ID Nº:31 La secuencia de aminoácidos de la SEQ ID Nº:31 es sustancialmente similar a la secuencia de aminoácidos de TIC901 entre las posiciones de aminoácidos ochenta y cinco (85) aproximadamente y doscientos setenta y cuatro (274) aproximadamente como se muestra en la SEQ ID Nº·4, que solamente contiene dos (2) aminoácidos diferentes de la secuencia análoga en SEQ ID Nº:4, correspondiente a una identidad del 98,9% aproximadamente entre la SEQ ID Nº:31 y la correspondiente secuencia en la SEQ ID Nº:4. La secuencia de aminoácidos de la SEQ ID Nº 31 es sustancialmente similar a la secuencia de aminoácidos de TIC1201 entre las posiciones de aminoácidos ochenta y cinco (85) aproximadamente y doscientos setenta y cuatro (274) aproximadamente que se muestra en la SEQ ID Nº:6, que solamente contiene trece (13) aminoácidos diferentes de la secuencia análoga en la SEQ ID Nº.6, correspondiente a una identidad del 93,2% aproximadamente entre la SEQ ID Nº:31 y la correspondiente secuencia en la SEQ ID Nº:6. La secuencia de aminoácidos de la SEQ ID Nº·31 es sustancialmente similar a la secuencia de aminoácidos de TIC417 entre las posiciones de aminoácidos ochenta y cinco (85) aproximadamente y doscientos

setenta y cuatro (274) aproximadamente que se muestra en la SEQ ID Nº·10, que solamente contiene treinta (30) aminoácidos diferentes de la secuencia análoga en la SEQ ID Nº.10, correspondiente a una identidad del 83,7% aproximadamente entre la SEQ ID Nº·31 y la correspondiente secuencia en la SEQ ID Nº·10. La secuencia de aminoácidos de la SEQ ID Nº.31 también es sustancialmente similar a la secuencia de aminoácidos de TIC401 entre las posiciones de aminoácidos ochenta y cinco (85) aproximadamente y doscientos setenta y cuatro (274) aproximadamente que se muestra en la SEQ ID Nº·8, que contiene cuarenta y un (41) aminoácidos diferentes de la secuencia análoga en la SEQ ID Nº:8, correspondiente a una identidad del 78,4% aproximadamente entre la SEQ ID Nº 31 y la correspondiente secuencia en la SEQ ID Nº·8 La secuencia de aminoácidos que se muestra en la SEQ ID Nº.31 también es sustancialmente similar a la secuencia de aminoácidos que se muestra en la SEQ ID Nº·33 (TIC431).

La presente invención también contempla un vector de expresión que comprende un polinucleótido de la presente invención. Así, en una realización el vector de expresión es una molécula de ADN aislada y purificada que comprende un promotor ligado operativamente a una región de codificación que codifica un polipéptido de la presente invención, donde dicha región de codificación está ligada operativamente a una región de terminación de la transcripción, con lo cual el promotor dirige la transcripción de la región de codificación La región de codificación puede incluir un segmento que codifica una toxina insecticida de *B thuringiensis* de la presente invención y un segmento que codifica un péptido de direccionamiento a cloroplastos o plástidos La molécula de ADN que comprende al vector de expresión también puede contener una secuencia de un intrón funcional ubicada ya sea 5' con respecto a la secuencia de codificación o aún dentro de la secuencia de codificación, y también puede contener una secuencia no traducida 5' (es decir, una UTR o 5'-UTR) ubicada entre el promotor y el punto de inicio de la traducción.

Tal como se utiliza en la presente y con referencia a los elementos

promotores, los términos "ligado operativamente" o "unido operativamente" indican que la secuencia de nucleótidos que contiene un promotor, es decir un elemento genético que funciona en una célula huésped particular dirigiendo el inicio de la transcripción, está conectada con la región de codificación de manera tal que la transcripción de dicha región de codificación es controlada y regulada sustancialmente por dicho promotor Los medios para ligar operativamente a un promotor con una región de codificación son bien conocidos en el arte. Los promotores que son funcionales en bacterias son bien conocidos en el arte. Los promotores ilustrativos y preferidos para las proteínas cristalinas de *B thuringiensis* incluyen los promotores de los genes *sigA*, *sigE* y *sigK*. Como alternativa, se pueden usar promotores nativos, mutados, heterólogos o recombinantes derivados de *Bacillus thuringiensis* o de otras especies de *Bacillus* para expresar las proteínas de la presente invención en una cepa de especies de *Bacillus*.

Cuando se utilizará la secuencia de nucleótidos que codifica todo o una parte insecticida de una proteína de la presente invención para transformar una planta, se selecciona un promotor que tenga la capacidad para dirigir la expresión de la secuencia de codificación en dicha especie de planta particular Los promotores que funcionan en diferentes especies vegetales también son bien conocidos en el arte Los promotores que son de utilidad en la expresión de polipéptidos en plantas son aquellos que son inducibles, virales, sintéticos o constitutivos como se describe en Odell *et al.* (*Nature* 313 810-812, 1985), y/o promotores regulados transitoriamente, espacialmente o espacio-temporalmente Los promotores preferidos incluyen los promotores CaMV35S mejorados, y el promotor FMV35S. Para un control óptimo de especies de gusanos de la raíz por expresión de las proteínas de la presente invención en plantas, es preferible alcanzar los niveles más altos de expresión de estas proteínas en las raíces de las plantas de maíz. Se han identificado numerosos promotores mejorados para raíces y son conocidos en el arte (Lu et al , *J. Plant Phys* , 2000, 156(2):277-283; Patente de EE.UU Nº 5 837.848; Patente de EE.UU.

Nº:6 489.542). Una regulación temporal o espacial sustancial se refiere a la expresión de un gen en una planta o tejido vegetal a partir de un promotor funcional en plantas Con referencia a la regulación temporal, el promotor puede ser regulado para expresarse solamente durante momentos específicos durante el crecimiento y desarrollo de la célula o tejido vegetal o aún toda la planta. Un promotor que permite expresar activamente a uno o más genes pero sólo durante la germinación de las semillas constituye un ejemplo de regulación temporal Otros ejemplos incluyen promotores que expresan activamente uno o más genes pero sólo durante los momentos en que la planta, célula vegetal o tejido vegetal es expuesto a determinadas intensidades de luz o durante la oscuridad total. Una regulación temporal sustancial se refiere a un promotor que se expresa activamente en un momento determinado pero que puede estar completamente suprimido, o no, en otros momentos, de modo tal que la expresión aún puede ser detectada monitoreando la presencia de algún indicador, tal como una enzima producida por una secuencia de codificación ligada a dicho promotor, o midiendo el incremento o la disminución de algún producto genético, tal como un ARNm producido en distintos momentos durante todo el período de crecimiento, diferenciación y desarrollo de la planta y/o como respuesta a diversos estímulos ambientales Una regulación espacial sustancial se refiere a la expresión de un gen ligado a un promotor que solamente permite la expresión durante el crecimiento y desarrollo de determinadas células o tejidos en una planta. Por ejemplo, un promotor del tapete sólo se expresará de una forma sustancialmente espacial durante el crecimiento y desarrollo de las flores De manera similar, un promotor específico de raíces o mejorado para raíces sólo se expresará de una forma sustancialmente espacial en las células de las raíces o en los tejidos radiculares. El término regulado sustancialmente de forma espacial también se refiere al nivel de expresión de un promotor específico de un tejido particular en dicho tejido particular y relacionado con niveles de expresión de dicho promotor o un promotor similar en otros tejidos, donde dicha expresión también

puede ser detectada en tejidos distintos de dicho tejido particular en el cual se prefiere la expresión del promotor, pero a niveles de expresión significativamente más bajos medidos por la producción de una enzima producida a partir de una secuencia de codificación ligada al promotor o por aparición de algunos productos genéticos detectables Los promotores también pueden estar regulados sustancialmente en el tiempo y sustancialmente en el espacio juntos y simultáneamente de una manera coordinadamente regulada. Otros promotores que específicamente se consideran dentro del alcance de la presente invención incluyen, sin limitaciones, el promotor de ubiquitina, el promotor del virus a ADN bacilliforme de caña de azúcar, el promotor de la subunidad grande de la ribulosa bifosfato carboxilasa, entre otros.

En la construcción de expresión de ADN también se pueden incluir las secuencias preferidas de intrones para lograr una expresión óptima de secuencias de nucleótidos no naturales en plantas monocotiledóneas Dicho intrón se encuentra típicamente cerca del extremo 5 del ARNm dentro o inmediatamente 3' con respecto a la secuencia no traducida. El intrón se puede obtener, por ejemplo, del conjunto de intrones formado por el intrón 70 de la proteína de shock de calor de maíz (HSP) (Patente de EE.UU. Nº.5.424.412; 1995), el intrón de *Act1* de arroz (McElroy *et al.*, *Plant Cell* 2 163-171, 1990), el intrón 1 Adh (Callis *et al.*, *Genes & Develop.* 1:1183-1200, 1987) o el intrón de la sacarosa sintetasa (Vasil *et al*, *Plant Phys* 91·1575-1579, 1989).

Otro elemento que funciona en la regulación o modulación de la expresión genética es la secuencia de ADN entre el sitio de inicio de la transcripción y el inicio de la secuencia de codificación, denominado secuencia directriz no traducida (UTL). Se han efectuado compilaciones de secuencias directrices con el fin predecir las secuencias óptimas o subóptimas y generar secuencias directrices "consenso" y preferidas (Joshi, *Nucl. Acids Res* 15·9627-9640, 1987). Se contempla que las secuencias directrices preferidas incluyen las que comprenden secuencias predichas

para dirigir la expresión óptima del gen estructural ligado, es decir, incluyen una secuencia directriz consenso preferida que incrementa o conserva la estabilidad del ARNm e impide un inicio inapropiado de la traducción La elección de dichas secuencias resultará conocida para los especialistas en el arte a la luz de la presente descripción. Las secuencias que derivan de genes de gran expresión en plantas, y en particular en maíz, son las más preferidas. Una secuencia directriz de particular utilidad es la directriz HSP70 de petunia.

Se pueden utilizar potenciadores de la transcripción o duplicaciones de potenciadores para incrementar la expresión. Estos potenciadores a menudo se encuentran 5′ con respecto al inicio de la transcripción en un promotor funcional en células eucariotas, pero a menudo se pueden insertar con orientación 5′ o 3′ directa o inversa con respecto a la secuencia de codificación. Los ejemplos de potenciadores incluyen los elementos del promotor CaMV 35S, el promotor de los genes de la octopina sintetasa (Ellis *et al.*, *EMBO Journal* 6:11-16, 1987), el gen de actina de arroz y un promotor de eucariotas no vegetales (por ejemplo, levadura, Ma *et al.*, *Nature* 334.631-633, 1988).

La ARN polimerasa transcribe una secuencia de codificación de ADN del genoma nuclear en el sitio donde se produce la poliadenilación La ARN polimerasa transcribe una secuencia de codificación de ADN del genoma nuclear en el sitio donde se produce la poliadenilación. Típicamente, las secuencias de ADN ubicadas unos cientos de pares de bases hacia el extremo 3' con respecto al sitio de poliadenilación sirven para terminar la transcripción. Dichas secuencias de ADN se denominan regiones de terminación de la transcripción en la presente Dichas regiones son necesarias para la poliadenilación eficiente del ARN mensajero transcripto (ARNm) nuclear. Para las secuencias de codificación introducidas en cloroplastos o plástidos, o en el genoma de cloroplastos o plástidos, la terminación de la transcripción del ARNm es similar a los métodos bien conocidos en el arte de la expresión genética en bacterias Por ejemplo, ya sea en una secuencia policistrónica

o monocistrónica, la transcripción se puede terminar con estructuras de bucles o estructuras similares a secuencias dependientes de *rho* bacterianas. Típicamente, las secuencias de ADN ubicadas unos cientos de pares de bases hacia el extremo 3' con respecto al sitio de poliadenilación sirven para terminar la transcripción. Dichas secuencias de ADN se denominan regiones de terminación de la transcripción en la presente. Dichas regiones son necesarias para una poliadenilación eficiente del ARN mensajero transcripto (ARNm) nuclear. Para las secuencias de codificación introducidas en cloroplastos o plástidos, o en el genoma de cloroplastos o plástidos, la terminación de la transcripción del ARNm es similar a los métodos bien conocidos en el arte de la expresión genética en bacterias. Por ejemplo, ya sea en una secuencia policistrónica o monocistrónica, la transcripción se puede terminar con estructuras de bucles o estructuras similares a secuencias dependientes de *rho* bacterianas.-{}-

Las construcciones de expresión incluirán típicamente una secuencia de codificación ilustrada en la presente invención o un derivado de la misma junto con una ADN secuencia 3' terminal que funciona como una señal para terminar la transcripción y, en las construcciones pretendidas para su expresión desde el genoma nuclear de la planta, permite la poliadenilación del extremo 3' del transcripto de ARN resultante Se considera que los elementos 3' más preferidos son los del gen de la nopalina sintetasa de *A tumefaciens* (extremo 3' *nos*), el terminador del transcripto T7 del gen de la octopina sintetasa de *A. tumefaciens* y la secuencia de terminación de la transcripción y poliadenilación no traducida 3' del gen E9 (E9 3') de la RUBISCO sintetasa de arveja. Estas y otras secuencias reguladoras 3' son bien conocidas en el arte.

Los vectores de transformación vegetales preferidos incluyen los que derivan de un plásmido Ti de *Agrobacterium tumefaciens*, así como los descriptos, por ejemplo, por Herrera-Estrella (*Nature* 303:209-213, 1983), Bevan (*Nature* 304:184-187,1983), Klee (*Bio/Technol* 3:637-642, 1985) y Solicitud de Patente de Europa Nº.

EP 0120516 (cada una incorporada específicamente en este documento a modo de referencia).

La presente invención describe secuencias de nucleótidos aisladas y purificadas que codifican proteínas insecticidas derivadas de especies de *Bacillus* y en particular de la especie *Bacillus thuringiensis*. En particular, en la presente se demuestra que las cepas 86833, EG2158, EG3618, EG6489, EG6561, EG6618 y EG4653 de *B. thuringiensis* producen una o más proteínas insecticidas solubles localizadas en los sobrenadantes de cultivo (véase la Tabla 1, excepto para EG4653 que se describirá con detalle en el Ejemplo 11).

Las cepas de *B. thuringiensis* y otras cepas bacterianas descritas en la presente se pueden cultivar usando medios de crecimiento convencionales y técnicas de fermentación estándar Las cepas de *B. thuringiensis* que contienen uno o más genes *tic*901, *tic*1201, *tic*407, *tic*417, *tic*431 o genes relacionados, se pueden fermentar como se describe en la presente hasta que las células cultivadas de *B. thuringiensis* alcanzan la etapa de su ciclo de crecimiento en la cual se producen las TIC901, TIC1201, TIC407, TIC417, TIC431 y/o proteínas relacionadas

Los cultivos se han depositado en condiciones que, durante la tramitación de la publicación de la presente solicitud y patentes, aseguren el posible acceso a los mismos Sin embargo, se debe entender que la asequibilidad a los depósitos no constituye licencia para practicar la presente invención, anulando derechos de patente otorgados por acción gubernamental.

Además, estos depósitos de los microorganismos de la presente invención se efectuaron de acuerdo con las disposiciones del "Tratado de Budapest sobre el reconocimiento internacional del depósito de microorganismos para el procedimiento de tramitación de patentes". Los depósitos del cultivo de la invención se guardarán y estarán disponibles para el público de acuerdo con las disposiciones del Tratado de Budapest, es decir, se guardarán con todo el cuidado necesario para mantenerlos en buenas condiciones y no contaminados durante un período de por lo menos cinco

años, después de la petición más reciente de actuación de una muestra del depósito y, en cualquier caso, durante un período de por lo menos 30 (treinta) años después de la fecha de depósito o durante la vigencia aplicable de cualquier patente que se pueda otorgar y que revele los cultivos. El depositante reconoce su deber de reponer los depósitos en caso de que, debido al estado de los mismos, el depositario no pudiera proporcionar una muestra que se le solicitara Todas las restricciones en cuanto a disponibilidad al público de los depósitos del cultivo de la invención se eliminarán de manera irrevocable al otorgarse la patente que los revela.

Se demuestra aquí que las proteínas TIC901, TIC1201, TIC407, TIC417, TIC431 y las proteínas relacionadas de la presente invención son producidas y secretadas hacia el medio de crecimiento por numerosas cepas de *Bacillus thuringiensis*. Las fermentaciones con las cepas de la presente invención se pueden continuar durante la etapa de esporulación cuando las proteínas cristalinas, si presentes, se forman junto con las esporas Las esporas y los desechos celulares se pueden separar del sobrenadante por centrifugación, y el medio de cultivo utilizado se pueden empelar para aislar las proteínas insecticidas de la presente invención. Los inventores ilustran aquí el método de precipitación por sulfato de amonio como un medio para concentrar y recolectar todas o la mayor parte de las proteínas presentes en el medio de cultivo usado y clarificado. Sin embargo, el especialista en el arte comprenderá que existen numerosas maneras diferentes para purificar y aislar las proteínas de la presente invención La filtración por gel y la cromatografía por exclusión de tamaño son dos medios disponibles para extraer las proteínas directamente del medio usado. El medio usado también puede ser desalado y el filtrado se puede usar para extraer las proteínas usando columnas de intercambio iónico. Además, se pueden emplear columnas de afinidad, que contienen anticuerpos que se unen específicamente a las proteínas TIC901, TIC1201, TIC407, TIC417, TIC431 o proteínas relacionadas para purificar las proteínas de la presente invención directamente a partir del medio

Las secuencias de aminoácidos de la presente invención fueron comparadas con las secuencias de aminoácidos presentes en las bases de datos de secuencias de proteínas disponibles comercialmente y no se identificaron homologías o similitudes significativas Sobre la base de este análisis, las proteínas TIC901, TIC1201, TIC407, TIC417 y TIC431, y secuencias relacionadas, aparecen como únicas y forman la base para establecer una clase nueva y separada de proteínas insecticidas de *Bacillus* porque no se ha observado que las proteínas de la presente invención exhiban ninguna relación significativa con otras proteínas insecticidas conocidas

Se pueden efectuar modificaciones y cambios en la estructura de los péptidos de la presente invención y en los segmentos de ADN que los codifica y aún obtener una molécula funcional que codifica una proteína o péptido con las características deseadas. Los péptidos, polipéptidos y proteínas equivalentes biológica y funcionalmente contemplados aquí deben presentar entre aproximadamente un 65 y aproximadamente un 70% o más de similitud de secuencias de aminoácidos, o entre aproximadamente un 80% o más de similitud de secuencias de aminoácidos, o entre aproximadamente un 90% o más de similitud de secuencias de aminoácidos, con la secuencia, o una porción correspondiente en la misma, de la secuencia de aminoácidos TIC901 fundamental que se muestra en la SEQ ID Nº.4, o la porción correspondiente en las secuencias de aminoácidos de TIC1201, TIC407, TIC417 y TIC431 que se muestran, respectivamente, en las SEQ ID Nº:6, SEQ ID Nº:8, SEQ ID Nº:10 y SEQ ID Nº:32 y secuencias relacionadas

De acuerdo con la presente invención, la referencia al gen *tic*901 y la toxina proteica codificada, no sólo incluye las secuencias de longitud completa descritas en la presente sino también fragmentos de estas secuencias, variantes naturales, mutantes y derivados recombinantes o manipulados genéticamente del gen *tic*901 que comprende la SEQ ID Nº 3 Las proteínas codificadas por estas secuencias deberían retener esencialmente las mismas propiedades insecticidas características,

o mejores, que las de la proteína TIC901 que comprende la SEQ ID Nº·4. Las proteínas de utilidad en la presente invención también pueden incluir proteínas de fusión capaces de retener propiedades insecticidas características que son esencialmente las mismas, o mejores, que las de la proteína TIC901. En algunos casos, la proteína de fusión puede contener, además de las propiedades insecticidas características de las proteínas que ilustran específicamente en la presente, otra actividad insecticida contribuida por la otra secuencia de aminoácidos del par de fusión. Como alternativa, el análisis cristalográfico de la proteína TIC901 o variantes insecticidas de la misma puede proveer un medio para determinar si la proteína sería un candidato para la construcción de una permuteína con la misma, o preferentemente mayor, actividad insecticida que la TIC901 nativa o una proteína relacionada, y que preferentemente presenta características mejoradas con relación a la expresión en una célula huésped preferida, tal como una célula vegetal Las mismas calidades y características se aplican también a *tic*1201, *tic*407 y *tic*417; no sólo a las secuencias de longitud completa descritas en la presente sino también a los fragmentos de estas secuencias, variantes naturales, mutantes y derivados recombinantes o manipulados genéticamente de estos genes que comprenden, respectivamente, las SEQ ID Nº 5, SEQ ID Nº.7, SEQ ID Nº:9 y SEQ ID Nº.32. Las proteínas codificadas por estas secuencias deben retener esencialmente las mismas, o mejores, propiedades insecticidas características que las de las proteínas TIC1201, TIC407, TIC417 y TIC431 que comprenden, respectivamente, las SEQ ID Nº:6, SEQ ID Nº.8, SEQ ID Nº:10 y SEQ ID Nº 32.

Será evidente para el especialista en el arte que las secuencias de nucleótidos que codifican toxinas inhibidoras de insectos, y en particular toxinas inhibidoras de coleópteros, se pueden identificar y obtener de varias maneras, como se describe en la presente. Las secuencias específicas y secuencias relacionadas ilustradas aquí se pueden obtener a partir de las formas aisladas depositadas un depósito de cultivos, como se describió previamente. Estas secuencias, o porciones insecticidas o

variantes de las mismas, también se pueden construir sintéticamente, por ejemplo, utilizando un sintetizador de secuencias de nucleótidos Las variaciones de estas secuencias de codificación se pueden construir fácilmente utilizando las técnicas estándar para obtener mutaciones puntuales También se pueden elaborar fragmentos de estas secuencias utilizando las exonucleasas o endonucleasas disponibles comercialmente de acuerdo con procedimientos estándar. Por ejemplo, se pueden utilizar enzimas, como la *Bal*31, o mutagénesis dirigida al sitio para recortar sistemáticamente nucleótidos de los extremos de dichas secuencias ilustradas aquí o de la secuencia de codificación de la proteína. También se pueden obtener secuencias de nucleótidos que codifican fragmentos de proteínas activas como insecticidas usando diversas enzimas de restricción, endonucleasas, métodos de amplificación térmica y semejantes. Las proteasas, tal como proteinasa K, tripsina, quimiotripsina, pepsina y semejantes se pueden utilizar para obtener directamente fragmentos activos de estas toxinas.

Se pueden derivar otras toxinas y las secuencias de nucleótidos que codifican dichas toxinas relacionadas con las toxinas y secuencias de codificación de la presente invención a partir del ADN obtenido de *B. thuringiensis*, *B. laterosporous*, *B sphaericus*, y formas aisladas de especies relacionadas de *Bacillus* usando las descripciones provistas en el arte en combinación con las secuencias de nucleótidos descritas en la presente. Dichas toxinas y secuencias de nucleótidos relacionadas con las toxinas y las secuencias de codificación de la presente invención se consideran equivalentes en la presente de las toxinas y secuencias de nucleótidos de la presente invención. El término "equivalente" significa que una proteína presenta las características de una o más de las proteínas descritas en la presente, incluyendo, sin limitaciones, una bioactividad inhibidora insecticida similar, rango de huéspedes para la bioactividad insecticida, presenta epitopes antigénicos similares que reaccionan de manera cruzada con anticuerpos generados contra TIC901, TIC1201, TIC407 y TIC417 y semejantes, e incluyendo las proteínas relacionadas,

presentan un tamaño similar con relación a TIC901 y de las proteínas relacionadas, presentan perfiles de expresión y características similares, presentan una tendencia de aislamiento hacia el entorno extracelular cuando se expresan en *Bacillus thuringiensis* o especies bacterianas relacionadas, y semejantes. La frase "presentan una tendencia de aislamiento hacia el entorno extracelular" pretende incluir la proteína TIC901 y las proteínas relacionadas incluyendo, sin limitaciones, TIC1201, TIC407, TIC417 y TIC431 y semejantes, que son producidas por la bacteria o célula huésped como una proteína precursora que contiene una secuencia de aminoácidos ligada a la proteína insecticida cuya función consiste en dirigir la proteína insecticida hacia el aparato secretor de dicha bacteria o célula huésped y que, cuando toma contacto con el aparato secretor, es clivado proteolíticamente por una peptidasa señal, liberando la proteína madura o insecticida hacia el entorno extracelular en el caso de un microbio Gram-positivo, al menos hacia el periplasma en el caso de un microbio Gram-negativo y hacia el retículo endoplasmático o vesículas secretoras o hacia un organela subcelular, tal como mitocondrias o cloroplastos o plástidos, en el caso de una célula huésped fúngica o vegetal o de otro eucariota. Las secuencias de nucleótidos codificadoras crípticas también se consideran dentro del alcance de la presente invención

Existen numerosos métodos para identificar la presencia y obtener toxinas insecticidas equivalentes relacionadas con los péptidos descritos en la presente Por ejemplo, se pueden usar anticuerpos para las toxinas insecticidas descritas y reivindicadas en la presente para identificar y aislar otras toxinas de una mezcla de proteínas Específicamente, se pueden generar anticuerpos contra las porciones de las toxinas que son más constantes dentro de la nueva clase de proteínas y más distintas de otras toxinas de *B. thuringiensis*. Estos anticuerpos se pueden usar luego para identificar específicamente toxinas equivalentes con la actividad característica mediante inmunoprecipitación, ensayos de inmunoabsorción ligada a enzimas (ELISA) o transferencia Western. Los anticuerpos para las toxinas descritas en la

presente, o para toxinas equivalentes, o fragmentos de estas toxinas, se pueden preparar fácilmente usando los procedimientos estándar del arte. Las secuencias de nucleótidos que codifican estas toxinas se pueden obtener luego a partir del microorganismo o de diversas fuentes adicionales

Otro método adicional para identificar las toxinas y los genes de la presente invención comprende el uso de sondas de oligonucleótidos Estas sondas son esencialmente secuencias de nucleótidos que se hibridizan bajo condiciones de hibridación severas con la secuencia de codificación TIC901 o una secuencia relacionada con la secuencia de codificación de TIC901. Como es de conocimiento general en el arte, si la molécula de la sonda y la molécula de la secuencia de ácido nucleico en una muestra se hibridizan formando un enlace suficientemente fuerte entre las dos moléculas, se puede suponer razonablemente que las dos moléculas exhiben una homología sustancial La unión a la sonda se detecta usando cualquiera de los numerosos medios conocidos en el arte incluyendo, sin limitaciones, fluorescencia, luminiscencia, isótopos, inmunología, espectroscopía de resonancia de plasmón de superficie y semejantes. Dicho análisis de la sonda provee un método rápido para identificar los genes que codifican las toxinas de la presente invención. Los segmentos de nucleótidos que se usan como sondas de acuerdo con la invención se pueden sintetizar utilizando sintetizadores de ADN con procedimientos estándar u otros medios conocidos en el arte Estas secuencias de nucleótidos también se pueden usar como cebadores para PCR para amplificar las secuencias de nucleótidos de la presente invención o porciones de las mismas

Los fragmentos y equivalentes de las proteínas relacionadas que retienen al menos la actividad insecticida de las toxinas ilustradas se encuentran dentro del alcance de la presente invención Además, dada la redundancia del código genético, es posible codificar una gran variedad de secuencias de ADN diferentes que codifican las secuencias de aminoácidos descritas en la presente. Es propio de la habilidad de un especialista en el arte crear estas secuencias de ADN alternativas

que codifican las mismas, o esencialmente las mismas, toxinas. Estas variantes de las secuencias de ADN se encuentran dentro del alcance de la presente invención.

Es un hecho bien conocido en el arte que se pueden sustituir determinados aminoácidos por otros aminoácidos en la estructura de la proteína sin una pérdida apreciable de la capacidad de unión interactiva con estructuras tales como, por ejemplo, las regiones de unión del antígeno de los anticuerpos o los sitios de unión en las moléculas de substrato. Dado que es la capacidad interactiva y la naturaleza de una proteína que define la actividad biológica funcional de dicha proteína, se pueden efectuar determinadas sustituciones en la secuencia de aminoácidos en la secuencia de la proteína, y por supuesto en la secuencia de codificación del ADN subyacente, y no obstante obtener una proteína con propiedades similares. Los inventores han contemplado de esta manera que se pueden efectuar varios cambios en las secuencias de los péptidos de las composiciones descriptas, o en las secuencias de ADN correspondientes que codifican dichos péptidos, sin una pérdida apreciable de su actividad o utilidad biológica Dichas sustituciones también se conocen en el arte como sustituciones conservadoras

Al efectuar dichos cambios, se puede considerar el índice hidropático de los aminoácidos. La importancia del índice de aminoácidos hidropáticos en conferirle la función biológica interactiva a una proteína es un hecho comprendido en el arte (Kyte y Doolittle, 1982) Se acepta que el carácter relativamente hidropático de los aminoácidos contribuye a la estructura secundaria de la proteína resultante, que a su vez define la interacción de la proteína con otras moléculas, por ejemplo, enzimas, substratos, receptores, ADN, anticuerpos, antígenos y semejantes.

A cada aminoácido se le asignó un índice hidropático sobre la base de su hidrofobicidad y características de carga (Kyte y Doolittle, 1982), ellos son: isoleucina (+4.5), valina (+4.2); leucina (+3.8), fenilalanina (+2.8), cisteína/cistina (+2.5); metionina (+1.9), alanina (+1 8), glicina (–0.4), treonina (–0 7), serina (–0 8); triptofano (–0 9); tirosina (–1 3), prolina (–1 6); histidina (–3 2), glutamato (–3 5),

glutamina (–3.5); aspartato (–3.5); asparagina (–3 5), lisina (–3 9) y arginina (–4.5).

Es un hecho conocido en el arte que determinados aminoácidos pueden sustituirse por otros aminoácidos que poseen un índice o calificación hidropático similar y aún dar como resultado una proteína con actividad biológica similar, es decir, que aún se puede obtener una proteína con funcionalidad biológica equivalente Al efectuar dichos cambios, se prefiere la substitución de aminoácidos cuyos índices hidropáticos se hallan entre ±2, se prefieren particularmente aquellos que se hallan entre ±1 y son más preferidos aún aquellos que se hallan entre ±0.5.

También se sabe en el arte que la substitución de aminoácidos similares se puede realizar eficazmente sobre la base de su hidrofilicidad. La mayor hidrofilicidad promedio local de una proteína, gobernada por la hidrofilicidad de los aminoácidos adyacentes, se correlaciona con una propiedad biológica de la proteína (Patente de EE.UU Nº:4.554.101).

Como se describió en detalle en la Patente de EE UU Nº·4.554.101, se han asignado los siguientes valores de hidrofilicidad a estos residuos aminoácidos. arginina (+3 0); lisina (+3 0); aspartato (+3 0 ± 1); glutamato (+3 0 ± 1); serina (+0 3), asparagina (+0.2), glutamina (+0.2), glicina (0), treonina (–0.4); prolina (–0.5 ± 1), alanina (–0.5); histidina (–0.5); cisteína (–1 0); metionina (–1 3); valina (–1.5); leucina (–1 8); isoleucina (–1.8), tirosina (–2.3); fenilalanina (–2 5), triptofano (–3 4)

Se debe tener en cuenta que es posible sustituir un aminoácido por otro que posee un valor de hidrofilicidad similar y aún obtener un equivalente biológico y, en particular, una proteína equivalente inmunológicamente. Entre dichos cambios, se prefiere la substitución de aminoácidos cuyos valores de hidrofilicidad se hallan entre ±2, se prefiere particularmente aquellos que se hallan entre ±1 y se prefiere aún más aquellos que se hallan entre ±0.5.

Por ello, como se mencionó anteriormente, las sustituciones de aminoácidos se basan en general en la similitud relativa de los sustituyentes de la cadena lateral

de los aminoácidos, por ejemplo, su hidrofobicidad, hidrofilicidad, carga, tamaño y semejantes. Las substituciones ilustrativas que tienen en cuenta varias de las características recién mencionadas son bien conocidas por los expertos en el arte e incluyen· arginina y lisina; glutamato y aspartato; serina y treonina; glutamina y asparagina, y valina, leucina e isoleucina.

Los péptidos, polipéptidos y proteínas biológicamente funcionales que son equivalentes de TIC901, TIC1201, TIC407, TIC417, TIC431 y proteínas relacionadas, y semejantes, incluyen secuencias de aminoácidos que contienen cambios de aminoácidos conservadores en la secuencia fundamental que se muestra en las SEQ ID Nº: 4, SEQ ID Nº:6, SEQ ID Nº·8, SEQ ID Nº·10 y SEQ ID Nº:33 En particular, con referencia a la SEQ ID Nº·4 entre el aminoácido 44 aproximadamente y el aminoácido 367 aproximadamente, con referencia a la SEQ ID Nº:6 entre el aminoácido 44 aproximadamente y el aminoácido 364 aproximadamente, con referencia a la SEQ ID Nº 8 entre el aminoácido 44 aproximadamente y el aminoácido 368 aproximadamente, con referencia a la SEQ ID Nº:10 entre el aminoácido 44 aproximadamente y el aminoácido 364 aproximadamente y con referencia a la SEQ ID Nº 33 entre el aminoácido 44 aproximadamente y el aminoácido 364 aproximadamente, para tales secuencias de aminoácidos, se puede sustituir uno o más aminoácidos en la secuencia fundamental por otro(s) aminoácido(s), cuya carga y polaridad sean similares a las del aminoácido nativo, es decir, una substitución de aminoácido conservadora, que da como resultado un cambio silente

Los sustitutos de un aminoácido dentro de la secuencia polipeptídica fundamental se pueden seleccionar entre otros miembros de la clase a la cual pertenecen los aminoácidos naturales. Los aminoácidos se pueden dividir en los siguientes cuatro grupos: (1) aminoácidos ácidos, (2) aminoácidos básicos, (3) aminoácidos de polaridad neutra y (4) aminoácidos neutros no polares. Los aminoácidos representativos de estos grupos incluyen, pero sin limitaciones: (1)

aminoácidos ácidos (cargados negativamente), tal como ácido aspártico y ácido glutámico, (2) aminoácidos básicos (cargados positivamente), tal como arginina, histidina y lisina; (3) aminoácidos de polaridad neutra, tal como glicina, serina, treonina, cisteína, cistina, tirosina, asparagina y glutamina; (4) aminoácidos neutros no polares (hidrofóbicos), tal como alanina, leucina, isoleucina, valina, prolina, fenilalanina, triptofano y metionina.

Los cambios de aminoácidos conservadores dentro de la secuencia polipeptídica fundamental se pueden efectuar sustituyendo un aminoácido dentro de uno de estos grupos por otro aminoácido perteneciente al mismo grupo Los equivalentes biológicamente funcionales de TIC901, TIC1201, TIC407, TIC417, TIC431 y semejantes y secuencias relacionadas, pueden tener 10 o menos cambios de aminoácidos conservadores, con mayor preferencia siete o menos cambios de aminoácidos conservadores y con mayor preferencia aún cinco o menos cambios de aminoácidos conservadores La secuencia de nucleótidos codificadora (gen, ADN plásmido, ADNc o ADN sintético) tendrá entonces sustituciones básicas correspondientes, permitiendo que codifique formas equivalentes biológicamente funcionales de las proteínas TIC901, TIC1201, TIC407, TIC417 y TIC431 y secuencias relacionadas

Las variantes de las secuencias de aminoácidos de las proteínas de la presente invención y las secuencias relacionadas se pueden obtener utilizando procedimientos bien conocidos en el arte. Por ejemplo, se puede obtener una variante de la secuencia de aminoácidos de la proteína TIC901 que no es secretada hacia el medio extracelular mediante mutagénesis con sulfonato de etilmetano (EMS) de una secuencia de nucleótidos que codifica TIC901. Los mutantes también se pueden construir usando luz ultravioleta y nitrosoguanidina con procedimientos bien conocidos en el arte, o por construcción de una secuencia de codificación que carece de toda o una parte de la secuencia de codificación que codifica la secuencia de aminoácidos del péptido señal

La mutagénesis específica del sitio es una técnica que es de utilidad en la preparación de péptidos individuales, o de proteínas o péptidos equivalentes biológicamente funcionales, mediante la mutagénesis específica del ADN subyacente. La técnica provee además la facilidad de preparar y probar variantes de secuencia, por ejemplo, incorporando una o más de las consideraciones anteriores, introduciendo uno o más cambios en la secuencia de nucleótidos en el ADN. La mutagénesis específica del sitio permite la producción de mutantes por medio del uso de secuencias específicas de oligonucleótidos que codifican la secuencia de ADN de la mutación deseada, así como una cantidad suficiente de nucleótidos adyacentes, para proveer una secuencia cebadora de tamaño y complejidad de secuencia suficiente como para formar un dúplex estable a ambos lados de la secuencia que se desea modificar. Típicamente, se prefiere un cebador de 17 a 25 nucleótidos de longitud aproximadamente, con al menos 5 a 10 residuos aproximadamente a ambos lados de la secuencia que se desea modificar

La preparación de variantes de la secuencia de los segmentos de ADN codificadores del péptido utilizando mutagénesis dirigida al sitio se provee como un medio para producir especies potencialmente útiles y no pretende ser limitativo ya que existen otras maneras para obtener variantes de secuencia de los péptidos y de las secuencias de ADN que las codifica. Por ejemplo, los vectores recombinantes que codifican la secuencia del péptido deseado pueden ser tratados con agentes mutagénicos, como hidroxilamina, para obtener variantes de secuencia.

La secuencia *tic*901 y las secuencias de nucleótidos codificadoras relacionadas aisladas de cepas de *B. thuringiensis* que se muestran aquí en las SEQ ID Nº:3, SEQ ID Nº:5, SEQ ID Nº:7, SEQ ID Nº:9 y SEQ ID Nº.32 y semejantes se pueden utilizar como sondas de hibridación para identificar y aislar variantes naturales de las mismas y secuencias de nucleótidos codificadoras relacionadas de otras cepas de *B. thuringiensis* o de otros microorganismos, tal como de especies microbianas, tal como *Clostridium*, *Bacillus*, *Xenorhabdus* y *Photorhabdus*. La

presente invención abarca las secuencias de nucleótidos de microorganismos, donde las secuencias de nucleótidos pueden ser aisladas por hibridación con toda, o parte, de la secuencia de nucleótidos de *Bacillus* de la invención Las proteínas codificadas por dichas secuencias de nucleótidos pueden ser evaluadas por su actividad insecticida La invención también abarca las proteínas codificadas por las secuencias de nucleótidos. Por ejemplo, una alineación de estas cuatro secuencias de nucleótidos provee regiones de identidad que serían de uso preferido como sondas y cebadores. Además, la alineación de las secuencias de aminoácidos codificadas por estas secuencias y las secuencias de nucleótidos codificadoras relacionadas provee información referida a las regiones de identidad y/o similitud sustancial entre las proteínas, y provee la base para la construcción de sondas y/o cebadores que se pueden usar para identificar estas y otras secuencias de nucleótidos relacionadas que codifican dichas proteínas insecticidas. Por ejemplo, las SEQ ID Nº:23-SEQ ID Nº:29 son representativas de dichas sondas y cebadores y se ilustran en la presente en el Ejemplo 10.

Los anticuerpos generados como respuesta a una prueba inmune con TIC901, TIC1201, TIC407, TIC417 y TIC431 y semejantes o proteínas relacionadas de la presente invención se pueden producir usando técnicas inmunológicas estándar para producir antisueros policlonales y, si se desea, inmortalizar las células productoras de anticuerpos del huésped inmunizado para la producción de fuentes de anticuerpos monoclonales. Las técnicas para producir anticuerpos contra cualquier sustancia de interés son bien conocidos, por ejemplo, como se describe en Harlow y Lane (1988) y como se describe en Goding (1986). Por ejemplo, se pueden utilizar anticuerpos que se unen a epitopes sobre o en TIC901 como sondas para identificar cepas de *B. thuringiensis* u otros microorganismos que producen variantes de TIC901 o proteínas relacionadas codificadas por variaciones de un gen *tic*901 o un gen relacionado. La presente invención abarca las proteínas insecticidas que presentan reacción cruzada con los anticuerpos generados contra una o más de las proteínas insecticidas de la

presente invención.

Los anticuerpos producidos en la presente invención también son de utilidad en inmunoensayos para determinar la cantidad o presencia de una proteína TIC901, TIC1201, TIC407, TIC417 Y TIC431 o una proteína relacionada en una muestra biológica. Dichos ensayos también son de utilidad en la producción con control de calidad de composiciones que contienen una o más de las proteínas de la presente invención o proteínas relacionadas. Además, los anticuerpos se pueden usar para evaluar la eficacia de la producción recombinante de una o más de las proteínas de la presente invención o una proteína relacionada, así como para examinar bibliotecas de expresión por la presencia de una secuencia de nucleótidos codificadora de una o más de las secuencias de codificación de las proteínas de la presente invención o proteínas relacionadas Los anticuerpos son de utilidad además como ligandos de afinidad para purificar y/o aislar cualquiera entre una o más de las proteínas de la presente invención y de las proteínas relacionadas Las proteínas de la presente invención y las proteínas que contienen epitopes antigénicos relacionados se pueden obtener por sobreexpresión de las longitudes completa o parcial de una secuencia que codifica toda o una parte de la proteína de la presente invención o una proteína relacionada en una célula huésped preferida

Los péptidos de la presente invención son de uso primario, aunque no exclusivo, en plantas, y en determinadas realizaciones preferidas, las secuencias de nucleótidos modificadas para codificar las proteínas de la presente invención en plantas están contenidas en uno o más vectores plásmido. Dichos vectores pueden contener diversos de elementos reguladores y otros elementos que permitirán una expresión óptima de las proteínas de la presente invención en células vegetales. Estos elementos adicional pueden incluir promotores, terminadores e intrones, descritos previamente. Se puede seleccionar cualquier vector que contenga la construcción de ADN y cualquier regulador u otros elementos del grupo formado por un cromosoma artificial de levadura, cromosoma artificial bacteriano, un plásmido o

un cósmido y semejantes. Además, los propios vectores de expresión pueden ser de diversas formas. Estas formas pueden diferir por muchas razones y probablemente estarán compuestas por diversos componentes dependiendo de si se pretende usarlos para transformar una planta monocotiledónea o una planta dicotiledónea

Los vectores considerados también dentro del alcance de la presente invención incluyen los vectores que contienen *tic*901, *tic*1201, *tic*407, *tic*417 o las composiciones de ácido nucleico relacionadas descritas previamente, así como cualquier otra construcción de ADN que además comprenda regiones de codificación de otras proteínas insecticidas derivadas de especies de *Bacillus* que se pueden expresar en plantas.

La secuencia de nucleótidos que codifica la proteína TIC901 (SEQ ID Nº: 3 que codifica la SEQ ID Nº.4) o que codifica una secuencia de péptidos relacionada tal como TIC1201 (SEQ ID Nº.5 que codifica la SEQ ID Nº:6), TIC407 (SEQ ID Nº·7 que codifica la SEQ ID Nº 8), TIC417 (SEQ ID Nº·9 que codifica la SEQ ID Nº:10) y TIC431 (SEQ ID Nº:32 que codifica la SEQ ID Nº:33) puede introducirse en diversos huéspedes de microorganismos sin experimentos excesivos, usando los procedimientos bien conocidos por los especialistas en la técnica de transformación de huéspedes adecuados y bajo condiciones que permitan un mantenimiento y expresión estable de la secuencia de nucleótidos introducida (Sambrook et al , 1989, Molecular Cloning A Laboratory Manual 2ª Ed., Cold Spring Harbor Press, Nueva York). Los huéspedes adecuados que permiten expresar las proteínas de la presente invención y las secuencias relacionadas incluyen *B thuringiensis* y otras especies de *Bacillus* tal como *Bacillus subtilis*, *Bacillus sphaericus*, *Bacillus laterosporous*, *Bacillus megaterium* o *Bacillus anthracis*. Los microorganismos genéticamente alteradas o manipulados que contienen un gen codificador de una o más de las proteínas de la presente invención, incluyendo TIC901, TIC1201, TIC407, TIC417 y TIC431 y semejantes, también pueden contener secuencias de nucleótidos que codifican otras proteínas de toxinas presentes en el mismo microorganismo; estas

secuencias de codificación pueden producir de forma concurrente proteínas insecticidas diferentes de las proteínas de la presente invención o proteínas relacionadas En particular, sería preferible producir dos o más proteínas insecticidas diferentes en una célula huésped, donde cada proteína es tóxica para la misma especie de insecto y cada proteína presenta un modo de acción diferente de los demás.

También se pueden emplear microorganismos colonizadores de plantas o de raíces como células huésped para la producción de una o más de las proteínas de la presente invención o proteínas relacionadas. Los microorganismos huésped ilustrativos para los genes de toxinas de *B. thuringiensis* incluyen el microbio colonizador de plantas *Clavibacter xyli* as descrito por Turner et al. (1993, *Endophytes. an alternative genome for crop improvement*; International Crop Science Congress, Ames, Iowa, EE.UU , 14-22 de julio, 1992., págs, 555-560) y las cepas de pseudomonas colonizadoras de raíces, descritas por Obukowicz et al (Patente de EE.UU. Nº·5.229 112)

Las secuencias de nucleótidos que codifican toxinas que se pueden obtener de las formas aisladas de la presente invención se pueden introducir en una amplia variedad de huéspedes microbianos o vegetales. La expresión del gen de la toxina da como resultado, directa o indirectamente, la producción y el mantenimiento intracelular del plaguicida Con huéspedes microbianos adecuados, por ejemplo, *Pseudomonas*, los microbios se pueden aplicar al sitio de la plaga, donde proliferarán y serán ingeridas por la plaga. El resultado es el control de la plaga observado por menores daños en las plantas, mayor rendimiento de las plantas, menor preponderancia de la plaga vegetal en el entorno local general del organismo transgénico que expresa la(s) proteína(s) de toxina(s), y la muerte o detención del crecimiento de la plaga vegetal, en general sin ningún efecto adicional sobre la flora microbiana que rodea a la planta o al organismo transgénico que expresa la(s) proteína(s) toxina(s), y sin ningún efecto adicional sobre el medio ambiente en

general. Como alternativa, el microbio que contiene al gen de la toxina puede ser tratado bajo condiciones que prolongan la actividad de la toxina y estabilizan a la célula. La célula tratada, que retiene la actividad tóxica, puede ser aplicado entonces al medio que rodea a la plaga de interés.

Cuando el gen de toxina de la presente invención, o una secuencia de nucleótidos codificadora relacionada, es introducido por medio de un vector adecuado en un huésped microbiano, y el huésped es aplicado al medio en estado vivo, resulta ventajoso usar determinados microbios huésped Por ejemplo, se pueden seleccionar huéspedes de microorganismos que ocupan el hábitat de la plaga. Los huéspedes de microorganismos también pueden vivir en simbiosis con una especie de plaga específica Estos microorganismos se seleccionan de modo tal que puedan competir exitosamente en el entorno particular con los microorganismos de tipo salvaje, proveer el mantenimiento y expresión estables del gen que expresa al polipéptido plaguicida y, si fuera conveniente, proveer una mayor protección al plaguicida de la degradación e inactivación ambiental.

Se conoce un gran número de microorganismos que viven en el hábitat de las plagas. Estos microorganismos incluyen bacterias, algas y hongos De particular interés son los microorganismos, tal como bacterias, por ejemplo, los géneros *Bacillus*, *Escherichia*, *Pseudomonas*, *Erwinia*, *Serratia*, *Klebsiella*, *Salmonella*, *Pasteurella*, *Xanthomonas*, *Streptomyces*, *Rhizobium*, *Rhodopseudomonas*, *Methylophilius*, *Agrobacterium*, *Acetobacter*, *Lactobacillus*, *Arthrobacter*, *Azotobacter*, *Leuconostoc* y *Alcaligenes*; los hongos, por ejemplo, de los géneros *Metarhizium*, *Bavaria*, *Saccharomyces*, *Cryptococcus*, *Kluyveromyces*, *Sporobolomyces*, *Rhodotorula* y *Aureobasidium*.

Existe una amplia variedad de medios para introducir un gen de toxina que codificador de una toxina en un microorganismo huésped bajo condiciones que permitan el mantenimiento y expresión estables del gen. Estos métodos son bien conocidos por los especialistas en el arte y se describen, por ejemplo, en la Patente

de EE UU. Nº:5.135.867.

Como se expresara precedentemente, las células de *B. thuringiensis* o células recombinantes que expresan una proteína de la presente invención o una proteína de toxina relacionada pueden ser tratadas para prolongar la actividad de la toxina y estabilizar la célula La microcápsula de plaguicida que se forma comprende una o más de las proteínas de la presente invención o una o más toxinas relacionadas dentro de una estructura que ha sido estabilizada y que funciona protegiendo la toxina o las toxinas cuando dicha microcápsula es aplicada al entorno de la plaga de interés Las células huésped adecuadas pueden incluir procariotas o eucariotas, pero habitualmente están limitadas a aquellas células que no produzcan sustancias tóxicas para los organismos superiores, como por ejemplo mamíferos. Sin embargo, los organismos que producen sustancias que son tóxicas para los organismos superiores se pueden utilizar cuando la toxina es inestable o cuando el nivel de aplicación es suficientemente bajo como para evitar toda posibilidad de toxicidad para un huésped mamífero. Como huéspedes son de particular interés los organismos procariotas y los eucariotas inferiores, como por ejemplo los hongos Las células de estos organismos habitualmente estarán intactos y se encontrarán sustancialmente en la forma proliferativa en el momento del tratamiento, en lugar de la forma de esporas, si bien es posible emplear esporas en algunos casos. Dichas microcápsulas también pueden contener una o más de las proteínas de la presente invención o una o más proteínas relacionadas junto con una o más composiciones de proteínas insecticidas no relacionas incluyendo, sin limitaciones, delta-endotoxinas tal como Cry1, Cry2, Cry3, Cry9, Cry22, ET70, TIC851 y/o toxinas binarias tal como ET80/76, ET33/34, PS149B1 y ET100/101 y semejantes, y proteínas insecticidas o complejos de proteínas insecticidas de organismos tan diversos como *Xenorhabdus* y/o *Photorhabdus*, o toxinas proteicas VIP, WAR y/o MIS y proteínas relacionadas

El tratamiento de las células microbianas, por ejemplo, un microbio que contiene una secuencia de nucleótidos o un segmento de nucleótidos de la presente

invención o una secuencia de codificación relacionada, puede ser por medios químicos o físicos o mediante una combinación de medios químicos y/o físicos, en tanto la técnica no afecte de forma deletérea las propiedades de la toxina, ni disminuya la capacidad celular para proteger la toxina. Los ejemplos de reactivos químicos son agentes halogenantes, en particular los halógenos con número atómico 17-80 Más particularmente, se puede utilizar yodo bajo condiciones suaves y durante el tiempo suficiente como para lograr los resultados deseados Otras técnicas adecuadas incluyen el tratamiento con aldehídos, tal como glutaraldehído; compuestos antiinfecciosos, tales como cloruro de zefirano y cloruro de cetilpiridina; alcoholes, tales como isopropilo y etanol; varios fijadores histológicos, tales como yodo Lugol, fijador de Bouin, diversos ácidos y fijador de Helly, (véase. Humason, 1967); o una combinación de agentes físicos (calor) y químicos que conservan y prolongan la actividad de la toxina producida en la célula cuando la célula se administra a un huésped animal. Los ejemplos de medios físicos son radiación de longitud de onda corta como la radiación gamma y radiación X, congelamiento, irradiación UV, liofilización y análogos. Los métodos de tratamiento de células microbianas se describen en las Patentes de EE UU. Nº 4.695 455 y 4.695 462.

Las células tendrán generalmente una mayor estabilidad estructural que aumentará la resistencia a las condiciones ambientales. Cuando el plaguicida es una proforma o una forma precursora, el método de tratamiento de las células debe seleccionarse para que el patógeno de interés no inhiba el procesamiento de dicha proforma en la forma madura del plaguicida Por ejemplo, el formaldehído entrecruzará las proteínas y podría inhibir el procesamiento de la proforma de un polipéptido plaguicida. El método de tratamiento de las célula retiene al menos una porción sustancial de la biodisponibilidad o bioactividad de la toxina.

Las características de interés particular en la selección de una célula huésped para fines productivos, incluyen la facilidad de introducción de una o más de las secuencias de codificación de la presente invención o una o más secuencias de

codificación relacionadas en el huésped, la disponibilidad de los sistemas de expresión, la eficacia de la expresión, la estabilidad del plaguicida en el huésped y la presencia de habilidades genéticas auxiliares. Las características de interés para el uso como plaguicida microencapsulada incluyen las cualidades protectoras para el plaguicida, como por ejemplo paredes celulares gruesas, la pigmentación y el empaquetamiento intracelular o formación de cuerpos de inclusión, supervivencia en entornos acuosos; la ausencia de toxicidad para mamíferos, la capacidad de atracción para ser ingeridas por las plagas, la facilidad de matar y fijar sin ocasionar daños en la toxina; y semejantes Otras consideraciones incluyen la facilidad de formulación y manejo, las de orden económico, de estabilidad con el almacenamiento y semejantes.

El huésped celular que contiene una secuencia de nucleótidos codificadora de una proteína de la presente invención o una proteína relacionada se puede cultivar en medio nutritivo conveniente, donde la construcción de ADN provee una ventaja selectiva, proporcionando un medio selectivo de modo que todas o sustancialmente todas las células retienen la secuencia de nucleótidos que codifican la proteína de la presente invención o una secuencia de codificación relacionada. Estas células se pueden cosechar luego de manera convencional. Como alternativa, las células pueden ser tratadas antes de la cosecha.

Las secuencias de codificación de la presente invención, incluyendo las que se muestra en las SEQ ID Nº 3, SEQ ID Nº:5, SEQ ID Nº.7, SEQ ID Nº 9 y SEQ ID Nº:32 y semejantes, se pueden usar como base para construir secuencias de nucleótidos modificadas para su incorporación en células vegetales Aún más preferible es la síntesis de una secuencia de nucleótidos no natural codificadora de una o más de las proteínas de la presente invención o una proteína insecticida relacionada o su equivalente, para su expresión en una célula vegetal, donde dicha síntesis de la secuencia de nucleótidos no natural se basa en la secuencia de aminoácidos de la proteína nativa sin referencia a la secuencia de nucleótidos nativa

de la cual se dedujo la secuencia de aminoácidos nativa La expresión de dichas secuencias en células vegetales permitiría obtener una planta compuesta de células más resistentes al ataque por insectos de especies de coleópteros y semejantes. La manipulación mediante ingeniería genética de plantas con secuencias modificadas que codifican una o más de las proteínas de la presente invención o una proteína relacionada o una secuencia de aminoácidos insecticida relacionada se puede llevar a cabo introduciendo el ADN para plantas deseado (la palabra 'para plantas' es sinónimo de la palabras 'modificada' o 'sintética') que contiene la secuencia de codificación en tejidos o células vegetales, usando moléculas de ADN de diversas formas y orígenes que son bien conocidas por los especialistas en la manipulación genética de plantas. Los ejemplos de técnicas para introducir ADN en un tejido vegetal se describen, al menos, en Perlak et al (1991)

El ADN que contiene un gen modificado que codifican una proteína de la presente invención, tal como una proteína TIC901, una proteína TIC1201, una proteína TIC407, una proteína TIC417 o una proteína TIC431 y semejantes, o una proteína insecticida relacionada, ligado operativamente a un promotor vegetal funcional, puede ser introducido directamente en las células o tejidos vegetales utilizando diversos medios incluyendo, sin limitaciones, métodos de transformación mediada por *Agrobacterium*, virus vegetales, electroporación, microinyección, infiltración bajo vacío, medios de fusión a liposomas y balísticos y semejantes. El promotor vegetal puede ser un promotor constitutivo, o el promotor puede ser un promotor regulado en el tiempo, en el espacio, químicamente, fotosintéticamente, térmicamente o artificialmente; un promotor específico de tejidos; o aún un promotor quimérico o híbrido obtenido a partir de partes de otros promotores vegetales funcionales. Por ejemplo, el promotor puede ser el promotor 35S del virus en mosaico de la coliflor (CaMV) o un derivado funcional en plantas del mismo.

Los genes y las secuencias de codificación bacterianas nativas a menudo son de expresión pobre en células vegetales transgénicas La utilización de codones en

plantas se parece más estrechamente a la de otros organismos superiores que a organismos unicelulares, tal como bacterias. En varios informes se describieron métodos para mejorar la expresión de genes recombinantes en plantas (Murray *et al.*, 1989, Nucleic Acid Research, Vol,17 477-498; Diehn *et al.*, 1998, *Plant Physiology*, 117 1433-1443, Rocher et al., 1998, *Plant Phys.* (Murray *et al.*, 1989, Nucleic Acid Research, Vol,17:477-498; Diehn *et al.*, 1998, *Plant Physiology*, 117:1433-1443; Rocher et al., 1998, *Plant Phys.* 117·1445-1461) En estos informes se describen diversos métodos para manipular las secuencias de codificación para que representen secuencias que se traduzcan con mayor eficacia sobre la base de las tablas de frecuencia de codones, mejoras en la preferencias de la base en la tercera posición de los codones, usando secuencias recombinantes que evitan las secuencias consensuadas de procesamiento de intrones o dominios ricos en A/T o de poliadenilación sospechadas En tanto estos métodos para la construcción de genes sintéticos son notables, los genes sintéticos de la presente invención para su expresión en plantas particulares se prepararon sustancialmente de acuerdo con el método de Brown et al (Patente de EE.UU N°·5 689 052)

El trabajo que se describe en la presente aprovecha los métodos para potenciar la expresión *in planta* de una o más de las proteínas de la presente invención y de las proteínas insecticidas relacionadas, que confieren resistencia a patógenos de coleópteros o de insectos lepidópteros o aún de nematodos, por incorporación o localización de secuencias de codificación en el genoma nuclear, de plástidos o cloroplastos de las plantas susceptibles. En la Patente de EE.UU. N°:5 500.365 y patentes relacionadas se describen métodos para sintetizar genes vegetales con el fin de alcanzar niveles de expresión óptimos de la proteína codificada por el gen sintetizado, no natural, sintético o artificial. Estos métodos se relacionan con la modificación de las secuencias de genes estructurales de Bt nativos para producir una secuencia de codificación que sea más "tipo planta" y por ello tenga mayor probabilidad de ser traducida y expresada por la planta, ya sea

monocotiledónea o dicotiledónea. Sin embargo, el método descrito en Brown et al (Patente de EE UU Nº:5 689 052) provee una mayor expresión de transgenes, preferentemente en plantas monocotiledóneas

Por consiguiente, se puede aumentar la cantidad de un gen o secuencia de nucleótidos o segmento de nucleótidos que codifica un polipéptido de interés, por ejemplo todo o una parte insecticida de una proteína TIC901, TIC1201, TIC407, TIC417, TIC431 o polipéptido relacionado, en plantas transformando dichas plantas usando los métodos de transformación mencionados previamente En particular, la transformación de organelas como cloroplastos o plástidos puede dar como resultado la presencia de hasta aproximadamente 10 000 copias de secuencias de codificación por célula en los tejidos que contienen estas estructuras de organelas subcelulares (McBride et al , WO 95/24492)

El ADN que codifica los péptidos de la presente invención y las proteínas relacionadas también se pueden introducir en plantas por transferencia un método de transferencia directa de ADN en polen como se describe en la literatura (Zhou *et al.*, 1983, *Mol Cell Biol* , 10:4529-4537; Hess, 1987, Intern Rev Cytol., 107 367). La expresión de secuencias de codificación de polipéptidos, es decir, *tic*901 y semejantes, se puede lograr por inyección del ADN en los órganos reproductores de una planta como se describe en la literatura (Pena *et al.*, 1987, *Nature*, 325:274) El ADN también puede ser inyectado directamente en las células de embriones inmaduros y en embriones desecados rehidratados como se describe en la literatura (Neuhaus *et al* , 1987, *Theor. Appl. Genet.*, 75.30).

Después de efectuar la distribución de las secuencias de nucleótidos que codifican las proteínas insecticidas de la presente invención o proteínas relacionadas exógenas en las células receptoras, el paso siguiente se refiere en general a la obtención de las células transformadas para su posterior cultivo y la regeneración en plantas, es decir, la selección de las células transformadas. Tal como se mencionó en la presente, con el fin de mejorar la capacidad de identificar transformantes,

puede resultar conveniente emplear un gen marcador seleccionable o rastreable como, o además de, el gen expresable de interés. En este caso, se evaluará en general la población potencialmente transformada exponiendo las células a uno o más agentes de selección, o se podrá examinar las células por la característica del gen marcador deseado

Un ejemplo de una realización de métodos para identificar células transformadas comprende la exposición de cultivos transformados a un agente selectivo, como un inhibidor metabólico, un antibiótico, herbicida o semejantes. Las células que han sido transformadas y que poseen un gen marcador integrado de forma estable que confiere resistencia al agente de selección utilizado, se dejarán crecer y se dividirán en cultivo. Las células sensibles no serán sometidas a cultivo posterior Un ejemplo de un gen marcador preferido confiere resistencia al herbicida glifosato. Cuando se utiliza este gen como marcador seleccionable, el cultivo putativo de células transformadas es tratado con glifosato. Con la exposición a glifosato, las células transgénicas que contienen una enzima GOX recombinante o una enzima EPSPS recombinante que no es sensible al glifosato estarán disponibles para realizar cultivos adicionales en tanto sucede lo contrario con las células sensibles, o no transformadas. (Patente de EE.UU. Nº:5.569.834). Otro ejemplo de un sistema de marcador seleccionable preferido es el sistema de resistencia de la neomicina fosfotransferasa (*npt*II) por medio del cual se confiere resistencia al antibiótico kanamicina como se describe en la Patente de los EE UU Nº:5 569 834. Nuevamente, después de la transformación con este sistema, las células transformadas estarán disponibles para el cultivo posterior luego del tratamiento con kanamicina, en tanto las células no transformadas no lo están. Aún otro sistema marcador seleccionable preferido comprende el uso de una construcción genética que confiere resistencia a paromomicina. El uso de este tipo de sistema marcador seleccionable se describe en la Patente de EE UU Nº:5 424.412. Hay otros marcadores seleccionables que son bien conocidos en el arte incluyendo, sin

limitaciones, marcadores de resistencia a antibióticos, tales como *npt*II, *tet*, *aad* y semejantes, *phn*O y otras diversas acetilasas (Patente de EE.UU. Nº:6.448.476), diversas estearasas (6 107.549), barnasa (Hartley, 1988), *J. Mol. Biol* 202 913), enzimas bacterianas que confieren actividad glifosato oxidasa a la célula transformada (*gox*) (Barry et al , 1992, Inhibitors of amino acid biosynthesis: *Strategies for imparting glyphosate tolerance to crop plants*. En: *Biosynthesis and Molecular Regulation of Amino Acids in Plants*. págs. 139-145. Singh, Flores y Shannon Eds., American Society of *Plant Physiologists*, Rockville, Md.) y semejantes

La selección transplastonómica (selección de eventos de transformación de plástidos o cloroplastos) se simplifica aprovechando la sensibilidad de los cloroplastos o plástidos a la espectinomicina, un inhibidor de la síntesis de proteínas en plástidos o cloroplastos, pero no la síntesis de proteínas por ribosomas citoplasmáticos codificados en el genoma nuclear. La espectinomicina impide la acumulación de las proteínas de cloroplastos necesarias para la fotosíntesis, de modo que las células vegetales transformadas resistentes a espectinomicina se pueden distinguir sobre la base de su diferencia en cuanto a color. las células transformadas resistentes son verdes, en tanto las células sensibles son blancas debido a la inhibición de la síntesis de proteínas de plástidos. La transformación de cloroplastos o plástidos con un gen *aad* bacteriano adecuado, o con un gene que codifica un ARN ribosómico funcional resistente a espectinomicina en plástidos o cloroplastos provee un medio de selección y mantenimiento de eventos transplastonómicos (Maliga, 1993, *Trends in Biotechnology* 11:101-106).

Se contempla además que las combinaciones de marcadores rastreables y seleccionables serán de utilidad en la identificación de células transformadas. En algunos tipos de células o tejidos el agente de selección, tal como glifosato o kanamicina, posiblemente no proporcione suficiente actividad para matar como para permitir un reconocimiento claro de las células transformadas o puede causar una inhibición sustancial no selectiva de transformantes y no transformantes por igual, lo

cual hace que la técnica de selección no sea eficaz Se propone que la selección con un compuesto inhibidor del crecimiento, tal como glifosato o AMPA (ácido amino-metilfosfónico) a concentraciones inferiores a las que causan un 100% de inhibición, seguido de examen del tejido en crecimiento por la expresión de un gen de un marcador rastreable, tal como kanamicina, permitirá recuperar transformantes de tipos de células o tejidos que no pueden ser seleccionados por sí solos. Se propone que las combinaciones de selección y examen/rastreo puede facilitar la identificación de transformantes en una variedad más amplia de tipos de células y tejidos.

El desarrollo o la regeneración de plantas a partir de protoplastos vegetales individuales o de diversos explantes es bien conocido en el arte (Weissbach y Weissbach, 1988) Este proceso de regeneración y crecimiento incluye típicamente los pasos de selección de células transformadas, cultivo de las células individualizadas a través de las etapas habituales desde el desarrollo embrionario hasta la etapa de plántula con raíz Los embriones y las semillas transgénicos se regeneran de manera similar. Los brotes transgénicos con raíces resultantes se plantan luego en un medio de crecimiento apropiado para plantas, tal como tierra

El desarrollo o la regeneración de plantas que contienen un gen extraño, exógeno que codifica todo o la parte insecticida de una proteína TIC901, TIC1201, TIC407, TIC417, TIC431 o un polipéptido relacionado introducido en el genoma vegetal por transformación con *Agrobacterium* de explantes foliares se pueden llevar a cabo utilizando métodos bien conocidos en el arte (Horsch *et al.*, 1985) En este procedimiento, los transformantes son cultivados en presencia de un agente de selección y en un medio para inducir la regeneración de brotes en la cepa vegetal que se está transformando como se describe en la literatura (Fraley *et al* , 1983). En particular, en la Patente de EE.UU. Nº·5.349.124 se describe con detalle la creación de células de lechuga transformadas genéticamente y de las plantas de lechuga obtenidas que expresan proteínas cristalinas híbridas que confieren actividad insecticida a dichas plantas contra larvas de coleópteros.

El procedimiento produce típicamente brotes dentro de los dos a cuatro meses y estos brotes son transferidos luego a un medio inductor de raíces apropiado que contiene al agente selectivo y un antibiótico para impedir el crecimiento bacteriano. Los brotes enraizados en presencia del agente selectivo para formar plántulas son transplantados luego a tierra u otro medio para permitir la formación de raíces. Estos procedimientos varían dependiendo de la cepa vegetal particular empleada, siendo dichas variaciones bien conocidas en el arte.

Preferentemente, las plantas regeneradas se autopolinizan con el fin de obtener plantas transgénicas homocigotas o bien, se cruza el polen obtenido de las plantas regeneradas con plantas cultivadas hasta semillas de líneas de importancia agronómica, con preferencia endogámicas. A la inversa, se utiliza el polen de plantas de dichas líneas importantes para polinizar las plantas regeneradas. La planta transgénica de la presente invención que codifica una proteína insecticida deseada de la presente invención o un polipéptido relacionado, se cultiva utilizando los métodos bien conocidos por los especialistas en el arte.

Una planta transgénica de la presente invención contiene al menos una región de codificación que codifica uno o más polipéptidos de la presente invención, o un polipéptido relacionado, tal como un fragmento insecticida de una proteína TIC901, una TIC1201, una TIC407, una TIC417 o una TIC431 o una quimera de estas proteínas. Una planta transgénica preferida es un segregante independiente y puede transmitir ese gen y su actividad a su progenie. Una planta transgénica más preferida es homocigota para ese gen y transmite dicho gen a toda su descendencia luego de apareamiento sexual Las semillas de la planta transgénica pueden ser sembradas en el campo o en invernadero y las plantas transgénicas resultantes sexualmente maduras se autopolinizan para generar plantas de líneas puras. La progenie de estas plantas se vuelven líneas puras que se evalúan por su mayor La progenie de estas plantas se vuelven líneas puras que se evalúan por la mayor expresión del transgen de *B thuringiensis*.

Se prefieren las plantas transgénicas que expresan más de un agente insecticida, siendo cada agente tóxico para una especie de plaga de insecto de interés, y mostrando cada agente insecticida un modo de acción separado o un medio diferente para introducir esporas en el epitelio intestinal de la plaga interés de insecto de interés, ya sea por unión a receptores separados e independientes o por formación de esporas que son evidentemente distintas de las esporas formadas por una de las otros agentes insecticidas presentes en la misma planta transgénica. Se puede considerar que dichas plantas pueden ser transformadas para expresar al menos dos o más de las proteínas de la presente invención o serán transformadas para expresar al menos una de las proteínas de la presente invención junto con una o más proteínas insecticidas no relacionadas incluyendo, a modo de ejemplo, proteínas de delta-endotoxinas tales como una más entre una toxina Cry1, Cry2, Cry3, Cry9, Cry22, ET70, TIC851 y/o cualquiera entre una o más de las toxinas binarias ET80/76, ET33/34, PS149B1 y ET100/101 y semejantes y fusiones, quimeras y variantes de las mismas, y proteínas insecticidas o complejos de proteínas insecticidas de dichos organismos diversos, tal como las proteínas de toxinas de *Xenorhabdus* y/o *Photorhabdus,* o VIP, WAR y/o MIS y proteínas relacionadas y uno o más ARN de cadena doble cuya expresión en la célula vegetal da como resultado la supresión de uno o más genes en una o más plagas de insectos de interés

Con el fin de identificar una planta transgénica que expresara niveles altos del polipéptido de la presente invención o una proteína relacionada a partir de una secuencia de nucleótidos preferida, es necesario examinar el evento transgénico seleccionado (generación R_0) por la actividad insecticida y/o expresión del gen de interés. Esto se puede llevar a cabo mediante diversos métodos bien conocidos por los especialistas en el arte, incluyendo sin limitaciones, los siguientes: 1) obtener muestras pequeñas de tejido de la planta transgénica R_0 y evaluar directamente en el tejido la actividad contra insectos susceptibles en paralelo con tejidos derivados de

una planta control negativa que no expresa dicha actividad. [Por ejemplo, se pueden identificar plantas de maíz transgénico R_0 que expresan un fragmento insecticida de una proteína TIC901 o una proteína relacionada, evaluando en tejidos de hoja o de raíz derivados de dichas plantas la actividad contra el escarabajo de papa de Colorado (CPB, *Leptinotarsa decemlineata*) y los gusanos de la raíz del maíz del sur (SCR, *Diabrotica undecimpunctata howardi*)]; 2) analizar extractos de proteínas por inmunoensayos ligados a enzimas (ELISA) específicos para el fragmento de proteína insecticida o proteína relacionada, o 3) efectuar una amplificación térmica con transcriptasa inversa (también conocida en el arte como RTPCR) para identificar los eventos que expresan la secuencia que codifica el fragmento de la proteína insecticida o una proteína relacionada.

Los agentes pesticidas de la presente invención pueden aplicarse solos o en combinación con otros agentes pesticidas a una semilla como un componente de un recubrimiento para semillas, o los agentes de la presente invención pueden producirse dentro de una semilla transgénica y combinarse con otros agentes pesticidas en la forma de un recubrimiento para semillas. Los métodos y las composiciones de recubrimientos de semillas que son conocidos en la técnica son de utilidad cuando se los modifica por la adición de una de las realizaciones de combinaciones de pesticidas de la presente invención. Dichos métodos de recubrimiento y los aparatos para su aplicación se describen, por ejemplo, en las Patentes de EE.UU. Nº: 5.918.413, 5.891.246, 5.554 445, 5 389 399, 5.107 787, 5 080 925, 4 759 945 y 4 465.017. Las composiciones de recubrimientos para semillas se describen, por ejemplo, en las Patentes de EE.UU. Nº: 5 939.356, 5.882.713, 5.876.739, 5.849.320, 5.834.447, 5.791.084, 5.661 103, 5.622.003, 5.580.544, 5.328 942, 5.300.127, 4.735 015, 4.634.587, 4.383 391, 4.372.080, 4 339 456, 4.272.417 y 4.245 432, entre otros.

Tal como se utiliza en la presente, el término "muestra biológica", o "muestra", pretende incluir ácidos nucleicos, polinucleótidos, ADN, ARN, ARNt, ADNc y

semejantes en una composición o fijados a un sustrato que permita someter a la muestra a un análisis con sondas moleculares o amplificación térmica usando sondas y/o cebadores de oligonucleótidos. Se considera que una planta o producto de una planta o el fruto o la semilla de una planta es una muestra biológica y todo extracto de dicha planta o producto vegetal, parte de planta, fruto o semilla también es considerado una muestra biológica. Como tales, las muestras biológicas pueden derivar de productos de importancia agronómica y comercial y/o de composiciones de materia que incluyen, a modo de ejemplo, alimentos para animales, productos comerciales y productos y subproductos de maíz pretendidos para consumo como alimento para humanos o para su uso en composiciones pretendidas para consumo humano incluyendo, a modo de ejemplo, harina de maíz, alimentos de maíz, jarabe de maíz, aceite de maíz, almidón de maíz, palomitas de maíz, tortas de maíz, cereales que contienen maíz y subproductos de maíz y semejantes considerados todos dentro del alcance de la presente invención si dichos productos y composiciones de materia contienen cantidades detectables de las secuencias de nucleótidos o de las proteínas pesticidas codificadas por las mismas, descritas en la presente.

Los siguientes ejemplos ilustran aún más las características de las secuencias de nucleótidos descritas en la presente y la actividad insecticida de las proteínas codificadas por las secuencias de nucleótidos descritas. Además, se describen asimismo métodos y procedimientos para la práctica de la invención. Sin embargo, los especialistas en el arte podrán apreciar, a la luz de la presente descripción, que es posible efectuar muchos cambios en las realizaciones específicas descritas y aún obtener resultados iguales o similares sin apartarse del espíritu y alcance de la invención

EJEMPLOS

Ejemplo 1

Este ejemplo ilustra la identificación de una proteína insecticida secretada

hacia el medio por la cepa EG2158 de *B. thuringiensis*.

La cepa EG2158 es una cepa de *B thuringiensis* de tipo salvaje aislada originalmente a partir de polvo de granos de soja (Donovan et al., 1988) Durante la esporulación, las células de EG2158 producen cristales con forma romboide compuestos de una proteína de 73 kDa identificada como Cry3C Las mezclas de cristales de esporas de EG2158 son tóxicas para las larvas del escarabajo de papa de Colorado. Cuando se clona *Bacillus megaterium* transformada con el gen *cry*3C, . presenta una toxicidad para larvas de CPB similar a la cepa EG2158 lo cual sugiere que la toxicidad específica de coleópteros de EG2158 se debe a la proteína cristalina Cry3C (Donovan et al , 1988). Por ello, resultó inesperado que el medio usado empleado para la fermentación de la cepa EG2158, que no contiene esporas o proteína cristalina Cry3C, fuera tóxica para CPB porque el medio usado procesado no contiene suficientes esporas o proteína cristalina Cry3 como para presentar una toxicidad insecticida mensurable.

Se cultivaron cepas de *B. thunngiensis* en 60 ml de medio de cultivo PYG con agitación durante la noche a 30 °C. El medio PYG consistía de 11,8 gramos de peptona, 23,6 gramos de extracto de levadura, 4 mililitros de glicerol, 19,4 gramos de K_2HPO_4 anhidro y 2,2 gramos de KH_2PO_4 anhidro por litro de agua desionizada. Los cultivos de *B. thuringiensis* fermentados durante la noche se centrifugaron a 11 000x g durante 30 minutos y los sobrenadantes libres de células fueron transferidos a frascos limpios. El sobrenadante también se denomina medio usado y/o medio usado clarificado o sobrenadante clarificado en la presente.

Los sobrenadantes se enfriaron a 4° C Se agregaron lentamente 34 gramos de sulfato de amonio más 1 mililitro de NaOH 1 M a 60 mililitros de sobrenadante bajo agitación Se centrifugaron las mezclas de sobrenadante saturado con sulfato de amonio para recolectar las proteínas precipitables en un pelet. Los pelets se disolvieron en 2 mililitros de Tris.HCl 20 mM, pH 7,5. Las suspensiones de Tris fueron transferidas a tubos para diálisis (6000 MWCO) y se dializaron a 4 °C contra

Tris-HCl 20 mM, pH 7,5 Las suspensiones dializadas se denominan dialisatos *am sulf* (ASD).

Los dialisatos am.sulf (ASD) fueron evaluados por su toxicidad para larvas del escarabajo de papa de Colorado (CPB, también conocido formalmente como *Leptinotarsa decemlineata*) Se aplicaron 50 µl de cada ASD tópicamente a 2 ml de dieta para insectos en el recipiente con la dieta Se trataron un total de dieciséis recipientes de dieta para cada dialisato. Se colocaron larvas CPB del primer instar en cada recipiente y se evaluó la mortalidad de los insectos después de 3 días Se completaron numerosas repeticiones de este procedimiento con diferentes cepas de *B. thuringiensis* El dialisato de la cepa EG2158 de *B thuringiensis* presentó toxicidad para las larvas CPB.

Ejemplo 2

Este ejemplo ilustra un medio para purificar una proteína insecticida, denominada TIC901 en la presente, a partir del medio de cultivo usado producido por fermentación de la cepa EG2158 de *Bacillus thuringiensis*.

Las proteínas presentes en el ASD EG2158 fueron evaluadas por electroforesis sobre gel de poliacrilamida-dodecilsulfato de sodio (SDS-PAGE). Se mezclaron 30 µl de dialisato con 15 µl de solución amortiguadora para solubilizar proteínas Laemmli (*Mol. Biol.*, 80 575-599; 1973) y se calentó a 100 °C durante 5 minutos. Se cargaron aproximadamente 25 µl de la mezcla sobre un gel de poliacrilamida-SDS 10% Las proteínas se visualizaron después de la electroforesis usando coloración con Coomassie Brilliant Blue Los resultados indicaron que el ASD de EG2158 contenía aproximadamente treinta proteínas cuyo tamaño variaba en un rango entre 20 kDa aproximadamente y 120 kDa aproximadamente

Se aplicaron 2 mililitros del ASD de EG2158 a una columna de dietilaminoetilo (DEAE) equilibrada con Tris-HCl 20 mM pH 7,5. Las proteínas eluyeron de la columna con un gradiente de 20 mililitros de NaCl (0 a 500 mM) en Tris-HCl 20 mM, pH 7,5. Se recolectaron fracciones de 1 mililitro de eluato y cada fracción se dializó

contra Tris-HCl 20 mM, pH 7,5. Las fracciones individuales fueron evaluadas después de la diálisis por su toxicidad para CPB usando un bioensayo similar al descrito en el Ejemplo 1 Las fracciones con toxicidad para las larvas CPB se agruparon y dializaron Las fracciones agrupadas se denominaron fracciones EG2158 tóxicas-DEAE Las proteínas en la fracción EG2158 tóxica-DEAE fueron analizadas por SDS-PAGE y coloración con Coomassie. Los resultados indicaron que la fracción EG2158 tóxica-DEAE contenía cinco proteínas primarias, así como varias proteínas menores de distintos tamaños.

Se aplicaron 2 mililitros de la fracción tóxica-DEAE de EG2158 a una columna de amonio cuaternario (QA) equilibrada con Tris-HCl 20 mM pH 7,5. Las proteínas eluyeron de la columna con 20 mililitros de un gradiente lineal de NaCl (0 a 500 mM) en Tris-HCl 20 mM, pH 7,5. Se agruparon las fracciones que presentaban la mayor toxicidad para CPB Las fracciones dializadas fueron evaluadas por su toxicidad para larvas CPB Se agruparon las fracciones que presentaban la mayor toxicidad para CPB. La fracción agrupada se denominó fracción QA-tóxica EG2158. Las proteínas en la fracción EG2158 tóxica-QA fueron analizadas por SDS-PAGE y coloración con Coomassie Los resultados indicaron que la fracción QA-tóxica de EG2158 contenía una proteína principal de 38 kDa aproximadamente, denominada TIC901, y varias otras especies proteicas menores de diversos tamaños. Los resultados sugieren que la proteína purificada de 38 kDa era responsable de la toxicidad para CPB que se observaba con las muestras

Ejemplo 3

Este ejemplo ilustra el aislamiento y la identificación de la secuencia de nucleótidos de EG2158 que codifica la proteína TIC901.

Las proteínas presentes en la fracción EG2158 tóxica-QA fueron separadas por tamaño por SDS-PAGE sin coloración con Coomassie. Las proteínas separadas fueron transferidas de SDS-PAGE a una membrana de difluoruro de polivinilideno (PVDF) por electrotransferencia La membrana PVDF se coloreó brevemente con

colorante Coomassie y la porción de la membrana que contenía la proteína TIC901 purificada se recortó con una hoja de afeitar limpia y fue sometida a secuenciación por degradación de Edmund automática. Los resultados indicaron que la proteína TIC901 contenía la secuencia de aminoácidos amino-terminal correspondiente a ^{NH_3}-V I G P Y A E S Y I D R V Q D-$^{COO-}$ que se muestra en la SEQ ID Nº·1

Es reconocido ampliamente que la mayoría de las proteínas producidas *in vivo* en sistemas bacterianos tienen un residuo metionina (M) en su extremo N-terminal El hecho que el extremo N-terminal de la proteína TIC901 no contiene un residuo metionina amino terminal, junto con el hecho que se encontró que la proteína se localizaba en el medio usado, sugieren que la proteína TIC901 podría formarse por digestión proteolítica de un péptido señal secretor de la región N-terminal de una proteína precursora más grande.

Sobre la base de la secuencia de aminoácidos parcial de la proteína TIC901 purificada sobre gel determinada por degradación de Edmund y basado en la preferencia de utilización de codones que presentan los genes nativos de *Bacillus thuringiensis* que codifican δ-endotoxinas, se diseñó una sonda de oligonucleótidos degenerada para su uso como sonda en la detección de secuencias de nucleótidos de la cepa EG2158 de *Bacillus thuringiensis* que pudieran codificar la proteína TIC901 purificada sobre gel. La sonda se identificó como WD444 y comprendía la secuencia que se muestra en la SEQ ID Nº·2 (5'-GTA ATT GGA CCA TAT GCA GAA TCA TAT ATT GAT XGA GTA CAA GA-3', donde X es A o C)

Se purificó ADN a partir de células de la cepa EG2158 de *B. thuringiensis* utilizando procedimientos estándar (por ejemplo, Sambrook et al., 1989). Una muestra del extracto de ADN fue digerida con la endonucleasa de restricción *Hind*III, sometida a electroforesis sobre gel de agarosa 0,8% en solución amortiguadora TAE, transferida a un filtro de nylon y analizada por transferencia Southern en pruebas con sondas utilizando una mezcla de sondas WD444 conjugadas con fosfatasa alcalina durante aproximadamente 16 horas a 40 °C en solución amortiguadora de

hibridación, con severidad baja a moderada, y lavadas a 40 °C en solución amortiguadora de lavado, posteriormente fue expuesta a reactivos de quimioluminiscencia y finalmente expuesta a una película (las soluciones amortiguadoras de hibridación y lavado fueron suministradas junto con el conjunto de elementos AMERSHAM/PHARMACIA BIOTECH, Número de catálogo RPN3690). Los resultados indicaron que la sonda aparentemente se hibridizaba específicamente con un solo fragmento *Hind*III de aproximadamente 8 pares de kilobases.

Se construyó una biblioteca de *E. coli* que contenía fragmentos de restricción *Hind*III EG2158 de aproximadamente 8 pares de kilobases en el plásmido pUC18. Se transfirieron colonias recombinantes a filtros de nylon y se desnaturalizaron con NaOH mediante procedimientos estándar (por ejemplo, Sambrook et al., 1989). Los filtros se incubaron en solución de hibridación con una mezcla de sondas WD444 marcadas. Las membranas se lavaron después de incubación con la sonda bajo condiciones similares a las descritas previamente, fueron expuestas a reactivos de quimioluminiscencia y expuestas a una película Se identificaron varias colonias que se hibridizaban específicamente con la mezcla de sondas WD444. Se extrajo ADN plasmídico de algunas de estas colonias y se analizó por transferencia Southern usando la mezcla de oligonucleótidos WD444 como sonda. Los resultados indicaron que los plásmidos consistían del vector pUC18 más un fragmento del inserto *Hind*III de aproximadamente 8 kb Se seleccionó un plásmido típico para su caracterización adicional y se denominó pEG1379 (y posteriormente pMON74007) La colonia purificada de la cepa de *E. coli* que contenía el pMON1379 se denominó EG12447 y se depositó en NRRL el 6 de febrero, 2002, y el depósito recibió el número de acceso NRRL B 30549

Ejemplo 4

Este ejemplo ilustra la expresión de una proteína insecticida de un fragmento *Hind*III de aproximadamente 8 pares de kilobases en el plásmido pEG1381 en una cepa acristalífera de *Bacillus thuringiensis*

Los genes nativos de *Bacillus thuringiensis* a menudo se expresan pobremente en células huésped que no son *Bacillus*, tal como *E coli*. El pEG518 se construyó como un vector ambivalente *Bacillus-E. coli* y es una quimera del replicón de *Bacillus* pMM101 (Norton et al., *Plasmid*, 13.211-214, 1985) y pUC18 (Yanisch-Perron et al, *Gene* 33.103-119, 1985) y por ello puede replicarse ya sea en *E coli* y bacterias Gram-negativas estrechamente relacionadas o en *Bacillus* y bacterias Gram-positivas estrechamente relacionadas El pEG518 confiere resistencia al antibiótico β-lactama a bacterias *E. coli* y bacterias relacionadas y resistencia a tetraciclina a bacterias *Bacillus* y bacterias relacionadas transformadas para contener derivados de este plásmido en particular. Se recortó el fragmento *Hind*III de 8 pares de kilobases presente en el plásmido pMON1379 y se insertó en el sitio *Hind*III en el plásmido pEG518 El plásmido resultante, pEG1381, se transformó por electroporación en la cepa EG10650 acristalífera y no insecticida de *Bacillus thuringiensis* (NRRL, Nº Acceso NRRL B-30217).

(Cepa EG10650 de *Bacillus thuringiensis*. La cepa EG10650 de *B. thuringiensis* es un derivado de la cepa EG10368 (Patente de EE.UU. Nº 5 759.538, 2 de junio, 1998) que es deficiente en las actividades de proteasas neutras y proteasas alcalinas y solamente contiene un elemento plasmídico extra-cromosómico conocido de 7,5 kb. Se introdujo una supresión en el gen de la proteasa alcalina en la cepa EG10368, dando como resultado la cepa EG EG10654 (NRRL, Nº Acceso NRRL B-21344) Se introdujo una supresión en el gen de la proteasa neutra en la cepa EG10368, dando como resultado la cepa EG10624 (NRRL, Nº Acceso NRRL B-21347) Estas supresiones se combinaron luego en una cepa para obtener la cepa EG10650, que no contenían ningún plásmido del orden de mega-Daltons y por ello carecía del potencial para expresar cualquiera de las proteínas insecticidas conocidas, así como también era deficiente en cuanto a la producción de ambas proteasas alcalina y neutra.)

Se seleccionó una sola colonia que representaba un único transformante, y se

denominó EG12450, sobre un medio que contenía tetraciclina para su posterior análisis.

Se aisló el plásmido pEG1381 de un cultivo de la cepa EG12450 cultivado durante la noche en presencia de tetraciclina. El ADN del pEG1381 fue digerido con *Hind*III y luego se comparó por electroforesis sobre gel de agarosa y coloración con bromuro de etidio con el fragmento *Hind*III presente en el pEG1379. Ambos plásmidos contenían un fragmento *Hind*III de aproximadamente 8 kb que aparentemente eran idénticos

Se cultivaron las cepas EG12450, EG2158 y EG10650 durante la noche en medio PYG. Se preparó un dialisato de sulfato de amonio (ASD) a partir de cada uno de los sobrenadantes de cultivo como se describió en el Ejemplo 1. Las muestras de ASD fueron aplicadas a una columna DEAE, se eluyeron las proteínas de la columna y se recolectaron fracciones como se describió en el Ejemplo 2 Las proteínas en las fracciones DEAE fueron analizadas por SDS-PAGE y coloreadas con Coomassie Brilliant Blue. Los resultados demostraron que las fracciones DEAE de la cepa acristalífera EG10650 control no contenían cantidades significativas de ninguna proteína correspondiente en tamaño a la proteína TIC901, aunque se detectó la presencia de algunas proteínas menores del tamaño aproximadamente de la proteína TIC901. Por el contrario, las fracciones DEAE obtenidas con los sobrenadantes de cultivo usados en los cuales había crecido la cepa EG2158, contenían una cantidad importante de una proteína correspondiente en tamaño a la proteína TIC901 Las fracciones DEAE obtenidas de los sobrenadantes de cultivo usados en los cuales había crecido la cepa recombinante EG12450 contenían cantidades significativas de una proteína correspondiente al tamaño de la proteína TIC901 producida por la cepa EG2158 Este resultado indicaba que el fragmento *Hind*III de 8 kb derivada de la cepa EG2158 presente en el plásmido recombinante pEG1381 contenía un gen nativo de *Bacillus thuringiensis* de una proteína que presentaba una masa aproximadamente igual a la de la proteína TIC901 aislada del

medio de fermentación usado de la cepa EG2158

Ejemplo 5

Este ejemplo ilustra la identificación de la secuencia de ADN de la secuencia de nucleótidos nativa que codifica la proteína TIC901 y la secuencia de aminoácidos TIC901 deducida.

El fragmento *Hind*III de aproximadamente 8 pares de kilobases presente en el plásmido pMON1378 fue sometido a análisis por dideoxi de la secuencia de nucleótidos. Parte de la secuencia de nucleótidos derivada de este fragmento de 8 kb se muestra en la SEQ ID Nº:5. Se identificó un marco de lectura abierto formado por 1101 nucleótidos dentro del fragmento *Hind*III de aproximadamente 8 pares de kilobases, comenzando con el codón ATG de metionina en la posición de nucleótido 153-155 y terminando con el codón TTG leucina en la posición de nucleótido 1251-1253, inmediatamente 5' con respecto a un codón de terminación TAG como se muestra en la SEQ ID Nº·5. El marco de lectura abierto identificado en esta secuencia de nucleótidos está compuesto por 429 residuos de adenosina (aproximadamente 39% A), 334 residuos de timidina (aproximadamente 30% T), 196 residuos de guanidina (aproximadamente 18% G) y 145 residuos de citosina (aproximadamente 13% C) y presenta aproximadamente un 69% de AT y aproximadamente un 31% de GC. Hay tres codones de metionina representados en los primeros ochenta y cuatro (84) nucleótidos del ORF. Además del codón ATG en las posiciones de nucleótidos 153-155 como se muestra en la SEQ ID Nº.3, hay un segundo codón ATG en-marco con el de la posición 153-155 ubicado en los nucleótidos 182-184 y aún un tercer codón ATG en-marco con estos dos codones ATG mencionados previamente en los nucleótidos 201-203. No se observó ninguna secuencia de unión a ribosomas consenso 5' con respecto al segundo o tercer codón ATG, sin embargo se identificó una secuencia de unión a ribosomas consenso 5' con respecto al primer codón ATG, centrado en aproximadamente 10 nucleótidos 5' con respecto al primer ATG. La secuencia de unión a ribosomas consenso comprende

los nucleótidos entre la posición 141 aproximadamente y la posición 147 aproximadamente como se muestra en la SEQ ID Nº:3

La secuencia de aminoácidos deducida de este marco de lectura abierto consiste de 367 aminoácidos como se muestra en la SEQ ID Nº:4. Se predijo que el marco de lectura abierto codifica una proteína precursora que tiene un peso molecular calculado de aproximadamente 41.492 Daltons Contiene 40 aminoácidos fuertemente básicos (lisina y arginina), 43 aminoácidos fuertemente ácidos (aspartato y glutamato), 113 aminoácidos que característicamente son de naturaleza hidrofóbica (alanina, isoleucina, leucina, fenilalanina, triptofano y valina) y 122 aminoácidos que característicamente son de naturaleza polar (asparagina, cisteína, glutamina, serina, treonina y tirosina). La proteína precursora exhibe un punto isoeléctrico calculado de 6,368 aproximadamente y una carga neta calculada de –2,102 aproximadamente a pH 7,0. Se predijo que los primeros treinta (30) residuos de aminoácidos codificados por el marco de lectura abierto comprenden al péptido señal amino terminal que es procesado, probablemente por una peptidasa señal de tipo II, para liberar la proteína insecticida hacia el espacio extracelular. Se predijo que estos primeros treinta (30) residuos de aminoácidos comprenden un péptido señal amino-terminal basado en el algoritmo de Nielsen et al (1997, *Protein Engineering* 10 1-6) El péptido señal predicho además de comprender de los residuos de aminoácidos entre las posiciones treinta y uno (31) y cuarenta y tres (43), como se muestra en la SEQ ID Nº.4, no está ligado ni asociado con la proteína madura localizada en el medio usado y se cree que es eliminado proteolíticamente durante la translocación de membrana, liberando la proteína madura insecticida hacia el espacio extracelular

Los nucleótidos 130 a 174, que se muestran en la SEQ ID Nº·3, corresponden más estrechamente a la secuencia de la sonda que se muestra en la SEQ ID Nº·2 Los nucleótidos de la SEQ ID Nº:2, cuando se leen en marco con la correspondiente secuencia de codificación de la secuencia de aminoácidos TIC901, corresponden a los codones que codifican los aminoácidos amino-terminales de la proteína

insecticida purificada a partir del medio usado de cultivos de la cepa EG2158 de *Bacillus thuringiensis*. Se esperaba que los aminoácidos codificados por el marco de lectura TIC901 en toda esta región de la secuencia de nucleótidos correspondiera a la secuencia de aminoácidos de los aminoácidos deducida degradación de Edmund de la proteína insecticida purificada a partir del medio usado de cultivos de la cepa EG2158 y que se muestra en la SEQ ID N°:1. La secuencia de aminoácidos que se muestra en la SEQ ID N° 1 es idéntica a la secuencia de aminoácidos deducida de esta región del marco de lectura abierto para TIC901, excepto que la posición de aminoácido 12 que se muestra en la SEQ ID N° 1 corresponde a un residuo arginina, en tanto el correspondiente residuo de aminoácido deducido de la secuencia del marco de lectura abierto en la SEQ ID N°·3 corresponde a un residuo isoleucina en la posición de aminoácido 55 como se muestra en la SEQ ID N° 4 Este resultado sugería ya sea que se había encontrado un error en el resultado del análisis de la secuencia Edmund de la proteína purificada del medio de cultivo usado de EG2158 y considerada como la proteína TIC901 o que había más de una proteína tipo TIC901 en la muestra concentrada y purificada que se envió para el análisis de secuencia Edmund Es un hecho bien conocido que es bastante común encontrar errores en la secuenciación de Edmund. Sin embargo, este resultado solo sugería la presencia de más de un gen que codificaba una proteína insecticida secretada extracelular en la cepa EG2158 de *Bacillus thuringiensis* En el Ejemplo 9 se describe un análisis completo de esta hipótesis.

La mayoría de las proteínas bacterianas liberadas hacia el espacio extracelular tienden a ser sintetizadas como precursores grandes que son procesados por el aparato de secreción que cliva una porción de la proteína precursora de la proteína madura que es liberada hacia el medio extracelular Es posible que la proteína insecticida presente en el medio usado de los cultivos de la cepa EG2158 sea sintetizada en dicha cepa EG2158 como una proteína precursora que es dirigida para su secreción por todo o una parte de los cuarenta y tres (43)

aminoácidos deducidos como parte del marco de lectura abierto que se muestra en la SEQ ID Nº.3. Las proteínas secretadas producidas por bacterias Gram-positivas y Gram-negativas que son dirigidas hacia el aparato de secreción de tipo II presentan típicamente una señal secretora o péptido señal que comprende entre aproximadamente dieciocho (18) y aproximadamente veintiocho (28) a treinta (30) o más aminoácidos de longitud. Como se indicó previamente, hay dos codones de metionina correspondientes a los residuos de aminoácidos dieciocho (18) y veintisiete (27) respectivamente, deducidos del marco de lectura abierto que se muestra en la SEQ ID Nº:3, 5' con respecto al residuo de valina (el codón en los nucleótidos 282-284 que se muestra en la SEQ ID Nº:3) deducida por degradación de Edmund como el aminoácido amino-terminal de la proteína TIC901 El residuo valina que se muestra en la SEQ ID Nº.4 correspondiente al residuo amino-terminal predicho por degradación de Edmund de la proteína secretada hallada en el medio usado de los cultivos de EG2158 se encuentra en el número de residuo cuarenta y cuatro (44). Es probable que los dos residuos metionina proximales en la secuencia de aminoácidos amino-terminal para el residuo valina en la posición 44 comprenden sitios de inicio de la traducción alternativos para la síntesis del precursor propuesto de la proteína TIC901 producida en EG2158. Sin embargo, hay secuencias Shine-Delgarno consenso muy pobres 5' con respecto a estos dos codones alternativos de inicio de la traducción. Además, dado que la proteína se encuentra en el medio usado, un análisis de la secuencia de aminoácidos predicha codificada por los nucleótidos que le siguen al codón MET en la posición 153-155 usando los algoritmos de predicción de péptidos señal bien conocidos en el arte, indica que los primeros treinta aminoácidos se ajustan bien dentro de los parámetros establecidos para secuencias de aminoácidos que funcionan como péptidos señal Por ello, se cree que el primer ATG en las posiciones de nucleótido 153-155 que se muestra en la SEQ ID Nº 3 es el codón de inicio de la transcripción utilizado realmente por la cepa EG2158 de *Bacillus thuringiensis* para expresar el precursor de la proteína

TIC901 que posteriormente es procesado en la forma madura de la proteína secretada de la célula.

El péptido señal predicho, como se indicó previamente, está compuesto de los residuos de aminoácidos uno (1) a treinta (30) como se muestra en la SEQ ID Nº:4 y presenta las características de un clásico péptido señal de tipo II, incluyendo una carga básica en el extremo amino-terminal, una región central alfa-helicoidal de carga neutra predicha aproximadamente entre el residuo de aminoácido seis (6) y aproximadamente el residuo de aminoácido veintidós (22) y una secuencia de reconocimiento de la señal de peptidasa II consenso Ala-Xaa-Ala en los residuos veintiocho (28) a treinta (30) que se podría reconocer como un sustrato de clivaje proteolítico para el clivaje del enlace peptídico entre los residuos de aminoácidos treinta (30) y treinta y uno (31) El extremo carboxilo-terminal del péptido señal se identificó como esencial como secuencia de reconocimiento para la señal de peptidasa responsable del clivaje proteolítico del péptido señal de la proteína precursora, que libera la proteína madura hacia el espacio extra citoplasmático, es decir, el periplasma o espacio extracelular de las bacterias Gram-negativas, o hacia el espacio extracelular de las bacteria Gram-positivas La secuencia de reconocimiento de la señal de peptidasa II consenso estereotípica es Ala-Xaa-Ala. La regla para el residuo Xaa es que típicamente no puede ser un residuo de aminoácido grande o con carga Por el contrario, los tres residuos adyacentes al residuo de aminoácido amino-terminal (+1) en la proteína madura TIC901 no se ajustan a la secuencia de reconocimiento de la señal peptidasa de tipo II consenso. En su lugar, la secuencia de aminoácidos del péptido señal TIC901 que puede ser reconocida por una señal de peptidasa alternativa o desconocida hasta el momento y/o no caracterizada es Ser-Gln-Gln (S-Q-Q), que es extremadamente poco característico de ninguna de las secuencias de reconocimiento de señales de peptidasa observadas previamente Como alternativa, podría haber alguna peptidasa no identificada hasta el momento capaz de reconocer la estructura general de los

aminoácidos treinta y uno (31) a cuarenta y tres (43) que se muestra en la SEQ ID N° 4 y de eliminar estos aminoácidos para liberar la proteína insecticida activa hacia el medio u otros espacios extracelulares circundantes Estos últimos trece aminoácidos de las posiciones treinta y uno (31) a cuarenta y tres (43) que se muestran en la SEQ ID N°.4 pueden servir para amortiguar la toxicidad de la proteína hasta que sea liberada de la célula

Ejemplo 6

Este ejemplo ilustra una comparación entre la actividad biológica insecticida de la proteína purificada de la cepa nativa EG2158 de *Bacillus thuringiensis* con la actividad de la proteína insecticida aislada de la cepa recombinante EG12450 de *Bacillus thuringiensis*.

Se obtuvieron muestras de medio usado de cultivos de las cepas EG12450, EG2158 y EG10650 y se concentraron las proteínas solubles presentes en cada una de las muestras y luego fueron sometidas a fraccionamiento con DEAE de acuerdo con el método descrito en el Ejemplo 3 de la presente. Se agruparon las fracciones que contienen la proteína correspondiente a TIC901 para cada cepa. Las fracciones agrupadas fueron comparadas en paralelo en un bioensayo contra el escarabajo de papa de Colorado. Las fracciones de DEAE del medio usado de EG12450 y EG2158 contenía 87 µg/µl y 22 µg/µl de proteína TIC901, respectivamente, medido por densitometría sobre gel usando cantidades conocidas de BSA como estándar La toxicidad para las larvas CPB se midió aplicando distintas cantidades de las fracciones a la superficie de la dieta para insectos en recipientes, como se describió en el Ejemplo 1. Los resultados se muestran en la Tabla 2.

Tabla 2. Porcentaje de mortalidad de larvas CPB.

	Cantidad de fracción DEAE aplicada				
Cepa	50µl	25µl	12,5µl	6,3µl	3,1µl
EG2158	97%	78%	47%	25%	19%

EG10650	3	0	6	3	9
EG12450	91	94	72	38	16

Los resultados que se muestran en la Tabla 2 demuestran que las fracciones DEAE que contienen la proteína TIC901 de las cepas EG2158 y EG12450 de *B. thuringiensis* son tóxicas para las larvas CPB Se usaron fracciones DEAE similares de la cepa EG10650 como control negativo y no eran tóxicas para las larvas CPB. No se calcularon los errores estándar de la media.

Ejemplo 7

Este ejemplo ilustra el aislamiento de un fragmento de restricción *NsiI-HincII* de 1,96 kb, que contenía toda la región de codificación de TIC901 del pEG1379, y un análisis de la actividad biológica inhibidora para los gusanos de la raíz del maíz de la proteína TIC901 aislada de cultivos usados de la cepa recombinante SIC8098 que expresa la proteína TIC901 a partir de un plásmido que contiene al fragmento *NisI-HincII*

Se subclonó un fragmento de restricción *NsiI-HincII* de 1,96 kb que contenía toda la región de codificación TIC901 del pEG1379 en los sitios de restricción compatibles (*PstI* y *SmaI*) en el vector ambivalente para *B. thuringiensis-E Coli*, pEG597 (Baum et al., 1990) El único marco de lectura abierto en este fragmento de restricción capaz de codificar una proteína del tamaño de la TIC901 es la secuencia de codificación que se muestra en la SEQ ID Nº 3 El plásmido recombinante resultante, denominado pIC17048, se introdujo en la cepa huésped acristalífera EG10650 de *B thuringiensis* por electroporación. La cepa recombinante de *B. thuringiensis* que contiene al pIC17048 se denominó cepa SIC8098.

Se cultivaron las cepas EG10650 y SIC8098 de *B thuringiensis* en 300 ml de caldo Terrific (en frascos de 1 l) durante 40 horas a 30 °C con agitación vigorosa Los cultivos usados se centrifugaron a 8 000 rpm y 4 °C durante 30 minutos en un rotor Sorvall GS3. Se analizaron alícuotas de los sobrenadantes de cultivo por electroforesis sobre gel de SDS-poliacrilamida. La cepa SIC8098 acumula niveles

altos de la proteína TIC901 en el sobrenadante del cultivo. La evaluación de las condiciones de crecimiento para EG2158 y para las cepas recombinantes de *B thuringiensis* que codifican TIC901 indicó que se acumulaba una cantidad sustancialmente mayor de la proteína TIC901 en el medio usado cuando las cepas fermentaban a 25 °C durante cuarenta y ocho (48) horas en lugar de las condiciones estándar de 30 °C durante dieciocho (18) horas. Esto sugiere que la proteína TIC901 no se produce óptimamente hasta las alcanzar las fases de crecimiento logarítmica tardía o estacionaria temprana

Los sobrenadantes de cultivo de EG10650 y SIC8098 se llevaron a un 85% de saturación con sulfato de amonio a 4 °C. Las proteínas precipitadas se recolectaron por centrifugación. Los precipitados de proteína fueron solubilizados en Tris-HCl 10 mM, EDTA 0,1 mM, Triton X-100 0,005% (pH 6,8) y dializados contra un exceso de volumen de 200 veces de la misma solución amortiguadora El material insoluble se agrupó en pelets por centrifugación y el dialisato clarificado que contenía la proteína TIC901 sustancialmente purificada se usó directamente en los bioensayos con insectos La cantidad de proteína TIC901 purificada en el dialisato se determinó por electroforesis sobre gel de SDS-poliacrilamida y densitometría, usando albúmina de suero bovino como estándar de proteínas Los bioensayos se condujeron en placas de 96 cavidades, donde cada cavidad contenía una dieta artificial para gusanos de la raíz. Se depositaron muestras que contenían toxina o muestras control sobre la dieta, proporcionando una capa superficial impregnada en toxina. Se colocó un promedio de un huevo de gusano de la raíz por cavidad. Se utilizó un promedio de veinticuatro (24) cavidades por tratamiento Las placas se sellaron y se incubaron durante una semana antes de caracterizar la mortalidad y masa de las larvas.

Se condujeron bioensayos contra larvas de gusanos de la raíz del maíz del sur y del oeste, usando la toxina Cry3Bb para gusanos de la raíz (variante de la secuencia de aminoácidos 11231, English et. al., Patente de EE.UU Nº:6.063.597) como control positivo. Los resultados del ensayo se muestran en las Tablas 3 y 4.

Tabla 3. Bioensayo con TIC901 vs gusanos de la raíz del maíz del oeste

Cepa de origen	Toxina mg/ml	Masa larvaria media	CI[1] 95%	Significancia[2]
EG10650	0	0,44	0,35-0,53	
Cry3Bb	0,25	0,18	0,13-0,22	*
Cry3Bb	0,50	0,15	0,10-0,20	*
Cry3Bb	1,0	0,13	0,10-0,16	*
TIC901	0,25	0,36	0,25-0,48	
TIC901	0,50	0,48	0,36-0,61	
TIC901	1,0	0,29	0,26-0,31	*
TIC901	2,0	0,27	0,23-0,32	*

[1] Intervalo de confianza del 95%

[2] Un asterisco (*) indica que los resultados eran significativamente diferentes de los obtenidos con el control negativo EG10650, prueba de Dunnett, alfa = 0,05

TABLA 4. BIOENSAYO CON TIC901 VS GUSANOS DE LA RAÍZ DEL MAÍZ DEL SUR

Cepa de origen	Toxina mg/ml	Masa larvaria media	CI[1] 95%	Significancia[2]
EG10650	0	0,64	0,55-0,72	
Cry3Bb	0,25	0,45	0,40-0,51	*
Cry3Bb	0,50	0,42	0,39-0,44	*
Cry3Bb	1,0	0,31	0,19-0,42	*
TIC901	0,25	0,38	0,31-0,45	*
TIC901	0,50	0,33	0,20-0,45	*
TIC901	1,0	0,26	0,14-0,38	*

[1] Intervalo de confianza del 95%

[2] Un asterisco (*) indica que los resultados eran significativamente diferentes de los obtenidos con el control negativo EG10650, prueba de Dunnett, alfa = 0,05

Los resultados demuestran que la TIC901 es activa contra ambas especies de

gusanos de la raíz, pero que presenta una mayor actividad contra la especie de gusanos de la raíz del maíz del sur.

Ejemplo 8

Este ejemplo ilustra la construcción de la secuencia de nucleótidos que codifica una proteína TIC901 para aumentar su expresión en plantas.

Se usó la secuencia de aminoácidos de la proteína TIC901 que se muestra en la SEQ ID Nº:4 para construir la secuencia de nucleótidos esencialmente de acuerdo con Fischhoff y Perlak, Patente de EE UU Nº:5.500.365, y esencialmente de acuerdo con Brown et al , Patente de EE UU. Nº.5.689.052. Brevemente, se obtuvo una tabla de frecuencia de codones derivada de más de 53.000 segmentos de secuencia de ADN genómico de maíz, arroz y trigo considerados codificadores de la secuencia de la proteína en dichas plantas. Se derivaron más de diez millones novecientos mil codones (10 900.000) de estos segmentos de secuencia. Después se seleccionaron codones para usarlos en la construcción de la secuencia no natural o sintética que codifica TIC901 para su uso en plantas La secuencia de codificación resultante está comprende una composición de GC que se asemeja a las secuencias presentes en plantas, y en particular y dado que se usaron plantas monocotiledóneas para derivar la tabla de utilización de codones, la secuencia de codificación se asemeja más a un gen de monocotiledóneas en cuanto a arquitectura que a un gen de dicotiledóneas y por ello se prefiere su uso para expresar la proteína TIC901 o variantes de la proteína en especies vegetales monocotiledóneas. Una de dichas secuencias de una secuencia de codificación no natural preferido para su uso en especies monocotiledóneas se muestra en la SEQ ID Nº:13, y la secuencia de aminoácidos de la proteína codificada se muestra en la SEQ ID Nº:14. El especialista comprenderá que la SEQ ID Nº:13 es una entre múltiples secuencias de codificación que podrían funcionar para expresar la proteína insecticida en una especie vegetal monocotiledónea. Aún más, el especialista comprenderá que también se encuentra dentro del alcance de la presente invención una secuencia de codificación sintética,

tal como la SEQ ID Nº.13, que codifica todo o una porción insecticida de la proteína TIC901 o una variante de la secuencia de aminoácidos insecticida de la misma formada por una secuencia distinta a la que se muestra en la SEQ ID Nº·4, pero que se hibridiza con la SEQ ID Nº.13 bajo condiciones de hibridación severas.

Ejemplo 9

Este ejemplo ilustra la identificación y caracterización de homólogos de *tic*901 y péptidos codificados por esta secuencia de nucleótidos homóloga que son sustancialmente similares y por consiguiente están relacionadas con la proteína TIC901

Tal como se muestra en la Tabla 1, se identificaron numerosas cepas que produjeron perfiles de proteína extracelular que presentaba actividad biológica insecticida para coleópteros. La secuencia de codificación de la proteína TIC901 se identificó, tal como se ilustró en los Ejemplos anteriores, a partir de una biblioteca de plásmidos de la cepa EG2158 mediante pruebas con sondas con la SEQ ID Nº·2, correspondiente a secuencias que codifican la SEQ ID Nº.1 y con preferencia por la utilización de codones de los genes nativos de *Bacillus thuringiensis* que codifican δ-endotoxinas. Después se usó la secuencia de codificación *tic*901 como sonda para identificar homólogos de *tic*901 en clones de bibliotecas de plásmidos genómicos preparados a partir de genomas de otras cepas de *Bacillus thuringiensis* que presentan proteínas secretadas con actividad insecticida cuyo perfil RFLP es diferente del perfil RFLP correspondiente a la cepa EG2158

Tal como se muestra en la Tabla 1, la cepa 86833 de *Bacillus thuringiensis* contenía una secuencia homóloga *tic*901 diferente de la secuencia de la cepa EG2158 El perfil de proteínas secretadas de la cepa 86833 contiene una banda importante a 38 kDa que es similar a la banda generada por la cepa EG2158. Estos hechos tomados en su conjunto sugieren que la cepa 86833 contenía un homólogo *tic*901. La secuencia N-terminal de la proteína de la cepa 86833 fue determinada por degradación de Edmund, y se identificó una secuencia de quince (15) aminoácidos

(los aminoácidos setenta y seis (76) a noventa y uno (91) que se muestran en la SEQ ID Nº·6, que también corresponde a los residuos 2-16 de la proteína madura TIC1201 predicho) La secuencia de aminoácidos no correspondía a ninguna secuencia de aminoácidos conocida previamente, pero presentaba un 87% de identidad con la correspondiente secuencia de aminoácidos amino-terminal de la proteína madura TIC901 predicha descrita previamente en la presente. Se identificó un clon plasmídico, el pJP263, derivado de una biblioteca de ADN Lambda ZAP Express preparada usando ADN total, digerido parcialmente con *Sau*IIIA, seleccionado por tamaño, de la cepa 86833 que contiene un inserto de aproximadamente 5 pares de kilobases utilizando pruebas con sondas esencialmente como se describió en el Ejemplo 2 con el ADN de *tic*901 marcado con digoxigenina bajo condiciones de amplificación térmica El plásmido pJP263 fue sometido a condiciones de amplificación térmica para recortar el inserto completo de la cepa 86833 que posteriormente se insertó en el vector ambivalente *E coli / Bacillus,* pEG597. El plásmido resultante pJP274-1 se usó para transformar la cepa acristalífera EG10650 y se seleccionó una forma aislada que contenía este plásmido y se denominó cepa recombinante SIC4018 de *B thuringiensis*. La cepa SIC4018 se fermentó como se describe en el Ejemplo 1 y se concentraron las proteínas presentes en el medio de cultivo utilizado y luego fueron evaluadas por su actividad insecticida en un bioensayo y analizado por SDS-PAGE El medio de cultivo usado dio positivo para la bioactividad insecticida para coleópteros y se identificó una proteína de aproximadamente 38 kDa mediante análisis por SDS PAGE presente en una cantidad sobreabundante en el medio de cultivo usado. El inserto en el pJP274-1 fue secuenciado utilizando el método con dideoxi de Sanger y se identificó un solo marco de lectura abierto que se muestra en la SEQ ID Nº.5 desde los nucleótidos cuatrocientos treinta y siete (437) hasta mil seiscientos veintiuno (1621). La secuencia de aminoácidos del precursor de la proteína insecticida deducida a partir del marco de lectura abierto que se muestra en la SEQ ID Nº:5 se denominó TIC1201

(SEQ ID Nº 6) También se caracterizó la secuencia de nucleótidos de un clon correspondiente al perfil por RFLP de tipo "B" de otra cepa, EG3618, y se determinó que contenía un ORF idéntico en cuanto a secuencia al que se muestra en la SEQ ID Nº:5

La cepa EG6618 presentaba un perfil RFLP de tipo "C" como se ilustra en la Tabla 1. Se identificó un clone plasmídico, pJP264, derivado de una biblioteca de ADN Lambda ZAP Express preparada usando ADN total parcialmente digerido con *Sau*IIIA, seleccionado por tamaño, de la cepa EG6618 que contenía un fragmento de aproximadamente 5 pares de kilobases, mediante pruebas con sondas esencialmente como se describió en el Ejemplo 2, con el ADN *tic*901 marcado con digoxigenina bajo condiciones de amplificación térmica. . El análisis de la secuencia de ADN del ADN insertado en el pJP264 indicó que había un marco de lectura abierto incompleto, o parcial, en este clon, que posiblemente codificaba la porción C-terminal de una proteína nueva y no caracterizada previamente. La alineación de este ORF parcial con el ORF que codifica la TIC901 que se muestra en la SEQ ID Nº:3 reveló que la secuencia de nucleótidos del ADN insertado en el pJP264 (nucleótidos 481 a 1302 en la SEQ ID Nº:7) codificaba un péptido, cuya secuencia de aminoácidos deducida era sustancialmente similar a la secuencia de aminoácidos correspondiente a los residuos de aminoácidos 96 a 367 de TIC901. La nueva secuencia de aminoácidos codificada por la secuencia de nucleótidos en el pJP264 presentaba una identidad del 78 por ciento aproximadamente con la secuencia análoga del péptido TIC901. Las secuencias de nucleótidos que codifican estos péptidos presentaron una identidad del 81 por ciento aproximadamente. La secuencia de codificación parcial contenida en el pJP264 corresponde a las posiciones de nucleótidos 481 a 1302 en la SEQ ID Nº:7 y es análoga de la secuencia de codificación TIC901 que se muestra en la SEQ ID Nº:3 entre las posiciones de nucleótidos 438 a 1253. La identidad sustancial entre los extremos carboxilo-terminales de cada una de estas dos proteínas sugería que el péptido

codificado por la secuencia en el pJP264 representaba un nuevo miembro del grupo de proteínas insecticidas relacionadas con la proteína TIC901. Por ello, esta proteína se denominó TIC407 y la secuencia de nucleótidos que codifica la proteína TIC407 se denominó *tic*407

Se consideró que existía un ORF de longitud completa para la TIC407 dentro del genoma de EG6618 Con el fin de identificar la secuencia de nucleótidos que equilibra la porción 5' de la secuencia de codificación TIC407, se digirió una muestra de ADN total de la cepa EG6618 de *Bacillus thuringiensis* con la enzima de restricción *Nde*I, cuyos productos fueron diluidos y circularizados por ligación para generar una población de templados para su uso una reacción de amplificación térmica inversa La enzima *Nde*I se seleccionó por (1) no corta la secuencia del en el pJP264, (2) a menudo corta dentro del genoma de *Bacillus thuringiensis* debido a la preponderancia de A y T en la secuencia de reconocimiento de la enzima de restricción y (3) el análisis por transferencia Southern del ADN de EG6618 indicaba que *Nde*I cortaba, razonablemente, cerca de los puntos terminales del ADN insertado en el plásmido pJP264 pero suficientemente lejos de dichos puntos terminales como para incorporar un ORF de longitud completa que codifica TIC407.

Se prepararon dos cebadores divergentes inversos para amplificación térmica basado en la secuencia de nucleótidos parcial *tic*407 presentes en el pJP264. El cebador prJWP151 (SEQ ID Nº:15 CCTTTGGCAGAAACTTTAACTCC) corresponde al complemento inverso de los nucleótidos 766-788 que se muestran en la SEQ ID Nº 7 El cebador prJWP152 (SEQ ID Nº:16 GTGTATTCTGGTACGCATGAC) corresponde a los nucleótidos 955-975 que se muestran en la SEQ ID Nº.7. Estos cebadores se incluyeron junto con las moléculas de templado circularlizados con *Nde*I descritas previamente y los reactivos necesarios, soluciones amortiguadoras y polimerasa *Pfu* en una reacción de amplificación térmica formada por un paso de desnaturalización de cinco minutos a 95 °C durante cinco minutos, seguido de treinta ciclos formados por un paso de alineación a menor temperatura de un minuto a 37

°C, un minuto paso de extensión de dos minutos a 55 °C y un paso de desnaturalización de un minuto a 95 °C, terminando con un minuto paso de extensión de cinco minutos a 55 °C y embebimiento a 4 °C hasta el posterior procesamiento de las muestras Se identificó un único producto de la amplificación térmica por electroforesis sobre gel, que se recortó y se clonó en un vector T para construir al plásmido pJP300-2. Se obtuvo la secuencia de nucleótidos de la secuencia del ADN insertado en el pJP300-2 y proporcionó suficiente secuencia de nucleótidos 5' y 3' con respecto a la secuencia anterior presente en el pJP264 como para identificar al marco de lectura abierto completo que codifica a la proteína insecticida TIC407, así como secuencias flanqueadoras 5' y 3'. La información de secuencia de nucleótidos obtenida del pJP300-2 se obtuvo por computadora usando software de análisis de secuencia porque la secuencia presente en el pJP300-2 se construyó usando amplificación térmica inversa de un fragmento de restricción que se circularizó a partir del genoma de la cepa EG6618 Por ello, para confirmar que la secuencia que se había obtenido era precisa y que de hecho codificaba la secuencia de aminoácidos TIC407, primero desde el marco de lectura abierto parcial identificado en el pJP264 y luego desde las secuencias parciales identificadas en el pJP300-2, se diseñaron dos nuevos cebadores de amplificación térmica que flanqueaban el marco de lectura abierto que codifica TIC407 (prJWP186 y prJWP183) para su uso en la amplificación de la secuencia de codificación completa de TIC407 a partir de ADN genómico de EG6618

La construcción del cebador prJWP186 (SEQ ID Nº:17 gccggatccCTAGCTGAATATGCAGTAGATAATG), se basó en parte en la secuencia identificada como 5' con respecto del marco de lectura abierto TIC407 en el pJP300-2. Los primeros nueve (9) nucleótidos de este cebador corresponden a la secuencia de nucleótidos que no está presente en la secuencia identificada en el pJP264 o bien en el pJP300-2, y también contiene una secuencia que incorpora un sitio de restricción *Bam*HI en la secuencia 5' con respecto del amplicón producido usando

estos cebadores Los veinticinco (25) nucleótidos terminales en cebador prJWP186 corresponden a la secuencia que se muestra en los nucleótidos veintiocho (28) a cincuenta y dos (52) en la SEQ ID Nº·7, que se encuentra 5' con respecto de la secuencia de inicio ATG propuesta para el ORF de TIC407 y permite extender al cebador hacia el ORF del TIC407.

La construcción del cebador prJWP183 (SEQ ID Nº·18 GTGGCACGTTTATAGGCCATTGTTC) se basó por completo en la secuencia deducida considerada dentro de la secuencia 3' con respecto al ORF TIC407 dentro del pJP300-2, y es el complemento inverso de la secuencia entre las posiciones de nucleótidos mil setecientos treinta y cinco (1.735) y mil setecientos cincuenta y nueve (1.759) que se muestra en la SEQ ID Nº·7 Aunque no se han incorporado secuencias para el sitio de restricción en este cebador, existe una secuencia de reconocimiento *Hind*III natural ubicada 3' con respecto al ORF TIC407 como se muestra en la SEQ ID Nº:7 desde los nucleótidos mil quinientos hasta mil quinientos cinco (1.500-1.505).

Se incluyeron los cebadores prJWP186 y prJWP183 TIC407 en una reacción de amplificación térmica junto con una muestra de ADN total de EG6618 y los reactivos necesarios para llevar a cabo la reacción Las condiciones de amplificación eran similares a los descritos previamente para la amplificación de la muestra circularizada con *Nde*I Se aisló un amplicón que comprende 1.732 nucleótidos correspondientes al ORF TIC407, incluyendo secuencia de nucleótidos flanqueadoras 5' y 3' del ORF. Este amplicón se clonó en un vector T para construir el pJP306-4 y se obtuvo la secuencia de nucleótidos completa del amplicón. Tal como se esperaba, la secuencia de nucleótidos del amplicón en el pJP306-4 correspondía a los nucleótidos veintiocho (28) a mil setecientos cincuenta y nueve (1 759) que se muestra en la SEQ ID Nº.7, y contenía un marco de lectura abierto que se muestra en la SEQ ID Nº 7 entre las posiciones de nucleótidos ciento noventa y seis (196) a mil doscientos noventa y nueve (1.299) que codifica una proteína

formada por 368 aminoácidos. La alineación de la secuencia de nucleótidos que codifica TIC407 con la secuencia que codifica TIC901 demostró que los dos ORF son aproximadamente un ochenta por ciento (80%) idénticas, sin embargo, las secuencias de aminoácidos deducidas a partir de los ORF solamente presentan un 74% de identidad aproximadamente.

Se alinearon las secuencias de aminoácidos de TIC901, TIC1201 y TIC407 para formar una secuencia de aminoácidos consenso Aún cuando las secuencias de aminoácidos primarias eran significativamente diferentes, eran fácilmente identificables las regiones de identidad sustancial Dos de las regiones, correspondientes a la secuencia de aminoácidos ASN-ASN-ASN-HIS-GLN-THR-ASN-ARG de las posiciones en la secuencia de aminoácidos 96-103 de la SEQ ID Nº:4 y la secuencia de aminoácidos GLN-LYS-PHE-ILE-TYR-PRO-ASN de las posiciones de aminoácidos 276-282 de la SEQ ID Nº·4, estaban un 100% conservadas en la secuencia y posición primarias dentro de cada una de las tres secuencias de proteínas insecticidas. Las secuencias de nucleótidos que codifican estas secuencias de proteínas en cada uno de los correspondientes marcos de lectura abiertos, es decir, como se muestra en cada una de las secuencias en las SEQ ID Nº·3, SEQ ID Nº.5 y SEQ ID Nº.7, (TIC901, TIC1201 y TIC407, respectivamente) también estaban sustancialmente conservadas, proporcionando un medio para identificar otras secuencias homólogas que codifican dichas proteínas insecticidas secretadas en otras cepas de *Bacillus* o *Bacillus thuringiensis*

Se prepararon dos secuencias de cebadores de amplificación térmica para su uso en una reacción de amplificación térmica destinada a identificar las secuencias de nucleótidos que codifican otros homólogos de las tres proteínas insecticidas TIC901, TIC1201 y TIC407. Las secuencias de los cebadores de amplificación se muestran aquí en la (1) SEQ ID Nº:11, una secuencia de un cebador de amplificación directo de veintiún mer denominada prJWP139, correspondiente a la secuencia de codificación que se muestra en la SEQ ID Nº.3 entre las posiciones de nucleótidos

438 a 458 excepto que los nucleótidos de la posición dieciocho (18) del oligonucleótido son redundantes por cuanto se puede incorporar un nucleótidos adenosina o timidina en dicha posición, dando como resultado dos posibles secuencias cebadoras y (2) SEQ ID Nº:12, una secuencia de un cebador de amplificación inversa de vente mer denominada prJWP143, correspondiente al complemento inverso de la secuencia que se muestra en la SEQ ID Nº:3 entre las posiciones de nucleótidos 998 a 978, excepto que los nucleótidos del oligonucleótido en la posición cuatro (4) y/o en la posición diez (10) es/son redundantes por cuanto se puede incorporar un nucleótido adenosina o timidina en cualquiera de dichas posiciones, dando como resultado cuatro secuencias de cebadores posibles. Las SEQ ID Nº 11 y SEQ ID Nº.12 comprenden una secuencia de nucleótidos con codones preferidos para su uso en las secuencias genéticas derivadas de *Bacillus thuringiensis* o otras cepas de especies de *Bacillus*, en las cuales los codones contienen A y/o T en la tercera posición para cada codón representado en dicha la secuencia.

Como se indicó previamente, la secuencia de aminoácidos identificada por degradación de Edmund de la proteína aislada de 38 kDa aproximadamente y purificada a partir del medio de cultivo usado en las fermentaciones con la cepa EG2158 era diferente en un aminoácido de la correspondiente secuencia deducida fundamentalmente del marco de lectura abierto para TIC901 que se muestra en la SEQ ID Nº:4. Se consideró que la diferencia en cuanto a secuencias de aminoácidos observada por la degradación de Edmund y la secuencia de aminoácidos deducida de TIC901 resultante del ORF *tic*901 podría deberse a la presencia de más de un gen en la cepa EG2158, donde cada uno de dichos genes codifica una proteína insecticida secretada diferente y cada proteína secretada tiene un tamaño de aproximadamente 38 kDa. De hecho, si hubiera dos o más genes en EG2158, cada uno tendría que estar presente en fragmentos *Hind*III casi idénticos o ambos se encontrarían en el mismo fragmento *Hind*III o uno podría ser suficientemente

diferente de la secuencia de codificación *tic*901 como para no hibridizarse con la secuencia de nucleótidos que comprende la SEQ ID Nº:2 bajo condiciones de hibridación severas. Si cada uno de los genes codificara realmente una proteína insecticida homóloga, o al menos una proteína que fuera suficientemente similar a la proteína codificada por la otra secuencia de modo que las secuencias conservadas identificadas previamente estuvieran en cada secuencia de codificación, entonces los cebadores redundantes, tal como los que se muestran en las SEQ ID Nº:11 y SEQ ID Nº:12, pueden permitir la identificación de los genes por amplificación de las secuencias de codificación relacionadas a partir de una muestra común en una reacción de amplificación térmica. Por ejemplo, se esperaría que los cebadores que comprenden las secuencias que se muestran en las SEQ ID Nº.11 y SEQ ID Nº:12 produzcan un amplicón de aproximadamente 560 pares de bases cuando se usan las secuencias de codificación TIC901, TIC1201 o TIC407 como templados en una muestra Con el fin de evaluar esta hipótesis, se usó ADN total EG2158 como templado en una reacción de amplificación térmica usando el juego de cebadores de amplificación definido en las SEQ ID Nº:11 y SEQ ID Nº:12 bajo condiciones de reacción estándar que comprenden 10-100 nanogramos de templado de ADN genómico bacteriano, 50 picomoles de cada cebador, solución amortiguadora para la reacción de amplificación ROCHE 1X que contiene una concentración final 1,5 mM del catión divalente Mg^{2+}, 0,2 mM de cada deoxinucleótido trifosfato (dNTP) y 2,5 unidades de TAQ polimerasa ROCHE por cada 50 µl de volumen de reacción, un ciclo de desnaturalización inicial a 94 °C por dos minutos, treinta (30) ciclos de amplificación que comprenden una fase de desnaturalización a 94 °C durante un minuto, una fase de alineamiento a una temperatura menor de 45 °C durante dos (2) minutos y una fase de extensión con cebador a 72 °C durante un (1) minuto, seguido de una extensión final con cebador a 72 °C y una fase de terminación de cinco (5) minutos, y embebimiento a 4 °C Además, también se usó el templado de ADN obtenido de las siguientes cepas en reacciones de amplificación térmica paralelas

junto con los cebadores de amplificación que se muestran en las SEQ ID Nº.11 y SEQ ID Nº:12: cepa EG10650, cepa EG4332, cepa EG5552, cepa EG5858, cepa EG6489, cepa EG6561, cepa EG3618, cepa EG6555, cepa EG6618 y cepa 86833. Las muestras de cada reacción de amplificación térmica fueron analizadas por electroforesis sobre gel Los resultados se muestran en la Tabla 5

Tabla 5. Amplicones producidos a partir de cepas de *B. thuringiensis*

ADN templado	Tipo de RFLP *Hindi*	Presencia del Amplicon[1]
EG10650	-	-
EG4332	-	-
EG5552	-	-
EG5858	-	-
EG2158	A	+
EG6489	A	+
EG6561	A	+
EG3618	B	+
EG6555	B	+
86833	B	+
EG6618	C	+

[1] Amplicón producido en una reacción de amplificación térmica usando el ADN templado de la cepa indicada, los cebadores de amplificación que comprenden las secuencias que se muestran en las SEQ ID Nº 11 y SEQ ID Nº·12 bajo condiciones estándar como se describió previamente

No se detectaron amplicones cuando se usó ADN de las cepas EG10650, EG4332, EG5552 o EG5858 como templado junto con este juego particular de cebadores. Este resultado sugiere que estas cepas no contienen secuencias de nucleótidos que codifican un homólogo de TIC901, TIC1201 y TIC407 Sin embargo, aparentemente cada una de las otras cepas evaluadas producía un solo amplicón de

aproximadamente 560 pares de bases de longitud dentro de los límites de resolución del sistema de geles utilizado. Los amplicones de cada una de las muestras que producían una reacción positiva bajo estas condiciones de amplificación fueron clonados en un vector T y se determinó la secuencia de nucleótidos para los amplicones clonados seleccionados aleatoriamente de cada reacción de amplificación térmica. Cada una de las cepas EG2158, EG6489 y EG6561 produjo un amplicón que presentaba una secuencia de nucleótidos idéntica a la correspondiente secuencia de codificación TIC901 Las cepas EG3618 y 86833 produjeron un amplicón que presentaban una secuencia de nucleótidos idéntica a la secuencia de codificación correspondiente a TIC1201 La cepa EG6555, que es una cepa con un perfil de RFLP *Hind*III de tipo "B" idéntico al observado con las cepas EG3618 y 86833, también produjeron un amplicón de 560 pares de bases. La cepa EG6555 no se había caracterizado por completo, sin embargo, se había demostrado previamente que generaba un RFLP de tipo "B", tal como se muestra en la Tabla 1, que corresponde al fenotipo RFLP que presenta la cepa 86833 a partir de la cual se aisló TIC1201. Por ello, se anticipó que el homólogo presente en la cepa EG6555 codificaba una proteína sustancialmente similar, si no idéntica, a TIC1201. El análisis de secuencia de un amplicón obtenido a partir de EG6555 indicó que era idéntico a la correspondiente secuencia obtenida con la cepa 86833 Como era esperado, la cepa EG6618 produjo un amplicón con una secuencia de nucleótidos idéntica a la correspondiente secuencia de codificación TIC407

Sorpresivamente, además de los amplicones que comprenden la correspondiente secuencia *tic*901, se derivaron amplicones de ADN total aislado de EG2158 que contenía una segunda secuencia de codificación que era diferente de la correspondiente secuencia de nucleótidos TIC901 identificada usando la sonda redundante que se muestra en la SEQ ID N°:2 Como se anticipara, un amplicón obtenido del ADN EG2158 también comprendía aproximadamente 560 nucleótidos, pero el análisis de la secuencia de ADN del clon dio como resultado la identificación

de un ORF (los nucleótidos 377 a 937 que se muestran en la SEQ ID Nº.9) que codifica la secuencia de un péptido (secuencia de aminoácidos 96 a 282 en la SEQ ID Nº·10) que era sustancialmente diferente de la correspondiente secuencia del péptido para TIC901. Este resultado sugería que había al menos una secuencia de nucleótidos adicional en la cepa EG2158 que codificaba una proteína homóloga de TIC901 que se podía considerar como otro miembro del grupo de proteínas relacionadas con TIC901, TIC1201 y TIC407.

Con el fin de caracterizar completamente esta segunda secuencia identificada a partir del ADN de la cepa EG2158, se diseñaron cebadores de amplificación térmica inversos basados en la segunda secuencia que específicamente podían amplificar las secuencias flanqueadoras 5' y 3' con respecto a la segunda secuencia. Se usó una amplificación térmica inversa del ADN de EG2158 para identificar el marco de lectura abierto completo que contenía esta segunda secuencia. Primero, se seleccionaron las enzimas que se podían usar en la digestión del ADN de EG2158 que (1) no efectuaban cortes dentro de la segunda secuencia identificada previamente y (2) frecuentemente efectuaban cortes en el ADN de *Bacillus* debido a la secuencia rica en AT contenida en la secuencia de reconocimiento de la enzima. Teniendo en cuenta estos parámetros, se seleccionaron las enzimas *Xba*I y *Spe*I, pero ninguna de las enzimas digería el ADN de EG2158, cuando se utilizaban solos, en fragmentos que se pudieran circularizar fácilmente para su uso como templados de amplificación térmica inversa. Sin embargo, dado que ambas enzimas producen extremos 5' superpuestos compatibles, la digestión doble del ADN de EG2158 usando ambas enzimas permitió obtener productos de un tamaño suficientemente pequeño que, cuando eran circularizados, se podían usar como templados para una amplificación térmica inversa Los cebadores de la amplificación inversa fueron diseñados sobre la base de la segunda secuencia identificada previamente, en particular las secuencias en las cuales la correspondiente secuencia *tic*901 (SEQ ID Nº 3) y la segunda secuencia exhibían diferencias significativas con la alineación.

El cebador prJWP155 (SEQ ID Nº:19) corresponde a y es el complemento inverso de la segunda secuencia que se muestra en la SEQ ID Nº 9 en los nucleótidos 454-476). El cebador prJWP156 (SEQ ID Nº:20) corresponde a la segunda secuencia que se muestra en la SEQ ID Nº:9 en los nucleótidos 692-717. Estos dos cebadores se usaron en una reacción de amplificación con ADN templado de EG2158, circularizado, de doble digestión, como se indicó previamente, bajo condiciones similares a las definidas previamente para la amplificación térmica inversa del ADN de EG6618 Se obtuvo un solo amplicón que se insertó en un vector T para construir el plásmido pJP290-1 y se determinó la secuencia de nucleótidos del amplicón insertado. La secuencia de la amplificación térmica inversa derivada usando este método se ensambló usando software de análisis de secuencias de ADN y se alineó con la segunda secuencia identificada previamente para construir la secuencia que se muestra en la SEQ ID Nº:9 que incluía el marco de lectura abierto completo que codifica la secuencia del péptido denominada TIC417, así como secuencias flanqueadoras 5' y 3'.

Para confirmar la secuencia, ensamblada en parte del segundo amplicón identificado a partir de ADN de EG2158 tal como se mostró previamente, y en parte de la reacción de amplificación térmica inversa usando los cebadores prJWP155 y 156, se diseñaron dos cebadores de amplificación térmica adicionales que eran complementarias de las secuencias que flanqueaban el marco de lectura abierto que codifica TIC417 (prJWP168 y prJWP170). Estos cebadores permitirían amplificar el marco de lectura abierto completo de TIC417 junto con una extensión corta de ADN 5' y 3' con respecto al ORF de TC417. El cebador prJWP168 (SEQ ID Nº.21) se construyó basado en parte en la secuencia identificada como 3' o como 3' con respecto al ORF de TIC417 y funciona extendiendo la polimerización de la secuencia de nucleótidos en el ORF de TIC417 Los primeros cinco nucleótidos de este cebador corresponden a nucleótidos que no están presentes en la secuencia cuya presencia fue identificada en el plásmido pJP290-1 y también sirven para incorporar un sitio de

restricción *Hind*III cuando se acopla con las bases 7-8 de la SEQ ID Nº.21. Los 39 nucleótidos terminales del prJWP168 corresponden al complemento inverso de los nucleótidos 1148 a 1186 como se muestra en la SEQ ID Nº:9. El cebador prJWP170 (SEQ ID Nº 22) se construyó basado en parte en la secuencia identificada como 5' o 5' con respecto al ORF de TIC417 y funciona extendiendo al cebador hacia el extremo 5' del ORF de TIC417. Los primeros 9 nucleótidos de este cebador corresponden a nucleótidos que no están presentes en la secuencia identificada en pJP290-1 y también sirven para incorporar una secuencia de reconocimiento del sitio de restricción *Bam*HI 5' con respecto al codón de inicio ATG predicho de TIC417 Los 30 nucleótidos terminales de este cebador corresponden a nucleótidos La proteína codificada por esta secuencia de codificación diferente se denominó TIC417. La secuencia de aminoácidos codificada por la porción clonada del amplicón TIC417 se muestra en la SEQ ID Nº:10, y la secuencia de codificación del amplicón clonado se muestra en la SEQ ID Nº:9 La secuencia de aminoácidos de TIC417 es un 85% idéntica aproximadamente a la correspondiente secuencia de TIC901 y TIC1201, y es un 76% idéntica aproximadamente a la correspondiente secuencia de aminoácidos de TIC407. Se cree que el fragmento de proteína correspondiente a la secuencia de codificación TIC417 revelada en el amplicón TIC417 derivado del ADN de EG2158 es una proteína insecticida similar a TIC901, TIC1201 y TIC407, correspondiente a la proteína que se expresa como una proteína precursora que contiene un péptido señal de aproximadamente 43 residuos de aminoácidos de longitud que termina con el clivaje de la secuencia del péptido señal consenso correspondiente a SER-GLN-GLN, que es procesado por una señal de peptidasa para liberar aproximadamente 38 kDa, más o menos 2 kDa aproximadamente, de proteína madura hacia el espacio extracelular. También se ha anticipado que la secuencia de aminoácidos madura amino-terminal TIC417 puede corresponder a la secuencia de aminoácidos que se muestra en la SEQ ID Nº:1, y que la proteína TIC417 presenta actividad biológica insecticida inhibidora de especies de coleópteros

contra gusanos de la raíz del maíz y escarabajos de papa de Colorado

Ejemplo 10

Este ejemplo ilustra el diseño y uso de sondas y cebadores degenerados para su uso en la identificación de secuencias de nucleótidos naturales que codifican todo o parte de las proteínas insecticidas derivadas de las proteínas de la presente invención u homólogos de las mismas

Según la alineación de las secuencias de aminoácidos de los precursores de las proteínas TIC901, TIC1201, TIC407 y TIC417, tal como se muestra en la Figura 1, se observaron numerosas regiones de identidad de secuencia de aminoácidos y se diseñaron secuencias de cebadores de oligonucleótidos degenerados basado en la correspondiente secuencia de nucleótidos de codificación nativa para cada proteína, teniendo en cuenta el perfil de las proteínas insecticidas conocidas de *Bacillus thuringiensis*. Si bien existen muchos segmentos de secuencias de aminoácidos conservadas entre los cuales se puede elegir debido a dicha alineación, tal como se muestra en la Figura 1, para diseñar cebadores degenerados, se seleccionaron tres segmentos según la posición de las secuencias cebadoras dentro de la región de codificación de las proteínas TIC conocidas y el tamaño único de los amplicones que se podrían generar en una reacción de amplificación térmica con estos cebadores.

Se identificó una primera región o segmento de secuencia de aminoácidos conservada correspondiente a los aminoácidos setenta y cinco (75) a ochenta y tres (83) como se muestra en la SEQ ID Nº:4, y la correspondiente secuencia de nucleótidos que codifica este segmento de aminoácidos de cada una de las secuencias que se muestran en la SEQ ID Nº.3, SEQ ID Nº 5, SEQ ID Nº:7 y SEQ ID Nº 9 cuando se alinean. Se utilizó este segmento de secuencia de nucleótidos consenso que codifica este segmento de secuencia de aminoácidos particular para determinar el alcance de las secuencias de nucleótidos que comprenden la secuencia de un cebador de amplificación directa que se muestra en la SEQ ID Nº 23

(prJWP200), así como otros dos cebadores degenerados formados por subconjuntos del cebador degenerado de la SEQ ID Nº·23 (SEQ ID Nº·24-25, correspondiente a los plásmidos prJWP201 y prJWP202, respectivamente).

La segunda región o segmento de secuencia de aminoácidos conservada entre las proteínas de la presente invención se ilustra con una secuencia de aminoácidos entre el aminoácido ciento cuarenta y siete (147) aproximadamente y ciento cincuenta y tres (153) aproximadamente como se muestra en la SEQ ID Nº.4. Después se alinearon las secuencias de nucleótidos que codifica este segmento de aminoácidos de cada una de las secuencias que se muestran en las SEQ ID Nº.3, SEQ ID Nº·5, SEQ ID Nº 7 y SEQ ID Nº·9 El segmento de la secuencia de nucleótidos consenso que codifica este segmento de secuencia de aminoácidos particular se utilizó para determinar el alcance de las secuencias de nucleótidos que comprenden la secuencia del cebador de amplificación directo que se muestra en la SEQ ID Nº 26 (prJWP203)

La tercera región o segmento de secuencia de aminoácidos conservada entre las proteínas de la presente invención se ilustra con una secuencia de aminoácidos entre el aminoácido ciento cuarenta y siete (275) aproximadamente y doscientos ochenta y tres (183) aproximadamente que se muestra en la SEQ ID Nº:4. Después se alinearon las secuencias de nucleótidos que codifica este segmento de aminoácidos de cada una de las secuencias que se muestran en las SEQ ID Nº 3, SEQ ID Nº 5, SEQ ID Nº 7 y SEQ ID Nº:9 Se utilizó este segmento de secuencia de nucleótidos consenso que codifica este segmento de secuencia de aminoácidos particular para determinar el alcance de las secuencias de nucleótidos que comprenden la secuencia de un cebador de amplificación inverso que se muestra en la SEQ ID Nº:27 (prJWP200), así como otros dos cebadores degenerados formados por subconjuntos del cebador degenerado de la SEQ ID Nº:27 (SEQ ID Nº·28-29, correspondiente a los plásmidos prJWP205 y prJWP206, respectivamente).

Se puede usar cualquier combinación de los cebadores prJWP200,

prJWP201, prJWP202 o prJWP203 con prJWP204, prJWP205 o prJWP206 juntos en una reacción de amplificación térmica con una secuencia de nucleótidos templado que codifica una proteína TIC901, TIC1201, TIC407, TIC417 o una proteína insecticida secretada relacionada o un homólogo de las mismas para producir un amplicón correspondiente al segmento de secuencia que se hibridiza con las correspondientes secuencias entre las posiciones comprendidas dentro de las secuencias de codificación TIC usadas para diseñar las secuencias de los cebadores prJWP200-206. Si se usaron cualquiera de los cebadores prJWP200-202 junto con cualquiera de los cebadores prJWP204-206 en una reacción de amplificación térmica con un templado apropiado en una muestra, la producción de un amplicón de un tamaño entre aproximadamente 590 y aproximadamente 650 pares de bases es característicamente diagnóstico de la presencia en la muestra de una o más de las secuencias de codificación que están relacionadas con las secuencias de la presente invención. En general, el amplicón de diagnóstico que emplea cebadores dentro de estos rangos comprenderá entre aproximadamente 617 y aproximadamente 626 pares de bases

El uso de cualquiera de los cebadores que comprenden las SEQ ID Nº:26 (prJWP203) junto con cualquiera de los cebadores que comprenden las SEQ ID Nº.27-29 (prJWP204-206) en una reacción de amplificación térmica con un templado apropiado en una muestra, permitirá producir un amplicón que comprende entre aproximadamente 390 y aproximadamente 415 pares de bases que característicamente es diagnóstico de la presencia en la muestra de una o más secuencias de codificación que están relacionadas con las secuencias de la presente invención En general, el amplicón de diagnóstico que emplea cebadores dentro de estos rangos comprenderá entre aproximadamente 400 y aproximadamente 410 pares de bases

Para ilustrar la utilidad de estos cebadores, se combinó el cebador prJWP200 (SEQ ID Nº 23) en una reacción de amplificación térmica con el cebador prJWP204

(SEQ ID Nº.27), cada uno a una concentración de al menos 1 pico-mol por micro-litro aproximadamente junto con solución amortiguadora de amplificación TAQ 1X, 0,2 molar de cada uno de los desoxinucleótidos trifosfato (dATP, dTTP, dCTP y dGTP), $MgCl_2$ 2 milimolar, 2 unidades de TAQ polimerasa y entre diez (10) y cien (100) nanogramos aproximadamente de una muestra de ADN genómico de la cepa EG2158 conocida por contener las secuencias de codificación para TIC901 y TIC417. Las condiciones de los ciclos de amplificación térmica consistieron de una desnaturalización inicial de 2 minutos aproximadamente a 94°C seguido de 35 ciclos de un paso de desnaturalización de 30 segundos a 94°C, un paso de alineación de 30 segundos a 50°C y un paso de extensión de 45 segundos a 72°C seguido de un paso de extensión final de 7 minutos a 72°C. La temperatura del paso de alineación se disminuyó en 0,3°C por cada ciclo sucesivo de modo que la temperatura de alineación final era de 39,8°C aproximadamente.

Se analizó una muestra de diez microlitros sobre un gel de agarosa TAE 1,2% y se comparó con un estándar de co-migración de 100 pares de bases, y se coloreó con bromuro de etidio. Los resultados indicaron que se había obtenido un segmento de amplicón correspondiente a 620 pares de bases aproximadamente. Se recortó la banda de 620 pares de bases aproximadamente y se recuperó el ADN para su clonación. Se aislaron varios clones independientes que representaban este segmento y se obtuvo la secuencia de nucleótidos para cada uno Tal como se esperaba, se identificó a una primera secuencia idéntica a la correspondiente secuencia de nucleótidos que codifica TIC901 (sin incluir las secuencias de los cebadores de cualquiera de los extremos del clon, entre la posición de nucleótido cuatrocientos dos (402) aproximadamente y la posición de nucleótido novecientos setenta y cuatro (974) aproximadamente que se muestra en la SEQ ID Nº:3), y se identificó una segunda secuencia idéntica a la correspondiente secuencia de nucleótidos que codifica TIC417 (sin incluir la secuencia de los cebadores de cualquiera de los extremos del clon, entre la posición de nucleótido cuatrocientos

sesenta y cuatro (464) aproximadamente y la posición de nucleótido mil treinta y seis (1.036) aproximadamente que se muestra en la SEQ ID N°.9).

Sorpresivamente, también se identificó una tercera secuencia (SEQ ID N°:30) que no correspondía idénticamente a ninguna de las secuencias de la presente. Sin embargo, dicha tercera secuencia era sustancialmente similar en cuanto a secuencia de nucleótidos a cada una de las secuencias que codifican una proteína insecticida secretada de la presente, incluyendo *TIC901, TIC1201, TIC407 y TIC417.* Tal como se indicó en los ejemplos anteriores, era bastante inesperado encontrar la secuencia de codificación *tic*417 en el ADN genómico de la cepa EG2158 Era aún más sorpresivo identificar aún una tercera secuencia de nucleótidos que probablemente correspondía a un segmento de nucleótidos que codificaba aún una tercera proteína insecticida secretada de la cepa EG2158 que era diferente de las que se describieron precedentemente en la presente, pero que era suficientemente similar en cuanto a secuencia como para ser clasificada como una de las especies pertenecientes al grupo de proteínas insecticidas secretadas codificadas por una secuencia de nucleótidos que se hibridiza con una o más de las secuencias de la presente, y es ilustrativa de la novedad y utilidad de las sondas y los cebadores de oligonucleótidos degenerados ilustrados aquí para su uso en la identificación de secuencias que codifican proteínas insecticidas secretadas y que se hibridizan bajo condiciones severas con las secuencias de codificación relacionadas *tic*901, *tic*1201 *tic*07 y *tic*417, descritas en este documento.

La secuencia de aminoácidos codificada por el marco de lectura abierto no interrumpido que se muestra en la SEQ ID N°:30, de la cual se han suprimido las secuencias de oligonucleótidos degeneradas de veintiséis mer de ambos extremos 5' y 3', se muestra en la SEQ ID N°:31. La secuencia de aminoácidos de la SEQ ID N°:31 es sustancialmente similar a la secuencia de aminoácidos de TIC901 entre las posiciones de aminoácidos ochenta y cinco (85) aproximadamente y doscientos setenta y cuatro (274) aproximadamente como se muestra en la SEQ ID N°:4, que

solamente contiene dos (2) aminoácidos diferentes de la secuencia análoga en SEQ ID Nº 4, correspondiente a una identidad del 98,9% aproximadamente entre la SEQ ID Nº:31 y la correspondiente secuencia en la SEQ ID Nº:4 La secuencia de aminoácidos de la SEQ ID Nº.31 es sustancialmente similar a la secuencia de aminoácidos de TIC1201 entre las posiciones de aminoácidos ochenta y cinco (85) aproximadamente y doscientos setenta y cuatro (274) aproximadamente que se muestra en la SEQ ID Nº.6, que solamente contiene trece (13) aminoácidos diferentes de la secuencia análoga en la SEQ ID Nº 6, correspondiente a una identidad del 93,2% aproximadamente entre la SEQ ID Nº:31 y la correspondiente secuencia en la SEQ ID Nº 6. La secuencia de aminoácidos de la SEQ ID Nº:31 es sustancialmente similar a la secuencia de aminoácidos de TIC417 entre las posiciones de aminoácidos ochenta y cinco (85) aproximadamente y doscientos setenta y cuatro (274) aproximadamente que se muestra en la SEQ ID Nº.10, que solamente contiene treinta (30) aminoácidos diferentes de la secuencia análoga en la SEQ ID Nº 10, correspondiente a una identidad del 83,7% aproximadamente entre la SEQ ID Nº.31 y la correspondiente secuencia en la SEQ ID Nº:10. La secuencia de aminoácidos de la SEQ ID Nº:31 también es sustancialmente similar a la secuencia de aminoácidos de TIC401 entre las posiciones de aminoácidos ochenta y cinco (85) aproximadamente y doscientos setenta y cuatro (274) aproximadamente que se muestra en la SEQ ID Nº 8, que contiene cuarenta y un (41) aminoácidos diferentes de la secuencia análoga en la SEQ ID Nº 8, correspondiente a una identidad del 78,4% aproximadamente entre la SEQ ID Nº.31 y la correspondiente secuencia en la SEQ ID Nº 8

Por el contrario, los resultados obtenidos cuando se utiliza un templado de ADN de las cepas EG5858, EG4332 y EG5552 junto con el juego de cebadores compuesto por las SEQ ID Nº 11 y SEQ ID Nº 12 (como se muestra en la Tabla 5, precedentemente), el ADN de estas cepas produce cada uno un amplicón, cuya secuencia parece corresponder a la secuencia de codificación TIC407, cuando se

utiliza el juego de cebadores correspondiente a los cebadores prJWP200 y prJWP204. Puede haber una secuencia de codificación correspondiente a un homólogo de TIC407 en un fragmento *Hind*III grande en estas cepas y por ello posiblemente no estuvieran disponibles para su identificación al conducir el examen inicial que se describió previamente, con una transferencia Southern que empleaba ADN de *tic*901 como sonda, cuyos resultados se muestran en la Tabla 1

Ejemplo 11

Este ejemplo ilustra la identificación de patrones RFLP únicos relacionados con TIC901 y secuencias de codificación de proteínas insecticidas de cepas de *Bacillus* que producen en sobrenadantes de cultivo proteínas extracelulares inmunológicamente relacionadas con TIC901 según se determinó usando método de detección con ELISA pero cuyo ADN no da como resultado la producción de productos de amplificación térmica detectables usando los métodos de detección por amplificación térmica estándar para *tic*901.

Los análisis iniciales que se expusieron en los ejemplos anteriores se basaron en la detección de actividad insecticida en sobrenadantes de cultivo luego de crecimiento durante la noche a 30 °C en medio PYG seguido de detección de las secuencias relacionadas con *tic*901 en cepas de *B. thuringiensis* cuando se efectuaron prueba con sondas utilizando el método de Southern empleando una secuencia específica de TIC901. Se analizó el ADN total de varios cientos de cepas de *B. thuringiensis*. Se generaron perfiles de RFLP y se compararon. La amplificación térmica de estos ADN usando cebadores específicos para *TIC901, TIC1201, TIC407 y TIC417* determinó que un veinte a veinticinco por ciento aproximadamente de las cepas de Bt contienen dichas secuencias relacionadas. Se considera que posiblemente no se pudieron detectar algunas de las secuencias relacionadas con TIC901 usando los métodos de detección con RFLP y de amplificación térmica. Sobre la base de los perfiles de proteínas extracelulares de las cepas de Bt se continuó con la fermentación hasta bien avanzada la fase log tardía y

pasando por las fases estacionaria y de esporulación, para lo cual sería de utilidad contar con un enfoque más directo usando métodos inmunológicos para detectar las cepas de Bt que producían proteínas relacionadas con TIC901

Se dispuso una colección de cultivos de Bt en un formato de 96 cavidades y se guardó como soluciones madre en glicerol a –80 °C y luego se emplearon como inóculo para fermentaciones a microescala transfiriendo muestras a bloques de cavidades profundas (DWB, QUIAGEN INC) Cada muestra se fermentó en 1 mililitro de caldo Terrific (TB) por cavidad. Los DWB se sellaron con cinta AIRPORE (QUIAGEN) y se incubaron a 30 °C por un lapso entre 20 y 45 horas con agitación en un incubador HiGro (GENOMIC SOLUTIONS/GENOMIC INSTRUMENTATION SERVICES) a 300 RPM Al final del tiempo de fermentación, se agregaron veinticinco unidades por mililitro de benzonasa (SIGMA ALDRICH) al caldo de cultivo para reducir la viscosidad y se incubó todo a 30 °C durante otra hora adicional. Se eliminaron los sólidos, incluyendo células y esporas, por centrifugación a 1900 x g en un rotor con canasta de oscilación libre a 4 °C durante treinta minutos y los sobrenadantes clarificados fueron transferidos en alícuotas de 200 microlitros a placas de 96 cavidades de poca profundidad a –80 °C y luego se procesaron mediante análisis por ELISA.

Se desarrolló un método por ELISA para la identificación de los caldos sobrenadantes que contienen las proteínas relacionadas con TIC901 Se recubrieron placas de 96 cavidades NUNC Immuno MaxSorb durante la noche a 4 °C con 100 microlitros por cavidad de un anticuerpo policlonal IgG anti-TIC901 de conejo purificado por afinidad inversa diluido 1 1000 en solución amortiguadora de recubrimiento que contiene carbonato de sodio 15 milimolar, bicarbonato de sodio 35 milimolar, NaCl 150 milimolar, pH 9,6 Las cavidades se lavaron tres veces con PBST (fosfato K/Na 50 mM, NaCl 150 mM, Tween 20 0,05%, pH 7,4) y se bloqueó el exceso de sitios de unión con 250 microlitros de una solución de BSA 1% en PBST durante una hora a temperatura ambiente Se descartó la solución amortiguadora de

bloqueo y cada cavidad se cargó con 200 microlitros de un sobrenadante de cultivo Bt diluido 1:75 (volumen en volumen) en PBST que contenía BSA 0,2%, seguido de incubación a 4 °C durante la noche. Cada cavidad se lavó tres veces con PBST seguido de adición de 200 microlitros por cavidad de una IgG anti-TIC901 policlonal biotinilado diluido 1:2000 en PBST que contiene BSA 0,2% Se incubó la solución de anticuerpo biotinilado durante dos horas a temperatura ambiente y se lavó tres veces como se describió previamente Los complejos proteicos relacionados con el anticuerpo-TIC901 biotinilado fueron detectados incubando 200 microlitros por cavidad de una estreptavidina conjugada con HRP diluida 1.3000 durante dos horas a temperatura ambiente, seguido de tres lavados en PBST y la adición de 100 microlitros por cavidad de un sustrato de peroxidasa TMP. Las reacciones se terminaron con la adición de 100 microlitros por cavidad de una solución de fosfato 3 M La absorbancia de las cavidades individuales se midió a una longitud de onda de 450 nanometros y se determinó que los sobrenadantes de la cepa Bt contenían proteína relacionada con TIC901 utilizando la absorbancia promedio de controles negativos más tres desviaciones estándar

Se identificaron numerosas cepas usando este método de ELISA que presentaba un resultado positivo por ELISA cuando se comparaban con los sobrenadantes obtenidos con los cultivos control negativos, pero que no daban como resultado la producción de un producto de amplificación térmica cuando se usaban pares de cebadores de amplificación específicos de TIC901, TIC1201, TIC407 o TIC417 Se obtuvo ADN genómico de una de dichas cepas positivas por ELISA/negativas por PCR denominada previamente cepa EG4653 de *Bacillus thuringiensis* y se construyó una biblioteca genómica de expresión de plásmidos en la cepa acristalífera EG10650. La biblioteca de plásmidos EG10650 derivada del ADN de la cepa EG4653 se transfirió a membranas y se efectuaron pruebas con sondas con el anticuerpo policlonal IgG anti-TIC901 de conejo descrito precedentemente. Los clones positivos fueron purificados y amplificados y se

secuenció el inserto de ADN presente en uno de dichos plásmidos. Se dedujo una proteína homóloga de TIC901 completa a partir de un solo marco de lectura abierto La proteína homóloga se denominó TIC431 y tiene la secuencia de aminoácidos que se muestra en la SEQ ID Nº:33. La SEQ ID Nº 32 consiste del marco de lectura abierto a partir del cual se dedujo la secuencia de aminoácidos TIC431 La alineación de la secuencia de aminoácidos TIC431 con las secuencias de aminoácidos de TIC901 y de las proteínas homólogas de TIC1201, 407 y 417 se muestra en la FIGURA 1. Una comparación de las proteínas maduras deducidas TIC431m a TIC901m, TIC1201m, TIC407m y TIC417m muestra que 431 presenta un 75% de identidad aproximadamente con 407, un 79% de identidad aproximadamente con 901, un 80% de identidad con 1201 aproximadamente y un 95% de identidad aproximadamente con 417 Por ello se prevé que podría haber otras proteínas homólogas de TIC431 con un 75 a 95% de identidad aproximadamente con las secuencias de aminoácidos TIC431. La proteína precursora deducida consiste de una secuencia predicha amino terminal de 30 aminoácidos aproximadamente coherente con un péptido señal que exhibe una secuencia de clivaje por una peptidasa señal de tipo II consenso. Además, existe la misma similitud, o sustancialmente similar, de secuencia de aminoácidos entre la secuencia del péptido señal y la secuencia de aminoácidos amino terminal deducida del péptido maduro que comienza en la posición de aminoácido 44 que se muestra en la SEQ ID Nº:33 Se espera que los cebadores para amplificación térmica que se muestran en las SEQ ID Nº:23-26 y SEQ ID Nº 27-29, coherentes con las secuencias de aminoácidos consenso codificadas por las secuencias de nucleótidos que se muestran en la SEQ ID Nº 32 entre la posición 223 aproximadamente y 249 aproximadamente o entre 439 aproximadamente y 456 aproximadamente o entre 823 aproximadamente y 846 aproximadamente produzcan amplicones que puedan ser diagnósticos de la presencia de la SEQ ID Nº.32, así como de otras secuencias de codificación de proteínas insecticidas homólogas de *tic*901.

El método ELISA acoplado con el método de amplificación térmica debería proveerle al especialista el medio necesario para identificar cualquier secuencia de nucleótidos codificante de una proteína insecticida derivada de una especie de *Bacillus* que presenta entre un 67% aproximadamente y un 99% aproximadamente o más de identidad de secuencia de aminoácidos con una de las secuencias de aminoácidos TIC901 o un homólogo de TIC901 descritas en la presente.

El especialista en el arte comprenderá que hay otras realizaciones posibles usando las sondas y los cebadores de la presente invención para identificar secuencias que codifican proteínas de la presente invención y para identificar genes de las proteínas de la presente invención y de proteínas relacionadas con las mismas

En resumen, en la memoria descriptiva anterior se describen las realizaciones preferidas de la presente invención. Los especialistas en el arte comprenderán que, sin apartarse del alcance y espíritu de la presente invención y sin experimentación innecesaria, es posible llevar a cabo la presente invención dentro de un amplio rango de parámetros equivalentes. En tanto la presente invención se describió con relación a las realizaciones específicas de la misma, se comprenderá que es factible efectuar modificaciones adicionales. La presente invención pretende abarcar cualquiera de los usos, variaciones o adaptaciones de la invención aplicando en general los principios de la misma. Es posible efectuar diversas permutaciones y combinaciones de los elementos provistos en todas las reivindicaciones que se detallan más adelante y que se encuentran dentro del alcance de esta invención Todas las publicaciones y patentes mencionadas en esta memoria descriptiva se incorporan en la presente a modo de referencia como si cada publicación o patente individual se incorporara especial e individualmente en la presente a modo de referencia.

REFERENCIAS

Barry et al., 1992, *Inhibitors of amino acid biosynthesis· Strategies for imparting glyphosate tolerance to crop plants.* En: *Biosynthesis and Molecular*

Regulation of Amino Acids in Plants. págs 139-145. Singh, Flores, and Shannon Eds., American Society of *Plant Physiologists*, Rockville, Md

Baum, J. A ; Coyle, D. M., Gilbert, M. P., Jany, C. S.; Gawron-Burke, C , 1990 Novel cloning vectors for *Bacillus thuringiensis*, 1983, *Applied and Environmental Microbiology* 56 (11):3420-3428. 3420-3428.

Bevan , *Nature* 304.184-187.

Callis *et al.*, 1987, *Genes & Develop.* 1:1183-1200.

Crickmore et al , 1998, *Microbiol Molecular Biol Rev* 62, págs 807-813

DeBarjac and Frachon, 1990, *Entomophaga* 35, págs. 233-240.

Diehn *et al.*, 1996, En· *Genetic Engineering*, Ed J.K Setlow, Plenum Press, Nueva York, NY, 18:83-99.

Diehn *et al.*, 1998, *Plant Physiology*, 117 1433-1443

Donovan et al , 1988, *Mol. Gen. Genet.* 214.365-372.

Ellis *et al* , 1987, *EMBO Journal* 6 11-16

Estruch *et al.*, 1996, *Proc. Natl. Acad. Sci* EE UU., 93.5389-5394.

Fraley et al., 1983, *Proc Natl Acad. Sci.* EE UU , 80 4803.

Goding, *Monoclonal Antibodies: Principles and Practice*, 2ª eds, Academic Press, NY, 1986

Harlow and Lane, 1988, *Antibodies· A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Hartley, J , 1988, *Mol. Biol.* 202:913. 913.

Herrera-Estrella, 1983, *Nature* 303 209-213

Hess, 1987, *Intern Rev. Cytol.*, 107:367. 367.

Höfte et al., 1989, *Microbiol. Rev* 53 242-255,

Horsch *et al.*, 1985, *Science*, 227:1229-1231.

Humason, Gretchen L , 1967, *Animal Tissue Techniques*, W. H. Freeman and Company.

Joshi, 1987, *Nucl. Acids Res* 15·9627-9640

Klee, 1985, *Bio/Technol.* 3 637-642

Kostıchka et al , 1996, *J. Bactenol.* 178:2141-2144.

Kyte and Doolittle, 1982, J *Mol. Biol* , 157:105-132.

Laemmlı , 1973, *Mol Biol* 80·575-599.

Lu et al., 2000, *J. Plant Phys.*, 156·277-283.

Ma *et al* , 1988, *Nature* 334:631-633

Maliga, 1993, *Trends ın Biotechnology* 11.101-106.

McElroy *et al.*, 1990, *Plant Cell* 2:163-171.

Murray et al , 1989, *Nucl. Acids Res.*, 17.477-498

Neuhaus *et al.,* 1987, *Theor. Appl. Genet.*, 75:30

Nielsen et al., 1997, *Protein Engineering* 10:1-6.

Norton et al , 1985, *Plasmid* 13:211-214

Odell *et al.*, 1985, *Nature* 313.810-812.

Pena *et al.*, 1987, *Nature*, 325:274.

Perlak et al., 1991, *Proc. Natl. Acad. Sci.* EE.UU , 88·3324-3328.

Sambrook et al , 1989, Molecular Cloning - A Laboratory Manual, 2ª Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, Nueva York.

Schnepf et al., 1998, *Microbiol. Molec Biol Rev* 62·775-806.

Turner et al., 1993, *Endophytes: an alternative genome for crop improvement*; International Crop Science Congress, Ames, Iowa, EE.UU., 14-22 julio, 1992, págs.

Vasıl *et al.*, 1989, *Plant Phys.* 91.1575-1579.

Yanısch-Perron et al, 1985, *Gene* 33 103-119.

Zhou *et al* , 1983, *Mol Cell Biol.*, 10·4529-4537

Patente de EE UU Nº:6 489.542

Patente de EE.UU. Nº:6.448.476

Patente de EE UU Nº·6 279 369

Patente de EE.UU. Nº.6.242.669

Patente de EE UU Nº·6 204 435

Patente de EE.UU. Nº.6 107.549

Patente de EE.UU. Nº.6.063.597

Patente de EE.UU. Nº:5 888.801

Patente de EE.UU Nº.5 872.212

Patente de EE.UU. Nº.5.877.012

Patente de EE UU Nº·5 866 326

Patente de EE.UU. Nº·5.849 870

Patente de EE.UU. Nº.5.837.848

Patente de EE UU Nº 5 759 538

Patente de EE.UU Nº:5.689 052

Patente de EE.UU. Nº.5.569.834

Patente de EE.UU. Nº:5.500.365

Patente de EE.UU. Nº.5 424.412

Patente de EE.UU Nº·5.349 124

Patente de EE.UU. Nº.5.229.112

Patente de EE.UU. Nº·5.135 867

Patente de EE.UU. Nº:5.080.897

Patente de EE.UU. Nº.4.935.353

Patente de EE.UU. Nº·4.695.462

Patente de EE.UU. Nº:4.695.455

Patente de EE UU. Nº:4 554.101

WO 96/10083.

WO 94/21795.

WO 95/24492

EPO 0 120 516.

REIVINDICACIONES

1. Un polinucleótido aislado que codifica una toxina insecticida de *Bacillus thuringiensis* o un fragmento insecticida de la misma, donde dicho polinucleótido se hibridiza, bajo condiciones seleccionadas del grupo formado por condiciones de hibridación severas y condiciones de hibridación específicas, con una o más de las secuencias de nucleótidos seleccionadas del grupo formado por las SEQ ID Nº:2, SEQ ID Nº.3 (tic901), SEQ ID Nº:5(tic1201), SEQ ID Nº:7 (tic407), SEQ ID Nº:9 (tic417), SEQ ID Nº:11, SEQ ID Nº.12, SEQ ID Nº:15, SEQ ID Nº 16, SEQ ID Nº:17, SEQ ID Nº:18, SEQ ID Nº:19, SEQ ID Nº:20, SEQ ID Nº:21, SEQ ID Nº 22, SEQ ID Nº:23, SEQ ID Nº:24, SEQ ID Nº:25, SEQ ID Nº 26, SEQ ID Nº:27, SEQ ID Nº:28, SEQ ID Nº:29, SEQ ID Nº:30 y SEQ ID Nº:32 o con el complemento de las mismas.

2. El polinucleótido aislado de la cláusula 1, donde dicha toxina es activa contra una plaga de insectos coleópteros.

3. El polinucleótido aislado acorde con la cláusula 2, donde dicha plaga de insectos coleópteros se selecciona del grupo formado por gusanos de la raíz del maíz y escarabajos de papa de Colorado.

4. El polinucleótido acorde con la cláusula 3, donde dichos gusanos de la raíz del maíz se seleccionan del grupo formado por gusanos de la raíz del maíz del oeste, gusanos de la raíz del maíz del sur o gusanos de la raíz del maíz del norte.

5. El polinucleótido acorde con la cláusula 1, donde dicha secuencia de nucleótidos comprende las SEQ ID Nº:3, SEQ ID Nº:5, SEQ ID Nº:7, SEQ ID Nº 9 y SEQ ID Nº.32.

6. Un polinucleótido que comprende una secuencia de nucleótidos que codifica una toxina, de 34 a 39 kDa aproximadamente, activa contra plagas de coleópteros, donde dicha secuencia de nucleótidos ha sido optimizada para su expresión en plantas y donde dicha toxina comprende una secuencia de aminoácidos seleccionada del grupo formado por las SEQ ID Nº:2, SEQ ID Nº:4, SEQ ID Nº 6, SEQ ID Nº.8, SEQ ID Nº 10, SEQ ID Nº:31 y SEQ ID Nº 33.

7 Una célula huésped transformada para contener un polinucleótido que codifica una proteína insecticida o un fragmento insecticida de la misma, donde dicho polinucleótido comprende la secuencia de nucleótidos que se selecciona del grupo de secuencias formado por las SEQ ID Nº.3, SEQ ID Nº·5, SEQ ID Nº·7, SEQ ID Nº·9, SEQ ID Nº.13, SEQ ID Nº 30 y SEQ ID Nº.32

8 La célula huésped de la cláusula 7, donde dicha célula huésped es una célula vegetal.

9. Una bacteria recombinante que comprende un polinucleótido aislado que codifica una proteína insecticida, donde dicha proteína comprende una secuencia de aminoácidos seleccionada del grupo formado por las SEQ ID Nº:2, SEQ ID Nº:4, SEQ ID Nº.6, SEQ ID Nº.8, SEQ ID Nº.10, SEQ ID Nº·31 y y SEQ ID Nº:33

10. Un método para controlar una plaga de insectos coleópteros, que comprende poner dicha plaga en contacto con una cantidad plaguicida de una proteína toxina de *Bacillus thuringiensis*, de 34 a 39 kDa aproximadamente, o un fragmento insecticida de la misma, donde dicha proteína toxina se selecciona del grupo formado por las SEQ ID Nº 2, SEQ ID Nº.4, SEQ ID Nº.6, SEQ ID Nº.8, SEQ ID Nº:10, SEQ ID Nº:31 y SEQ ID Nº 33

11 Un método para controlar una plaga por insectos coleópteros, que comprende poner dicha plaga en contacto con una cantidad eficaz como pesticida de una proteína de toxina de *Bacillus thuringiensis* de 34 aproximadamente a 39 kDa aproximadamente, o un fragmento insecticida de la misma, donde dicha proteína de la toxina es codificada por una secuencia de nucleótidos que es o que se hibridiza bajo condiciones severas con una secuencia de nucleótidos que comprende al menos 18 nucleótidos consecutivos seleccionados del grupo formado por las SEQ ID Nº.2, SEQ ID Nº:3, SEQ ID Nº·5, SEQ ID Nº:7, SEQ ID Nº:9, SEQ ID Nº.30 y SEQ ID Nº.32.

12 Una proteína insecticida aislada derivada de *Bacillus thuringiensis* que comprende entre 34 kDa aproximadamente y 39 kDa aproximadamente y que está

codificada por una secuencia de nucleótidos que es, o que se hibridiza bajo condiciones severas con, una secuencia de nucleótidos que comprende al menos 18 nucleótidos consecutivos seleccionados del grupo formado por las SEQ ID Nº·2, SEQ ID Nº 3, SEQ ID Nº 5, SEQ ID Nº·7, SEQ ID Nº·9, SEQ ID Nº:30 y SEQ ID Nº:32.

13. Una proteína insecticida aislada que deriva de *Bacillus thuringiensis*, que comprende entre 34 kDa y 39 kDa aproximadamente y está codificada por una secuencia de nucleótidos que comprende, o que se hibridiza bajo condiciones severas con, al menos 18 nucleótidos consecutivos seleccionadas del grupo formado por las SEQ ID Nº 1, SEQ ID Nº 3, SEQ ID Nº 5, SEQ ID Nº 7, SEQ ID Nº.9, SEQ ID Nº:11 y SEQ ID Nº.32 y donde dicha bacteria *Bacillus thuringiensis* se selecciona del grupo formado por EG2158, EG6489, EG6561, EG3618, EG6555, EG6618, 86833 y EG4653.

14. Un método para detectar una primera secuencia de nucleótidos que codifica una proteína insecticida, que comprende hibridizar dicha primera secuencia de nucleótidos con una segunda secuencia de nucleótidos que se selecciona del grupo formado por las SEQ ID Nº.2, SEQ ID Nº.3 (tic901), SEQ ID Nº:5(tic1201), SEQ ID Nº 7 (tic407), SEQ ID Nº.9 (tic417), SEQ ID Nº.11, SEQ ID Nº:12, SEQ ID Nº:15, SEQ ID Nº:16, SEQ ID Nº·17, SEQ ID Nº·18, SEQ ID Nº·19, SEQ ID Nº·20, SEQ ID Nº:21, SEQ ID Nº·22, SEQ ID Nº.23, SEQ ID Nº.24, SEQ ID Nº 25, SEQ ID Nº 26, SEQ ID Nº:27, SEQ ID Nº·28, SEQ ID Nº:29, y SEQ ID Nº:30 y SEQ ID Nº:32, o con el complemento de las mismas bajo condiciones de hibridación severas.

15. Una secuencia de nucleótidos que codifica una proteína insecticida o un fragmento insecticida o un homólogo de la misma, donde dicha proteína insecticida o fragmento insecticida u homólogo de la misma se selecciona del grupo de secuencias formado por las SEQ ID Nº 1, SEQ ID Nº:4, SEQ ID Nº:6, SEQ ID Nº:8, SEQ ID Nº:10, SEQ ID Nº.31 y SEQ ID Nº:33

16. Un método para detectar una primera secuencia de nucleótidos que codifica una proteína insecticida, que comprende los pasos de

(a) producir un amplicón a partir de un templado que comprende dicha primera secuencia de nucleótidos en una muestra usando cebadores de amplificación que comprenden las secuencias que se muestran en las SEQ ID Nº 13 y SEQ ID Nº·14;

(b) aislar y purificar dicho amplicón;

(c) aislar y purificar dicha primera secuencia de nucleótidos completa que codifica dicha proteína insecticida basado en la secuencia aislada y purificada del amplicón, y

(d) expresar dicha proteína insecticida en una célula huésped a partir de dicha primera secuencia de nucleótidos,

donde dicha primera secuencia de nucleótidos se hibridiza con una segunda secuencia de nucleótidos seleccionada del grupo formado por las SEQ ID Nº 1, SEQ ID Nº 3, SEQ ID Nº 5, SEQ ID Nº·7, SEQ ID Nº·9, SEQ ID Nº:11, SEQ ID Nº:30 y SEQ ID Nº·32 o complementos de las mismas, bajo condiciones de hibridación severas.

17. El método acorde con la cláusula 16, donde dicha célula huésped se selecciona del grupo de células que comprende una célula vegetal, una célula bacteriana, una célula fúngica, una célula de insecto y una célula de mamífero.

18. El método acorde con la cláusula 17, donde dicha célula huésped es una célula vegetal seleccionada del grupo de células vegetales que comprende una célula vegetal monocotiledónea y una célula vegetal dicotiledónea.

19. El método acorde con la cláusula 18, donde dicha célula vegetal monocotiledónea se selecciona del grupo de células vegetales que comprende una célula vegetal de maíz, una célula vegetal de trigo, una célula vegetal de arroz, una célula vegetal de avena, una célula vegetal de cebolla y una célula vegetal de pasto.

20. El método acorde con la cláusula 18, donde dicha célula vegetal dicotiledónea se selecciona del grupo de células vegetales que comprende una célula vegetal de algodón, una célula vegetal de canola, una célula vegetal de soja, una célula vegetal de árboles frutales, una célula vegetal de quimbombó, una célula vegetal de

pimiento, una célula vegetal de una planta ornamental, una célula vegetal de girasol, una célula vegetal de pepino y una célula vegetal de melón

21. Una molécula aislada de ácido nucleico, que comprende una secuencia de polinucleótidos que codifica una proteína toxina, donde dicha proteína toxina comprende una secuencia que presenta al menos un 70% de identidad de secuencia aproximadamente con una secuencia de nucleótidos seleccionada del grupo de secuencias de aminoácidos formado por las SEQ ID Nº:2, SEQ ID Nº:4, SEQ ID Nº.6, SEQ ID Nº:8, SEQ ID Nº:10, SEQ ID Nº:31 y SEQ ID Nº:33, o una variante activa en coleópteros o una porción de la misma.

22. Una molécula aislada de ácido nucleico que comprende una secuencia de polinucleótidos que codifica una proteína tóxica para plagas de insectos coleópteros, donde dicha secuencia de polinucleótidos se hibridiza con una secuencia de ácido nucleico seleccionada del grupo formado por las SEQ ID Nº:2, SEQ ID Nº.3 (tic901), SEQ ID Nº·5(tic1201), SEQ ID Nº:7 (tic407), SEQ ID Nº:9 (tic417), SEQ ID Nº.11, SEQ ID Nº·12, SEQ ID Nº:15, SEQ ID Nº:16, SEQ ID Nº·17, SEQ ID Nº 18, SEQ ID Nº.19, SEQ ID Nº.20, SEQ ID Nº.21, SEQ ID Nº:22, SEQ ID Nº.23, SEQ ID Nº:24, SEQ ID Nº:25, SEQ ID Nº:26, SEQ ID Nº:27, SEQ ID Nº·28, SEQ ID Nº:29, SEQ ID Nº·30 y SEQ ID Nº:32, o con el complemento de las mismas, bajo condiciones de hibridación severas

23. El ácido nucleico aislado de la cláusula 1 ó 5, donde dicha secuencia de polinucleótidos se aísla de una bacteria.

24. El ácido nucleico aislado de la cláusula 6, donde dicha secuencia de polinucleótidos se aísla de *Bacillus thuringiensis.*

25 Una proteína toxina sustancialmente purificada, donde dicha proteína toxina comprende una secuencia que tiene al menos un 70% de identidad de secuencia aproximadamente con una secuencia de aminoácidos seleccionada del grupo formado por las SEQ ID Nº·2, SEQ ID Nº.4, SEQ ID Nº:6, SEQ ID Nº·8, SEQ ID Nº·10, SEQ ID Nº.31 y SEQ ID Nº:33

26. La proteína toxina sustancialmente purificada de la cláusula 25, donde dicha proteína toxina comprende la SEQ ID Nº· 4.

27. Una construcción de ADN recombinante que comprende una secuencia de polinucleótidos que codifica una proteína insecticida, donde dicha secuencia de polinucleótidos es al menos un 70% idéntica aproximadamente a la SEQ ID Nº:13, donde dicha proteína insecticida se selecciona del grupo formado por todo o un fragmento insecticida de una proteína de las SEQ ID Nº·4, SEQ ID Nº:6, SEQ ID Nº·6, SEQ ID Nº:10, SEQ ID Nº:31 y SEQ ID Nº:33.

28. La construcción de ADN recombinante de la cláusula 27, donde dicha secuencia de polinucleótidos se muestra en la SEQ ID Nº·3.

29. La construcción de ADN recombinante de la cláusula 28, que además comprende un promotor o una región promotora parcial ligado operativamente a dicha secuencia de polinucleótidos.

30. La construcción de ADN recombinante de la cláusula 29, donde dicho promotor es un promotor constitutivo, un promotor inducible o un promotor específico de tejidos.

31 La construcción de ADN recombinante de la cláusula 30, donde dicho promotor es el promotor 35S del virus en mosaico de la coliflor (CaMV).

32 La construcción de ADN recombinante de la cláusula 31, que además comprende un intrón.

33. La construcción de ADN recombinante de la cláusula 32, donde dicho intrón es el intrón hsp70.

34. La construcción de ADN recombinante de la cláusula 33, que además comprende una región no traducida 3' ligada operativamente.

35. Una célula huésped recombinante transformada con una secuencia de polinucleótidos de la cláusula 1 o de la cláusula 27 que codifican una proteína insecticida, donde dicha célula huésped recombinante expresa dicha proteína que es al menos un 70% similar aproximadamente en cuanto a secuencia de aminoácidos a

una proteína insecticida seleccionada del grupo formado por las SEQ ID Nº.4, SEQ ID Nº:6, SEQ ID Nº:8, SEQ ID Nº.10, SEQ ID Nº 31 y SEQ ID Nº:33

36. La célula huésped recombinante de la cláusula 35, donde dicha célula huésped recombinante es una célula vegetal.

37. Un método para expresar una proteína insecticida en una célula vegetal que comprende los pasos de.

(a) insertar en el genoma de una célula vegetal una secuencia de ácido nucleico que comprende:

(i) un promotor funcional en dicha célula vegetal ligado operativamente a,

(ii) un ADN estructural que codifica una proteína seleccionada del grupo formado por las SEQ ID Nº 4, SEQ ID Nº 6, SEQ ID Nº.8, SEQ ID Nº:10, SEQ ID Nº 31, SEQ ID Nº:33 y proteínas insecticidas que son al menos un 70% similares aproximadamente a dicho grupo de proteínas y

(iii) una secuencia de ADN no traducida 3' cuya función en la célula vegetal causa la terminación de la transcripción y poliadenilación,

(b) obtener dicha célula vegetal transformada para contener la secuencia de ácido nucleico del paso (a), y

(c) regenerar a partir de la célula vegetal transformada una planta genéticamente transformada que expresa una cantidad insecticida de dicha proteína insecticida.

38. Una semilla producida por la planta transformada de la cláusula 37.

39. Una planta cultivada a partir de la semilla de la cláusula 38

40. Una muestra biológica derivada de una planta, tejido o semilla, donde dicha muestra comprende una secuencia de nucleótidos que codifica una proteína insecticida seleccionada del grupo formado por las SEQ ID Nº 4, SEQ ID Nº:6, SEQ ID Nº.8, SEQ ID Nº:10, SEQ ID Nº:31 y SEQ ID Nº.33.

41. La muestra biológica de la cláusula 40, que comprende una planta, tejido o semilla de una planta transgénica que codifica dicha proteína

42. La muestra biológica de la cláusula 41, donde dicha muestra se selecciona del

grupo formado por un extracto que se puede obtener de la planta transgénica y donde dicho extracto comprende una secuencia de nucleótidos que codifica un fragmento insecticida de una proteína seleccionada del grupo formado por las SEQ ID Nº:4, SEQ ID Nº.6, SEQ ID Nº:8, SEQ ID Nº:10, SEQ ID Nº:31 y SEQ ID Nº:33.

43. La muestra biológica de la cláusula 42, donde dicha muestra se selecciona del grupo formado por harina de maíz, alimento de maíz, jarabe de maíz, aceite de maíz, almidón de maíz y cereales elaborados total o parcialmente con subproductos de maíz

44. Un extracto derivado de una planta, tejido o semilla de maíz transgénico, que comprende una secuencia de nucleótidos que codifica un fragmento insecticida de una proteína seleccionada del grupo formado por las SEQ ID Nº 4, SEQ ID Nº 6, SEQ ID Nº.8, SEQ ID Nº:10, SEQ ID Nº.31 y SEQ ID Nº.33.

45 El extracto de la cláusula 44, donde dicha secuencia se puede detectar en dicho extracto usando un método de amplificación de ácidos nucleicos o de hibridación de ácidos nucleicos.

46 El extracto de la cláusula 45, que comprende una planta, tejido o semilla de una planta de maíz transgénico que expresa dicha proteína.

47. El extracto de la cláusula 46, donde dicha muestra se selecciona del grupo formado por harina de maíz, alimento de maíz, jarabe de maíz, aceite de maíz, almidón de maíz y cereales elaborados total o parcialmente con subproductos de maíz.

PROTEÍNAS INSECTICIDAS SECRETADAS DE *BACILLUS THURINGIENSIS* Y USOS DE LAS MISMAS

RESUMEN

Aislamiento y caracterización de secuencias de nucleótidos que codifican nuevas proteínas insecticidas secretadas hacia el espacio extracelular de *Bacillus thuringiensis* y cepas relacionada Las proteínas se aíslan a partir de sobrenadantes de cultivos de *Bacillus thuringiensis* y cepas relacionada y presentan actividad insecticida contra insectos coleópteros incluyendo escarabajo de papa de Colorado (*Lymantria dispar*) y gusanos de la raíz de maíz del sur (*Diabrotica undecimpunctata*) Se describen proteínas insecticidas codificadas por secuencias de nucleótidos que se hibridizan con las secuencias de nucleótidos aisladas y caracterizadas También se describen métodos para obtener y usar las células transgénicas y plantas que contienen la nueva secuencia de nucleótidos de la invención.

SEQUENCE LISTING

<110> Monsanto Technology LLC

Baum, James A.
Donovan, Judith C.
Donovan, William P.
Engleman, James T.
Krasomil-Osterfeld, Karina
Pitkin, John W.
Roberts, James K.

<120> Insecticidal Proteins Secreted From Bacillus Species and Uses Therefor

<130> 38-21(52806)PCT

<150> US 60/485,483

<151> 2003-07-07

<160> 33

<170> PatentIn version 3.1

<210> 1

<211> 15

<212> PRT

<213> Bacillus thuringiensis

<400> 1

Val Ile Gly Pro Tyr Ala Glu Ser Tyr Ile Asp Arg Val Gln Asp
1 5 10 15

<210> 2

<211> 45

<212> DNA

<213> Artificial Sequence

<220>

<223> tic gene probe encoding SEQ ID NO 1

<400> 2
gtaattggac catatgcaga atcatatatt gatacgagta caaga 45

<210> 3

<211> 1253

<212> DNA

<213> Bacillus thuringiensis

<220>

<221> CDS

<222> (153)..(1253)

<223>

<400> 3
aattatgatt ttaatattct tatgttattc ctataatata caataaaagc ataattatcc 60

ttcatattat gtttataaat ttaataaaat acataaaaat agagtgttat aatatttttg 120

aaagcgttat caagagtgat ggagggataa tt atg aaa aat aga ttt tca aaa 173
Met Lys Asn Arg Phe Ser Lys
1 5

gtg gca tta tgc acc gta ccg att tta atg gtt tct aca ttc gcc agt 221
Val Ala Leu Cys Thr Val Pro Ile Leu Met Val Ser Thr Phe Ala Ser
10 15 20

tca agc atg tca gct ttt gct gca gaa gcc aaa tca cca gat tta aat 269
Ser Ser Met Ser Ala Phe Ala Ala Glu Ala Lys Ser Pro Asp Leu Asn
25 30 35

gta tct caa caa gta ata ggt ccc tat gcc gaa tct tat att gat att 317
Val Ser Gln Gln Val Ile Gly Pro Tyr Ala Glu Ser Tyr Ile Asp Ile
40 45 50 55

gtg cag gat aga atg aaa caa agg gat aag gga tca aaa tta act ggt 365
Val Gln Asp Arg Met Lys Gln Arg Asp Lys Gly Ser Lys Leu Thr Gly
60 65 70

aaa cca ata aat atg caa gaa caa ata ata gat ggg tgg ttt cta gct 413
Lys Pro Ile Asn Met Gln Glu Gln Ile Ile Asp Gly Trp Phe Leu Ala
75 80 85

aga ttt tgg ata ttt aag gat caa aac aat aac cat cag aca aat aga 461
Arg Phe Trp Ile Phe Lys Asp Gln Asn Asn Asn His Gln Thr Asn Arg
90 95 100

ttt ata tcc tgg ttt aaa gat aat att gct agt tca aaa ggg tat aat 509
Phe Ile Ser Trp Phe Lys Asp Asn Ile Ala Ser Ser Lys Gly Tyr Asn
105 110 115

agt att gcg gag caa atg ggt tta aaa ata gaa gca gaa aac gat atg 557
Ser Ile Ala Glu Gln Met Gly Leu Lys Ile Glu Ala Glu Asn Asp Met

```
120                 125                 130                 135

gat gta aca aat ata gat tat aca tct aag aca ggc gat acc att tat      605
Asp Val Thr Asn Ile Asp Tyr Thr Ser Lys Thr Gly Asp Thr Ile Tyr
                140                 145                 150

aat ggt att tca gaa ttg aaa aat tat aca gga tca act caa aag atg      653
Asn Gly Ile Ser Glu Leu Lys Asn Tyr Thr Gly Ser Thr Gln Lys Met
            155                 160                 165

aaa aca gat agt ttt caa aga gat tat aca aaa tca gaa tct act tca      701
Lys Thr Asp Ser Phe Gln Arg Asp Tyr Thr Lys Ser Glu Ser Thr Ser
        170                 175                 180

gta act aat gga tta caa tta gga ttt aaa gtt gct gct aaa gga gta      749
Val Thr Asn Gly Leu Gln Leu Gly Phe Lys Val Ala Ala Lys Gly Val
    185                 190                 195

gtt gct ttg gct ggg gca gac ttt gaa acc agt gtt act tat aat cta      797
Val Ala Leu Ala Gly Ala Asp Phe Glu Thr Ser Val Thr Tyr Asn Leu
200                 205                 210                 215

tca act act aca act gaa aca aat aca ata tca gac aag ttt act gtc      845
Ser Thr Thr Thr Thr Glu Thr Asn Thr Ile Ser Asp Lys Phe Thr Val
                220                 225                 230

cca tct caa gaa gtt aca ttg cct cca gga cat aaa gcg ata gtg aaa      893
Pro Ser Gln Glu Val Thr Leu Pro Pro Gly His Lys Ala Ile Val Lys
            235                 240                 245

cat gat tta aga aaa atg gtt tat tct ggt act cat gat cta aag ggt      941
His Asp Leu Arg Lys Met Val Tyr Ser Gly Thr His Asp Leu Lys Gly
        250                 255                 260

gat tta att gtg agt ttt aat gat aaa gag att gta caa aaa ttt att      989
Asp Leu Ile Val Ser Phe Asn Asp Lys Glu Ile Val Gln Lys Phe Ile
    265                 270                 275

tat cca aat tat aga gaa att aat tta tct gat atc cgt gaa act atg     1037
Tyr Pro Asn Tyr Arg Glu Ile Asn Leu Ser Asp Ile Arg Glu Thr Met
280                 285                 290                 295

att gaa att gat gaa tgg aat cat gta aac cct gtg aat ttt tat gaa     1085
Ile Glu Ile Asp Glu Trp Asn His Val Asn Pro Val Asn Phe Tyr Glu
                300                 305                 310

tta gtt ggg gtc aaa aat cat ata aaa aat ggt gaa act ttg tat ata     1133
Leu Val Gly Val Lys Asn His Ile Lys Asn Gly Glu Thr Leu Tyr Ile
            315                 320                 325

gat act cca gct aaa ttt atg ttt aat ggt gct aat cca tat tat aga     1181
Asp Thr Pro Ala Lys Phe Met Phe Asn Gly Ala Asn Pro Tyr Tyr Arg
        330                 335                 340

gca aca ttt aca gaa tac gac ggg aat aat aat cct gtt caa aca aag     1229
Ala Thr Phe Thr Glu Tyr Asp Gly Asn Asn Asn Pro Val Gln Thr Lys
    345                 350                 355

gta tta agt gaa aac ttt aaa ttg                                    1253
Val Leu Ser Glu Asn Phe Lys Leu
360                 365
```

<210> 4

<211> 367

<212> PRT

5187 Sequence Listing

<213> Bacillus thuringiensis

<400> 4

```
Met Lys Asn Arg Phe Ser Lys Val Ala Leu Cys Thr Val Pro Ile Leu
1               5                  10                  15

Met Val Ser Thr Phe Ala Ser Ser Ser Met Ser Ala Phe Ala Ala Glu
            20                  25                  30

Ala Lys Ser Pro Asp Leu Asn Val Ser Gln Gln Val Ile Gly Pro Tyr
        35                  40                  45

Ala Glu Ser Tyr Ile Asp Ile Val Gln Asp Arg Met Lys Gln Arg Asp
    50                  55                  60

Lys Gly Ser Lys Leu Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile
65                  70                  75                  80

Ile Asp Gly Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn
                85                  90                  95

Asn Asn His Gln Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Ile
            100                 105                 110

Ala Ser Ser Lys Gly Tyr Asn Ser Ile Ala Glu Gln Met Gly Leu Lys
        115                 120                 125

Ile Glu Ala Glu Asn Asp Met Asp Val Thr Asn Ile Asp Tyr Thr Ser
    130                 135                 140

Lys Thr Gly Asp Thr Ile Tyr Asn Gly Ile Ser Glu Leu Lys Asn Tyr
145                 150                 155                 160

Thr Gly Ser Thr Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr
                165                 170                 175

Thr Lys Ser Glu Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Phe
            180                 185                 190

Lys Val Ala Ala Lys Gly Val Val Ala Leu Ala Gly Ala Asp Phe Glu
        195                 200                 205

Thr Ser Val Thr Tyr Asn Leu Ser Thr Thr Thr Thr Glu Thr Asn Thr
    210                 215                 220

Ile Ser Asp Lys Phe Thr Val Pro Ser Gln Glu Val Thr Leu Pro Pro
225                 230                 235                 240

Gly His Lys Ala Ile Val Lys His Asp Leu Arg Lys Met Val Tyr Ser
```

245 250 255

Gly Thr His Asp Leu Lys Gly Asp Leu Ile Val Ser Phe Asn Asp Lys
260 265 270

Glu Ile Val Gln Lys Phe Ile Tyr Pro Asn Tyr Arg Glu Ile Asn Leu
275 280 285

Ser Asp Ile Arg Glu Thr Met Ile Glu Ile Asp Glu Trp Asn His Val
290 295 300

Asn Pro Val Asn Phe Tyr Glu Leu Val Gly Val Lys Asn His Ile Lys
305 310 315 320

Asn Gly Glu Thr Leu Tyr Ile Asp Thr Pro Ala Lys Phe Met Phe Asn
325 330 335

Gly Ala Asn Pro Tyr Tyr Arg Ala Thr Phe Thr Glu Tyr Asp Gly Asn
340 345 350

Asn Asn Pro Val Gln Thr Lys Val Leu Ser Glu Asn Phe Lys Leu
355 360 365

<210> 5

<211> 1621

<212> DNA

<213> Bacillus thuringiensis

<220>

<221> CDS

<222> (530)..(1621)

<223>

<400> 5

actagttttt tcataatggc ataagcggga tgatgatctt gtttttttacg atgttcaata 60

tccaatgtgt gcctttctat atcaatcgca cgatataaat aacactattt ccctttgaat 120

tttatatagg tttcatctaa ttttcaagac atgtggttgt tttgcgtttt cttcttccaa 180

atttgataaa tcaagctccc atattcatga atccagcgca taatgattgt gggatgaact 240

gaaacatcac gatagcttaa agcaaaacga caatagtagc ggacggctac cataataata 300

tcttgtttga actgtttccc tttaaaatat cacatttgtg attctcctcg atgctttttt 360

tagagtgtag cttcatctag aacactttgc aatagaacca ttcctttgat atacaattaa 420

accacattta tccttcatgg aatgtttata tattaaagaa tataaaaaaa catacgatgt 480

```
tataattaat ttgaaagcgt taacaaaaat gaatgatgga gggataatt atg aaa tac    538
                                                      Met Lys Tyr
                                                      1

aag ttt tca aaa gtc gtt aag tgt act tta cca gct tta atg att act      586
Lys Phe Ser Lys Val Val Lys Cys Thr Leu Pro Ala Leu Met Ile Thr
    5                  10                  15

aca ttc gtt act cca agt atg gca gtt ttt gcc gca gaa acc aag tcg      634
Thr Phe Val Thr Pro Ser Met Ala Val Phe Ala Ala Glu Thr Lys Ser
20                 25                  30                  35

cca aat cta aat gca tct caa caa gca ata act cca tat gct gaa tct      682
Pro Asn Leu Asn Ala Ser Gln Gln Ala Ile Thr Pro Tyr Ala Glu Ser
                40                  45                  50

tat att gat act gtt caa gat aga atg aaa caa aga gat agg gaa tca      730
Tyr Ile Asp Thr Val Gln Asp Arg Met Lys Gln Arg Asp Arg Glu Ser
            55                  60                  65

aaa cta act ggt aaa cca ata aat atg caa gaa caa ata ata gat gga      778
Lys Leu Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile Ile Asp Gly
        70                  75                  80

tgg ttt tta gct aga ttc tgg ata ttt aaa gat caa aat aac aat cat      826
Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn Asn Asn His
    85                  90                  95

caa aca aat aga ttt ata tcc tgg ttt aaa gat aat ctt gct agt tcg      874
Gln Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Leu Ala Ser Ser
100                 105                 110                 115

aag ggg tat gac agt ata gca gaa caa atg ggc tta aaa ata gaa gca      922
Lys Gly Tyr Asp Ser Ile Ala Glu Gln Met Gly Leu Lys Ile Glu Ala
                120                 125                 130

tta aat gat atg gat gta aca aat att gat tat aca tct aaa aca ggt      970
Leu Asn Asp Met Asp Val Thr Asn Ile Asp Tyr Thr Ser Lys Thr Gly
            135                 140                 145

gat acc ata tat aat gga att tct gaa cta aca aat tat aca gga aca     1018
Asp Thr Ile Tyr Asn Gly Ile Ser Glu Leu Thr Asn Tyr Thr Gly Thr
        150                 155                 160

acc caa aaa atg aaa acc gat agt ttt caa aga gat tat aca aaa tct     1066
Thr Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr Thr Lys Ser
    165                 170                 175

gaa tcc act tca gta aca aat ggg tta caa tta gga ttt aaa gtt gct     1114
Glu Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Phe Lys Val Ala
180                 185                 190                 195

gct aag gga gta gtt gca tta gca ggt gca gat ttt gaa aca agt gtt     1162
Ala Lys Gly Val Val Ala Leu Ala Gly Ala Asp Phe Glu Thr Ser Val
                200                 205                 210

acc tat aat tta tca tct act aca act gaa aca aat aca ata tcg gat     1210
Thr Tyr Asn Leu Ser Ser Thr Thr Thr Glu Thr Asn Thr Ile Ser Asp
            215                 220                 225

aag ttt act gtt cca tct caa gaa gtt aca tta tcc cca gga cat aaa     1258
Lys Phe Thr Val Pro Ser Gln Glu Val Thr Leu Ser Pro Gly His Lys
        230                 235                 240

gca gtg gtg aaa cat gat ttg aga aaa atg gtg tat ttt ggg act cat     1306
Ala Val Val Lys His Asp Leu Arg Lys Met Val Tyr Phe Gly Thr His
    245                 250                 255
```

```
gat tta aag ggt gat tta aaa gta ggt ttt aat gat aaa gag att gta     1354
Asp Leu Lys Gly Asp Leu Lys Val Gly Phe Asn Asp Lys Glu Ile Val
260                 265                 270                 275

caa aaa ttt att tat cca aat tat aga tca att gat tta tct gat att     1402
Gln Lys Phe Ile Tyr Pro Asn Tyr Arg Ser Ile Asp Leu Ser Asp Ile
                280                 285                 290

cgt aaa aca atg att gaa att gat aaa tgg aat cat gta aat acc att     1450
Arg Lys Thr Met Ile Glu Ile Asp Lys Trp Asn His Val Asn Thr Ile
            295                 300                 305

gac ttt tat caa tta gtt gga gtt aaa aat cat ata aaa aat ggt gat     1498
Asp Phe Tyr Gln Leu Val Gly Val Lys Asn His Ile Lys Asn Gly Asp
        310                 315                 320

act tta tat ata gat acc ccg gcc gaa ttt aca ttt aat gga gct aat     1546
Thr Leu Tyr Ile Asp Thr Pro Ala Glu Phe Thr Phe Asn Gly Ala Asn
    325                 330                 335

cca tat tat aga gca aca ttt aca gaa tac gac gag aac gga aat cct     1594
Pro Tyr Tyr Arg Ala Thr Phe Thr Glu Tyr Asp Glu Asn Gly Asn Pro
340                 345                 350                 355

gtt caa aca aag att tta agt gga aat                                 1621
Val Gln Thr Lys Ile Leu Ser Gly Asn
                360
```

<210> 6

<211> 364

<212> PRT

<213> Bacillus thuringiensis

<400> 6

```
Met Lys Tyr Lys Phe Ser Lys Val Val Lys Cys Thr Leu Pro Ala Leu
1               5                   10                  15

Met Ile Thr Thr Phe Val Thr Pro Ser Met Ala Val Phe Ala Ala Glu
            20                  25                  30

Thr Lys Ser Pro Asn Leu Asn Ala Ser Gln Gln Ala Ile Thr Pro Tyr
        35                  40                  45

Ala Glu Ser Tyr Ile Asp Thr Val Gln Asp Arg Met Lys Gln Arg Asp
    50                  55                  60

Arg Glu Ser Lys Leu Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile
65                  70                  75                  80

Ile Asp Gly Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn
                85                  90                  95

Asn Asn His Gln Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Leu
            100                 105                 110
```

Ala Ser Ser Lys Gly Tyr Asp Ser Ile Ala Glu Gln Met Gly Leu Lys
115 120 125

Ile Glu Ala Leu Asn Asp Met Asp Val Thr Asn Ile Asp Tyr Thr Ser
130 135 140

Lys Thr Gly Asp Thr Ile Tyr Asn Gly Ile Ser Glu Leu Thr Asn Tyr
145 150 155 160

Thr Gly Thr Thr Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr
165 170 175

Thr Lys Ser Glu Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Phe
180 185 190

Lys Val Ala Ala Lys Gly Val Val Ala Leu Ala Gly Ala Asp Phe Glu
195 200 205

Thr Ser Val Thr Tyr Asn Leu Ser Ser Thr Thr Thr Glu Thr Asn Thr
210 215 220

Ile Ser Asp Lys Phe Thr Val Pro Ser Gln Glu Val Thr Leu Ser Pro
225 230 235 240

Gly His Lys Ala Val Val Lys His Asp Leu Arg Lys Met Val Tyr Phe
245 250 255

Gly Thr His Asp Leu Lys Gly Asp Leu Lys Val Gly Phe Asn Asp Lys
260 265 270

Glu Ile Val Gln Lys Phe Ile Tyr Pro Asn Tyr Arg Ser Ile Asp Leu
275 280 285

Ser Asp Ile Arg Lys Thr Met Ile Glu Ile Asp Lys Trp Asn His Val
290 295 300

Asn Thr Ile Asp Phe Tyr Gln Leu Val Gly Val Lys Asn His Ile Lys
305 310 315 320

Asn Gly Asp Thr Leu Tyr Ile Asp Thr Pro Ala Glu Phe Thr Phe Asn
325 330 335

Gly Ala Asn Pro Tyr Tyr Arg Ala Thr Phe Thr Glu Tyr Asp Glu Asn
340 345 350

Gly Asn Pro Val Gln Thr Lys Ile Leu Ser Gly Asn
355 360

<210> 7

<211> 1552

5187 Sequence Listing

<212> DNA

<213> Bacillus thuringiensis

<220>

<221> CDS

<222> (169)..(1272)

<223>

<400> 7

```
ctagctgaat atgcagtaga taatgatttt gactatacta atgagatttt ttatgtaata     60

aaccatattt atccttaaat aaatgttcac atatttaaaa agaataaact accatgcggt    120

gctagaatat aaatgagagc gctaacaaaa aataatggag ggataatc atg aaa tac     177
                                                      Met Lys Tyr
                                                      1

aaa tca tca aaa gta gca ata tgt act tta tca gct tta atg ctt tca      225
Lys Ser Ser Lys Val Ala Ile Cys Thr Leu Ser Ala Leu Met Leu Ser
    5                  10                  15

aca att ggt act tcg agt atg tcc act ttt gct gca gaa aca aca tta      273
Thr Ile Gly Thr Ser Ser Met Ser Thr Phe Ala Ala Glu Thr Thr Leu
20                 25                  30                  35

cca ggt caa act ctt aag gaa caa tca ata acc cca cgt gca gaa tct      321
Pro Gly Gln Thr Leu Lys Glu Gln Ser Ile Thr Pro Arg Ala Glu Ser
                40                  45                  50

tat att gat att gta caa gat aga atg aaa caa agg gat ata gaa tcg      369
Tyr Ile Asp Ile Val Gln Asp Arg Met Lys Gln Arg Asp Ile Glu Ser
            55                  60                  65

aaa cgt act ggt aaa ccg att aat atg caa gaa caa ata ata gat gga      417
Lys Arg Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile Ile Asp Gly
        70                  75                  80

tgg ttt tta gca aga ttc tgg ata ttt aaa gat caa aat aat aac cat      465
Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn Asn Asn His
    85                  90                  95

caa aca aat aga ttc ata aca tgg ttt aaa aat aat gtt gcc agc tca      513
Gln Thr Asn Arg Phe Ile Thr Trp Phe Lys Asn Asn Val Ala Ser Ser
100                 105                 110                 115

aaa ggt tat gag ggt att gca gaa caa atg ggt ttg aaa ata gaa tcg      561
Lys Gly Tyr Glu Gly Ile Ala Glu Gln Met Gly Leu Lys Ile Glu Ser
                120                 125                 130

atg agt gat atg aat gta tcg aat ata aat tat aca ggt aaa aag ggt      609
Met Ser Asp Met Asn Val Ser Asn Ile Asn Tyr Thr Gly Lys Lys Gly
            135                 140                 145

gat act ata tat aat ggc gtt tcg gaa tta gaa aat aaa atg gga aca      657
Asp Thr Ile Tyr Asn Gly Val Ser Glu Leu Glu Asn Lys Met Gly Thr
        150                 155                 160

cct caa aaa atg aaa tca gat agt ttt caa aga gat tat acc aaa tct      705
Pro Gln Lys Met Lys Ser Asp Ser Phe Gln Arg Asp Tyr Thr Lys Ser
```

165 170 175

caa tca acc tca gta aca aat ggg tta caa tta gga gtt aaa gtt tct 753
Gln Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Val Lys Val Ser
180 185 190 195

gcc aaa ggt acg gtt gtc tta gga gag gca agc ctt gaa aca agc gtt 801
Ala Lys Gly Thr Val Val Leu Gly Glu Ala Ser Leu Glu Thr Ser Val
200 205 210

act tat aat tta tcg tct act gca act gaa aca gat aca aca tcg gac 849
Thr Tyr Asn Leu Ser Ser Thr Ala Thr Glu Thr Asp Thr Thr Ser Asp
215 220 225

aag ttt act gtc cca tcc caa gaa gtt aca tta cca cca gga cat aaa 897
Lys Phe Thr Val Pro Ser Gln Glu Val Thr Leu Pro Pro Gly His Lys
230 235 240

gca gta att aag cat gat tta aga aaa atg gtg tat tct ggt acg cat 945
Ala Val Ile Lys His Asp Leu Arg Lys Met Val Tyr Ser Gly Thr His
245 250 255

gac tta aag ggg gat tta aaa gta gct ttt aac gat aaa gca att gta 993
Asp Leu Lys Gly Asp Leu Lys Val Ala Phe Asn Asp Lys Ala Ile Val
260 265 270 275

caa aaa ttt att tat cca aat tat aga tct ata aat tta tct gat att 1041
Gln Lys Phe Ile Tyr Pro Asn Tyr Arg Ser Ile Asn Leu Ser Asp Ile
280 285 290

cgt aaa aca atg aaa gaa att gat gaa tgg aat cat gta aaa ccc att 1089
Arg Lys Thr Met Lys Glu Ile Asp Glu Trp Asn His Val Lys Pro Ile
295 300 305

gat ttt tat caa ctg gtt gga ata aaa aat cat ata aaa aat ggg gat 1137
Asp Phe Tyr Gln Leu Val Gly Ile Lys Asn His Ile Lys Asn Gly Asp
310 315 320

acc tta tat ata gag act cca gct aaa ttt att ttt aat gga gct aat 1185
Thr Leu Tyr Ile Glu Thr Pro Ala Lys Phe Ile Phe Asn Gly Ala Asn
325 330 335

gta tat tat aga gca act ttt aca gaa tat gat aag gat gga aaa cct 1233
Val Tyr Tyr Arg Ala Thr Phe Thr Glu Tyr Asp Lys Asp Gly Lys Pro
340 345 350 355

gtt caa ttc aac aaa ttt tta agt gaa aat tac aag tta tagaggaagt 1282
Val Gln Phe Asn Lys Phe Leu Ser Glu Asn Tyr Lys Leu
360 365

aaagatgccg tagtgagatc gtttcacagc tactgagtat tcaaataata cacgggaaaa 1342

ttcaccttcc tggaaggacg gatttacttt ttttacggag gaacttgttt tatacatcaa 1402

aatgtttttt tatgaggttt gtgtattctt atttgagcct ggaacggaac cattttgagt 1462

aagcttaatt tgacttggaa atgtattttt attaccttat tacgtgaaca atggcctata 1522

aacgtgccac acaggaatgg gaggacgagt 1552

<210> 8

<211> 368

<212> PRT

<213> Bacillus thuringiensis

<400> 8

```
Met Lys Tyr Lys Ser Ser Lys Val Ala Ile Cys Thr Leu Ser Ala Leu
1               5                  10                  15

Met Leu Ser Thr Ile Gly Thr Ser Ser Met Ser Thr Phe Ala Ala Glu
            20                  25                  30

Thr Thr Leu Pro Gly Gln Thr Leu Lys Glu Gln Ser Ile Thr Pro Arg
        35                  40                  45

Ala Glu Ser Tyr Ile Asp Ile Val Gln Asp Arg Met Lys Gln Arg Asp
    50                  55                  60

Ile Glu Ser Lys Arg Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile
65                  70                  75                  80

Ile Asp Gly Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn
                85                  90                  95

Asn Asn His Gln Thr Asn Arg Phe Ile Thr Trp Phe Lys Asn Asn Val
            100                 105                 110

Ala Ser Ser Lys Gly Tyr Glu Gly Ile Ala Glu Gln Met Gly Leu Lys
        115                 120                 125

Ile Glu Ser Met Ser Asp Met Asn Val Ser Asn Ile Asn Tyr Thr Gly
    130                 135                 140

Lys Lys Gly Asp Thr Ile Tyr Asn Gly Val Ser Glu Leu Glu Asn Lys
145                 150                 155                 160

Met Gly Thr Pro Gln Lys Met Lys Ser Asp Ser Phe Gln Arg Asp Tyr
                165                 170                 175

Thr Lys Ser Gln Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Val
            180                 185                 190

Lys Val Ser Ala Lys Gly Thr Val Val Leu Gly Glu Ala Ser Leu Glu
        195                 200                 205

Thr Ser Val Thr Tyr Asn Leu Ser Ser Thr Ala Thr Glu Thr Asp Thr
    210                 215                 220

Thr Ser Asp Lys Phe Thr Val Pro Ser Gln Glu Val Thr Leu Pro Pro
225                 230                 235                 240

Gly His Lys Ala Val Ile Lys His Asp Leu Arg Lys Met Val Tyr Ser
                245                 250                 255
```

```
Gly Thr His Asp Leu Lys Gly Asp Leu Lys Val Ala Phe Asn Asp Lys
            260                 265                 270

Ala Ile Val Gln Lys Phe Ile Tyr Pro Asn Tyr Arg Ser Ile Asn Leu
        275                 280                 285

Ser Asp Ile Arg Lys Thr Met Lys Glu Ile Asp Glu Trp Asn His Val
    290                 295                 300

Lys Pro Ile Asp Phe Tyr Gln Leu Val Gly Ile Lys Asn His Ile Lys
305                 310                 315                 320

Asn Gly Asp Thr Leu Tyr Ile Glu Thr Pro Ala Lys Phe Ile Phe Asn
                325                 330                 335

Gly Ala Asn Val Tyr Tyr Arg Ala Thr Phe Thr Glu Tyr Asp Lys Asp
            340                 345                 350

Gly Lys Pro Val Gln Phe Asn Lys Phe Leu Ser Glu Asn Tyr Lys Leu
        355                 360                 365
```

<210> 9

<211> 1378

<212> DNA

<213> Bacillus thuringiensis

<220>

<221> CDS

<222> (215)..(1306)

<223>

<400> 9

```
cagtggatag gaatttgttt tcgtgctagg tatcaattta atttgttcta taagataagt     60

gaagtacgat caaaatgaat acttttgtgt attagatcaa taggtaaaat aataataaat    120

tttatatttg aaccttaaaa aattatttaa tcaaatcttt ttcactttaa aaacaaaata    180

tccagaaaaa acaatagtta acggagggat aata atg aaa tac aag tca tca aaa    235
                                       Met Lys Tyr Lys Ser Ser Lys
                                       1               5

gta gca atg tgt aca tta tca gct tta atg ctt tcg aca atc gcc act      283
Val Ala Met Cys Thr Leu Ser Ala Leu Met Leu Ser Thr Ile Ala Thr
        10                  15                  20

cca agt ata tct gtt ttc gct gca gaa aca act tcg tca cat gcg gtt      331
Pro Ser Ile Ser Val Phe Ala Ala Glu Thr Thr Ser Ser His Ala Val
    25                  30                  35
```

```
act aat cag caa aca att acg cag cgt gca gaa tct tat att gat att      379
Thr Asn Gln Gln Thr Ile Thr Gln Arg Ala Glu Ser Tyr Ile Asp Ile
40                  45                  50                  55

gtg cac aat aga atg aaa caa aga gat att gaa tca aaa atg aca ggt      427
Val His Asn Arg Met Lys Gln Arg Asp Ile Glu Ser Lys Met Thr Gly
                60                  65                  70

aaa tcc att aat atg caa gaa caa ata att gat gga tgg ttt tta gct      475
Lys Ser Ile Asn Met Gln Glu Gln Ile Ile Asp Gly Trp Phe Leu Ala
            75                  80                  85

aga ttt tgg ata ttt aag gat caa aat aat agt cac caa aca aat aga      523
Arg Phe Trp Ile Phe Lys Asp Gln Asn Asn Ser His Gln Thr Asn Arg
        90                  95                  100

ttt att tca tgg ttt aag gat aat ttg gct agc cca gga ggg tat gat      571
Phe Ile Ser Trp Phe Lys Asp Asn Leu Ala Ser Pro Gly Gly Tyr Asp
    105                 110                 115

agt atc gct gaa cag atg ggc cta aaa gta gca gca tta aat gat atg      619
Ser Ile Ala Glu Gln Met Gly Leu Lys Val Ala Ala Leu Asn Asp Met
120                 125                 130                 135

gat ata tca aat gta aat tat act tct aag aca ggg gat act ata tat      667
Asp Ile Ser Asn Val Asn Tyr Thr Ser Lys Thr Gly Asp Thr Ile Tyr
                140                 145                 150

aat ggt gtt tca gaa tta aaa aat atc aca gga aca act caa aaa atg      715
Asn Gly Val Ser Glu Leu Lys Asn Ile Thr Gly Thr Thr Gln Lys Met
            155                 160                 165

aaa aca gat agt ttt caa aga gat tat aca aaa tcc cag tca act tca      763
Lys Thr Asp Ser Phe Gln Arg Asp Tyr Thr Lys Ser Gln Ser Thr Ser
        170                 175                 180

atc acc aat gga tta caa tta gga ttt aaa gtt tca gct aaa gga ata      811
Ile Thr Asn Gly Leu Gln Leu Gly Phe Lys Val Ser Ala Lys Gly Ile
    185                 190                 195

gtg gcc tta gcc ggt gcg gat ttt gaa gca agt gta aac tat aat tta      859
Val Ala Leu Ala Gly Ala Asp Phe Glu Ala Ser Val Asn Tyr Asn Leu
200                 205                 210                 215

tcc act acc gca act gaa acc aat aca ata tct gat aaa ttt acc gtt      907
Ser Thr Thr Ala Thr Glu Thr Asn Thr Ile Ser Asp Lys Phe Thr Val
                220                 225                 230

cct tca caa gaa gtc aca tta gcg cca gga cat aag gcg atc gta aaa      955
Pro Ser Gln Glu Val Thr Leu Ala Pro Gly His Lys Ala Ile Val Lys
            235                 240                 245

cat agt ttg aag aaa atg gta tac tct gga acg cat gat tta aaa gga     1003
His Ser Leu Lys Lys Met Val Tyr Ser Gly Thr His Asp Leu Lys Gly
        250                 255                 260

gat tta aca att act ttt aat gat aag gat tta gtt caa aaa ttt att     1051
Asp Leu Thr Ile Thr Phe Asn Asp Lys Asp Leu Val Gln Lys Phe Ile
    265                 270                 275

tat cca aat tat aaa gct att gat tta tct aat att cgt aaa gca atg     1099
Tyr Pro Asn Tyr Lys Ala Ile Asp Leu Ser Asn Ile Arg Lys Ala Met
280                 285                 290                 295

aca gaa att gat gaa tgg aat cat gta aaa cct acc gat ttc tat caa     1147
Thr Glu Ile Asp Glu Trp Asn His Val Lys Pro Thr Asp Phe Tyr Gln
                300                 305                 310
```

```
tta gtt ggg aat aaa aat tat ata aaa aac ggg gac act tta tac atc     1195
Leu Val Gly Asn Lys Asn Tyr Ile Lys Asn Gly Asp Thr Leu Tyr Ile
            315                 320                 325

gaa aca cct gct aaa ttc act ttg aat gga ggc aac cct tat tat aca     1243
Glu Thr Pro Ala Lys Phe Thr Leu Asn Gly Gly Asn Pro Tyr Tyr Thr
        330                 335                 340

gca acc ttt acg gaa tat gat gaa aat gga aat caa gtc aaa aca aag     1291
Ala Thr Phe Thr Glu Tyr Asp Glu Asn Gly Asn Gln Val Lys Thr Lys
    345                 350                 355

cgt tta aat aac aaa taagttactt aaaggtaatt cattaacaat gtatccatta     1346
Arg Leu Asn Asn Lys
360

tataattaat ttataaaaat aatgttttaa aa                                1378
```

<210> 10

<211> 364

<212> PRT

<213> Bacillus thuringiensis

<400> 10

```
Met Lys Tyr Lys Ser Ser Lys Val Ala Met Cys Thr Leu Ser Ala Leu
1               5                   10                  15

Met Leu Ser Thr Ile Ala Thr Pro Ser Ile Ser Val Phe Ala Ala Glu
            20                  25                  30

Thr Thr Ser Ser His Ala Val Thr Asn Gln Gln Thr Ile Thr Gln Arg
        35                  40                  45

Ala Glu Ser Tyr Ile Asp Ile Val His Asn Arg Met Lys Gln Arg Asp
    50                  55                  60

Ile Glu Ser Lys Met Thr Gly Lys Ser Ile Asn Met Gln Glu Gln Ile
65                  70                  75                  80

Ile Asp Gly Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn
                85                  90                  95

Asn Ser His Gln Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Leu
            100                 105                 110

Ala Ser Pro Gly Gly Tyr Asp Ser Ile Ala Glu Gln Met Gly Leu Lys
        115                 120                 125

Val Ala Ala Leu Asn Asp Met Asp Ile Ser Asn Val Asn Tyr Thr Ser
    130                 135                 140

Lys Thr Gly Asp Thr Ile Tyr Asn Gly Val Ser Glu Leu Lys Asn Ile
```

145 150 155 160

Thr Gly Thr Thr Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr
165 170 175

Thr Lys Ser Gln Ser Thr Ser Ile Thr Asn Gly Leu Gln Leu Gly Phe
180 185 190

Lys Val Ser Ala Lys Gly Ile Val Ala Leu Ala Gly Ala Asp Phe Glu
195 200 205

Ala Ser Val Asn Tyr Asn Leu Ser Thr Thr Ala Thr Glu Thr Asn Thr
210 215 220

Ile Ser Asp Lys Phe Thr Val Pro Ser Gln Glu Val Thr Leu Ala Pro
225 230 235 240

Gly His Lys Ala Ile Val Lys His Ser Leu Lys Lys Met Val Tyr Ser
245 250 255

Gly Thr His Asp Leu Lys Gly Asp Leu Thr Ile Thr Phe Asn Asp Lys
260 265 270

Asp Leu Val Gln Lys Phe Ile Tyr Pro Asn Tyr Lys Ala Ile Asp Leu
275 280 285

Ser Asn Ile Arg Lys Ala Met Thr Glu Ile Asp Glu Trp Asn His Val
290 295 300

Lys Pro Thr Asp Phe Tyr Gln Leu Val Gly Asn Lys Asn Tyr Ile Lys
305 310 315 320

Asn Gly Asp Thr Leu Tyr Ile Glu Thr Pro Ala Lys Phe Thr Leu Asn
325 330 335

Gly Gly Asn Pro Tyr Tyr Thr Ala Thr Phe Thr Glu Tyr Asp Glu Asn
340 345 350

Gly Asn Gln Val Lys Thr Lys Arg Leu Asn Asn Lys
355 360

<210> 11

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> probe sequence, or amplification primer sequence for use with pri

mer as set forth in SEQ ID NO 12, corresponding to CDS as set for th in SEQ ID NO 3 from 438-458, biased toward codons preferred in Bacillus species genes containing A or T in 3rd position

<400> 11
aataataatc atcaaacwaa t 21

<210> 12

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> probe sequence, or amplification primer sequence for use with SEQ ID NO 11 corresponding to SEQ ID NO 3 from nucleotide position 978- 998, biased toward codons preferred in Bacillus species genes in which A or T is in 3rd position

<400> 12
attwggataw ataaattttt g 21

<210> 13

<211> 1101

<212> DNA

<213> Artificial Sequence

<220>

<223> coding sequence preferred for use in monocot species encoding a Bt TIC901 amino acid sequence variant

<220>

<221> CDS

<222> (1)..(1101)

<223>

<400> 13
atg aag aac cgc ttc agc aag gtc gcc ctc tgc acg gtg cct atc ctc 48
Met Lys Asn Arg Phe Ser Lys Val Ala Leu Cys Thr Val Pro Ile Leu
1 5 10 15

atg gtt tct acg ttc gcg tcc agc tcg atg tcc gcg ttc gca gcg gag 96
Met Val Ser Thr Phe Ala Ser Ser Ser Met Ser Ala Phe Ala Ala Glu
20 25 30

gcc aaa agt cct gac ttg aac gtg tcc caa cag gtc ata ggc cct tac 144
Ala Lys Ser Pro Asp Leu Asn Val Ser Gln Gln Val Ile Gly Pro Tyr
35 40 45

gca gaa tct tac atc gac atc gtc cag gac aga atg aag cag aga gac 192

```
Ala Glu Ser Tyr Ile Asp Ile Val Gln Asp Arg Met Lys Gln Arg Asp
    50                  55                  60

aaa gga tcc aaa ctc act ggc aaa ccc atc aac atg caa gag cag atc      240
Lys Gly Ser Lys Leu Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile
65                  70                  75                  80

atc gat ggg tgg ttt ctc gca cga ttc tgg att ttc aag gat cag aac      288
Ile Asp Gly Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn
                85                  90                  95

aat aac cac cag aca aac agg ttc atc tca tgg ttt aag gat aac atc      336
Asn Asn His Gln Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Ile
            100                 105                 110

gcc tca tct aag gga tac aac tca ata gcc gaa cag atg ggc ctc aaa      384
Ala Ser Ser Lys Gly Tyr Asn Ser Ile Ala Glu Gln Met Gly Leu Lys
        115                 120                 125

atc gaa gca gag aat gat atg gac gtg aca aat atc gac tac act agt      432
Ile Glu Ala Glu Asn Asp Met Asp Val Thr Asn Ile Asp Tyr Thr Ser
    130                 135                 140

aag acc gga gac aca atc tac aac ggc att tcg gaa ctt aaa aac tat      480
Lys Thr Gly Asp Thr Ile Tyr Asn Gly Ile Ser Glu Leu Lys Asn Tyr
145                 150                 155                 160

acg ggc agc acc cag aaa atg aag acc gat agc ttt caa agg gac tac      528
Thr Gly Ser Thr Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr
                165                 170                 175

aca aaa tcc gag tcg acc tcc gtg acc aat ggc ctc cag ctg ggc ttc      576
Thr Lys Ser Glu Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Phe
            180                 185                 190

aag gtg gca gca aag ggc gtc gtc gct tta gcc ggc gca gac ttc gag      624
Lys Val Ala Ala Lys Gly Val Val Ala Leu Ala Gly Ala Asp Phe Glu
        195                 200                 205

act tcg gtg acc tac aat ctg tct aca act acg act gag acg aac aca      672
Thr Ser Val Thr Tyr Asn Leu Ser Thr Thr Thr Thr Glu Thr Asn Thr
    210                 215                 220

att tcc gac aag ttt acg gtt ccg tct cag gag gtt acg ttc cct cca      720
Ile Ser Asp Lys Phe Thr Val Pro Ser Gln Glu Val Thr Phe Pro Pro
225                 230                 235                 240

ggc cac aag gca atc gtc aag cac gac ctg agg aaa atg gtc tac agc      768
Gly His Lys Ala Ile Val Lys His Asp Leu Arg Lys Met Val Tyr Ser
                245                 250                 255

ggc acc cat gat ctc aaa ggc gac ctc atc gtg tcg ttc aac gac aag      816
Gly Thr His Asp Leu Lys Gly Asp Leu Ile Val Ser Phe Asn Asp Lys
            260                 265                 270

gag ata gtc cag aag ttc atc tac cca aat tac cgc gac atc aac ctc      864
Glu Ile Val Gln Lys Phe Ile Tyr Pro Asn Tyr Arg Asp Ile Asn Leu
        275                 280                 285

agt gac atc cga gag acc atg atc gag atc gac gag tgg aac cac gtg      912
Ser Asp Ile Arg Glu Thr Met Ile Glu Ile Asp Glu Trp Asn His Val
    290                 295                 300

aac cct gtc aat ttc tac gaa ctc gta gga gtt aag aac cac atc aag      960
Asn Pro Val Asn Phe Tyr Glu Leu Val Gly Val Lys Asn His Ile Lys
305                 310                 315                 320

aac ggt gaa aca ttg tac atc gac acg ccg gct aag ttc atg ttc aac     1008
```

Asn Gly Glu Thr Leu Tyr Ile Asp Thr Pro Ala Lys Phe Met Phe Asn
325 330 335

gga gcg aat cct tac tat cga gct acc ttc acg gag tac gat ggc aac 1056
Gly Ala Asn Pro Tyr Tyr Arg Ala Thr Phe Thr Glu Tyr Asp Gly Asn
340 345 350

aac aat cct gtt cag acc aag gtg ttg agt gag aat ttc aag ctg 1101
Asn Asn Pro Val Gln Thr Lys Val Leu Ser Glu Asn Phe Lys Leu
355 360 365

<210> 14

<211> 367

<212> PRT

<213> Artificial Sequence

<220>

<223> coding sequence preferred for use in monocot species encoding a B t TIC901 amino acid sequence variant

<400> 14

Met Lys Asn Arg Phe Ser Lys Val Ala Leu Cys Thr Val Pro Ile Leu
1 5 10 15

Met Val Ser Thr Phe Ala Ser Ser Ser Met Ser Ala Phe Ala Ala Glu
20 25 30

Ala Lys Ser Pro Asp Leu Asn Val Ser Gln Gln Val Ile Gly Pro Tyr
35 40 45

Ala Glu Ser Tyr Ile Asp Ile Val Gln Asp Arg Met Lys Gln Arg Asp
50 55 60

Lys Gly Ser Lys Leu Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile
65 70 75 80

Ile Asp Gly Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn
85 90 95

Asn Asn His Gln Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Ile
100 105 110

Ala Ser Ser Lys Gly Tyr Asn Ser Ile Ala Glu Gln Met Gly Leu Lys
115 120 125

Ile Glu Ala Glu Asn Asp Met Asp Val Thr Asn Ile Asp Tyr Thr Ser
130 135 140

Lys Thr Gly Asp Thr Ile Tyr Asn Gly Ile Ser Glu Leu Lys Asn Tyr
145 150 155 160

Thr Gly Ser Thr Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr
165 170 175

Thr Lys Ser Glu Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Phe
180 185 190

Lys Val Ala Ala Lys Gly Val Val Ala Leu Ala Gly Ala Asp Phe Glu
195 200 205

Thr Ser Val Thr Tyr Asn Leu Ser Thr Thr Thr Thr Glu Thr Asn Thr
210 215 220

Ile Ser Asp Lys Phe Thr Val Pro Ser Gln Glu Val Thr Phe Pro Pro
225 230 235 240

Gly His Lys Ala Ile Val Lys His Asp Leu Arg Lys Met Val Tyr Ser
245 250 255

Gly Thr His Asp Leu Lys Gly Asp Leu Ile Val Ser Phe Asn Asp Lys
260 265 270

Glu Ile Val Gln Lys Phe Ile Tyr Pro Asn Tyr Arg Asp Ile Asn Leu
275 280 285

Ser Asp Ile Arg Glu Thr Met Ile Glu Ile Asp Glu Trp Asn His Val
290 295 300

Asn Pro Val Asn Phe Tyr Glu Leu Val Gly Val Lys Asn His Ile Lys
305 310 315 320

Asn Gly Glu Thr Leu Tyr Ile Asp Thr Pro Ala Lys Phe Met Phe Asn
325 330 335

Gly Ala Asn Pro Tyr Tyr Arg Ala Thr Phe Thr Glu Tyr Asp Gly Asn
340 345 350

Asn Asn Pro Val Gln Thr Lys Val Leu Ser Glu Asn Phe Lys Leu
355 360 365

<210> 15

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer coupled with prJPW152 SEQ ID NO 16 in thermal amplification reactions

<400> 15
cctttggcag aaactttaac tcc 23

<210> 16

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer coupled with prJPW151 SEQ ID NO 15 in thermal amplification reactions

<400> 16
gtgtattctg gtacgcatga c 21

<210> 17

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer coupled with prJPW183 SEQ ID NO 18 in thermal amplification reactions

<400> 17
gccggatccc tagctgaata tgcagtagat aatg 34

<210> 18

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer coupled with prJPW186 SEQ ID NO 17 in thermal amplification reactions

<400> 18
gtggcacgtt tataggccat tgttc 25

<210> 19

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer coupled with prJPW156 SEQ ID NO 20 in thermal amplification reactions

<400> 19
cttttaggcc catctgttca gcg 23

<210> 20

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer coupled with prJPW155 SEQ ID NO 19 in thermal amplification reactions

<400> 20
gccttagccg gtgcggattt tgaagc 26

<210> 21

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer coupled with prJPW170 SEQ ID NO 22 in thermal amplification reactions

<400> 21
ggagcttatt tgttatttaa acgctttgtt ttgacttgat ttcc 44

<210> 22

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide primer coupled with prJPW168 SEQ ID NO 21 in thermal amplification reactions

<400> 22
gccggatccc agtggatagg aatttgtttt cgtgctagg 39

<210> 23

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> universal forward amplification primer similar to SEQ ID NO:24 and SEQ ID NO:25 that, when used in a thermal amplification reaction with any of SEQ ID NO:SEQ ID NO:27-29 and template DNA homologous to tic901, 1201, 407, 417, or 431 and the like result in amplicon of from about 600 to about 650 base pairs

<400> 23
aayatgcarg arcarathat hgaygg 26

<210> 24

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> universal forward amplification primer similar to SEQ ID NO:23 and SEQ ID NO:25 that, when used in a thermal amplification reaction with any of SEQ ID NO:27-29 and template DNA homologous to tic901, 1201, 407, 417, or 431 and the like result in amplicon of from about 600 to about 650 base pairs

<400> 24
aayatgcarg arcarathat hga 23

<210> 25

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> universal forward amplification primer similar to SEQ ID NO:23 and SEQ ID NO:24 that, when used in a thermal amplification reaction with any of SEQ ID NO:27-29 and template DNA homologous to tic901, 1201, 407, 417, or 431 and the like result in amplicon of from about 600 to about 650 base pairs

<400> 25
aayatgcarg arcarathat 20

<210> 26

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> universal forward amplification primer that, when used in a therm al amplification reaction with any of SEQ ID NO:27-29 and templat e DNA homologous to tic901, 1201, 407, 417, 431 and the like resu lt in amplicon of from about 395 to about 425 base pairs

<220>

<221> misc_feature

<222> (3)..(3)

<223> inosine

<220>

<221> misc_feature

<222> (9)..(9)

<223> inosine

<400> 26
ggngayacna thtayaaygg 20

<210> 27

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> universal reverse amplification primer similar to SEQ ID NO:28 an d SEQ ID NO:29

<220>

<221> misc_feature

<222> (6)..(6)

<223> inosine

<220>

<221> misc_feature

<222> (24)..(24)

<223> inosine

<400> 27
tarttnggrt adatraaytt ytgnac 26

<210> 28

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> universal reverse amplification primer similar to SEQ ID NO:27 and SEQ ID NO:29

<220>

<221> misc_feature

<222> (6)..(6)

<223> inosine

<400> 28
tarttnggrt adatraaytt ytg 23

<210> 29

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> universal reverse amplification primer similar to SEQ ID NO:27 and SEQ ID NO:28

<220>

<221> misc_feature

<222> (18)..(18)

<223> inosine

<400> 29
ggrtadatra ayttytgnac 20

<210> 30

<211> 570

<212> DNA

<213> Bacillus thuringiensis

<220>

<221> CDS

<222> (1)..(570)

<223>

<400> 30

```
ttt tta gct aga ttt tgg ata ttt gag gat caa aat aat agt cac caa        48
Phe Leu Ala Arg Phe Trp Ile Phe Glu Asp Gln Asn Asn Ser His Gln
1               5                   10                  15

aca aat aga ttt att tca tgg ttt aag gat aat att gct agt tca aaa        96
Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Ile Ala Ser Ser Lys
            20                  25                  30

ggg tat aat agt att gcg gag caa atg ggt tta aaa ata gaa gca gaa       144
Gly Tyr Asn Ser Ile Ala Glu Gln Met Gly Leu Lys Ile Glu Ala Glu
        35                  40                  45

aac gat atg gat gta aca aat ata gat tat aca tct aag aca ggc gat       192
Asn Asp Met Asp Val Thr Asn Ile Asp Tyr Thr Ser Lys Thr Gly Asp
    50                  55                  60

acc att tat aat ggt att tca gaa ttg aaa aat tat aca gga tca act       240
Thr Ile Tyr Asn Gly Ile Ser Glu Leu Lys Asn Tyr Thr Gly Ser Thr
65                  70                  75                  80

caa aag atg aaa aca gat agt ttt caa aga gat tat aca aaa tca gaa       288
Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr Thr Lys Ser Glu
                85                  90                  95

tct act tca gta act aat gga tta caa tta gga ttt aaa gtt gct gct       336
Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Phe Lys Val Ala Ala
            100                 105                 110

aaa gga gta gtt gct ttg gct ggg gca gac ttt gaa acc agt gtt act       384
Lys Gly Val Val Ala Leu Ala Gly Ala Asp Phe Glu Thr Ser Val Thr
        115                 120                 125

tat aat cta tca act act aca act gaa aca aat aca ata tca gac aag       432
Tyr Asn Leu Ser Thr Thr Thr Thr Glu Thr Asn Thr Ile Ser Asp Lys
    130                 135                 140

ttt act gtc cca tct caa gaa gtt aca ttg cct cca gga cat aaa gcg       480
Phe Thr Val Pro Ser Gln Glu Val Thr Leu Pro Pro Gly His Lys Ala
145                 150                 155                 160

ata gtg aaa cat gat tta aga aaa atg gtt tat tct ggt act cat gat       528
Ile Val Lys His Asp Leu Arg Lys Met Val Tyr Ser Gly Thr His Asp
                165                 170                 175

cta aag ggt gat tta att gtg agt ttt aat gat aaa gag att               570
Leu Lys Gly Asp Leu Ile Val Ser Phe Asn Asp Lys Glu Ile
```

180 185 190

<210> 31

<211> 190

<212> PRT

<213> Bacillus thuringiensis

<400> 31

Phe Leu Ala Arg Phe Trp Ile Phe Glu Asp Gln Asn Asn Ser His Gln
1 5 10 15

Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Ile Ala Ser Ser Lys
20 25 30

Gly Tyr Asn Ser Ile Ala Glu Gln Met Gly Leu Lys Ile Glu Ala Glu
35 40 45

Asn Asp Met Asp Val Thr Asn Ile Asp Tyr Thr Ser Lys Thr Gly Asp
50 55 60

Thr Ile Tyr Asn Gly Ile Ser Glu Leu Lys Asn Tyr Thr Gly Ser Thr
65 70 75 80

Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr Thr Lys Ser Glu
85 90 95

Ser Thr Ser Val Thr Asn Gly Leu Gln Leu Gly Phe Lys Val Ala Ala
100 105 110

Lys Gly Val Val Ala Leu Ala Gly Ala Asp Phe Glu Thr Ser Val Thr
115 120 125

Tyr Asn Leu Ser Thr Thr Thr Thr Glu Thr Asn Thr Ile Ser Asp Lys
130 135 140

Phe Thr Val Pro Ser Gln Glu Val Thr Leu Pro Pro Gly His Lys Ala
145 150 155 160

Ile Val Lys His Asp Leu Arg Lys Met Val Tyr Ser Gly Thr His Asp
165 170 175

Leu Lys Gly Asp Leu Ile Val Ser Phe Asn Asp Lys Glu Ile
180 185 190

<210> 32

<211> 1095

<212> DNA

<213> Bacillus thuringiensis

<220>

<221> CDS

<222> (1)..(1092)

<223> sequence encoding TIC431 precursor amino acid sequence

<400> 32

```
atg aaa tac aag tct tca aaa gta gca atg tgt aca tta tcg gct tta       48
Met Lys Tyr Lys Ser Ser Lys Val Ala Met Cys Thr Leu Ser Ala Leu
1               5                  10                  15

atg ctt tcg aca atc gcc act cca agt ata tct gtt ttc gct gct gaa       96
Met Leu Ser Thr Ile Ala Thr Pro Ser Ile Ser Val Phe Ala Ala Glu
            20                  25                  30

aca act gca tca cat aag gtt act aat cag caa aca att gca cag cgt      144
Thr Thr Ala Ser His Lys Val Thr Asn Gln Gln Thr Ile Ala Gln Arg
        35                  40                  45

gca gaa tct tat atc gat att gtg cat aat aga atg aaa aaa cga gat      192
Ala Glu Ser Tyr Ile Asp Ile Val His Asn Arg Met Lys Lys Arg Asp
    50                  55                  60

att gaa tca aaa atg aca ggt aaa cct att aat atg caa gaa caa ata      240
Ile Glu Ser Lys Met Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile
65                  70                  75                  80

att gat gga tgg ttt tta gct aga ttt tgg ata ttc aag gac caa aat      288
Ile Asp Gly Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn
                85                  90                  95

aat agt cac caa aca aat aga ttt att tca tgg ttt aaa gat aat tta      336
Asn Ser His Gln Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Leu
            100                 105                 110

gct agt cca gga ggg tat aat agt atc gct aaa caa atg ggg tta aaa      384
Ala Ser Pro Gly Gly Tyr Asn Ser Ile Ala Lys Gln Met Gly Leu Lys
        115                 120                 125

ata gaa gta tta aat gat atg gat ata tca aat gta aat tat act tct      432
Ile Glu Val Leu Asn Asp Met Asp Ile Ser Asn Val Asn Tyr Thr Ser
    130                 135                 140

aag aca ggg gat act ata tat aat ggt gtt tcc gaa tta aaa aat atc      480
Lys Thr Gly Asp Thr Ile Tyr Asn Gly Val Ser Glu Leu Lys Asn Ile
145                 150                 155                 160

aca ggt aca act caa aaa atg aaa aca gat agt ttt caa aga gat tat      528
Thr Gly Thr Thr Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr
                165                 170                 175

aca aaa tca cag tca act tca atc acc aat gga tta caa tta gga ttt      576
Thr Lys Ser Gln Ser Thr Ser Ile Thr Asn Gly Leu Gln Leu Gly Phe
            180                 185                 190

aaa gtt tct gcc aaa ggg gtg ata gct tta gca gga gca gac ttc gaa      624
Lys Val Ser Ala Lys Gly Val Ile Ala Leu Ala Gly Ala Asp Phe Glu
        195                 200                 205
```

```
gca agt gtc aac tat aat tta tcc act acc gca act gaa acc aat ata      672
Ala Ser Val Asn Tyr Asn Leu Ser Thr Thr Ala Thr Glu Thr Asn Ile
    210                 215                 220

ata tct gat aaa ttt acc gtt cct tca caa gaa gtt aca tta gcg cca      720
Ile Ser Asp Lys Phe Thr Val Pro Ser Gln Glu Val Thr Leu Ala Pro
225                 230                 235                 240

gga cat aag gcg atc gta aaa cat agt tta aag aaa atg gta tac tcc      768
Gly His Lys Ala Ile Val Lys His Ser Leu Lys Lys Met Val Tyr Ser
                245                 250                 255

gga acg cat gat tta aaa gga gat tta aca att act ttt aat gat aag      816
Gly Thr His Asp Leu Lys Gly Asp Leu Thr Ile Thr Phe Asn Asp Lys
            260                 265                 270

gat tta gtt caa aaa ttt att tat cca aat tat aaa gct att gat tta      864
Asp Leu Val Gln Lys Phe Ile Tyr Pro Asn Tyr Lys Ala Ile Asp Leu
        275                 280                 285

tct aat att cgt aaa gca ctg act gaa att gat gaa tgg aat cat gta      912
Ser Asn Ile Arg Lys Ala Leu Thr Glu Ile Asp Glu Trp Asn His Val
    290                 295                 300

aaa cct acc gat ttc tat caa tta gtt ggg aac aaa aat tat ata aaa      960
Lys Pro Thr Asp Phe Tyr Gln Leu Val Gly Asn Lys Asn Tyr Ile Lys
305                 310                 315                 320

aac ggg gac act tta tac atc gaa aca cct gct aaa ttc act ttg aat     1008
Asn Gly Asp Thr Leu Tyr Ile Glu Thr Pro Ala Lys Phe Thr Leu Asn
                325                 330                 335

gga gga aac cct tat tat aca gca acc ttt acg gaa tat gat gaa agt     1056
Gly Gly Asn Pro Tyr Tyr Thr Ala Thr Phe Thr Glu Tyr Asp Glu Ser
            340                 345                 350

gga aat caa gtc aaa aca aag cat tta agt gtc aaa taa                 1095
Gly Asn Gln Val Lys Thr Lys His Leu Ser Val Lys
        355                 360
```

<210> 33

<211> 364

<212> PRT

<213> Bacillus thuringiensis

<400> 33

```
Met Lys Tyr Lys Ser Ser Lys Val Ala Met Cys Thr Leu Ser Ala Leu
1               5                   10                  15

Met Leu Ser Thr Ile Ala Thr Pro Ser Ile Ser Val Phe Ala Ala Glu
            20                  25                  30

Thr Thr Ala Ser His Lys Val Thr Asn Gln Gln Thr Ile Ala Gln Arg
        35                  40                  45

Ala Glu Ser Tyr Ile Asp Ile Val His Asn Arg Met Lys Lys Arg Asp
    50                  55                  60
```

```
Ile Glu Ser Lys Met Thr Gly Lys Pro Ile Asn Met Gln Glu Gln Ile
65                  70                  75                  80

Ile Asp Gly Trp Phe Leu Ala Arg Phe Trp Ile Phe Lys Asp Gln Asn
                85                  90                  95

Asn Ser His Gln Thr Asn Arg Phe Ile Ser Trp Phe Lys Asp Asn Leu
            100                 105                 110

Ala Ser Pro Gly Gly Tyr Asn Ser Ile Ala Lys Gln Met Gly Leu Lys
        115                 120                 125

Ile Glu Val Leu Asn Asp Met Asp Ile Ser Asn Val Asn Tyr Thr Ser
    130                 135                 140

Lys Thr Gly Asp Thr Ile Tyr Asn Gly Val Ser Glu Leu Lys Asn Ile
145                 150                 155                 160

Thr Gly Thr Thr Gln Lys Met Lys Thr Asp Ser Phe Gln Arg Asp Tyr
                165                 170                 175

Thr Lys Ser Gln Ser Thr Ser Ile Thr Asn Gly Leu Gln Leu Gly Phe
            180                 185                 190

Lys Val Ser Ala Lys Gly Val Ile Ala Leu Ala Gly Ala Asp Phe Glu
        195                 200                 205

Ala Ser Val Asn Tyr Asn Leu Ser Thr Thr Ala Thr Glu Thr Asn Ile
    210                 215                 220

Ile Ser Asp Lys Phe Thr Val Pro Ser Gln Glu Val Thr Leu Ala Pro
225                 230                 235                 240

Gly His Lys Ala Ile Val Lys His Ser Leu Lys Lys Met Val Tyr Ser
                245                 250                 255

Gly Thr His Asp Leu Lys Gly Asp Leu Thr Ile Thr Phe Asn Asp Lys
            260                 265                 270

Asp Leu Val Gln Lys Phe Ile Tyr Pro Asn Tyr Lys Ala Ile Asp Leu
        275                 280                 285

Ser Asn Ile Arg Lys Ala Leu Thr Glu Ile Asp Glu Trp Asn His Val
    290                 295                 300

Lys Pro Thr Asp Phe Tyr Gln Leu Val Gly Asn Lys Asn Tyr Ile Lys
305                 310                 315                 320

Asn Gly Asp Thr Leu Tyr Ile Glu Thr Pro Ala Lys Phe Thr Leu Asn
                325                 330                 335
```

```
Gly Gly Asn Pro Tyr Tyr Thr Ala Thr Phe Thr Glu Tyr Asp Glu Ser
            340                 345                 350

Gly Asn Gln Val Lys Thr Lys His Leu Ser Val Lys
        355                 360
```

```
                10                  20                  30                  40
     ------------------+-------------------+-------------------+-------------------+
1    M K N R F S K V A L C T V P I L M V S T F A S S S M S A F A A E A K S P D L N V  TIC901p
1    M K Y K F S K V V K C T L P A L M I T T F V T P S M A V F A A E T K S P N L N A  TIC1201p
1    M K Y K S S K V A I C T L S A L M L S T I G T S S M S T F A A E T T L P G Q T L  TIC407p
1    M K Y K S S K V A M C T L S A L M L S T I A T P S I S V F A A E T T S S H A V T  TIC417p
1    M K Y K S S K V A M C T L S A L M L S T I A T P S I S V F A A E T T A S H K V T  TIC431p
     M K Y K S S K V A X C T L S A L M L S T I A T P S M S V F A A E T T S P X L N X  Consenso
     ------------------+-------------------+-------------------+-------------------+

                50                  60                  70                  80
     ------------------+-------------------+-------------------+-------------------+
41   S Q Q V I G P Y A E S Y I D I V Q D R M K Q R D K G S K L T G K P I N M Q E Q I  TIC901p
41   S Q Q A I T P Y A E S Y I D T V Q D R M K Q R D R E S K L T G K P I N M Q E Q I  TIC1201p
41   K E Q S I T P R A E S Y I D I V Q D R M K Q R D I E S K R T G K P I N M Q E Q I  TIC407p
41   N Q Q T I T Q R A E S Y I D I V H N R M K Q R D I E S K M T G K S I N M Q E Q I  TIC417p
41   N Q Q T I A Q R A E S Y I D I V H N R M K K R D I E S K M T G K P I N M Q E Q I  TIC431p
     S Q Q X I T P R A E S Y I D I V Q D R M K Q R D I E S K L T G K P I N M Q E Q I  Consenso
     ------------------+-------------------+-------------------+-------------------+

                90                  100                 110                 120
     ------------------+-------------------+-------------------+-------------------+
81   I D G W F L A R F W I F K D Q N N N H Q T N R F I S W F K D N I A S S K G Y N S  TIC901
81   I D G W F L A R F W I F K D Q N N N H Q T N R F I S W F K D N L A S S K G Y D S  TIC1201
81   I D G W F L A R F W I X K D Q N N N H Q T N R F I T W F K N N V A S S K G Y E G  TIC407
81   I D G W F L A R F W I F K D Q N N S H Q T N R F I S W F K D N L A S P G G Y D S  TIC417
81   I D G W F L A R F W I F K D Q N N S H Q T N R F I S W F K D N L A S P G G Y N S  TIC431p
     I D G W F L A R F W I F K D Q N N N H Q T N R F I S W F K D N L A S S K G Y D S  Consenso
     ------------------+-------------------+-------------------+-------------------+

                130                 140                 150                 160
     ------------------+-------------------+-------------------+-------------------+
121  I A E Q M G L K I E A E N D M D V T N I D Y T S K T G D T I Y N G I S E L K N Y  TIC901
121  I A E Q M G L K I E A L N D M D V T N I D Y T S K T G D T I Y N G I S E L T N Y  TIC1201
121  I A E Q M G L K I E S M S D M N V S N I N Y T G K K G D T I Y N G V S E L E N K  TIC407
121  I A E Q M G L K V A A L N D M D I S N V N Y T S K T G D T I Y N G V S E L K N I  TIC417
121  I A K Q M G L K I E V L N D M D I S N V N Y T S K T G D T I Y N G V S E L K N I  TIC431p
     I A E Q M G L K I E A L N D M D V S N I N Y T S K T G D T I Y N G V S E L K N Y  Consenso
     ------------------+-------------------+-------------------+-------------------+
```

Figura 1A

308

```
                  170                 180                 190                 200
                  +                   +                   +                   +
161  T G S T Q K M K T D S F Q R D Y T K S E S T S V T N G L Q L G F K V A A K G V V  TIC901
161  T G T T Q K M K T D S F Q R D Y T K S E S T S V T N G L Q L G F K V A A K G V V  TIC1201
161  M G T P Q K M K S D S F Q R D Y T K S Q S T S V T N G L Q L G V K V S A K G T V  TIC407
161  T G T T Q K M K T D S F Q R D Y T K S Q S T S I T D G L Q L G F K V S A K G I V  TIC417
161  T G T T Q K M K T D S F Q R D Y T K S Q S T S I T N G L Q L G F K V S A K G V I  TIC431p
     T G T T Q K M K T D S F Q R D Y T K S Q S T S V T N G L Q L G F K V S A K G V V  Consenso
                  +                   +                   +                   +

                  210                 220                 230                 240
                  +                   +                   +                   +
201  A L A G A D F E T S V T Y N L S T T T T E T N T I S D K F T V P S Q E V T L P P  TIC901
201  A L A G A D F E T S V T Y N L S S T T T E T N T I S D K F T V P S Q E V T L S P  TIC1201
201  V L G E A S L E T S V T Y N L S S T A T E T D T T S D K F T V P S Q E V T L P P  TIC407
201  A L A G A D F E A S V N Y N L S T T A T E T N T I S D K F T V P S Q E V T L A P  TIC417
201  A L A G A D F E A S V N Y N L S T T A T E T N I I S D K F T V P S Q E V T L A P  TIC431p
     A L A G A D F E T S V T Y N L S T T A T E T N T I S D K F T V P S Q E V T L P P  Consenso
                  +                   +                   +                   +

                  250                 260         270             280
                  +                   +           +               +
241  G H K A I V K H D L R K M V Y S G T H D L K G D L I V S F N D K E I V Q K F I Y  TIC901
241  G H K A V V K H D L R K M V Y F G T H D L K G D L K V G F N D K E I V Q K F I Y  TIC1201
241  G H K A V I K H D L R K M V Y S G T H D L K G D L K V A F N D K A I V Q K F I Y  TIC407
241  G H K A I V K H S L K K M V Y S G T H D L K G D L T I T F N D K D L V Q K F I Y  TIC417
241  G H K A I V K H S L K K M V Y S G T H D L K G D L T I T F N D K D L V Q K F I Y  TIC431p
     G H K A I V K H D L R K M V Y S G T H D L K G D L K V X F N D K E I V Q K F I Y  Consenso
                  +                   +                   +                   +

                  290                 300                 310                 320
                  +                   +                   +                   +
281  P N Y R E I N L S D I R E T M I E I D E W N H V N P V N F Y E L V G V K N H I K  TIC901
281  P N Y R S I D L S D I R K T M I E I D K W N H V N T I D F Y Q L V G V K N H I K  TIC1201
281  P N Y R S I N L S D I R K T M K E I D E W N H V K P I D F Y Q L V G I K N H I K  TIC407
281  P N Y K A I D L S N I R K A M T E I D E W N H V K P T D F Y Q L V G N K N Y I K  TIC417
281  P N Y K A I D L S N I R K A L T E I D E W N H V K P T D F Y Q L V G N K N Y I K  TIC431p
     P N Y R S I D L S D I R K T M I E I D E W N H V N P I D F Y Q L V G V K N H I K  Consenso
                  +                   +                   +                   +
```

Figura 1B

309

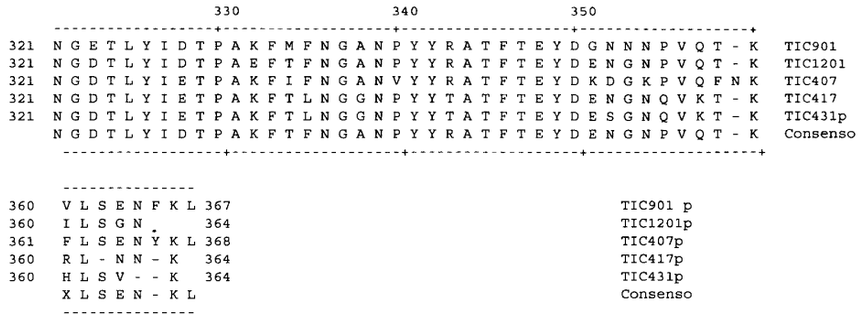

330
340
350
321 N G E T L Y I D T P A K F M F N G A N P Y Y R A T F T E Y D G N N N P V Q T - K TIC901
321 N G D T L Y I D T P A E F T F N G A N P Y Y R A T F T E Y D E N G N P V Q T - K TIC1201
321 N G D T L Y I E T P A K F I F N G A N V Y Y R A T F T E Y D K D G K P V Q F N K TIC407
321 N G D T L Y I E T P A K F T L N G G N P Y Y T A T F T E Y D E N G N Q V K T - K TIC417
321 N G D T L Y I E T P A K F T L N G G N P Y Y T A T F T E Y D E S G N Q V K T - K TIC431p
N G D T L Y I D T P A K F T F N G A N P Y Y R A T F T E Y D E N G N P V Q T - K Consenso
360 V L S E N F K L 367 TIC901 p
360 I L S G N 364 TIC1201p
361 F L S E N Y K L 368 TIC407p
360 R L - N N - K 364 TIC417p
360 H L S V - - K 364 TIC431p
X L S E N - K L Consenso

Figura 1C

FOLIOS 311 – 313 TARJETERO, FOLIO 314 PUBLICAR. 2006/06/04.

Industria y Comercio

FORMATO DE DIGITALIZACION
RELACION DE ANEXOS

Royal Technologies S.A.

Item		Expediente No 06001259	No de Folio
		No de Unidades	√
1	CD		
2	DVD		
3	REVISTA		
4	LIBRO		
5	PERIODICO		
6	DISQUETES		
7	SOBRE DE MANILA		
8	SOBRE BLANCO TAMAÑO CARTA		
9	SOBRE BLANCO TAMAÑO OFICIO X	1	
10	OTROS		

Observaciones:

215
312

SUPERINDUSTRIA Y COMERCIO
Radicacion : 06001259 00000000 Folios: 314
Fecha (AHD): 2006-01-06 15:32:18
Tramite : 011 PCT-PATENT 379 FASE NACI 411 PRESENTAC
Dependencia: 2020 DIVISION DE NUEVAS CREACIONES

REQUISITOS NECESARIOS PARA PRESENTACION Y AGILIZACION DEL TRAMITE

	USUARIO	RADICADOR
PATENTE DE INVENCIÓN MODELO DE UTILIDAD	( )	( )
- Indicación de que se solicita la concesión de una patente.	( )	( )
- Datos de identificación del solicitante o de la persona que se presenta la solicitud.	( )	( )
- Descripción de la invención.	( )	( )
- Dibujos de ser estos pertinentes.	( )	( )
- Comprobante de Pago de las tasas establecidas.		
COMPLETA ( ) INCOMPLETA	( )	( )
DISEÑO INDUSTRIAL ( )		
- Indicación de que se solicita el registro de Diseño Industrial	( )	( )
- Datos de identificación del solicitante o de la persona que presenta	( )	( )
- la solicitud.		
- Representación gráfica o fotográfica del Diseño Industrial o muestra del Material que Incorpora el diseño.	( )	( )
Comprobante de pago de las tasas establecidas	( )	( )
COMPLETA ( ) INCOMPLETA	( )	( )
ESQUEMA DE TRAZADO ( )		
- Indicación de que solicita el registro de un Esquema de trazado.	( )	( )
- Datos de identificación del solicitante o de la persona que presenta La solicitud.	( )	( )
Representación gráfica del esquema de trazado	( )	( )
Comprobante de pago de las tasas establecidas.	( )	( )
COMPLETA ( ) INCOMPLETA	( )	( )

Firma Responsable:

Fecha:

Industria y Comercio
SUPERINTENDENCIA

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO
DIVISION DE NUEVAS CREACIONES
EXTRACTO PARA PUBLICACIÓN
SOLICITUD DE PATENTE DE INVENCIÓN PCT

(21) N° de solicitud 06-1259	(51) Int Cl. C12N15/32, 1/21, C07K14/325, C12Q1/02, A01N63/02, A01H37/18
22) Fecha de solicitud 2006-01-06	(71)Solicitante(s) MONSANTO TECHNOLOGY, LLC
(30) Prioridad (31) No. Prioridad 32) Fecha (33) País 60/485.483 07/07/2003 US	(72) Inventor(es) JAMES A. BAUM; JUDITH C. DONOVAN; WILLIAM P. DONOVAN, JAMES T. ENGLEMAN; KARINA KRASOMIL-OSTERFELD; JOHN W. PITKIN; JAMES K. ROBERTS
	(74) Apoderado DRA. MARGARITA CASTELLANOS A.
(85) Fecha límite inicio fase nacional 07/02/2006 (86) Datos relativos a la presentación de la solicitud PCT Fecha de presentación de la solicitud 06/07/2004 No de solicitud PCT/US2004/021692	(87) Publicación internacional Fecha 03/03/2005 N° publicación WO 2005/019414

(54) Título **PROTEINAS INSECTICIDAS SECRETADAS DE BACILLUS THURINGIENSIS Y USOS DE LAS MISMAS**

Reivindicación(es)

1. Un polinucleótido aislado que codifica una toxina insecticida de *Bacillus thuringiensis* o un fragmento insecticida de la misma, donde dicho polinucleótido se hibridiza, bajo condiciones seleccionadas del grupo formado por condiciones de hibridación severas y condiciones de hibridación específicas, con una o más de las secuencias de nucleótidos seleccionadas del grupo formado por las SEQ ID N° 2, SEQ ID N° 3 (tic901), SEQ ID N°.5(tic1201), SEQ ID N° 7 (tic407), SEQ ID N°.9 (tic417), SEQ ID N° 11, SEQ ID N° 12, SEQ ID N° 15, SEQ ID N° 16, SEQ ID N° 17, SEQ ID N° 18, SEQ ID N° 19, SEQ ID N° 20, SEQ ID N° 21, SEQ ID N°.22, SEQ ID N° 23, SEQ ID N° 24, SEQ ID N° 25, SEQ ID N° 26, SEQ ID N° 27, SEQ ID N°.28, SEQ ID N°.29, SEQ ID N° 30 y SEQ ID N° 32 o con el complemento de las mismas

6. Un polinucleótido que comprende una secuencia de nucleótidos que codifica una toxina, de 34 a 39 kDa aproximadamente, activa contra plagas de coleópteros, donde dicha secuencia de nucleótidos ha sido optimizada para su expresión en plantas y donde dicha toxina comprende una secuencia de aminoácidos seleccionada del grupo formado por las SEQ ID N° 2, SEQ ID N° 4, SEQ ID N° 6, SEQ ID N° 8, SEQ ID N° 10, SEQ ID N° 31 y SEQ ID N° 33

14. Un método para detectar una primera secuencia de nucleótidos que codifica una proteína insecticida, que comprende hibridizar dicha primera secuencia de nucleótidos con una segunda secuencia de nucleótidos que se selecciona del grupo formado por las SEQ ID N° 2, SEQ ID N° 3 (tic901), SEQ ID N°.5(tic1201), SEQ ID N°.7 (tic407), SEQ ID N° 9 (tic417), SEQ ID N° 11, SEQ ID N° 12, SEQ ID N° 15, SEQ ID N°.16, SEQ ID N° 17, SEQ ID N° 18, SEQ ID N°.19, SEQ ID N° 20, SEQ ID N° 21, SEQ ID N° 22, SEQ ID N° 23, SEQ ID N°.24, SEQ ID N° 25, SEQ ID N° 26, SEQ ID N° 27, SEQ ID N° 28, SEQ ID N° 29, y SEQ ID N°.30 y SEQ ID N°.32, o con el complemento de las mismas bajo condiciones de hibridación severas

15. Una secuencia de nucleótidos que codifica una proteína insecticida o un fragmento insecticida o un homólogo de la misma, donde dicha proteína insecticida o fragmento insecticida u homólogo de la misma se selecciona del grupo de secuencias formado por las SEQ ID N° 1, SEQ ID N° 4, SEQ ID N° 6, SEQ ID N° 8, SEQ ID N° 10, SEQ ID N° 31 y SEQ ID N° 33

16. Un método para detectar una primera secuencia de nucleótidos que codifica una proteína insecticida, que comprende los pasos de
(a) producir un amplicón a partir de un templado que comprende dicha primera secuencia de nucleótidos en una muestra usando cebadores de amplificación que comprenden las secuencias que se muestran en las SEQ ID N° 13 y SEQ ID N° 14,
(b) aislar y purificar dicho amplicón,
(c) aislar y purificar dicha primera secuencia de nucleótidos completa que codifica dicha proteína insecticida basado en la secuencia aislada y purificada del amplicón, y
(d) expresar dicha proteína insecticida en una célula huésped a partir de dicha primera secuencia de nucleótidos,
donde dicha primera secuencia de nucleótidos se hibridiza con una segunda secuencia de nucleótidos seleccionada del grupo formado por las SEQ ID N° 1, SEQ ID N°.3, SEQ ID N° 5, SEQ ID N°.7, SEQ ID N° 9, SEQ ID N°.11, SEQ ID N° 30 y SEQ ID N° 32 o complementos de las mismas, bajo condiciones de hibridación severas
Un método para expresar una proteína insecticida en una célula vegetal que comprende los pasos de
(a) insertar en el genoma de una célula vegetal una secuencia de ácido nucleico que comprende:
(i) un promotor funcional en dicha célula vegetal ligado operativamente a,
(ii) un ADN estructural que codifica una proteína seleccionada del grupo formado por las SEQ ID N° 4, SEQ ID N° 6, SEQ ID N°.8, SEQ ID N° 10, SEQ ID N° 31, SEQ ID N° 33 y proteínas insecticidas que son al menos un 70% similares aproximadamente a dicho grupo de proteínas y
(iii) una secuencia de ADN no traducida 3' cuya función en la célula vegetal causa la terminación de la transcripción y poliadenilación,
(b) obtener dicha célula vegetal transformada para contener la secuencia de ácido nucleico del paso (a), y
(c) regenerar a partir de la célula vegetal transformada una planta genéticamente transformada que expresa una cantidad insecticida de dicha proteína insecticida

375
314

REPUBLICA DE COLOMBIA

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

Bogotá,
2020

Doctor(a)
CASTELLANOS ABONDANO MARGARITA

REFERENCIA	Radicación No.	06 1259
	Solicitud internacional	US2004/021692
	Fecha inicio fase nacional	07/02/2006
	Trámite	11
	Evento	379
	Actuación	430
	Oficio No.	6526

Como resultado del examen de forma, realizado con fundamento en el artículo 38 de la Decisión 486 de la Comisión de la Comunidad Andina, artículo 27 del Tratado de Cooperación en Materia de Patentes (PCT), regla 51 bis del Reglamento y Circular Única de la Superintendencia de Industria y Comercio, se le comunica que:

1. Si fuere el caso, el solicitante debe adjuntar la traducción de los anexos del reporte de examen preliminar internacional (Art. 36.2.b) del tratado PCT.

2. La solicitud no contiene la copia del documento en que consta la cesión del derecho a la patente del inventor al solicitante o a su causante. (Art. 26-k) Decisión 486.

3. La solicitud no contiene el poder debidamente otorgado, con el lleno de los requisitos exigidos por los artículos 63 y subsiguientes del Código de Procedimiento Civil. Adicionalmente, deberá presentar el documento que acredita la existencia y representación legal de la sociedad solicitante, cuando ésta fuere persona jurídica.

En cumplimiento de lo dispuesto por el artículo 39 de la Decisión 486, se le requiere para que dentro del término de dos meses siguientes a la fecha de notificación del presente oficio, complete los requisitos anteriormente enunciados. En caso de no dar cumplimiento al presente requerimiento, la solicitud se considerará abandonada y perderá su prelación.

Notifíquese el presente oficio conforme a lo dispuesto en el numeral 5.2, literal e), del capítulo quinto del título primero de la Circular Unica No.10 de 2001.

ALIX CARMENZA CESPEDES DE VERGEL
Jefe de la división de Nuevas Creaciones

El anterior requerimiento fue notificado por fijación en lista de fecha 16 JUN. 2006
jramos