



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 337 952**

51 Int. Cl.:
C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **07799738 .5**

96 Fecha de presentación : **20.07.2007**

97 Número de publicación de la solicitud: **2046991**

97 Fecha de publicación de la solicitud: **15.04.2009**

54 Título: **Método para identificar genes novedosos.**

30 Prioridad: **21.07.2006 US 832423 P**

45 Fecha de publicación de la mención BOPI:
30.04.2010

45 Fecha de la publicación del folleto de la patente:
30.04.2010

73 Titular/es:
PIONEER HI-BRED INTERNATIONAL Inc.
7100 N.W. 62nd Avenue
Johnston, Iowa 50131-1014, US
E.I. DU PONT DE NEMOURS AND COMPANY

72 Inventor/es: **Abad, Andre R.;**
Dong, Hua;
Lo, Susan B.;
Mccutchen, Billy F. y
Shi, Xiaomei

74 Agente: **Ungría López, Javier**

ES 2 337 952 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Método para identificar genes novedosos.

5 **Campo de la invención**

La presente invención se refiere a métodos y composiciones para identificar genes novedosos que son homólogos a genes conocidos, particularmente genes Cry de *Bacillus thuringiensis* (*Bt*).

10 **Antecedentes de la invención**

Las plagas de insectos son un factor muy importante en la pérdida de los cultivos agrícolas mundiales. Por ejemplo, el daño por la alimentación del gusano del maíz y el daño por el picudo del algodón pueden ser económicamente devastadores para los productores agrícolas. La pérdida de cultivos relacionada con plagas de insectos solamente por el gusano del maíz ha alcanzado mil millones de dólares anuales.

Tradicionalmente, los métodos principales para afectar a las poblaciones de plagas de insectos, tales como poblaciones de gusano del maíz, son rotación de cultivos y la aplicación de pesticidas químicos sintéticos de amplio espectro. Sin embargo, los consumidores y reguladores gubernamentales por igual están siendo cada vez más conscientes de los riesgos ambientales asociados con la producción y el uso de pesticidas químicos sintéticos. Debido a tales preocupaciones, los reguladores han bloqueado o limitado el uso de algunos de los pesticidas más peligrosos. Por tanto, existe un interés sustancial en desarrollar alternativas a pesticidas químicos tradicionales que presenten un menor riesgo de polución y riesgos ambientales y proporcionen una mayor especificidad diana de lo que es característico de los insecticidas químicos de amplio espectro tradicionales.

Se conoce que ciertas especies de microorganismos del género *Bacillus* poseen actividad pesticida contra una gran variedad de plagas de insectos que incluyen *Lepidoptera*, *Diptera*, *Coleoptera*, *Hemiptera* y otros. *Bacillus thuringiensis* (*Bt*) y *Bacillus papilliae* están entre los agentes de biocontrol más exitosos descubiertos hasta la fecha. La patogenia de insectos se ha atribuido a las cepas de: *B. larvae*, *B. lentimorbus*, *B. papilliae*, *B. sphaericus*, *Bt* (Harwood, ed. (1989) *Bacillus* (Plenum Press), pág. 306) y *B. cereus* (Publicación Internacional N° WO 96/10083). La actividad pesticida parece estar concentrada en inclusiones proteicas cristalinas parasporales, aunque también se han aislado proteínas pesticidas del estadio de crecimiento vegetativo de *Bacillus*. Se han aislado y caracterizado varios genes que codifican estas proteínas pesticidas (véase, por ejemplo, las Patentes de Estados Unidos N° 5.366.892 y 5.840.868).

Los pesticidas microbianos, particularmente los obtenidos de cepas de *Bacillus*, han desempeñado un papel importante en la agricultura como alternativas al control químico de plagas. Las proteínas pesticidas aisladas de cepas de *Bt*, conocidas como δ -endotoxinas o toxinas Cry, se producen inicialmente en una forma de protoxina inactiva. Estas protoxinas se convierten proteolíticamente en una toxina activa mediante la acción de proteasas en el intestino del insecto. Véase, Rukmini *et al.* (2000) *Biochimie* 82: 109-116; Oppert (1999) *Arch. Insect Biochem. Phys.* 42: 1-12; y Carroll *et al.* (1997) *J. Invertebrate Pathology* 70: 41-49. La activación proteolítica de la toxina puede incluir la eliminación de péptidos N- y C-terminales de la proteína así como escisión interna de la proteína. Una vez activada, la toxina Cry se une con alta afinidad a receptores sobre células epiteliales en el intestino del insecto, creando de este modo canales permeables en la membrana celular, lisis del intestino del insecto y muerte posterior del insecto por inanición y septicemia. Véase, por ejemplo, Li *et al.* (1991) *Nature*. 353: 815-821.

Recientemente, científicos agrícolas han desarrollado plantas de cultivo con resistencia a insecto mejorada modificando genéticamente plantas de cultivo con genes pesticidas para producir proteínas pesticidas de *Bacillus*. Por ejemplo, las plantas de maíz y algodón modificadas genéticamente para producir toxinas Cry (véase, por ejemplo, Aronson (2002) *Cell Mol. Life Sci.* 59(3): 417-425; Schnepf *et al.* (1998) *Microbiol. Mol. Biol. Rev.* 62(3): 775-806) ahora se usan ampliamente en la agricultura americana y han proporcionado al agricultor una alternativa respetuosa con el medio ambiente a métodos tradicionales de control de insectos. Además, se han desarrollado patatas modificadas genéticamente para contener toxinas Cry pesticidas. Estos éxitos con la ingeniería genética han llevado a los investigadores a buscar genes pesticidas novedosos, particularmente genes Cry. Por lo tanto, en la técnica se necesitan nuevos métodos para identificar de forma eficaz genes pesticidas novedosos, incluyendo los que son homólogos a genes Cry conocidos, así como los que representan familias novedosas de genes Cry.

Sumario de la invención

La presente invención se refiere a métodos y composiciones para identificar genes novedosos. Los métodos descritos en este documento permiten la exploración rápida y eficaz de un gran número de secuencias de nucleótidos para identificar genes novedosos potenciales de una diversidad de organismos. Los presentes métodos para identificar genes novedosos permiten la identificación de genes que son homólogos a genes conocidos así como genes completamente novedosos que pueden ser miembros de familias actualmente no identificadas de genes de interés, incluyendo genes pesticidas. En ciertos aspectos de la invención, los métodos permiten la identificación de genes pesticidas novedosos que son homólogos a genes pesticidas conocidos, incluyendo, por ejemplo, genes de toxina Cry de *Bt*.

Los métodos de la invención comprenden diseñar sistemáticamente cebadores oligonucleotídicos que son específicos para regiones de homología (es decir, secuencias distintivas) dentro de un grupo diana de genes conocidos de interés (por ejemplo, genes pesticidas) y realizar una primera ronda de amplificación por PCR de material de ácido nucleico de un organismo de interés. La primera ronda de PCR tiene por objeto amplificar genes tanto conocidos como novedosos que contengan la secuencia distintiva. Si se detectan productos de PCR en la primera ronda de PCR, se obtiene una segunda muestra de material de ácido nucleico del organismo y se somete a una segunda ronda de PCR usando un segundo conjunto de cebadores oligonucleotídicos que son específicos para secuencias distintivas en el grupo diana de genes. Los productos de PCR de la segunda ronda de PCR se separan por electroforesis en gel de agarosa y los ácidos nucleicos aislados resultantes se clonan en vectores de clonación, particularmente vectores de clonación bacterianos. Después, los vectores de clonación se introducen por transformación en células hospedadoras competentes tales como células bacterianas. El material de ácido nucleico aislado de colonias de células hospedadoras individuales se analiza por análisis de hibridación de transferencia puntual usando sondas oligonucleotídicas marcadas que son específicas para todos los genes conocidos en el grupo diana. La etapa de análisis por transferencia puntual del método de la invención tiene por objeto identificar y eliminar genes conocidos del grupo diana de una consideración adicional. Los productos de PCR amplificados en la segunda ronda de PCR que no se detectan por análisis de transferencia puntual comprenden supuestos genes novedosos (por ejemplo, genes pesticidas novedosos) o fragmentos de los mismos. Estos ácidos nucleicos se someten a análisis de secuencia adicional para confirmar la novedad y para determinar secuencias de nucleótidos. Los supuestos genes novedosos se expresan y las proteínas recombinantes se ensayan para evaluar la actividad biológica, tal como actividad pesticida, cuando se usan los métodos de la invención para identificar genes pesticidas novedosos. Los métodos de la invención son susceptibles de forma adicional a automatización y exploración de alto rendimiento.

Las composiciones de la invención incluyen polinucleótidos aislados novedosos y variantes y fragmentos de los mismos, que comprenden genes novedosos, que incluyen, por ejemplo, genes pesticidas novedosos. También se proporcionan polipéptidos codificados por los polinucleótidos de la invención. Los genes pesticidas novedosos (por ejemplo, genes de toxina Cry de *Bt*) identificados por los métodos descritos en este documento se usan en la protección de plantas contra plagas, particularmente plagas de insectos y daños relacionados con plagas.

Breve descripción de los dibujos

La Figura 1 proporciona una representación esquemática del diseño de los cebadores oligonucleotídicos para el uso en la primera y segunda ronda de PCR y sondas oligonucleotídicas para análisis de transferencia puntual, como se describe con detalle más adelante en este documento.

Descripción detallada de la invención

La presente invención se refiere a métodos y composiciones para identificar genes novedosos, particularmente genes pesticidas novedosos, más particularmente genes de toxina Cry de *Bt* novedosos. Los métodos de la invención permiten la exploración rápida y eficaz de un gran número de secuencias de nucleótidos para identificar supuestos genes novedosos que son homólogos a genes conocidos. Como se usa en este documento, la expresión “grupo diana de genes” se refiere a cualquier colección de genes conocidos de cualquier organismo de interés que comprende regiones de homología. Un “grupo diana” puede comprender en algunas realizaciones una colección de genes pesticidas conocidos, más particularmente un grupo de genes de toxina Cry de *Bt* conocidos. En general, los métodos de la invención comprenden tres etapas distintas para identificar genes novedosos de un grupo diana: una primera ronda de PCR, más particularmente PCR en tiempo real, una segunda ronda de PCR y una etapa de análisis de transferencia puntual. En realizaciones particulares, se realiza una primera ronda de amplificación por PCR de material de ácido nucleico de un organismo de interés y se pretende amplificar genes tanto conocidos como novedosos que comprenden una secuencia distintiva dirigida. Se pretende que una “secuencia distintiva” signifique una región de homología que está presente dentro de todos los miembros del grupo diana de genes de interés. Si se detectan productos de PCR en la primera ronda de PCR, se obtiene una segunda ronda de PCR de una segunda muestra de material de ácido nucleico del organismo y se somete a una ronda adicional de amplificación por PCR. La segunda ronda de PCR tiene por objeto amplificar genes tanto conocidos como novedosos que contienen secuencias distintivas dirigidas particulares. Los productos de PCR de la segunda ronda generalmente se aíslan para análisis adicional. La tercera etapa comprende realizar análisis de transferencia puntual de los productos de PCR individuales aislados en la segunda ronda de PCR. La etapa de análisis de transferencia puntual se realiza con sondas oligonucleotídicas que son específicas para genes conocidos en el grupo diana y, por lo tanto, se pretende detectar y eliminar genes conocidos de una consideración adicional. Los productos de PCR de la segunda ronda de PCR que no se detectan mediante análisis de transferencia puntual comprenden supuestos genes novedosos (por ejemplo, genes pesticidas novedosos) o fragmentos de los mismos y se someten a análisis de secuencia adicional para confirmar la novedad. Las secuencias de supuestos genes novedosos se determinan y estas moléculas de ácido nucleico y las proteínas codificadas por las mismas se usan en bioensayos para evaluar la actividad biológica, tal como, por ejemplo, actividad pesticida.

Más particularmente, los métodos para identificar genes novedosos comprenden diseñar sistemáticamente cebadores oligonucleotídicos para regiones de homología (es decir, secuencias distintivas) dentro del grupo diana de genes conocidos, tal como un grupo diana de genes pesticidas de *Bt* y usar estos cebadores en una primera ronda de amplificación por PCR de material de ácido nucleico de un organismo de interés. En algunos aspectos de la invención, el organismo de interés es un microorganismo, más particularmente, una cepa de *Bt*. Los cebadores diseñados para la primera ronda de amplificación por PCR tienen por objeto amplificar genes tanto conocidos como novedosos que

contengan las secuencias distintivas dirigidas, como se describe más adelante con más detalle. Si se detectan productos de PCR en la primera ronda de PCR, se obtiene una segunda muestra de material de ácido nucleico del organismo y se somete a una segunda ronda de amplificación por PCR.

Los cebadores oligonucleotídicos usados en la segunda ronda de PCR también se diseñan para amplificar genes tanto conocidos como novedosos (por ejemplo, genes pesticidas) que contienen secuencias distintivas dirigidas. Los cebadores oligonucleotídicos usados en la segunda ronda de PCR se diseñan generalmente para generar productos de PCR de una longitud particular (por ejemplo, de aproximadamente 500 pares de bases (pb) a aproximadamente 800 pb de longitud, particularmente de aproximadamente 600 pb a aproximadamente 750 pb, más particularmente de aproximadamente 650 pb a aproximadamente 700 pb). Los productos de PCR de la longitud esperada que se generan durante la segunda ronda de PCR se aíslan, por ejemplo, mediante electroforesis en gel de agarosa. Por lo tanto, la segunda ronda de PCR permite la amplificación de genes conocidos y novedosos, o fragmentos de los mismos, que contienen secuencias distintivas particulares y permite el aislamiento de estas moléculas de ácido nucleico para análisis adicional.

Los productos de PCR de la segunda ronda de amplificación que tienen la longitud esperada comprenden generalmente fragmentos de genes conocidos o novedosos. Estos fragmentos de ácidos nucleicos se clonan en vectores de clonación (por ejemplo, vectores de clonación bacterianos). Los insertos de vector de clonación (es decir, los productos de PCR de la segunda ronda de PCR) comprenden genes conocidos y potencialmente novedosos del grupo diana, o probables fragmentos de los mismos, y se usan para transformar células hospedadoras competentes, particularmente células bacterianas tales como, por ejemplo, células de *E. coli*. En aspectos particulares de la invención, se aísla material de ácido nucleico (por ejemplo, ADN plasmídico) de colonias de células hospedadoras (por ejemplo, bacterianas) individuales y se analizan por análisis de transferencia puntual usando sondas oligonucleotídicas marcadas específicas para todos los genes conocidos dentro del grupo diana. En realizaciones particulares, las sondas oligonucleotídicas se diseñan para ser complementarias a fragmentos de los productos de PCR generados durante la segunda ronda de amplificación, como se describe más adelante en este documento. El análisis de transferencia puntual con sondas oligonucleotídicas que son específicas para todos los genes conocidos dentro del grupo diana permite la identificación de moléculas de ácido nucleico que comprenden genes novedosos (por ejemplo, genes de toxina Cry de *Bt* conocidos de un grupo diana particular). Los ácidos nucleicos que contienen genes conocidos se eliminan de una consideración adicional. Los ácidos nucleicos que no se detectan mediante análisis de transferencia puntual comprenden supuestos genes novedosos, o fragmentos de los mismos, y se someten a análisis de secuencia adicional y ensayos de actividad biológica. En realizaciones particulares de la invención, los métodos se usan para identificar genes pesticidas novedosos, particularmente genes de toxina Cry de *Bt* novedosos y, por consiguiente, los supuestos genes pesticidas novedosos se analizan adicionalmente para actividad pesticida.

En ciertos aspectos de la invención, los productos de PCR generados en la segunda ronda de PCR que no se detectan mediante análisis de transferencia puntual, como se ha descrito anteriormente, se secuencian y comparan con secuencias conocidas de bases de datos públicas para evaluar la novedad. Si las comparaciones de secuencias indican que el producto de PCR contiene un gen potencialmente novedoso, tal como un gen pesticida novedoso (por ejemplo, un gen de toxina Cry de *Bt* novedoso), se obtiene la secuencia de longitud completa usando, por ejemplo, el Kit Universal GenomeWalker (Becton Dickinson Bioscience, Inc.). La secuencia resultante también se compara frente a secuencias en bases de datos públicas para verificar adicionalmente la novedad. En realizaciones particulares, los genes novedosos se clonan en vectores de expresión y las proteínas codificadas de este modo se ensayan para actividad biológica, tal como actividad pesticida en el caso de supuestos genes pesticidas novedosos.

El método descrito en la reivindicación 1 se refiere a identificar genes pesticidas, más particularmente genes pesticidas de toxina Cry de *Bt*. Aunque los métodos de la invención se describen en este documento más adelante para la identificación de genes pesticidas, tales métodos se pueden usar para identificar genes novedosos que son homólogos a cualquier grupo de genes conocidos (es decir, un grupo diana de interés) de cualquier organismo de interés. La descripción de la identificación de genes pesticidas novedosos tiene por objeto ser únicamente ilustrativa y no es limitante.

Los métodos de la invención se pueden usar para identificar genes novedosos que son homólogos a genes Cry conocidos mientras que también identifican genes que comparten poca homología con genes Cry identificados previamente y que en realidad pueden representar familias novedosas de genes pesticidas de *Bt*. En una realización de la invención, el material de ácido nucleico aislado de cepas *Bt* de interés se somete a una primera ronda de PCR, generalmente, PCR en tiempo real, usando al menos un conjunto de cebadores oligonucleotídicos degenerados que son específicos para una región de homología (es decir, una secuencia distintiva) presente en todos los miembros de un grupo diana de genes pesticidas. Como se usa en este documento, “grupo diana de genes pesticidas” se refiere a cualquier colección de genes pesticidas conocidos que comprenden regiones de homología. Los miembros del grupo diana de genes pesticidas se alinean para diseñar cebadores oligonucleotídicos. Como se ha indicado anteriormente, una región de homología que está presente dentro de todos los miembros del grupo diana de genes pesticidas se denomina una “secuencia distintiva”. Las secuencias distintivas dentro de las secuencias de nucleótidos del grupo diana de genes pesticidas sirven como la base para diseñar cebadores oligonucleotídicos para el uso en la primera y segunda ronda de PCR, como se describe con más detalle más adelante. En ciertos aspectos de la invención, el grupo diana comprende todos los genes Cry pesticidas conocidos que son activos contra insectos del orden Coleoptera (es decir, genes Cry activos en Coleópteros). En otras realizaciones, el grupo diana comprende, por ejemplo, todos los genes de *Bt* que tienen actividad pesticida contra insectos de los órdenes Lepidoptera y Coleoptera, excluyendo los genes Cry

que son activos contra insectos del orden Diptera. El grupo diana de genes pesticidas se selecciona y define por el investigador al comienzo de la investigación para genes pesticidas novedosos.

Los cebadores oligonucleotídicos específicos para el grupo diana de genes pesticidas se mezclan con una primera muestra de material de ácido nucleico de un microorganismo de interés y una ADN polimerasa en condiciones que son adecuadas para la amplificación por PCR. Los métodos de la presente invención comprenden además realizar una primera ronda de PCR y detectar la presencia o ausencia de productos de amplificación por PCR. En realizaciones particulares, la primera ronda de PCR comprende realizar PCR en tiempo real cuantitativa usando un colorante verde SYBR[®] para detectar la presencia de productos de PCR.

Si se detectan productos de PCR en la primera ronda de PCR, se obtiene una segunda muestra de ácido nucleico del microorganismo de interés y se somete a una segunda ronda de PCR. Los cebadores oligonucleotídicos usados en la segunda ronda de PCR también son específicos para secuencias distintivas dentro de las secuencias de nucleótidos del grupo diana de genes pesticidas, como se ha descrito anteriormente. En general, los cebadores oligonucleotídicos inversos usados en la primera ronda de PCR se usan para generar los cebadores directos para la segunda ronda de PCR, sirviendo de este modo como un puente entre la primera y la segunda ronda de PCR. Los cebadores inversos usados en la segunda ronda de PCR se diseñan para dirigir una secuencia distintiva diferente que se localiza típicamente 3' con respecto a la secuencia distintiva usada para diseñar los cebadores inversos para la primera ronda de PCR. Los cebadores oligonucleotídicos para la segunda ronda de PCR se diseñan adicionalmente para producir productos de PCR de una longitud particular, específicamente de aproximadamente 500 pb a aproximadamente 800 pb, particularmente de aproximadamente 600 pb a aproximadamente 750 pb, más particularmente de aproximadamente 650 pb a aproximadamente 700 pb. Las reacciones de PCR de la segunda ronda de amplificación por PCR se pueden separar mediante electroforesis en gel de agarosa y los productos de PCR resultantes que comprenden los fragmentos de ácido nucleico de la longitud esperada se ligan en vectores de clonación, particularmente en vectores de clonación bacterianos. Después, los vectores se introducen por transformación en células hospedadoras competentes, por ejemplo, células bacterianas tales como células de *E. coli*.

Los ejemplos de células hospedadoras adecuadas incluyen, pero sin limitación, células bacterianas, células fúngicas, células vegetales (dicotiledóneas y monocotiledóneas) y células animales. En realizaciones particulares, las células hospedadoras son células bacterianas. En la técnica se conocen bien los vectores de clonación para el suministro de polinucleótidos a una diversidad de células hospedadoras. También se conocen bien en la técnica los métodos para clonar moléculas de ácido nucleico en vectores y para transformar células hospedadoras. Para descripciones generales de sistemas y métodos de clonación, empaquetamiento y expresión véase Gilman y Smith (1979) *Gene* 8: 81-97; Roberts *et al.* (1987) *Nature* 328: 731-734; Berger y Kimmel (1989) *Guide to Molecular Cloning Techniques, Methods in Enzymology*, Vol. 152 (Academic Press, Inc., San Diego, California); Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, Vols. 1-3 (2a ed; Cold Spring Harbor Laboratory Press, Plainview, New York); y Ausubel *et al.*, eds. (1994) *Current Protocols in Molecular Biology*. Current Protocols (Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., New York; 1994 Suplemento.

El material de ácido nucleico, por ejemplo, preparaciones de plásmidos, de colonias bacterianas individuales que comprenden los productos de PCR de la segunda ronda de PCR se somete a análisis adicional para identificar supuestos genes pesticidas novedosos. En una realización particular, el ADN plasmídico de las colonias individuales se analiza mediante análisis de transferencia puntual usando sondas oligonucleotídicas marcadas que se diseñan para detectar todos los genes pesticidas conocidos dentro del grupo diana. Las sondas oligonucleotídicas se diseñan típicamente para ser complementarias a un fragmento de los productos de PCR generados durante la segunda ronda de amplificación por PCR. Las sondas se diseñan de tal forma que se identificará cualquier ácido nucleico que contenga un gen pesticida conocido dentro del grupo diana (es decir, "positivos a transferencia puntual"). Cualquier ácido nucleico que no se detecta mediante estas sondas usando análisis de transferencia puntual (es decir, "negativos a transferencia puntual") contienen supuestos genes pesticidas novedosos o probablemente fragmentos de los mismos y se analiza adicionalmente para evaluar la novedad. Los fragmentos de los supuestos genes pesticidas novedosos identificados de acuerdo con los presentes métodos se secuencian y someten a comparación de secuencia con genes pesticidas conocidos para evaluar la novedad. Tales análisis de secuencia se conocen bien en la técnica. Las secuencias de nucleótidos novedosas se analizan adicionalmente para obtener supuestos genes pesticidas. En algunas realizaciones, las moléculas de ácido nucleico que comprenden supuestos genes pesticidas novedosos se clonan en vectores de expresión y los polipéptidos codificados por estos genes se ensayan para actividad pesticida usando ensayos convencionales, tales como los que se describen más adelante en este documento.

Los anteriores métodos descritos para la identificación de genes pesticidas novedosos también se pueden usar para identificar genes novedosos de otros grupos diana de interés. Cuando los métodos de la invención se usan para identificar genes no pesticidas, particularmente genes de toxina Cry no de *Bt*, el material de partida de ácido nucleico se puede obtener de un organismo de interés diferente. Sin embargo, las demás etapas del método, concretamente el diseño de cebador sistemático (descrito más adelante en este documento), la primera ronda de PCR, la segunda ronda de PCR y el análisis de transferencia puntual se realizan esencialmente de la misma manera, sin tener en cuenta el grupo diana de genes de interés.

Aunque no se desea limitarse a ningún mecanismo, los cebadores oligonucleotídicos usados en la primera y segunda ronda de amplificación por PCR se diseñan para y probablemente permitir la amplificación de genes tanto conocidos como novedosos que contienen secuencias distintivas, como se ha definido anteriormente en este documento. Por el

contrario, las sondas oligonucleotídicas usadas en la tercera etapa de la invención, típicamente análisis de transferencia puntual, se seleccionan para detectar específicamente solamente genes conocidos. Por tanto, los organismos, tales como microorganismos, particularmente células de *Bt*, que comprenden material de ácido nucleico que se amplifica durante la primera y segunda ronda de PCR pero que no se detecta durante la etapa de análisis de transferencia puntual pueden comprender un gen novedoso.

En aspectos particulares de la invención, el diseño de al menos un par de cebadores oligonucleotídicos para el uso en la primera ronda de amplificación por PCR que se ha descrito en este documento anteriormente comprende diseñar cebadores oligonucleotídicos degenerados por un proceso multi-etapa. En ciertas realizaciones, se prepara un alineamiento de secuencias de nucleótidos para un grupo diana de genes. Por ejemplo, un grupo diana de genes pesticidas puede comprender todos los genes Cry conocidos que tengan actividad pesticida contra insectos del orden Coleópteros (es decir, genes activos en Coleópteros). Los genes dentro de un grupo diana comparten bloques de homología, denominados en este documento secuencias distintivas. Las secuencias distintivas sirven como el punto de partida para el diseño de cebador oligonucleotídico, como se describe con detalle más adelante. Aunque una secuencia distintiva es un bloque de nucleótidos que está conservado en todos los miembros del grupo diana de genes, puede haber cierta divergencia en la secuencia distintiva de gen a gen dentro del grupo diana. Como resultado, puede no ser posible diseñar un único conjunto de cebadores oligonucleotídicos que será específico para todos los genes en el grupo diana. Por lo tanto, se puede usar una mezcla de cebadores oligonucleotídicos para abarcar todas las posibles variaciones de la secuencia distintiva que aparece en el grupo diana. La utilización de una mezcla de cebadores oligonucleotídicos se usa particularmente cuando, debido a variaciones de secuencia de la secuencia distintiva dentro del grupo diana de genes, es difícil o imposible desarrollar un conjunto de cebadores que sea específico para una secuencia distintiva dentro de todo el grupo diana. Cuando es posible, se diseña y usa un único conjunto de cebadores que es específico para tantos genes dentro del grupo diana como sea posible. En ciertos aspectos de la invención para la identificación de genes novedosos de toxina Cry de *Bt*, los cebadores oligonucleotídicos usados en la primera ronda de PCR se diseñan para dirigir secuencias distintivas en el “dominio 1” de un grupo diana de genes Cry de *Bt* conocidos y los usados en la segunda ronda de PCR son específicos para secuencias en el “dominio 2”.

El diseño de cebadores oligonucleotídicos degenerados para el uso en la primera y segunda ronda de PCR implica la exploración de los alineamientos de secuencias de nucleótidos del grupo diana de los genes para identificar varias regiones de homología que son puntos de partida apropiados para el diseño de cebador descrito más adelante. Estas regiones de homología se denominan “secuencias distintivas”. Se selecciona una longitud de cebador inicial, donde la longitud de cebador inicial está entre aproximadamente 15 pares de bases (pb) y aproximadamente 30 pb, por ejemplo, 15 pb, 16 pb, 17 pb, 18 pb, 19 pb, 20 pb, 21 pb, 22 pb, 23 pb, 24 pb, 25 pb, 26 pb, 27 pb, 28 pb, 29 pb y 30 pb. Después se realiza una primera ronda de exploración para un cebador oligonucleotídico visualizando una ventana inicial de nucleótidos contiguos dentro de una de las secuencias distintivas. La ventana inicial comienza en el extremo 5' de la secuencia distintiva seleccionada y es equivalente en longitud a la longitud de cebador inicial. La secuencia de nucleótidos dentro de la ventana inicial se revisa para determinar si posee las siguientes características requeridas de secuencia. Una secuencia de nucleótidos apropiada para un cebador para el uso en la primera o segunda ronda de PCR:

- 1) no tiene cuatro o más restos de nucleótidos idénticos contiguos;
- 2) tiene no más de dos restos de guanina o citosina dentro de los últimos cinco restos del extremo 3' de la secuencia de nucleótidos;
- 3) tiene una temperatura de fusión entre aproximadamente 50°C y 65°C, más particularmente aproximadamente 54°C +/- 2°C;
- 4) no forma estructuras de horquilla o dímero;
- 5) está presente en al menos una de las secuencias de nucleótidos del grupo de genes diana (es decir, el alineamiento que se ha descrito anteriormente); y
- 6) no está conservada entre secuencias de nucleótidos de genes de grupo no diana.

Para aumentar la diversidad de los cebadores oligonucleotídicos, se permite que un resto de nucleótido sea n, donde n es A, T, C o G. Una secuencia de nucleótidos en la ventana inicial se selecciona para el uso como un cebador oligonucleotídico si están presentes todas las anteriores características de secuencia. Si la secuencia de nucleótidos dentro de la ventana inicial no posee todas estas características de secuencia, se selecciona una ventana adyacente de nucleótidos contiguos moviendo la ventana inicial un par de bases hacia el extremo 3' de la secuencia distintiva. La secuencia de nucleótidos dentro de la ventana adyacente se revisa como se ha descrito anteriormente y se selecciona para el uso como un cebador oligonucleotídico si están presentes todas las características de secuencia. Se realizan rondas adicionales de exploración si es necesario para identificar una secuencia de nucleótidos que satisfaga los anteriores requisitos. Se selecciona un oligonucleótido dentro de la secuencia distintiva que tenga las características requeridas y se usa como un cebador oligonucleotídico en la primera o segunda ronda de PCR. Se diseñan cebadores tanto directos como inversos como se ha descrito anteriormente. Además, los cebadores directo e inverso usados en la primera ronda de PCR se diseñan de tal forma que son complementarios a secuencias de nucleótidos dentro de los genes en el grupo diana que están separados de aproximadamente 50 pb a aproximadamente 150 pb. Los cebadores directo e inverso de la segunda ronda de amplificación por PCR se diseñan generalmente de tal forma que son complementarios

a secuencias de nucleótidos dentro de los genes del grupo diana que están separados de aproximadamente 500 pb a aproximadamente 800 pb.

Como se usa anteriormente en este documento, una secuencia de nucleótidos está “presente” en al menos una de las secuencias de nucleótidos del grupo diana de genes si se encuentra la secuencia de nucleótidos idéntica en la secuencia de nucleótidos de al menos un miembro del grupo diana, con la advertencia de que se permite que un resto de nucleótido sea cualquier nucleótido (es decir, $n = A, T, C \text{ o } G$). La expresión “grupo no diana de genes pesticidas” se refiere a todos los genes pesticidas dentro de una familia particular de genes pesticidas, excluyendo los genes pesticidas que se han seleccionado como el grupo diana. Por ejemplo, si el grupo diana comprende todos los genes Cry de *Bt* que son activos en Coleópteros, el correspondiente grupo no diana de genes pesticidas comprende todos los genes de *Bt* excepto los que son activos contra insectos del orden Coleoptera. De forma similar, un “gen no diana” o “grupo no diana de genes” se refiere a todos los genes dentro de una familia particular de genes pesticidas, excluyendo los genes que se han seleccionado como el grupo diana. Una secuencia de nucleótidos “no está conservada entre secuencias nucleótidos de grupos no diana de genes” si difiere de todas las secuencias de nucleótidos dentro del grupo no diana en al menos dos restos de nucleótidos. En ciertos aspectos de la invención, la determinación de si una secuencia de nucleótidos dentro de una ventana particular de nucleótidos contiguos no está conservada entre genes de grupo no diana comprende buscar la secuencia de longitud completa de cada gen del grupo no diana de genes. En algunas realizaciones, la secuencia de longitud completa de cada gen del grupo no diana de genes se busca de forma exhaustiva usando la secuencia de nucleótidos dentro de la ventana como un término de búsqueda de cadena. Es decir, si una secuencia de nucleótidos dentro de una ventana aparece en cualquier lugar en un gen de grupo no diana o si una secuencia de nucleótidos con menos de 2 diferencias de resto de nucleótido aparece en cualquier lugar en un gen de grupo no diana, entonces esa secuencia de nucleótidos particular dentro de la ventana no se seleccionará como un cebador oligonucleotídico.

Como se ha indicado anteriormente, los cebadores inversos usados en la primera ronda de PCR se usan típicamente para generar los cebadores directos para la segunda ronda de PCR. Los cebadores inversos para la segunda ronda de PCR se diseñan de acuerdo con los métodos que se han descrito anteriormente usando una secuencia distintiva diferente como el punto de partida para el diseño de cebador, específicamente uno que esté 3' con respecto a la secuencia distintiva usada para diseñar los cebadores oligonucleotídicos para la primera ronda de PCR. En la Figura 1 se presenta un esquema de diseño de cebador ilustrativo para la primera y segunda ronda de PCR.

Debido a que las secuencias distintivas dentro del grupo diana de genes típicamente no serán idénticas entre todos los miembros, generalmente se usará una mezcla de cebadores oligonucleotídicos tanto en la primera como en la segunda ronda de PCR para explicar estas variaciones de secuencia. Cuando se usan mezclas de cebadores oligonucleotídicos en las reacciones de PCR de la invención, los cebadores se diseñarán adicionalmente de tal forma que todos los cebadores tendrán temperaturas de fusión idénticas o prácticamente idénticas. En algunas realizaciones, la temperatura de fusión para cebadores oligonucleotídicos usados en la primera y segunda ronda de PCR será aproximadamente $54^{\circ}\text{C} \pm 2^{\circ}\text{C}$.

Un “gen pesticida” se refiere a una secuencia de nucleótidos que codifica un polipéptido que muestra actividad pesticida. Como se usa en este documento, la expresión “actividad pesticida” se refiere a la capacidad de una sustancia, tal como un polipéptido, de inhibir el desarrollo, la alimentación o la reproducción de una plaga de insecto y/o para destruir la plaga de insecto. Se desea que un “polipéptido pesticida” o una “toxina de insecto” signifique una proteína que tenga actividad pesticida. La actividad pesticida se puede medir mediante ensayos rutinarios conocidos en la técnica. Tales ensayos incluyen, pero sin limitación, mortalidad de plaga, pérdida de peso de plaga, repulsión de plaga, atracción de plaga y otros cambios conductuales y físicos de una plaga después de la alimentación y exposición a la sustancia durante un período de tiempo apropiado. Los procedimientos generales incluyen la adición del compuesto experimental u organismo a la fuente de alimentación en un recipiente cerrado. En la técnica se conocen bien los ensayos para evaluar la actividad pesticida. Véase, por ejemplo, las Patentes de Estados Unidos N° 6.570.005 y 6.339.144.

El estadio de desarrollo preferido para ensayar actividad pesticida son las larvas o las formas inmaduras de un insecto de interés. Los insectos se pueden criar en oscuridad total, de aproximadamente 20°C a aproximadamente 30°C y con una humedad relativa de aproximadamente el 30% a aproximadamente el 70%. Se pueden realizar bioensayos como se describe en Czaplá y Lang (1990) J. Econ. Entomol. 83 (6): 2480-2485. Los métodos para criar larvas de insectos y realizar bioensayos se conocen bien por el experto en la materia.

En algunas realizaciones de la invención, el grupo diana de interés son genes pesticidas que comprenden genes de toxina Cry de *Bt* o un subconjunto específico de genes de *Bt*, tales como, por ejemplo, genes Cry de *Bt* activos en Coleópteros. Se desea que un gen de “*Bt*” o “*Bacillus thuringiensis*” signifique la clase más amplia de genes encontrados en diversas cepas de *Bt* que codifican toxinas de *Bt*, que incluyen toxinas tales como, por ejemplo, toxinas Cry (cristal) (es decir, δ -endotoxinas) y toxinas Cyt (citotóxicas). La “toxina Cry” y la “toxina Cyt” incluyen polipéptidos pesticidas que son homólogos a proteínas Cry o Cyt conocidas, respectivamente. Los genes Cry incluyen secuencias de nucleótidos que codifican cualquier polipéptido clasificado como una toxina Cry, por ejemplo, Cry1, Cry2, Cry3, Cry7, Cry8 y Cry9. Véase, Crickmore *et al.* (1998) Microbiol. Molec. Biol. Rev. 62: 807-813 y Crickmore *et al.* (2004) *Bacillus Thuringiensis Toxin Nomenclature* en lifesci.sussex.ac.uk/Home/Neil_Crickmore/B.thuringiensis. Las toxinas de *Bt* son una familia de proteínas pesticidas que se sintetizan como protoxinas y cristalizan como inclusiones parasporales. Cuando se ingiere por una plaga de insecto, la estructura microcristalina se disuelve por el pH alcalino

del intestino medio del insecto y la protoxina se escinde por proteasas del intestino del insecto para generar la toxina activa. La toxina de *Bt* activada se une a receptores en el epitelio del intestino del insecto, provocando lesiones de membrana e hinchazón asociada y lisis del intestino del insecto. La muerte del insecto se produce por inanición y septicemia. Véase, por ejemplo, Li *et al.* (1991) *Nature* 353: 815-821.

La forma de protoxina de las toxinas Cry contiene un segmento formador cristalino. Una comparación de las secuencias de aminoácidos de toxinas Cry activas de diferentes especificidades pone de manifiesto adicionalmente cinco bloques de secuencia altamente conservados. Estructuralmente, las toxinas Cry comprende tres dominios distintos que son, del extremo N al C: un agrupamiento de siete hélices alfa implicadas en la formación de poro (denominadas “dominio 1”), tres láminas beta anti-paralelas implicadas en la unión celular (denominadas “dominio 2”) y un sándwich beta (denominado “dominio 3”). La localización y propiedades de estos dominios se conocen por los expertos en la materia. Véase, por ejemplo, Li *et al.* (1991), anteriormente, y Morse *et al.* (2001) *Structure* 9: 409-417.

El sistema de nomenclatura de toxina de *Bt* original clasificó las toxinas basándose en los perfiles de actividad pesticida. Este sistema se ha sustituido con una nueva nomenclatura que se basa solamente en la identidad de secuencia de aminoácidos. En este sistema, las toxinas Cry y Cyt se han agrupado en clases o familias basándose en la identidad de secuencia de aminoácidos y el nombre de la toxina proporciona información con respecto a su homología con otras secuencias. Por tanto, por ejemplo, las toxinas Cry2Aa, Cry2Ab y Cry2Ac, que son miembros de la familia Cry2, comparten una identidad de secuencia de aminoácidos de aproximadamente el 80%. De forma similar, las toxinas de la familia Cry8 Cry8Aa y Cry8Ba comparten una identidad de secuencia de aminoácidos de aproximadamente el 65%. Véase Crickmore *et al.* (1998), anteriormente.

Los cebadores oligonucleotídicos específicos para secuencias distintivas dentro de un grupo diana de interés, tal como un grupo diana de genes pesticidas, usados tanto en la primera como en la segunda ronda de PCR y diseñados de acuerdo con los métodos de este documento, se diseñan de forma general para tener un punto de fusión térmico (T_m) o temperatura entre aproximadamente 50°C y 65°C. En realizaciones particulares, los cebadores oligonucleotídicos tienen una T_m entre aproximadamente 52°C y 56°C, más particularmente de aproximadamente 54°C. Se han utilizado varias fórmulas para determinar la T_m . Se puede usar cualquier fórmula para calcular la T_m para practicar los presentes métodos. Por ejemplo, un algoritmo clásico para la determinación de T_m basado en la termodinámica de vecino más cercano es el siguiente:

$$T_m = EH^\circ / (ES^\circ + (R \times \ln(Ct))) - 273,15 + 16,6 \log [X]$$

donde EH° y ES° son la entalpía y entropía para la formación de hélice, respectivamente; R es la constante de gas molar (1,987 (cal) (K^{-1}) (mol^{-1})); Ct es la concentración (de cebador) de cadena total; y X es la concentración salina. Rychlik *et al.* (1990) *Nucleic Acid Res.* 18 (21): 6409-6412. Además, en algunas realizaciones, la T_m de un cebador oligonucleotídico se calcula usando la siguiente fórmula:

$$T_m = (EH^\circ / [ES^\circ + (R \times \ln(Ct))]) - 273,15 + 16,6 \log ([X]) \times 1,144 - 14,964$$

donde EH° (entalpía) = $\sum \Delta H$, ES° (entropía) = $\sum \Delta S + 0,368 \times 19 \times 1,585$; R (constante de gas molar) = 1,987; Ct (concentración de cebador total) = $\log(0,0000005/4) \times 1000$; y X (concentración de sal [K^+]) = 0,05.

Un experto en la materia reconocerá que los cebadores oligonucleotídicos usados para practicar los métodos de la invención son cebadores oligonucleotídicos emparejados de tal forma que hay dos cebadores individuales por par (es decir, un cebador directo y un cebador inverso). Uno de los cebadores en cada par es complementario (es decir, capaz de hibridar) con una porción de la cadena 5' de una secuencia distintiva del grupo diana de genes (cebador directo), mientras que el otro es complementario a una porción de la cadena 3' de una secuencia distintiva (cebador inverso). Los cebadores oligonucleotídicos se diseñan de tal forma que una polimerasa adecuada copiará la secuencia de cada cadena 3' a cada cebador para producir copias amplificadas (es decir, el “producto de amplificación por PCR” o “producto de PCR”). Los presentes métodos utilizan al menos un par de cebadores oligonucleotídicos para la amplificación por PCR. En ciertos aspectos de la invención se usa una mezcla de pares de cebadores oligonucleotídicos que comprenden 2, 3, 4, 5, 10, 20, 30, 40, 50 o más pares de cebadores. En la técnica se conocen bien los métodos para diseñar cebadores oligonucleotídicos, incluyendo cebadores oligonucleotídicos degenerados, específicos para secuencias de nucleótidos particulares de interés (por ejemplo, secuencias distintivas).

Los cebadores oligonucleotídicos de la presente invención tendrán una longitud adecuada para permitir la amplificación de genes novedosos, tales como genes pesticidas novedosos. Los cebadores individuales de cada par comprenderán típicamente entre aproximadamente 15 pb y aproximadamente 30 pb, más particularmente entre aproximadamente 20 pb y aproximadamente 25 pb. La distancia entre los cebadores individuales en un par de cebadores oligonucleotídicos también será suficiente para producir productos de PCR de una longitud detectable. Por tanto, en la primera ronda de PCR, los cebadores directo e inverso se seleccionan de tal forma que son complementarios a secuencias de nucleótidos dentro de las secuencias de nucleótidos para miembros del grupo diana de genes que típicamente están separados entre aproximadamente 50 pb a aproximadamente 150 pb, más particularmente separados aproximadamente 100 pb. En la segunda ronda de PCR, los cebadores directo e inverso generalmente serán complementarios a secuencias de nucleótidos dentro del grupo diana que están separados entre aproximadamente 500 pb a

aproximadamente 800 pb, particularmente separados de aproximadamente 600 pb a aproximadamente 750 pb, más particularmente separados de aproximadamente 600 a aproximadamente 650 pb.

El material de ácido nucleico para el uso en los presentes métodos se puede obtener mediante cualquier método de cualquier organismo de interés. Los organismos de interés incluyen, por ejemplo, microorganismos (más particularmente cepas de *Bt*), plantas, animales, hongos, bacterias e insectos. El material de ácido nucleico puede comprender, por ejemplo, ADN plasmídico preparado a partir de un organismo de interés, tal como una cepa de *Bt*. En algunas realizaciones, la obtención de material de ácido nucleico comprende aislar ADN de un organismo de interés, particularmente un microorganismo de interés. El material de ácido nucleico puede comprender, por ejemplo, ADN genómico. En aspectos particulares de la invención, el material de ácido nucleico comprende una biblioteca de plásmido generada a partir de cepas de *Bt*. Cuando se realizan múltiples rondas de amplificación por PCR, se puede obtener una nueva muestra de material de ácido nucleico del organismo y usar para cada ronda de PCR. Por tanto, por ejemplo, se puede preparar una nueva preparación de plásmido de ADN a partir de una cepa de *Bt* para el uso en cada ronda de PCR.

La amplificación de ácidos nucleicos mediante PCR es una técnica de biología molecular fundamental. Los métodos para realizar PCR se conocen bien en la técnica y se pueden realizar con instrumentación que está disponible en el mercado. Véase, por ejemplo, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2a ed., Cold Spring Harbor Laboratory Press, Plainview, New York); Innis *et al.*, eds. (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic Press, New York); Innis y Gelfand, eds. (1995) *PCR Strategies* (Academic Press, New York); e Innis y Gelfand, eds. (1999) *PCR Methods Manual* (Academic Press, Nueva York). En resumen, la PCR permite la amplificación rápida y eficaz de material de ácido nucleico (por ejemplo, ADN de un gen de interés) que comprende una secuencia diana de interés. El material de ácido nucleico a amplificar, los cebadores oligonucleotídicos y una ADN polimerasa termoestable (por ejemplo, polimerasa de Taq) se mezclan en condiciones adecuadas para amplificación por PCR. La mezcla de reacción de PCR comprende además suficientes cantidades de los cuatro desoxinucleósidos trifosfato y cloruro de magnesio. Los componentes de reacción individuales para PCR están disponibles en el mercado y se ofrecen por varias compañías (por ejemplo, Roche Diagnostics, Qiagen, Promega, Stratagene, etc.). También están disponibles mezclas de reacción preparadas previamente o “mezclas madre” a las que se tiene que añadir solamente el material de ácido nucleico y los cebadores oligonucleotídicos. La PCR se realiza durante al menos un tiempo suficiente para permitir la producción de copias de secuencias de ácido nucleico entre cebadores oligonucleotídicos en una cantidad detectable.

En realizaciones particulares, los métodos de la invención comprenden realizar una primera ronda de PCR, particularmente PCR en tiempo real, más particularmente, PCR en tiempo real cuantitativa. La PCR en tiempo real permite la detección de productos de PCR en estadios tempranos de la reacción de amplificación. Específicamente, en la PCR en tiempo real, la cuantificación de productos de PCR depende de los pocos ciclos en los que la cantidad de material de ácido nucleico se amplifica logarítmicamente hasta que se alcanza una meseta. Durante la fase exponencial, la cantidad de material de ácido nucleico diana debe duplicarse cada ciclo y no existe desviación debido a reactivos limitantes. En la técnica se conocen bien los métodos y la instrumentación para realizar PCR en tiempo real. Véase, por ejemplo, Bustin (2000) *J. Molec. Endocrinol.* 25: 169-193; Freeman *et al.* (1999) *Biotechniques* 112: 124-125; Halford (1999) *Nat. Biotechnol.* 17: 835; y Heid *et al.* (1996) *Genome Res.* 6(10): 986-994. En ciertos aspectos de la invención, la primera ronda de amplificación por PCR comprende realizar PCR en tiempo real.

Como se usa en este documento, “detectar” los productos de amplificación por PCR comprende cualquier método para detectar la presencia, ausencia o cantidad de ácidos nucleicos amplificados por las etapas de PCR de la presente invención. Los métodos de detección pueden proporcionar información cualitativa o cuantitativa con respecto al nivel de amplificación. Tales métodos para detectar los productos de amplificación por PCR se conocen bien en la técnica e incluyen, por ejemplo, electroforesis en gel de agarosa teñido con bromuro de etidio, transferencia de Southern/hibridación de sonda y ensayos de fluorescencia.

Están disponibles muchos colorantes y sondas diferentes para controlar la PCR y detectar productos de PCR. Por ejemplo, los productos de PCR generados por amplificación por PCR en tiempo real se pueden detectar usando una diversidad de colorantes fluorescentes y sondas oligonucleotídicas marcadas covalentemente con moléculas fluorescentes. Tales entidades fluorescentes son capaces de indicar la presencia de productos de PCR y proporcionar una señal relacionada con la cantidad de productos de PCR. Además, mediante el uso del control de fluorescencia continua de los productos de PCR, se puede determinar el punto en el cual se detecta la señal sobre el fondo (Ct; umbral de ciclo) y está en la fase exponencial. Cuanto más abundante sea la secuencia de ácido nucleico de molde, antes se alcanza la Ct.

Se pueden usar colorantes específicos de ADN bicatenario para detectar la formación de producto de PCR en cualquier amplificación por PCR sin la necesidad de sintetizar sondas específicas de secuencia. Tales colorantes se unen específicamente a ADN bicatenario (ADNbc) e incluyen, pero sin limitación, Verde SYBR[®], SYBR Gold[®] y bromuro de etidio. “Verde SYBR[®]” se refiere a cualquiera de los colorantes fluorescentes de Verde SYBR[®] disponibles en el mercado, incluyendo Verde SYBR[®] I y Verde SYBR[®] II. Con colorantes de ADNbc, la especificidad del producto se puede aumentar por análisis de curvas de fusión o adquiriendo la fluorescencia a una alta temperatura en la que se han fundido los productos no específicos. Véase Ririe *et al.* (1997) *Anal. Biochem.* 245: 154-160; Morrison *et al.* (1998) *Bio Techniques* 24: 954-962.

Las sondas oligonucleotídicas también se pueden marcar covalentemente con moléculas fluorescentes y usar para detectar productos de PCR. Los cebadores de horquilla (cebadores Sunrise®), las sondas de horquilla (Molecular Beacons®) y las sondas de exonucleasa (sondas TaqMan®) son oligonucleótidos fluorescentes doblemente marcados que se pueden controlar durante la PCR. Estas sondas dependen de la inactivación de fluorescencia de un fluoróforo por un inactivador en el mismo oligonucleótido. La fluorescencia aumenta cuando tiene lugar la hibridación o la hidrólisis por exonucleasa.

Los productos de PCR también se pueden detectar usando dos oligonucleótidos, marcado cada uno con una sonda fluorescente. La hibridación de estos oligonucleótidos con un ácido nucleico diana acerca las dos sondas fluorescentes para permitir que tenga lugar la transferencia de energía de resonancia. Véase, por ejemplo, Wittwer *et al.* (1997) *Bio Techniques* 22: 130-138. Los expertos en la materia conocen bien los pares de fluoróforos aceptables para el uso como pares de transferencia de energía de resonancia fluorescentes e incluyen, pero sin limitación, fluoresceína/rodamina, ficoeritrina/Cy7, fluoresceína/Cy5, fluoresceína/Cy5.5, fluoresceína/LC Red 640 y fluoresceína/LC Red 705.

En ciertos aspectos de la invención, se usa un colorante fluorescente Verde SYBR® para detectar productos de PCR, más particularmente, productos de PCR en tiempo real generados durante la primera ronda de PCR. Como se ha descrito anteriormente, Verde SYBR® es un colorante fluorescente que se une al surco menor de ADNbc. Cuando se une el colorante Verde SYBR® al ADNbc, aumenta la intensidad de la emisión de fluorescencia. Por tanto, cuando se producen más productos de PCR bicatenarios, también aumenta la señal de fluorescencia de Verde SYBR®. En otros aspectos de la invención, se usa un ensayo de nucleasa 5' para controlar la PCR, particularmente PCR en tiempo real y para detectar productos de amplificación por PCR. En el ensayo de nucleasa 5', se añade una sonda oligonucleotídica denominada una sonda TaqMan® a la mezcla de reactivo de PCR. La sonda TaqMan® comprende un colorante indicador de fluorescencia de alta energía en el extremo 5' (por ejemplo, FAM) y un colorante inactivador de baja energía en el extremo 3' (por ejemplo, TAMRA). Cuando la sonda está intacta, la emisión fluorescente del colorante indicador se suprime por la estrecha proximidad del inactivador. La sonda TaqMan® se diseña adicionalmente para hibridar con una secuencia específica de molde entre los cebadores directo e inverso y, por lo tanto, la sonda se une al material de ácido nucleico de molde en la ruta de la polimerasa. La amplificación por PCR da como resultado la escisión y liberación del colorante indicador de la sonda que contiene inactivador por la actividad nucleasa de la polimerasa. Por tanto, la señal de fluorescencia generada por el colorante indicador liberado es proporcional a la cantidad del producto de PCR. En la técnica se conocen bien los métodos y la instrumentación (por ejemplo, ABI Prism 7700 Detector; Perkin Elmer/Applied Biosystems Division) para realizar PCR en tiempo real usando sondas Verde SYBR® o TaqMan®. En realizaciones particulares, los productos de PCR de la primera ronda de amplificación por PCR se detectan usando Verde SYBR®.

Como se ha indicado anteriormente, los productos de PCR generados durante la segunda ronda de PCR generalmente se separan mediante electroforesis en gel de agarosa. Las moléculas de ácido nucleico de la longitud esperada se aíslan y se someten a análisis de transferencia puntual para eliminar genes conocidos en el grupo diana de una consideración adicional.

El "análisis de transferencia puntual" o la "hibridación de transferencia puntual" es un método convencional en el campo de la biología molecular. En general, la hibridación por transferencia puntual comprende inmovilizar material de ácido nucleico, por ejemplo, sobre una membrana de nitrocelulosa o nylon. El material de ácido nucleico inmovilizado se expone a una sonda oligonucleotídica marcada en condiciones adecuadas para hibridación y se detecta la presencia o ausencia de sonda unida. Las sondas oligonucleotídicas de la invención se pueden marcar con un marcador radiactivo o no radiactivo para facilitar la detección de la unión de la sonda. En la técnica están disponibles diversos marcadores radiactivos y no radiactivos. Tales marcadores incluyen, por ejemplo, digoxigenina (DIG), biotina, moléculas fluorescentes y tritio (³H). En la técnica se conocen bien los métodos para producir sondas oligonucleotídicas marcadas para el uso en análisis de transferencia puntual.

Las sondas oligonucleotídicas usadas para análisis de transferencia puntual en los métodos de la invención son específicas para todos los genes conocidos (por ejemplo, genes pesticidas) dentro del grupo diana. Las sondas se diseñan para ser complementarias a fragmentos de los productos de PCR generados durante la segunda ronda de PCR. Un esquema del diseño de sonda oligonucleotídica para la etapa de análisis por transferencia puntual de la presente invención se proporciona en la Figura 1. En realizaciones particulares, se usa una mezcla de sondas oligonucleotídicas que son específicas para todos los genes conocidos en el grupo diana. El diseño de una mezcla de sondas oligonucleotídicas, donde cada sonda es específica para un gen dentro del grupo diana, se usa particularmente cuando, debido a diferencias de secuencia, es difícil desarrollar una única sonda que sea específica para todo un grupo diana. Cuando es posible, se diseña y usa un único conjunto de sondas que es específico para tantos genes (por ejemplo, genes pesticidas) dentro del grupo diana como sea posible. Además, cuando se usa más de una sonda oligonucleotídica, las sondas se pueden incubar con una única membrana de transferencia puntual como una mezcla de sondas o, alternativamente, se pueden preparar múltiples membranas e incubarse por separado con las sondas individuales. Las sondas oligonucleotídicas de transferencia puntual tendrán típicamente de aproximadamente 20 pb a aproximadamente 40 pb de longitud, particularmente de aproximadamente 25 pb a aproximadamente 35 pb, más particularmente de aproximadamente 30 pb a aproximadamente 35 pb. Además, las sondas oligonucleotídicas usadas para el análisis de transferencia puntual se diseñarán típicamente para tener una T_m de al menos aproximadamente 70°C, particularmente al menos aproximadamente 75°C, más particularmente al menos aproximadamente 80°C. Cuando se usa una mezcla de sondas oligonucleotídicas, cada sonda se diseñará para tener aproximadamente la misma T_m.

El experto en la materia entenderá que los métodos o cualquiera de las etapas en este documento, para identificar genes novedosos, incluyendo genes pesticidas novedosos, más particularmente genes de toxina Cry de *Bt* novedosos, se pueden implementar de un modo automatizado, semi-automatizado o manual. Los métodos descritos en este documento se pueden usar en ensayos de exploración de alto rendimiento.

Las composiciones de la invención incluyen polinucleótidos aislados y variantes y fragmentos de los mismos, que comprenden genes novedosos. Tales genes novedosos se identifican usando los métodos de la presente invención. Se proporcionan adicionalmente las secuencias de aminoácidos que comprenden polipéptidos codificados por las moléculas de ácido nucleico de la invención. Las moléculas de ácido nucleico novedosas y polipéptidos pesticidas identificados mediante los métodos proporcionados en este documento se usan, por ejemplo, para proteger plantas contra daños relacionados con plagas.

La invención incluye composiciones de polinucleótidos o proteínas aislados o sustancialmente purificados. Un polinucleótido o proteína “aislado” o “purificado” o porción biológicamente activa del mismo, está sustancial o esencialmente libre de componentes que normalmente acompañan o interaccionan con el polinucleótido o la proteína cuando se encuentra en su entorno natural. Por tanto, un polinucleótido o proteína aislado o purificado está sustancialmente libre de otro material celular o medio de cultivo cuando se produce mediante técnicas recombinantes o sustancialmente libre de precursores químicos u otros agentes químicos cuando se sintetiza químicamente. De forma óptima, un polinucleótido “aislado” está libre de secuencias (de forma óptima, secuencias que codifican proteína) que flanquean de forma natural el polinucleótido (es decir, secuencias localizadas en los extremos 5' y 3' del polinucleótido) en el ADN genómico del organismo del que se obtiene el polinucleótido. Por ejemplo, en diversas realizaciones, el polinucleótido aislado puede contener menos de aproximadamente 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0,5 kb o 0,1 kb de secuencia de nucleótido que flanquea de forma natural el polinucleótido en el ADN genómico de la célula de la que se obtiene el polinucleótido. Una proteína que está sustancialmente libre de material celular incluye preparaciones de proteína que tienen menos de aproximadamente el 30%, 20%, 10%, 5% o el 1% (en peso seco) de proteína contaminante. Cuando la proteína de la invención o porción biológicamente activa de la misma se produce recombinantemente, de forma óptima el medio de cultivo representa menos de aproximadamente el 30%, 20%, 10%, 5% o el 1% (en peso seco) de precursores químicos o agentes químicos que no son proteína de interés.

Como se usa en este documento, “ácido nucleico” incluye la referencia a un polímero de desoxirribonucleótido o ribonucleótido en forma mono- o bicatenaria y, a menos que se limite de otro modo, abarca análogos conocidos (por ejemplo, péptido ácido nucleico) que tienen la naturaleza esencial de nucleótidos naturales porque hibridan con ácidos nucleicos monocatenarios de un modo similar a nucleótidos de origen natural.

El uso del término “oligonucleótido” o “polinucleótido” no tiene por objeto limitar la presente invención a polinucleótidos que comprenden ADN. Los expertos en la materia reconocerán que los oligonucleótidos y polinucleótidos pueden comprender ribonucleótidos y combinaciones de ribonucleótidos y desoxirribonucleótidos. Tales desoxirribonucleótidos y ribonucleótidos incluyen tanto moléculas de origen natural como análogos sintéticos. Los oligonucleótidos y polinucleótidos de la invención también incluyen todas las formas de secuencias que incluyen, pero sin limitación, formas monocatenarias, formas bicatenarias y similares.

Los términos “polipéptido”, “péptido” y “proteína” se usan de forma intercambiable en este documento para referirse a un polímero de restos aminoacídicos. Los términos se aplican a polímeros de aminoácidos en los que uno o más restos aminoacídicos son un análogo químico artificial de un aminoácido correspondiente de origen natural así como a polímeros de aminoácido de origen natural.

Como se usa en este documento, una “secuencia de longitud completa” con referencia a un polinucleótido especificado o su proteína codificada significa que tiene toda la secuencia de ácido nucleico o toda la secuencia de aminoácidos de una secuencia nativa. Se desea que una “secuencia nativa” signifique una secuencia endógena, es decir, una secuencia no modificada encontrada en el genoma de un organismo. Un polinucleótido de longitud completa codifica la forma de longitud completa de la proteína especificada.

Como se usa en este documento, los términos “que codifica” o “codificado” cuando se usan en el contexto de un ácido nucleico especificado significan que el ácido nucleico comprende la información requerida para dirigir la traducción de la secuencia de nucleótidos en una proteína especificada. La información por la que está codificada una proteína se especifica mediante el uso de codones. Una molécula de ácido nucleico que codifica una proteína puede comprender secuencias no traducidas (por ejemplo, intrones) dentro de regiones traducidas de la molécula de ácido nucleico o puede carecer de tales secuencias no traducidas intermedias (por ejemplo, como en ADNc).

Los fragmentos y las variantes de los polinucleótidos descritos y las proteínas codificadas por los mismos también están incluidos en la presente invención. Se desea que un “fragmento” signifique una porción del polinucleótido o una porción de la secuencia de aminoácidos y, por tanto, la proteína codificada de este modo. Los fragmentos de un polinucleótido pueden codificar fragmentos de proteínas que conservan la actividad biológica de la proteína nativa y, por tanto, poseen, por ejemplo, actividad pesticida. Alternativamente, los fragmentos de un polinucleótido que son útiles como sondas de hibridación generalmente no codifican proteínas de fragmento que conservan actividad biológica. Por tanto, los fragmentos de un polinucleótido pueden variar de al menos aproximadamente 20 nucleótidos, aproximadamente 50 nucleótidos, aproximadamente 100 nucleótidos y hasta el polinucleótido de longitud completa que codifica las proteínas de la invención.

Un fragmento de un polinucleótido de la invención que codifica una porción biológicamente activa de una proteína codificará al menos 15, 25, 30, 50, 100, 150, 200 ó 250 aminoácidos contiguos o hasta el número total de aminoácidos presente en una proteína de longitud completa de la invención, tal como una proteína pesticida. Los fragmentos de un polinucleótido que son útiles como sondas de hibridación o cebadores para PCR generalmente no necesitan codificar una porción biológicamente activa de la proteína.

Por tanto, un fragmento de un polinucleótido puede codificar una porción biológicamente activa de una proteína o puede ser un fragmento que se puede usar como una sonda de hibridación o un cebador para PCR usando los métodos que se describen más adelante. Una porción biológicamente activa de una proteína se puede preparar aislando una porción de uno de los polinucleótidos de la invención, expresando la porción codificada de la proteína (por ejemplo, mediante expresión recombinante *in vitro*) y evaluando la actividad biológica de la porción codificada de la proteína. Los polinucleótidos que son fragmentos de una secuencia de nucleótidos identificada mediante los métodos de este documento comprenden al menos 16, 20, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 800, 900, 1.000, 1.100, 1.200, 1.300 ó 1.400 nucleótidos contiguos o hasta el número de nucleótidos presente en un polinucleótido de longitud completa descrito en este documento.

Se desea que las “variantes” signifiquen secuencias sustancialmente similares. Para polinucleótidos, una variante comprende una delección y/o adición de uno o más nucleótidos en uno o más sitios internos dentro del polinucleótido nativo y/o una sustitución de uno o más nucleótidos en uno o más sitios en el polinucleótido nativo. Como se usa en este documento, un polinucleótido o polipéptido “nativo” comprende una secuencia de nucleótidos o secuencia de aminoácidos de origen natural, respectivamente. Para polinucleótidos, las variantes conservativas incluyen las secuencias que, debido a la degeneración del código genético, codifican la secuencia de aminoácidos de uno de los polipéptidos de la invención. Se pueden identificar variantes alélicas de origen natural como éstas con el uso de técnicas de biología molecular bien conocidas tales como, por ejemplo, con la reacción en cadena de la polimerasa (PCR) y las técnicas de hibridación como se resumen más adelante. Los polinucleótidos variantes también incluyen polinucleótidos obtenidos sintéticamente, tales como los generados, por ejemplo, mediante el uso de mutagénesis dirigida pero que todavía codifican una proteína biológicamente activa de la invención (por ejemplo, una proteína pesticida). Generalmente, las variantes de un polinucleótido particular de la invención tendrán una identidad de secuencia de al menos aproximadamente el 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% con la del polinucleótido particular como se determina por programas de alineamiento de secuencia y parámetros descritos a lo largo de este documento.

Las variantes de un polinucleótido particular de la invención (es decir, el polinucleótido de referencia) también se pueden evaluar mediante comparación del porcentaje de identidad de secuencia entre el polipéptido codificado por un polinucleótido variante y el polipéptido codificado por el polinucleótido de referencia. Por tanto, por ejemplo, se describe un polinucleótido aislado que codifica un polipéptido con un porcentaje dado de identidad de secuencia con un polipéptido de la invención. Se puede calcular el porcentaje de identidad de secuencia entre dos polipéptidos cualesquiera usando programas de alineamiento de secuencias y los parámetros descritos a lo largo de este documento. Cuando se evalúa cualquier par dado de polinucleótidos de la invención mediante comparación del porcentaje de identidad de secuencia compartido por los dos polipéptidos que codifican, el porcentaje de identidad de secuencia entre los dos polipéptidos codificados es una identidad de secuencia de al menos aproximadamente el 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más.

Una proteína “variante” quiere decir una proteína obtenida de la proteína nativa por delección o adición de uno o más aminoácidos en uno o más sitios internos en la proteína nativa y/o sustitución de uno o más aminoácidos en uno o más sitios en la proteína nativa. Las proteínas variantes abarcadas por la presente invención son biológicamente activas, es decir, continúan poseyendo la actividad biológica deseada de la proteína nativa, por ejemplo, actividad pesticida como se describe en este documento. Tales variantes se pueden producir, por ejemplo, a partir de polimorfismo genético o por manipulación humana. Las variantes biológicamente activas de una proteína nativa de la invención tendrán una identidad de secuencia de al menos aproximadamente el 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más con la secuencia de aminoácidos para la proteína nativa, como se determina por programas de alineamiento de secuencia y parámetros descritos a lo largo de este documento. Una variante biológicamente activa de una proteína de la invención puede diferir de esa proteína en tan sólo 1-15 restos aminoacídicos, tan sólo 1-10, tal como 6-10, tan sólo 5, tan sólo 4, 3, 2 o incluso 1 resto aminoacídico.

Las proteínas de la invención se pueden alterar de diversos modos que incluyen sustituciones, delecciones, truncamientos e inserciones de aminoácidos. Los métodos para tales manipulaciones se conocen de forma general en la técnica. Por ejemplo, se pueden preparar variantes de secuencias de aminoácidos y fragmentos de proteínas pesticidas u otras mediante mutaciones en el ADN. Los métodos para mutagénesis y alteraciones de polinucleótidos se conocen bien en la técnica. Véase, por ejemplo, Kunkel (1985) *Proc. Natl. Acad. Sci. USA* 82: 488-492; Kunkel *et al.* (1987) *Methods in Enzymol.* 154: 367-382; Patente de Estados Unidos N° 4.873.192; Walker y Gaastra, eds. (1983) *Techniques in Molecular Biology* (MacMillan Publishing Company, New York) y las referencias citadas en esos documentos. Se pueden encontrar orientaciones para sustituciones de aminoácidos apropiadas que no afecten a la actividad biológica de la proteína de interés en el modelo de Dayhoff *et al.* (1978) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, D.C.). Pueden ser óptimas sustituciones conservativas, tales como intercambiando un aminoácido por otro que tenga propiedades similares.

Por tanto, los polinucleótidos de la invención incluyen tanto las secuencias de origen natural como formas mutantes. Así mismo, las proteínas de la invención incluyen tanto proteínas de origen natural como variaciones y formas modificadas de las mismas. Tales variantes continuarán poseyendo la actividad biológica deseada, por ejemplo, actividad pesticida. Obviamente, las mutaciones que se realizarán en el ADN que codifica la variante no tienen que poner la secuencia fuera de la fase de lectura y, de forma óptima, no crearán regiones complementarias que podrían producir una estructura de ARNm secundaria. Véase, la Publicación de Solicitud de Patente EP N° 75.444.

No se espera que las delecciones, inserciones y sustituciones de las secuencias de proteína incluidas en este documento produzcan cambios radicales en las características de la proteína. Sin embargo, cuando es difícil predecir el efecto exacto de la sustitución, delección o inserción antes de hacer esto, el experto en la materia entenderá que el efecto se evaluará mediante ensayos de exploración rutinarios. Por ejemplo, la actividad de variantes de proteínas pesticidas novedosas se puede evaluar ensayando la actividad pesticida. Véase, por ejemplo, las Patentes de Estados Unidos N° 6.570.005 y 6.339.144.

Los polinucleótidos y las proteínas variantes también incluyen secuencias y proteínas obtenidas de un procedimiento mutagénico y recombinogénico tal como barajado de ADN. Con tal procedimiento, una o más secuencias codificantes de proteína diferentes se pueden manipular para crear un nuevo polipéptido que posea las propiedades deseadas, tales como actividad pesticida. De este modo, se generan bibliotecas de polinucleótidos recombinantes a partir de una población de polinucleótidos de secuencia relacionada que comprenden regiones de secuencia que tienen una sustancial identidad de secuencia y que se pueden recombinar de manera homóloga *in vitro* o *in vivo*. Por ejemplo, usando esta estrategia, los motivos de secuencia que codifican un dominio de interés se pueden barajar entre el gen de la invención (por ejemplo, un gen de toxina Cry de *Bt* novedoso) y otros genes relacionados conocidos para obtener un nuevo gen que codifique una proteína con una propiedad de interés mejorada, tal como actividad pesticida aumentada. Las estrategias para tal barajado de ADN se conocen en la técnica. Véase, por ejemplo, Stemmer (1994) Proc. Natl. Acad. Sci. USA 91: 10747-10751; Stemmer (1994) Nature 370: 389-391; Cramer *et al.* (1997) Nature Biotech. 15: 436-438; Moore *et al.* (1997) J. Mol. Biol. 272: 336-347; Zhang *et al.* (1997) Proc. Natl. Acad. Sci. USA 94: 4504-4509; Cramer *et al.* (1998) Nature 391: 288-291; y las Patentes de Estados Unidos N° 5.605.793 y 5.837.458.

Los polinucleótidos de la invención se pueden usar para aislar secuencias correspondientes de otros organismos, particularmente otros microorganismos. De este modo se pueden usar métodos tales como PCR, hibridación y similares para identificar tales secuencias basándose en su homología de secuencia con las secuencias indicadas en este documento. Las secuencias aisladas basándose en su identidad de secuencia con todas las secuencias indicadas en este documento o con variantes y fragmentos de las mismas están incluidas por la presente invención. Tales secuencias incluyen secuencias que son ortólogos de las secuencias descritas. Con "ortólogo" se quiere decir genes obtenidos de un gen ancestral común y que se encuentran en diferentes especies como resultado de la formación de especies. Los genes encontrados en diferentes especies se consideran ortólogos cuando sus secuencias de nucleótidos y/o sus secuencias de proteína codificadas comparten una identidad de secuencia de al menos el 60%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94, 95%, 96%, 97%, 98%, 99% o más. Las funciones de ortólogos con frecuencia están altamente conservadas entre especies. Por tanto, los polinucleótidos aislados que codifican un polipéptido con una actividad biológica de interés y que hibridan en condiciones rigurosas con una secuencia descrita en este documento, o con variantes o fragmentos de las mismas, se incluyen en la presente invención.

En una estrategia de PCR, se pueden diseñar cebadores oligonucleotídicos para el uso en reacciones de PCR para amplificar las secuencias de ADN correspondientes a partir de ADNc o ADN genómico extraído de cualquier organismo de interés. Los métodos para diseñar cebadores de PCR y clonación por PCR se conocen de forma general en la técnica y se describen en Sambrook *et al.* (1989) Molecular Cloning: A Laboratory Manual (2a ed., Cold Spring Harbor Laboratory Press, Plainview, New York). Véase también Innis *et al.*, eds. (1990) PCR Protocols: A Guide to Methods and Applications (Academic Press, New York); Innis y Gelfand, eds. (1995) PCR Strategies (Academic Press, New York); e Innis y Gelfand, eds. (1999) PCR Methods Manual (Academic Press, New York). Los métodos conocidos de PCR incluyen, pero sin limitación, métodos que usan cebadores emparejados, cebadores anidados, cebadores específicos únicos, cebadores degenerados, cebadores específicos de gen, cebadores específicos de vector, cebadores parcialmente desapareados y similares.

En las técnicas de hibridación, todo o parte de un polinucleótido conocido se usa como una sonda que hibrida selectivamente con otros polinucleótidos correspondientes presentes en una población de fragmentos de ADN genómico o fragmentos de ADNc clonados (es decir, bibliotecas genómicas o de ADNc) de un organismo seleccionado. Las sondas de hibridación pueden ser fragmentos de ADN genómico, fragmentos de ADNc, fragmentos de ARN u otros oligonucleótidos y se pueden marcar con un grupo detectable tal como ³²P o cualquier otro marcador detectable. Por tanto, por ejemplo, se pueden preparar sondas para la hibridación marcando oligonucleótidos sintéticos basándose en los polinucleótidos pesticidas de la invención. Los métodos para la preparación de sondas para la hibridación y para la construcción de bibliotecas de ADNc y genómicas se conocen de forma general en la técnica y se describen en Sambrook *et al.* (1989) Molecular Cloning: A Laboratory Manual (2a ed., Cold Spring Harbor Laboratory Press, Plainview, Nueva York).

Aunque la presente invención proporciona métodos más eficaces para identificar genes novedosos que comparten regiones homólogas (es decir, secuencias distintivas) con cualquier grupo diana de genes conocidos de interés, par-

ticularmente genes pesticidas novedosos, más particularmente genes de toxina Cry de *Bt* novedosos, el experto en la materia reconocerá que también se pueden usar métodos convencionales conocidos en la técnica para identificar secuencias que son homólogas a los polinucleótidos descritos en este documento. Por ejemplo, todo un polinucleótido descrito en este documento, o una o más porciones del mismo, se puede usar como una sonda capaz de hibridar específicamente con polinucleótidos correspondientes y ARN mensajero. Para conseguir la hibridación específica en una diversidad de condiciones, tales sondas incluyen secuencias que son únicas entre las secuencias polinucleotídicas y tienen de forma óptima al menos aproximadamente 10 nucleótidos de longitud y de forma más óptima, al menos aproximadamente 20 nucleótidos de longitud. Tales sondas se pueden usar para amplificar los polinucleótidos correspondientes (por ejemplo, polinucleótidos pesticidas) de un organismo seleccionado por PCR. Esta técnica se puede usar para aislar secuencias codificantes adicionales a partir de un organismo deseado o como un ensayo de diagnóstico para determinar la presencia de secuencias codificantes en un organismo. Las técnicas de hibridación incluyen la exploración de hibridación de genotecas de ADN sembradas (placas o colonias; véase, por ejemplo, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2a ed., Cold Spring Harbor Laboratory Press, Plainview, New York).

La hibridación de tales secuencias se puede realizar en condiciones rigurosas. Las “condiciones rigurosas” o las “condiciones de hibridación rigurosas” quiere decir condiciones en las que una sonda hibridará con su secuencia diana hasta un grado detectablemente mayor que con otras secuencias (por ejemplo, al menos 2 veces sobre el fondo). Las condiciones rigurosas son dependientes de secuencia y serán diferentes en circunstancias diferentes. Mediante el control de la rigurosidad de las condiciones de hibridación y/o lavado, se pueden identificar secuencias diana que son complementarias al 100% con la sonda (sondaje homólogo). Alternativamente, se pueden ajustar condiciones de rigurosidad para permitir cierto desapareamiento en secuencias de tal forma que se detectan menores grados de similitud (sondaje heterólogo). Generalmente, una sonda tiene menos de aproximadamente 1000 nucleótidos de longitud, de forma óptima menos de 500 nucleótidos de longitud.

Típicamente, las condiciones rigurosas serán aquellas en las que la concentración salina es menor de ión Na aproximadamente 1,5 M, típicamente una concentración de ión Na de aproximadamente 0,01 a 1,0 M (u otras sales) a pH de 7,0 a 8,3 y la temperatura es al menos aproximadamente 30°C para sondas cortas (por ejemplo, de 10 a 50 nucleótidos) y al menos aproximadamente 60°C para sondas largas (por ejemplo, más de 50 nucleótidos). También se pueden conseguir condiciones rigurosas con la adición de agentes desestabilizantes tales como formamida. Las condiciones de rigurosidad baja ilustrativas incluyen hibridación con una solución tampón de formamida del 30 al 35%, NaCl 1 M, SDS (dodecil sulfato sódico) al 1% a 37°C y un lavado en SSC de 1x a 2x (SSC 20X = NaCl 3,0 M/citrato trisódico 0,3 M) de 50 a 55°C. Las condiciones de rigurosidad moderada ilustrativas incluyen hibridación en formamida del 40 al 45%, NaCl 1,0 M, SDS al 1% a 37°C y un lavado en SSC de 0,5 X a 1 X de 55 a 60°C. Las condiciones de rigurosidad alta ilustrativas incluyen hibridación en formamida al 50%, NaCl 1 M, SDS al 1% a 37°C y un lavado en SSC 0,1 X de 60 a 65°C. Opcionalmente, los tampones de lavado pueden comprender SDS de aproximadamente el 0,1% a aproximadamente el 1%. La duración de la hibridación generalmente es menor de aproximadamente 24 horas, habitualmente de aproximadamente 4 a aproximadamente 12 horas. La duración del tiempo de lavado será al menos un período de tiempo suficiente para alcanzar el equilibrio.

La especificidad es típicamente la función de lavados post-hibridación, siendo los factores críticos la fuerza iónica y la temperatura de la solución de lavado final. Para híbridos de ADN-ADN, la T_m se puede aproximar a partir de la ecuación de Meinkoth y Wahl (1984) *Anal. Biochem.* 138: 267-284: $T_m = 81,5^\circ\text{C} + 16,6 (\log M) + 0,41 (\% \text{ GC}) - 0,61 (\% \text{ form}) - 500/L$, donde M es la molaridad de cationes monovalentes, % GC es el porcentaje de nucleótidos guanosina y citosina en el ADN, % form es el porcentaje de formamida en la solución de hibridación y L es la longitud del híbrido en pares de bases. La T_m es la temperatura (con fuerza iónica y pH definidos) a la que el 50% de una secuencia diana complementaria híbrida con una sonda perfectamente emparejada. La T_m se reduce aproximadamente 1°C por cada 1% de desapareamiento; por tanto, la T_m , las condiciones de hibridación y/o lavado se pueden ajustar para hibridar con secuencias de la identidad deseada. Por ejemplo, si se buscan secuencias con $\geq 90\%$ de identidad, la T_m puede disminuir 10°C. De forma general, se seleccionan condiciones rigurosas para que estén 5°C por debajo de la T_m para la secuencia específica y su complemento a una fuerza iónica y pH definidos. Sin embargo, las condiciones muy rigurosas pueden utilizar una hibridación y/o lavado 1, 2, 3 ó 4°C por debajo de la T_m ; las condiciones moderadamente rigurosas pueden utilizar una hibridación y/o lavado 6, 7, 8, 9 ó 10°C por debajo de la T_m ; las condiciones de rigurosidad baja pueden utilizar una hibridación y/o lavado 11, 12, 13, 14, 15 ó 20°C por debajo de la T_m . Mediante el uso de la ecuación, las composiciones de hibridación y lavado y la T_m deseada, los expertos en la materia entenderán que se describen de forma inherente variaciones en la rigurosidad de las soluciones de hibridación y/o lavado. Si el grado deseado de desapareamiento da como resultado una T_m menor de 45°C (solución acuosa) o 32°C (solución de formamida), es óptimo aumentar la concentración de SSC de tal forma que se pueda usar una mayor temperatura. Se encuentra una guía extensa para la hibridación de ácidos nucleicos en Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes*, Parte I, Capítulo 2 (Elsevier, New York); y Ausubel *et al.*, eds. (1995) *Current Protocols in Molecular Biology*, Capítulo 2 (Greene Publishing and Wiley-Interscience, New York). Véase Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2a ed., Cold Spring Harbor Laboratory Press, Plainview, New York).

Se usan las siguientes expresiones para describir las relaciones de secuencia entre dos o más polinucleótidos o polipéptidos: (a) “secuencia de referencia”, (b) “ventana de comparación”, (c) “identidad de secuencia” y (d) “porcentaje de identidad de secuencia”.

- 5 (a) Como se usa en este documento, “secuencia de referencia” es una secuencia definida usada como una base para la comparación de secuencia. Una secuencia de referencia puede ser un subconjunto o la totalidad de una secuencia especificada; por ejemplo, como un segmento de un ADNc o secuencia génica de longitud completa o el ADNc o la secuencia génica completa.
- 10 (b) Como se usa en este documento, “ventana de comparación” hace referencia a un segmento contiguo y especificado de una secuencia polinucleotídica, en el que la secuencia polinucleotídica en la ventana de comparación puede comprender adiciones o deleciones (es decir, huecos) en comparación con la secuencia de referencia (que no comprende adiciones o deleciones) para el alineamiento óptimo de los dos polinucleótidos. De forma general, la ventana de comparación tiene al menos 20 nucleótidos contiguos de longitud y, opcionalmente, puede tener 30, 40, 50, 100 o más. Los expertos en la materia entienden que para evitar una similitud alta con una secuencia de referencia debido a la inclusión de huecos en la secuencia polinucleotídica se introduce típicamente una penalización por hueco y se resta del número de coincidencias.
- 15

Los métodos de alineamiento de secuencias para la comparación se conocen bien en la técnica. Por tanto, se puede conseguir la determinación del porcentaje de identidad de secuencia entre dos secuencias usando un algoritmo matemático. Los ejemplos no limitantes de tales algoritmos matemáticos son el algoritmo de Myers y Miller (1988) CABIOS 4: 11-17; el algoritmo de alineamiento local de Smith *et al.* (1981) Adv. Appl. Math. 2: 482; el algoritmo de alineamiento global de Needleman y Wunsch (1970) J. Mol. Biol. 48: 443-453; el método de búsqueda de alineamiento local de Pearson y Lipman (1988) Proc. Natl. Acad. Sci. 85: 2444-2448; el algoritmo de Karlin y Altschul (1990) Proc. Natl. Acad. Sci. USA 87:2264, modificado como en Karlin y Altschul (1993) Proc. Natl. Acad. Sci. USA 90: 5873-5877.

Se pueden utilizar implementaciones informáticas de estos algoritmos matemáticos para la comparación de secuencias para determinar la identidad de secuencia. Tales implementaciones incluyen, pero sin limitación: CLUSTAL en el programa PC/Gene (disponible en Intelligenetics, Mountain View, California); el programa ALIGN (Versión 2.0) y GAP, BESTFIT, BLAST, FASTA y TFASTA en el programa de software GCG Wisconsin Genetics Versión 10 (disponible en Accelrys Inc., 9685 Scranton Road, San Diego, California, EE.UU). Se pueden realizar alineamientos usando estos programas usando los parámetros por defecto. El programa CLUSTAL se describe bien por Higgins *et al.* (1988) Gene 73: 237-244 (1988); Higgins *et al.* (1989) CABIOS 5: 51-153; Corpet *et al.* (1988) Nucleic Acids Res. 16: 10881-90; Huang *et al.* (1992) GABIOS 8: 155-65; y Pearson *et al.* (1994) Meth. Mol. Biol. 24: 307-331. El programa ALIGN está basado en el algoritmo de Myers y Miller (1988) anteriormente. Se pueden usar una tabla de resto de peso PAM120, una penalización por longitud de hueco de 12 y una penalización por hueco de 4 con el programa ALIGN cuando se comparan secuencias de aminoácidos. Los programas BLAST de Altschul *et al.* (1990) J. Mol. Biol. 215: 403 están basados en el algoritmo de Karlin y Altschul (1990) anteriormente. Se pueden realizar búsquedas de nucleótidos con BLAST con el programa BLASTN, puntuación = 100, longitud de palabra = 12, para obtener secuencias de nucleótidos homólogas a una secuencia de nucleótidos que codifica una proteína de la invención. Las búsquedas de proteínas BLAST se pueden realizar con el programa BLASTX, puntuación = 50, longitud de palabra = 3, para obtener secuencias de aminoácidos homólogas a una proteína o polipéptido de la invención. Para obtener alineamientos con huecos con fines de comparación, se puede utilizar BLAST con huecos (en BLAST 2.0) como se describe en Altschul *et al.* (1997) Nucleic Acids Res. 25: 3389. Alternativamente, se puede usar PSI-BLAST (en BLAST 2.0) para realizar una búsqueda iterada que detecte relaciones distantes entre moléculas. Véase Altschul *et al.* (1997) anteriormente. Cuando se utiliza BLAST, BLAST con huecos, PSI-BLAST, se pueden usar los parámetros por defecto de los respectivos programas (por ejemplo, BLASTN para secuencias de nucleótidos, BLASTX para proteínas). Véase www.ncbi.nlm.nih.gov. El alineamiento también se puede realizar manualmente por inspección.

A menos que se indique de otro modo, los valores de identidad/similitud de secuencia proporcionados en este documento se refieren al valor obtenido usando GAP versión 10 usando los siguientes parámetros: % de identidad y % de similitud para una secuencia de nucleótidos usando peso de hueco de 50 y longitud de hueco de 3 y la matriz de puntuación nwsgapdna.cmp; % identidad y % de similitud para una secuencia de aminoácidos usando el peso de hueco de 8 y longitud de hueco de 2 y la matriz de puntuación BLOSUM62; o cualquier programa equivalente de los mismos. Se pretende que “programa equivalente” signifique cualquier programa de comparación de secuencia que, para dos secuencias cualesquiera dadas en cuestión, genera un alineamiento que tenga coincidencias de nucleótidos o restos aminoácidos idénticas y un porcentaje de identidad de secuencia idéntico cuando se comparan con el alineamiento correspondiente generado por GAP, versión 10.

GAP usa el algoritmo de Needleman y Wunsch (1970) J. Mol. Bio/. 48: 443-453, para encontrar el alineamiento de dos secuencias completas que maximiza el número de coincidencias y minimiza el número de huecos. GAP considera todos los posibles alineamientos y posiciones de hueco y crea el alineamiento con el mayor número de bases emparejadas y el menor número de huecos. Permite la proporción de una penalización por creación de hueco y una penalización por extensión de hueco en unidades de bases emparejadas. GAP tiene que beneficiarse de un número de coincidencias con penalización por creación de hueco para cada hueco que inserta. Si se elige una penalización por extensión de hueco mayor que cero, además, GAP tiene que aprovecharse por cada hueco insertado de la longitud del hueco por la penalización por extensión de hueco. Los valores de penalización por creación de hueco y valores de

penalización por extensión de hueco por defecto en la versión 10 del programa de software GCG Wisconsin Genetics para las secuencias de proteínas son, respectivamente, 8 y 2. Para las secuencias de nucleótidos, la penalización por creación de hueco por defecto es 50, mientras que la penalización por extensión de hueco por defecto es 3. Las penalizaciones por creación de hueco y extensión de hueco se pueden expresar como un número entero seleccionado entre el grupo de números enteros que consiste en 0 a 200. Por tanto, por ejemplo, las penalizaciones por creación de hueco y extensión de hueco pueden ser 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65 o superior.

GAP presenta un miembro de la familia de mejores alineamientos. Puede haber muchos miembros de esta familia, pero ningún otro miembro tiene una mejor calidad. GAP presenta cuatro figuras de mérito para alineamientos: Calidad, Proporción, Identidad y Similitud. La calidad es la métrica maximizada para alinear las secuencias. La proporción es la calidad dividida por el número de bases en el segmento más corto. El porcentaje de identidad es el porcentaje de los símbolos que en realidad se emparejan. El porcentaje de similitud es el porcentaje de los símbolos que son similares. Los símbolos entre huecos se ignoran. Una similitud se puntúa cuando el valor de la matriz de puntuación para un par de símbolos es mayor o igual a 0,50, el umbral de similitud. La matriz de puntuación usada en la versión 10 del programa de software GCG Wisconsin Genetics es BLOSUM62 (véase Henikoff y Henikoff (1989) Proc. Natl. Acad. Sci. USA 89: 10915).

(c) Como se usa en este documento, “identidad de secuencia” o “identidad” en el contexto de dos polinucleótidos o secuencias polipeptídicas hace referencia a los restos en las dos secuencias que son iguales cuando se alinean para correspondencia máxima a lo largo de una ventana de comparación especificada. Cuando se usa el porcentaje de identidad de secuencia con referencia a proteínas, se reconoce que las posiciones de restos que no son idénticas con frecuencia difieren por sustituciones de aminoácidos conservativas, donde se sustituyen restos aminoácidos por otros restos aminoácidos con propiedades químicas similares (por ejemplo, carga o hidrofobicidad) y, por lo tanto, no cambian las propiedades funcionales de la molécula. Cuando las secuencias difieren en sustituciones conservativas, el porcentaje de identidad de secuencia se puede ajustar al alza para corregir la naturaleza conservativa de la sustitución. Se dice que las secuencias que difieren en tales sustituciones conservativas tienen “similitud de secuencia” o “similitud”. Los expertos en la materia conocen bien los medios para realizar este ajuste. Típicamente, esto implica puntuar una sustitución conservativa como un desapareamiento parcial en lugar de completo, aumentando de este modo el porcentaje de identidad de secuencia. Por tanto, por ejemplo, cuando se da a un aminoácido idéntico una puntuación de 1 y a una sustitución no conservativa se da una puntuación de cero, se da a una sustitución conservativa una puntuación entre cero y 1. Se calcula la puntuación de sustituciones conservativas, por ejemplo, como se implementa en el programa PC/GENE (Intelligenetics, Mountain View, California).

(d) Como se en usa en este documento, “porcentaje de identidad de secuencia” significa el valor determinado comparando dos secuencias alineadas de forma óptima a lo largo de una ventana de comparación, donde la porción de la secuencia polinucleotídica en la ventana de comparación puede comprender adiciones o deleciones (es decir, huecos) en comparación con la secuencia de referencia (que no comprende adiciones o deleciones) para el alineamiento óptimo de las dos secuencias. El porcentaje se calcula determinando el número de posiciones en las que aparece la base de ácido nucleico o el resto aminoácido idéntico en ambas secuencias para dar el número de posiciones coincidentes, dividiendo el número de posiciones coincidentes por el número total de posiciones en la ventana de comparación y multiplicando el resultado por 100 para proporcionar el porcentaje de identidad de secuencia.

Los métodos de la presente invención se pueden usar para identificar genes novedosos que comparten regiones de homología con cualquier grupo diana de genes conocidos. En una realización, los presentes métodos se usan para identificar genes pesticidas novedosos que sean eficaces contra una diversidad de plagas. Para los propósitos de la presente invención, las plagas incluyen, pero sin limitación, insectos, hongos, bacterias, nematodos, ácaros, patógenos protozoarios, trematodos hepáticos parasitarios de animales y similares. Las plagas de interés particular son plagas de insectos, particularmente plagas de insectos que provocan un daño significativo a plantas agrícolas. Las plagas de insectos incluyen insectos seleccionados entre los órdenes Coleoptera, Diptera, Hymenoptera, Lepidoptera, Mallophaga, Homoptera, Hemiptera, Orthoptera, Thysanoptera, Dermaptera, Isoptera, Anoplura, Siphonaptera, Trichoptera, etc., particularmente Coleoptera y Lepidoptera. Las plagas de insectos de la invención para los principales cultivos incluyen: Maíz: *Ostrinia nubilalis*, gusano barrenador del maíz; *Agrotis ipsilon*, gusano cortador grasiento; *Helicoverpa zea*, gusano del elote del maíz; *Spodoptera frugiperda*; gusano cogollero; *Diatraea grandiosella*, barrenador grande del maíz; *Elasmopalpus lignosellus*, gusano picador de la hoja; *Diatraea saccharalis*, gusano minador de la caña de azúcar; *Diabrotica virgifera*, gusano de la raíz del maíz; *Diabrotica longicornis barberi*, gusano del maíz del norte; *Diabrotica undecimpunctata howardi*, gusano del raíz del sur; *Melanotus spp.*, doradilla; *Cyclocephala borealis*, escarabajo enmascarado nortño (gusano blanco); *Cyclocephala immaculata*, escarabajo enmascarado sureño (gusano blanco); *Popillia japonica*, escarabajo japonés; *Chaetocnema pulicaria*, escarabajo negro del maíz; *Sphenophorus maidis*, gorgojo del maíz; *Rhopalosiphum maidis*, pulgón verde del maíz; *Anuraphis maidiradicis*, pulgón de la raíz del maíz; *Blissus leucopterus leucopterus*, chinche de los cereales; *Melanoplus femurrubrum*, saltamontes de patas rojas; *Melanoplus sanguinipes*, saltamontes migratorio; *Hylemya platura*, mosca del frijol; *Agromyza parvicornis*, minador de las hojas del maíz; *Anaphothrips obscurus*, trips de la hierba; *Solenopsis milesta*, hormiga ladrona; *Tetranychus urticae*, arañuela roja; Sorgo: *Chilo partellus*, barrenador del tallo; *Spodoptera frugiperda*, gusano cogollero; *Helicoverpa zea*, gusano del elote del maíz; *Elasmopalpus lignosellus*, gusano picador de la hoja; *Feltia subterranea*, gusano trozador; *Phyllophaga crinita*, gusano blanco; *Eleodes*, *Conoderus* y *Aeolus spp.*, doradilla; *Oulema melanopus*, criacero de los cereales; *Chaetocnema pulicaria*, escarabajo negro del maíz; *Sphenophorus maidis*, gorgojo del

maíz; *Rhopalosiphum maidis*; pulgón verde del maíz; *Sipha flava*, pulgón amarillo de la caña de azúcar; *Blissus leucopterus leucopterus*, chinche de los cereales; *Contarinia sorghicola*, cecidomía del sorgo; *Tetranychus cinnabarinus*, araña roja del clavel; *Tetranychus urticae*, arañuela roja; Trigo: *Pseudaletia unipunctata*, oruga militar, *Spodoptera frugiperda*, gusano cogollero; *Elasmopalpus lignosellus*, gusano picador de la hoja; *Agrotis orthogonia*, agrotis occidental; *Elasmopalpus lignosellus*, gusano picador de la hoja; *Oulema melanopus*, criacero de los cereales; *Hypera punctata*, fitonoma del trébol, *Diabrotica undecimpunctata howardi*, gusano del maíz del sur; gusano ruso del trigo; *Schizaphis graminum*, pulgón verde; *Macrosiphum avenae*, pulgón del trigo; *Melanoplus femurrubrum*, saltamontes de patas rojas; *Melanoplus differentialis*, saltamontes diferencial; *Melanoplus sanguinipes*, saltamontes migratorio; *Mayetiola destructor*, mosca de Hesse; *Sitodiplosis mosellana*, mosquito del trigo; *Meromyza americana*, gusano del tallo del trigo; *Hylemya coarctata*, mosca gris del trigo; *Frankliniella fusca*, trips del tabaco; *Cephus cinctus*, gusano esqueletizador de la hoja; *Aceria tulipae*, ácaro blanco del bulbo; Girasol: *Suleima helianthana*, polilla de la yema del girasol; *Homoeosoma electellum*, polilla de girasol; *Zygogramma exclamationis*, escarabajo del girasol; *Bothyrus gibbosus*, escarabajo de la zanahoria; *Neolasiptera murtfeldtiana*, mosquita de la semilla del girasol; Algodón: *Heliothis virescens*, gusano bellotero; *Helicoverpa zea*, gusano de la cápsula, *Spodoptera exigua*, gardama de la remolacha; *Pectinophora gossypiella*, gusano rosado del algodón; *Anthonomus grandis*, picudo del algodonero; *Aphis gossypii*, piojo del algodón; *Pseudatomoscelis seriatus*, pulga saltona del algodón; *Trialeurodes abutilonea*, mosca blanca de ala bandeada; *Lygus lineolaris*, chinche manchador; *Melanoplus femurrubrum*, saltamontes de patas rojas; *Melanoplus differentialis*, saltamontes diferencial; *Thrips tabaci*, trips de cebolla; *Frankliniella fusca*, trips del tabaco; *Tetranychus cinnabarinus*, araña roja del clavel; *Tetranychus urticae*, arañuela roja; Arroz: *Diatraea saccharalis*, gusano minador de la caña de azúcar; *Spodoptera frugiperda*, gusano cogollero; *Helicoverpa zea*, gusano del elote del maíz; *Colaspis brunnea*, colapsa de uva; *Lissorhoptus oryzophilus*, picudo acuático, *Sitophilus oryzae*, gorgojo del grano; *Nephotettix nigropictus*, pulga saltona del arroz; *Blissus leucopterus leucopterus*, chinche de los cereales; *Acrosternum hilare*, chinche hedionda verde; Soja: *Pseudoplusia includens*, falso medidor; *Anticarsia gemmatilis*, oruga de las leguminosas; *Plathypena scabra*, gusano verde del trébol; *Ostrinia nubilalis*, gusano barrenador del maíz; *Agrotis ipsilon*, gusano cortador grasiento, *Spodoptera exigua*, gardama de la remolacha; *Heliothis virescens*, gusano bellotero; *Helicoverpa zea*, gusano de la cápsula; *Epilachna varivestis*, escarabajo mejicano del frijol; *Myzus persicae*, pulgón verde del melocotonero; *Empoasca fabae*, pulga saltona de la patata; *Acrosternum hilare*, chinche hedionda verde; *Melanoplus femurrubrum*, saltamontes de patas rojas; *Melanoplus differentialis*, saltamontes diferencial; *Hylemya platura*, mosca del frijol; *Sericothrips variabilis*, trips de soja; *Thrips tabaci*, trips de cebolla; *Tetranychus turkestanii*, araña amarilla, *Tetranychus urticae*, arañuela roja; Cebada: *Ostrinia nubilalis*, gusano barrenador del maíz; *Agrotis ipsilon*, gusano cortador grasiento; *Schizaphis graminum*, pulgón verde; *Blissus leucopterus leucopterus*, chinche de los cereales; *Acrosternum hilare*, chinche hedionda verde; *Euschistus servus*, chinche hedionda marrón; *Delia platura*, mosca del frijol; *Mayetiola destructor*, mosca de Hesse; *Petrobia latens*, ácaro del ajo; Colza: *Brevicoryne brassicae*, pulgón de la col; *Phyllotreta cruciferae*, pulguilla; *Mamestra configurata*, gusano de las crucíferas; *Plutella xylostella*, polilla de la col; *Delia ssp.*, mosca de la semilla.

Los nematodos incluyen nematodos parasitarios tales como nematodos agalladores, de quistes y de lesión, incluyendo *Heterodera spp.*, *Meloidogyne spp.* y *Globodera spp.*; particularmente miembros de los nematodos de quistes, que incluyen, pero sin limitación, *Heterodera glycines* (nematodo de quiste de soja); *Heterodera schachtii* (nematodo de quiste de remolacha); *Heterodera avenae* (nematodo de quiste de cereales); y *Globodera rostochiensis* y *Globodera pallida* (nematodos de quiste de patata). Los nematodos de lesión incluyen *Pratylenchus spp.*

Como se usa en este documento, el término planta incluye células vegetales, protoplastos vegetales, cultivos tisulares de células vegetales a partir de los cuales se puede regenerar una planta, callos vegetales, macollas de planta y células vegetales que están intactos en plantas o partes de plantas tales como embriones, polen, óvulos, semillas, hojas, flores, ramas, frutas, granos, espigas, zuros, chalas, tallos, raíces, puntas de raíz, anteras y similares. Se pretende que grano signifique la semilla madura producida por cultivadores comerciales con fines diferentes de desarrollar o reproducir la especie. En el alcance de la invención también se incluyen la progenie, variantes, y mutantes de las plantas regeneradas, suponiendo que estas partes comprendan los polinucleótidos introducidos.

Aunque los presentes métodos se pueden usar para identificar genes novedosos que son homólogos a cualquier grupo diana de genes conocidos, la presente invención, por ejemplo, se puede usar para identificar genes pesticidas novedosos que codifican polipéptidos que protegen cualquier especie vegetal contra daño relacionado con plagas, incluyendo, pero sin limitación, monocotiledóneas y dicotiledóneas. Los ejemplos de especies vegetales de interés incluyen, pero sin limitación, maíz (*Zea mays*), *Brassica sp.* (por ejemplo, *B. napus*, *B. rapa*, *B. juncea*), particularmente las especies de *Brassica* útiles como fuentes de aceite de semilla, alfalfa (*Medicago sativa*), arroz (*Oryza sativa*), centeno (*Secale cereale*), sorgo (*Sorghum bicolor*, *Sorghum vulgare*), mijo (por ejemplo, mijo perla (*Pennisetum glaucum*), mijo común (*Panicum miliaceum*), panizo (*Setaria italica*), mijo africano (*Eleusine coracana*), girasol (*Helianthus annuus*), cártamo (*Carthamus tinctorius*), trigo (*Triticum aestivum*), soja (*Glycine max*), tabaco (*Nicotiana tabacum*), patata (*Solanum tuberosum*), cacahuetes (*Arachis hypogaea*), algodón (*Gossypium barbadense*, *Gossypium hirsutum*), batata (*Ipomoea batatas*), mandioca (*Manihot esculenta*), café (*Coffea spp.*), coco (*Cocos nucifera*), piña (*Ananas comosus*), árboles cítricos (*Citrus spp.*), cacao (*Theobroma cacao*), té (*Camellia sinensis*), plátano (*Musa spp.*), aguacate (*Persea americana*), higo (*Ficus casica*), guayaba (*Psidium guajava*), mango (*Mangifera indica*), aceituna (*Olea europaea*), papaya (*Carica papaya*), anacardo (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almendra (*Prunus amygdalus*), remolacha azucarera (*Beta vulgaris*), caña de azúcar (*Saccharum spp.*), avena, cebada, verduras, plantas ornamentales y coníferas.

Las verduras incluyen tomates (*Lycopersicon esculentum*), lechuga (por ejemplo, *Lactuca sativa*), judías verdes (*Phaseolus vulgaris*), alubias de lima (*Phaseolus limensis*), guisantes (*Lathyrus spp.*) y miembros del género *Cucumis* tales como pepino (*C. sativus*), cantalupo (*C. cantalupensis*) y melón bordado (*C. melo*). Las plantas ornamentales incluyen azalea (*Rhododendron spp.*), hortensia (*Macrophylla hydrangea*), hibisco (*Hibiscus rosasanensis*), rosas (*Rosa spp.*), tulipanes (*Tulipa spp.*), narcisos (*Narcissus spp.*), petunias (*Petunia hybrida*), clavel (*Dianthus caryophyllus*), flor de pascua (*Euphorbia pulcherrima*) y crisantemo.

Las coníferas que se pueden emplear al llevar a la práctica la presente invención incluyen, por ejemplo, pinos tales como pino bronco (*Pinus taeda*), pino antillano (*Pinus elliotii*), pino ponderosa (*Pinus ponderosa*), pino torcido (*Pinus contorta*) y pino de Monterrey (*Pinus radiata*); abeto Douglas (*Pseudotsuga menziesii*), tsuga de Alberta (*Tsuga canadensis*); picea de Sitka (*Picea glauca*); secuoya roja (*Sequoia sempervirens*); abetos verdaderos tales como abeto común (*Abies amabilis*) y abeto balsámico (*Abies balsamea*) y cedros tales como libocedro (*Thuja plicata*) y ciprés de Notka (*Chamaecyparis nootkatensis*). En realizaciones específicas, las plantas de la presente invención son plantas de cultivo (por ejemplo, maíz, alfalfa, girasol, Brassica, soja, algodón, cártamo, cacahuete, sorgo, trigo, mijo, tabaco, etc.). En otras realizaciones, son óptimas las plantas de maíz y soja y, en otras realizaciones más, son óptimas las plantas de maíz.

Otras plantas de interés incluyen plantas de grano que proporcionan semillas de interés, plantas de semillas oleaginosas y plantas leguminosas. Las semillas de interés incluyen semillas de grano, tales como maíz, trigo, cebada, arroz, sorgo, centeno, etc. Las plantas de semillas oleaginosas incluyen algodón, soja, cártamo, girasol, Brassica, maíz, alfalfa, palma, coco, etc. Las plantas leguminosas incluyen judías y guisantes. Las judías incluyen algarrobilla, algarrobo, alhova, soja, habas, chícharo salvaje, soja verde, alubia de lima, habichuelas, lentejas, garbanzos, etc.

Los artículos “un” y “una” se usan en este documento para referirse a más de uno (es decir, a al menos uno) del objeto gramatical del artículo. A modo de ejemplo, “un elemento” significa uno o más elementos.

Los siguientes ejemplos se proporcionan a modo de ilustración, no a modo de limitación.

Parte experimental

Ejemplo 1

Identificación de Genes Pesticidas Novedosos

Aislamiento de ADN de Plásmido de *Bt*

Se sembraron en estrías soluciones madre de glicerol de varias cepas de *Bt* sobre placas de agar LB. Al día siguiente, se inoculó una única colonia de cada cepa en 2 ml de medio TB por pocillo de una placa de 48 pocillos. Las placas se incubaron durante una noche a 28°C y 250 rpm. Las células se recogieron por centrifugación a 6.000 xg durante 10 minutos a temperatura ambiente. Los sedimentos celulares se resuspendieron por agitación vortical en tampón de suspensión P1 (Qiagen). Las células se lisaron y neutralizaron con tampones P2 y P3, respectivamente, y los lisados se transfirieron a TurboFilters (Qiagen) con vacío aplicado. Los filtrados se unieron a placas QIAprep y se lavaron con tampones PB y PE (Qiagen). Las preparaciones plasmídicas se eluyeron con tampón EB y se recogieron en placas de 96 pocillos.

Diseño de Cebador Oligonucleotídico Degenerado para la Primera Ronda de PCR

Para identificar genes de *Bt* novedosos, tanto los que son homólogos a genes Cry conocidos así como genes pesticidas que representan familias de genes Cry novedosos, se diseñaron cebadores oligonucleotídicos para regiones de alta homología dentro de un grupo diana de genes de interés de *Bt* conocidos. En el presente ejemplo, el grupo diana comprendía genes Cry conocidos que tienen actividad pesticida contra insectos de los órdenes Leptidoptera y Coleoptera, pero no los genes Cry que son activos en dípteros. Específicamente, las secuencias de nucleótidos para todos los genes Cry de *Bt* conocidos del grupo diana se recogieron de la base de datos pública y se generó un alineamiento de estas secuencias. Varias regiones de ADN a lo largo de las secuencias de nucleótidos que eran apropiadas para los requisitos estrictos de diseño del cebador se localizaron en todos los genes de *Bt* elegidos. Esas regiones se denominaron “secuencia distintiva” para genes de *Bt* insecticidas ya que unas pocas secuencias de ADN (de 17 a 24 nucleótidos contiguos) estaban presentes en todos los genes de *Bt* insecticidas conocidos.

Se seleccionó una longitud de cebador inicial para dar una T_m de 54°C y se visualizó una ventana de nucleótidos contiguos que comenzaba en el extremo 5' de la secuencia distintiva seleccionada. Específicamente se revisó la secuencia de nucleótidos dentro de la ventana para determinar si estaban presentes las siguientes características de secuencia:

- 1) no tiene cuatro o más restos de nucleótidos idénticos contiguos;
- 2) tiene no más de dos restos de guanina o citosina dentro de los últimos cinco restos del extremo 3' de la secuencia de nucleótidos;

ES 2 337 952 T3

3) tiene una temperatura de fusión T_m fijada en $54^\circ\text{C} \pm 2^\circ\text{C}$

4) no forma estructuras de horquilla o dímero;

5) está presente en al menos una de las secuencias de nucleótidos del grupo diana de genes pesticidas (es decir, el alineamiento); y

6) no está conservada entre secuencias de nucleótidos de genes pesticidas que no son de grupo diana.

10 Para aumentar la diversidad dentro del cebador, se permitió que un par de bases se fuera n, donde n se seleccionó entre el grupo que consiste en adenina, timina, citosina y guanina.

15 Si estaban presentes todas las características de secuencia, la secuencia de nucleótidos dentro de la ventana de nucleótidos se seleccionó para el uso como un cebador oligonucleotídico para la primera ronda de PCR. Si la secuencia de nucleótidos dentro de la ventana no poseía las características requeridas de secuencia, entonces se seleccionó una ventana adyacente de nucleótidos contiguos moviendo 1 pb más cerca del extremo 3' de la secuencia distintiva y se repitió el proceso. Se diseñaron un cebador oligonucleotídico tanto directo como inverso de acuerdo con los presentes métodos. Además, se diseñaron cebadores directo e inverso de tal forma que eran complementarios a secuencias de nucleótidos en los genes pesticidas de interés que están separados de aproximadamente 50 pb a aproximadamente 150 pb. En la Figura 1 se proporciona un esquema de la metodología de diseño de cebador general para la primera ronda de PCR.

Primera Ronda de Amplificación por PCR: Etapa con Verde SYBR®

25 Se realizó una primera ronda de amplificación por PCR de una primera muestra de material de ácido nucleico aislado de una cepa de *Bt* usando los cebadores oligonucleotídicos diseñados como se ha descrito anteriormente. Específicamente, se amplificaron por PCR las preparaciones de plásmido de *Bt* en placas de 96 pocillos en las siguientes condiciones de reacción:

30 Cantidad de ADN de molde: 100 ng

Cantidad de cebador: 7,5 nmol ($5 \mu\text{M} \times 1,5 \mu\text{l}$)

Volumen de mezcla de reacción: 25 μl

35 Activación con ADN polimerasa AmpGTag® Gold: 95°C durante 10 min

Ciclo de PCR (40 ciclos): 95°C durante 15 s, 60°C 1 min

40 Los productos de PCR de la primera ronda de amplificación se detectaron usando colorante fluorescente Verde SYBR® y el Sistema de Detección de Secuencia 7700 ABI Prism de acuerdo con los métodos conocidos en la técnica. Se usó una preparación de plásmido de la cepa DP 1218-1 que comprende el gen Cry8Bb1 como un control positivo. (Véase el documento WO 0 234 774). Usando las condiciones de PCR que se han descrito anteriormente, la preparación de plásmido 1218-1 produjo una curva patrón para amplificación por PCR en el Sistema de Detección de Secuencia 7700 ABI Prism y se obtuvo un valor de Ct de aproximadamente 13 para el control positivo. Se ensayó un control negativo que comprendía solamente la mezcla de reacción de PCR sin ADN de molde y se generó un valor de Ct de aproximadamente 35. Se seleccionaron preparaciones de plásmido de *Bt* que produjeron un valor de Ct menor de 16 para análisis adicional y se denominaron un positivo a Verde SYBR®.

Segunda Ronda de PCR

50 Se usaron todos los cebadores inversos del conjunto de cebador Verde SYBR® (es decir, los cebadores oligonucleotídicos inversos usados en la primera ronda de PCR) para generar los cebadores directos para la segunda ronda de PCR (es decir, el molde inverso de los cebadores de la primera ronda). Estos cebadores funcionaron como el puente entre la etapa de Verde SYBR® (es decir, primera ronda de PCR) y la segunda ronda de PCR. Los cebadores inversos para el uso en la segunda ronda de PCR se diseñaron esencialmente como se ha descrito anteriormente para los cebadores oligonucleotídicos de primera ronda. La T_m de cebador de PCR se mantuvo a $54^\circ\text{C} \pm 2^\circ\text{C}$ y se diseñó para generar un fragmento de aproximadamente 650 pb a aproximadamente 700 pb. Se proporcionó un esquema de la metodología de diseño de cebador general para la segunda ronda de PCR en la Figura 1.

60 Se aisló ADN plasmídico de las cepas de *Bt* identificadas como positivos a Verde SYBR® en la primera ronda de PCR y después se sometió a una segunda ronda de PCR. Las condiciones de PCR para la segunda ronda fueron las siguientes, usando el kit para PCR Qiagen Multiplex y las preparaciones de plásmido de *Bt* que se han descrito anteriormente:

65 0,5 μg de ADN

ES 2 337 952 T3

Programa:

95°C 15 min

94°C 30 s

54°C 1,5 min

72°C 1,5 min

35X de la etapa 2 a la etapa 4

72° C 10 min

4°C de forma indefinida

Las reacciones de PCR de la segunda se analizaron con electroforesis en gel de agarosa al 1,0% y después, los fragmentos esperados de 650 pb a 700 pb se clonaron en vectores de clonación bacterianos usando un kit de Vector Romo (Invitrogen). Después de la ligación, los productos se introdujeron por transformación en células competentes de *E. coli* Top 10. Se preparó ADN plasmídico de colonias bacterianas individuales y se analizó por análisis de transferencia puntual, como se describe más adelante.

Análisis de Transferencia Puntual

Para eliminar los genes de *Bt* conocidos del análisis y para identificar genes pesticidas novedosos que comprenden las secuencias distintivas usadas en la primera y segunda ronda de PCR se realizaron análisis de transferencia puntual. Específicamente, el ADN plasmídico aislado de las colonias bacterianas individuales se transfirió a una membrana de nylon cargada positivamente (Roche). Se diseñaron sondas específicas para todos los genes pesticidas dentro del grupo diana para que estuvieran dentro del fragmento de secuencia esperado de ADN generado durante la segunda ronda de PCR. En la Figura 1 se proporciona un esquema de la metodología de diseño de sonda general para la etapa de transferencia puntual. Todas las sondas se diseñaron para tener una T_m de aproximadamente de $74^\circ\text{C} \pm 2^\circ\text{C}$. En las tres etapas (es decir, la primera ronda de PCR, la segunda ronda de PCR y el análisis de transferencia puntual), la T_m de los cebadores/sondas oligonucleotídicos se fijó de tal forma que se pudo usar en cada etapa una mezcla de cebadores/sondas. Las sondas oligonucleotídicas se marcaron usando el kit de marcado de extremo 3' de oligonucleótido DIG (Roche) y se usaron para explorar la transferencia puntual para genes de *Bt* conocidos. Cada sonda se ensayó individualmente y en una mezcla de sondas para garantizar la especificidad y validez de cada sonda.

Todas las preparaciones plasmídicas caracterizadas como positivas para genes de *Bt* conocidos por análisis de transferencia puntual se eliminaron del análisis posterior. Las preparaciones plasmídicas que eran negativas cuando se analizaron mediante transferencia puntual se sometieron a un análisis de secuencia adicional, como se describe más adelante, para evaluar la novedad.

Análisis de Secuencia

Se secuenciaron los ácidos nucleicos generados durante la segunda ronda de PCR (es decir, fragmentos de 650 pb a 700 pb) y caracterizados como "negativos" por análisis de transferencia puntual. Los resultados de secuencia de estos ácidos nucleicos se compararon frente a secuencias de nucleótidos disponibles en bases de datos públicas usando BLAST. Si el análisis de secuencia indicó un gen de *Bt* potencialmente novedoso, se obtuvo la secuencia de nucleótidos para el gen de longitud completa usando el kit universal GenomeWalker (Becton Dickinson Bioscience). La secuencia de nucleótidos del supuesto gen pesticida novedoso de longitud completa se analizó adicionalmente como se ha descrito anteriormente para confirmar la novedad. Se identificaron genes pesticidas novedosos, tales como los indicados en las SEC ID N°: 1, 3 y 5 (y los polipéptidos codificados de este modo indicados en las SEC ID N°: 2, 4 y 6, respectivamente) por los presentes métodos. Los genes pesticidas novedosos se ensayaron para actividad pesticida, como se describe más adelante.

Bioensayos

Los genes pesticidas novedosos se clonaron en vectores de expresión y ensayaron para actividad pesticida contra plagas de insecto del maíz. Tales métodos se conocen de forma general en la técnica. En la técnica se conocen métodos para ensayar actividad pesticida frente a Coleópteros y se describen, por ejemplo, en la Publicación de Solicitud de Patente de Estados Unidos N° 2002/0151709. Los ensayos para actividad pesticida contra Lepidópteros se describen, por ejemplo, en la Publicación de Solicitud de Patente de Estados Unidos N° 2005/0138684.

Resultados

Los resultados de los bioensayos se presentan en la Tabla 1 y 2

TABLA 1

Genes pesticidas novedosos con actividad contra Lepidópteros

	GS001 (SEC ID N°: 3)	GS021 (SEC ID N°: 1)
<i>Ostrinia nubilalis</i> (ECB)	+	+
<i>Helicoverpa zea</i> (CEW)	+	+
<i>Agrotis ipsilon</i> (BCW)	+	+
<i>Spodoptera frugiperda</i> (FAW)	-	-

TABLA 2

Gen pesticida novedoso con actividad contra Coleópteros

	GS028 (SEC ID N°: 5)
<i>Diabrotica virgifera</i> LeConte (WCRW)	+
<i>Diabrotica undecimpunctata</i> (SCRW)	-
<i>Leptinotarsa decemlineata</i> (CPB)	-

Todas las publicaciones y solicitudes de patentes mencionadas en la memoria descriptiva son indicativas del nivel de los expertos en la materia a los que se refiere esta invención.

Aunque la anterior invención se ha descrito con cierto detalle a modo de ilustración y ejemplo con propósitos de la claridad de la comprensión, será evidente que se puedan practicar ciertos cambios y modificaciones dentro del alcance de las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un método para identificar genes pesticidas, comprendiendo el método:

- a) diseñar al menos un par de cebadores oligonucleotídicos para el uso en una primera ronda de PCR que sea específico para un grupo diana de genes pesticidas, comprendiendo el par de cebadores un cebador directo y un cebador inverso, donde cada cebador se dirige a una secuencia distintiva presente en las secuencias de nucleótidos del grupo diana;
- b) obtener una primera muestra de material de ácido nucleico de un microorganismo de interés;
- c) mezclar la primera muestra de material de ácido nucleico con el al menos un par de cebadores oligonucleotídicos para el uso en la primera ronda de PCR y una ADN polimerasa termoestable en condiciones que son adecuadas para la amplificación por PCR;
- d) realizar una primera ronda de PCR y detectar los productos de amplificación por PCR, determinando de este modo si se producen productos de PCR en la primera ronda de PCR;
- e) obtener una segunda muestra de material de ácido nucleico del microorganismo si se detectan productos de PCR en la primera ronda de PCR;
- f) diseñar al menos un par de cebadores oligonucleotídicos para el uso en una segunda ronda de PCR que sea específico para el grupo diana de genes pesticidas, comprendiendo el par de cebadores un cebador directo y un cebador inverso, donde cada cebador se dirige a una secuencia distintiva presente en las secuencias de nucleótidos del grupo diana;
- g) mezclar la segunda muestra de material de ácido nucleico con el al menos un par de cebadores oligonucleotídicos para el uso en la segunda ronda de PCR y una ADN polimerasa termoestable en condiciones que son adecuadas para la amplificación por PCR y realizar una segunda ronda de PCR;
- h) separar cualquier producto de amplificación por PCR producido en la segunda ronda de PCR usando electroforesis en gel de agarosa y aislar los fragmentos de ácido nucleico para el análisis posterior, donde los fragmentos de ácido nucleico pueden comprender un supuesto fragmento génico pesticida novedoso;
- i) clonar cada fragmento de ácido nucleico en un vector de clonación;
- j) transformar las células hospedadoras con los vectores de clonación, donde los vectores de clonación comprenden los fragmentos de ácido nucleico aislados en la etapa (h);
- k) preparar muestras de ácido nucleico de colonias hospedadoras individuales que comprenden un vector de clonación;
- l) someter las muestras de ácido nucleico de las colonias hospedadoras individuales a análisis de transferencia puntual usando sondas marcadas que son específicas para todos los genes pesticidas conocidos del grupo diana, donde un fragmento de ácido nucleico aislado en la etapa (h) que no se detecta durante la etapa de análisis de transferencia puntual comprende un supuesto fragmento génico pesticida novedoso; y
- m) analizar el supuesto fragmento génico pesticida novedoso.

2. El método de la reivindicación 1, en el que el microorganismo de interés comprende una cepa de *Bacillus thuringiensis*.

3. El método de la reivindicación 1, en el que el grupo diana de genes pesticidas comprende genes Cry de *Bacillus thuringiensis*.

4. El método de la reivindicación 3, en el que el grupo diana comprende genes Cry de *Bacillus thuringiensis* que tienen actividad pesticida contra insectos del orden Coleoptera.

5. El método de la reivindicación 1, en el que la primera ronda de PCR comprende realizar PCR en tiempo real cuantitativa.

6. El método de la reivindicación 5, en el que la primera ronda de PCR se realiza en presencia de una entidad fluorescente, siendo capaz la entidad fluorescente de indicar la presencia de productos de PCR y proporcionar una señal relacionada con la cantidad de productos de PCR.

7. El método de la reivindicación 1, en el que las sondas marcadas que son específicas para todos los genes pesticidas conocidos del grupo diana usadas para el análisis de transferencia puntual se diseñan para ser específicas para una región presente en los fragmentos de ácido nucleico generados durante la segunda ronda de PCR.

5 8. El método de la reivindicación 1, en el que

(a) las sondas marcadas usadas para el análisis de transferencia puntual tienen una temperatura de fusión térmica (T_m) de 70°C a 85°C; o

10 (b) en el que la T_m para el al menos un par de cebadores oligonucleotídicos usado en la primera y segunda ronda de PCR es de 50°C a 65°C.

9. El método de la reivindicación 3, en el que

15 (a) el al menos un par de cebadores oligonucleotídicos usado en la primera ronda de PCR se diseña para ser específico para una secuencia de nucleótidos presente en el dominio 1 de los genes Cry de *Bacillus thuringiensis*; o

20 (b) el al menos un par de cebadores oligonucleotídicos usados en la segunda ronda de PCR se diseña para ser específico para una secuencia de nucleótidos presente en el dominio 2 de los genes Cry de *Bacillus thuringiensis*.

10. El método de la reivindicación 1, en el que el análisis del supuesto fragmento génico pesticida novedoso comprende análisis de secuencia de nucleótidos, comprendiendo dicho análisis de secuencia de nucleótidos secuenciar el ácido nucleico que comprende un supuesto fragmento génico pesticida novedoso y comparar la secuencia de nucleótidos del supuesto fragmento génico pesticida novedoso con todos los genes pesticidas conocidos, determinando de este modo si el fragmento es novedoso.

11. El método de la reivindicación 10, que comprende además secuenciar el supuesto gen pesticida novedoso de longitud completa si se determina que el fragmento es novedoso y que comprende opcionalmente además clonar el gen pesticida novedoso en un vector de clonación y evaluar la actividad pesticida del polipéptido codificado por el gen pesticida novedoso.

12. El método de la reivindicación 1, en el que el al menos un par de cebadores oligonucleotídicos específico para un grupo diana de genes pesticidas usado en la primera ronda de PCR comprende cebadores que son específicos para regiones de homología compartidas por todos los miembros del grupo diana.

13. El método de la reivindicación 12, en el que los cebadores oligonucleotídicos inversos de la primera ronda de PCR se usan para generar los cebadores directos para la segunda ronda de PCR.

14. El método de la reivindicación 1, en el que los cebadores oligonucleotídicos directo e inverso usados en la primera ronda de PCR son complementarios a secuencias de nucleótidos dentro del grupo diana de genes pesticidas que están separados entre 50 pares de bases (pb) y 150 pb.

15. El método de la reivindicación 1, en el que los cebadores oligonucleotídicos usados en la segunda ronda de PCR se diseñan para generar fragmentos de 600 pb a 750 pb de longitud.

16. El método de la reivindicación 1, en el que

50 A. el diseño de al menos un par de cebadores oligonucleotídicos para el uso en la primera ronda de PCR que sea específico para el grupo diana de genes pesticidas comprende:

a) preparar un alineamiento de todas las secuencias de nucleótidos del grupo diana;

55 b) identificar secuencias distintivas dentro de las secuencias de nucleótidos del grupo diana de genes pesticidas, donde una secuencia distintiva comprende una región de homología entre todos los miembros del grupo diana;

60 c) seleccionar una longitud de cebador inicial, donde la longitud de cebador inicial tiene entre aproximadamente 15 pb y 30 pb;

65 d) realizar una primera ronda de exploración para una secuencia de cebador oligonucleotídico, comprendiendo la exploración visualizar una ventana inicial de nucleótidos contiguos de una secuencia distintiva dentro de las secuencias de nucleótidos del grupo diana, donde la ventana inicial se inicia en el extremo 5' de la secuencia de nucleótidos de la secuencia distintiva;

- e) determinar si la secuencia de nucleótidos dentro de la ventana inicial tiene las siguientes características de secuencia (i) - (vi):

- i) no tiene cuatro o más restos de nucleótidos idénticos contiguos;
- ii) tiene no más de dos restos de guanina o citosina dentro de los últimos cinco restos del extremo 3' de la secuencia de nucleótidos;
- iii) tiene una T_m entre aproximadamente 50°C y 65°C;
- iv) no forma estructuras de horquilla o dímero;
- v) está presente en al menos una de las secuencias de nucleótidos del grupo diana de genes pesticidas (es decir, el alineamiento que se ha descrito anteriormente); y
- vi) no está conservada entre secuencias de nucleótidos de genes pesticidas de grupo no diana;

donde se permite que un resto de nucleótido dentro de la secuencia de nucleótidos que se está revisando sea n, donde n es cualquier nucleótido seleccionado entre el grupo que consiste en adenina, timina, guanina y citosina;

- f) seleccionar la secuencia de nucleótidos dentro de la ventana inicial para el uso como un cebador oligonucleotídico si están presentes todas las características de secuencia de la etapa (e);
- g) seleccionar una ventana adyacente de nucleótidos contiguos moviendo la primera ventana hacia el extremo 3' de la secuencia distintiva dentro de las secuencias de nucleótidos del grupo diana un par de bases si la secuencia de nucleótidos dentro de la ventana inicial no tiene todas las características de secuencia de la etapa (e), donde la ventana adyacente es equivalente en longitud a la longitud de cebador inicial;
- h) repetir las etapas (e) - (g) con la ventana adyacente hasta que se identifique una secuencia de nucleótidos con todas las características de secuencia o hasta que se explore toda la secuencia distintiva para el grupo diana; y
- i) seleccionar una segunda secuencia distintiva dentro de las secuencias de nucleótidos del grupo diana de genes pesticidas y realizar rondas adicionales de exploración que comprenden repetir las etapas (c) a (h) usando la segunda secuencia distintiva si no se identifica ninguna secuencia de nucleótidos con todas las características explorando la primera secuencia distintiva; o

- B. el diseño de al menos un par de cebadores oligonucleotídicos para el uso en la segunda ronda de PCR que sea específico para el grupo diana de genes pesticidas comprende:

- a) usar un cebador oligonucleotídico inverso de la primera ronda de PCR para generar un cebador oligonucleotídico directo en la segunda ronda de PCR;
- b) preparar un alineamiento de todas las secuencias de nucleótidos del grupo diana de genes pesticidas para diseñar un cebador oligonucleotídico inverso para el uso en la segunda ronda de PCR;
- c) identificar secuencias distintivas dentro de las secuencias de nucleótidos del grupo diana, donde una secuencia distintiva comprende una región de homología entre todos los miembros del grupo diana y donde la secuencia distintiva usada para diseñar el cebador inverso para la segunda ronda de PCR se localiza 3' con respecto a la secuencia distintiva usada para diseñar el cebador oligonucleotídico inverso usado en la primera ronda de PCR;
- d) realizar las etapas (c) a (i) de la reivindicación 16A hasta que se identifique una secuencia de nucleótidos con todas las características de secuencia y seleccionar la secuencia de nucleótidos para el uso como un cebador inverso en la segunda ronda de PCR.

17. El método de la reivindicación 16, en el que:

- A. en la parte (A) de la reivindicación 16, el diseño de al menos un par de cebadores oligonucleotídicos para el uso en la primera ronda de PCR que sea específico para el grupo diana de genes pesticidas comprende diseñar una mezcla de pares de cebadores oligonucleotídicos degenerados de acuerdo con el método descrito en la reivindicación 16A, de tal forma que cada par de cebadores oligonucleotídicos sea específico para tantos genes pesticidas dentro del grupo diana como sea posible; o

- B. en la parte (B) de la reivindicación 16, el diseño de al menos un par de cebadores oligonucleotídicos para el uso en la segunda ronda de PCR que sea específico para el grupo diana de genes pesticidas comprende diseñar una mezcla de pares de cebadores oligonucleotídicos degenerados de acuerdo con el método descrito en la reivindicación 16, parte (B), de tal forma que cada par de cebadores oligonucleotídicos sea específico para tantos genes pesticidas dentro del grupo diana como sea posible.

18. El método de la reivindicación 17, parte (A), en el que la mezcla de pares de cebadores oligonucleotídicos diseñados de acuerdo con la reivindicación 17, parte (A), se usa en la primera ronda de PCR; o de la reivindicación 17, parte (B), en el que la mezcla de pares de cebadores oligonucleotídicos diseñados de acuerdo con la reivindicación 17, parte (B), se usa en la segunda ronda de PCR.

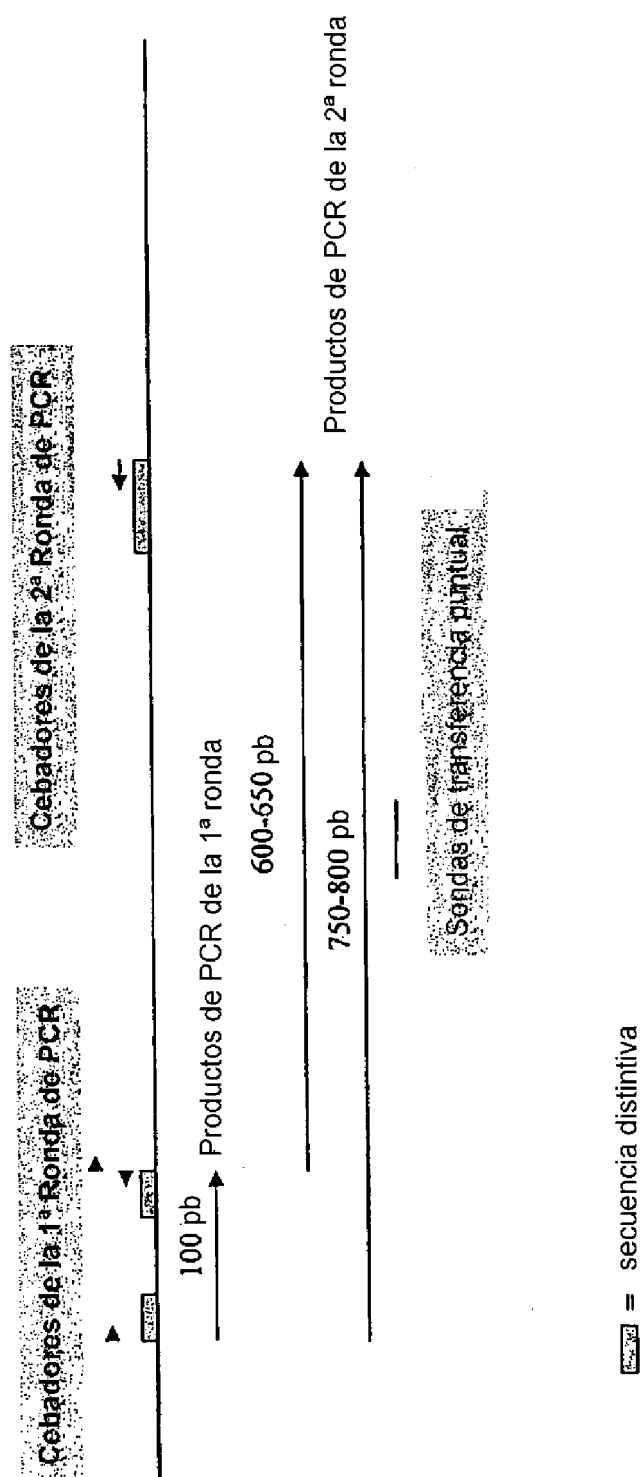
19. El método de 8(b) o 16(B), en el que la T_m es de 52°C a 56°C.

20. El método de la reivindicación 1, en el que una mezcla de pares de cebadores oligonucleotídicos degenerados diseñados de acuerdo con la reivindicación 17, parte (A), se usa en la primera ronda de PCR y en el que una mezcla de pares de cebadores oligonucleotídicos degenerados diseñados de acuerdo con la reivindicación 17, parte (B), se usa en la segunda ronda de PCR.

21. Un método para identificar genes novedosos que comparten homología con un grupo diana de genes conocidos, comprendiendo el método:

- a) diseñar al menos un par de cebadores oligonucleotídicos para el uso en una primera ronda de PCR que sea específico para el grupo diana de genes, comprendiendo el par de cebadores un cebador directo y un cebador inverso, donde cada cebador se dirige a una secuencia distintiva presente en las secuencias de nucleótidos del grupo diana;
- b) obtener una primera muestra de material de ácido nucleico de un organismo de interés;
- c) mezclar la primera muestra de material de ácido nucleico con el al menos un par de cebadores oligonucleotídicos para el uso en la primera ronda de PCR y una ADN polimerasa termoestable en condiciones que son adecuadas para la amplificación por PCR;
- d) realizar una primera ronda de PCR y detectar productos de amplificación por PCR, determinando de este modo si se producen productos de PCR en la primera ronda de PCR;
- e) obtener una segunda muestra de material de ácido nucleico del organismo si se detectan productos de PCR en la primera ronda de PCR;
- f) diseñar al menos un par de cebadores oligonucleotídicos para el uso en una segunda ronda de PCR que sea específico para el grupo diana de genes, comprendiendo el par de cebadores un cebador directo y un cebador inverso, donde cada cebador se dirige a una secuencia distintiva presente en las secuencias de nucleótidos del grupo diana;
- g) mezclar la segunda muestra de material de ácido nucleico con el al menos un par de cebadores oligonucleotídicos para el uso en la segunda ronda de PCR y una ADN polimerasa termoestable en condiciones que son adecuadas para la amplificación por PCR y realizar una segunda ronda de PCR;
- h) separar cualquier producto de amplificación por PCR producido en la segunda ronda de PCR usando electroforesis en gel de agarosa y aislar los fragmentos de ácido nucleico para el análisis posterior, donde los fragmentos de ácido nucleico pueden comprender un supuesto fragmento génico novedoso que comparte homología con los genes del grupo diana;
- i) clonar cada fragmento de ácido nucleico en un vector de clonación;
- j) transformar células hospedadoras con los vectores de clonación, donde los vectores de clonación comprenden los fragmentos de ácido nucleico aislados en la etapa (h);
- k) preparar muestras de ácido nucleico de colonias hospedadoras individuales que comprenden un vector de clonación;
- l) someter las muestras de ácido nucleico de las colonias hospedadoras individuales a análisis de transferencia puntual usando sondas marcadas que son específicas para todos los genes conocidos del grupo diana, donde un fragmento de ácido nucleico aislado en la etapa (h) que no se detecta durante la etapa de análisis de transferencia puntual comprende un supuesto fragmento génico novedoso que comparte homología con los genes del grupo diana; y
- m) analizar el supuesto fragmento génico novedoso.

FIG. 1



ES 2 337 952 T3

LISTA DE SECUENCIAS

<110> Abad, Andre R.
Dong, Hua
Lo, Susan B.
McCutchen, Billy F.
Shi, Xiaomei

<120> Métodos para identificar genes novedosos

<130> 035718/330895

<150> 60/832.423

<151> 21-07-2006

<160> 6

<170> FastSEQ para windows versión 4.0

<210> 1

<211> 2214

<212> ADN

<213> *Bacillus thuringiensis*

<400> 1

```

atg gag aat aat att caa aat caa tgc gtt cct tat aat tgt cta agc 48
acc cct gag aaa ata cta tta gat gag gaa aga att gag act ggg aat 96
aca tca atc gat ctt tct ttg tgc ctt gtg agc ctt ctt tta ggt gaa 144
ttc gtc cct ggt gcg tca ttt gta cta ggt cta att gat ata ata tgg 192
gga ttt gca ggt ccc tct caa tgg gag gca ttt ctg gta cag att gaa 240
cag tta att gac gaa aga ata ggt cag ttc gca agg aat caa gca att 288
tct aga tta gaa ggg cta agc aat ctc tat caa ata tac gca gaa gat 336
ttt aca cag tgg gaa gca gat ccc gat aat cca gca tta aga gaa gag 384
atg cgt act caa ttc aat gat atg aac agt gct ctt aca acc gct att 432
cct ctt ttg gca gtt caa aac tat caa att cct ctt tta tca gta tat 480
gtt caa gct gca aat tta cat tta tca gtt ttg aga gat gtt tca gtg 528
ttt gga caa agt tgg gga ttt gat gcg gcg act att aat agt cgt tat 576
aat gat tta act agg ctt att agc agc tat aca gat cat gta gta aga 624
tgg tat gat aca gga tta gac cgt ttg cga ggc tct act tac caa gac 672
tgg ttt aga tac aac cga ttt aga gaa tta aca ttg act gca tta 720
gat atc gtt gct ctt ttc cca aac tat gat atc aaa atg tat cca atc 768
caa ccc gtt agc caa cta aca agg gaa gtt tat acg gac cca tta att 816
aat ttc aat ccg cag tta cag tct gta gct caa ttg cct act ttt aac 864
gtt atg gaa agt aac gca att aga aac cct cat tta gtt gac ttc ttg 912
aat aac ctt aga att ttt aca gat tgg ttt agt gtc gga cgg cac tat 960
tat tgg gga gga cat cga gtg att tcc aaa cgt gta gga gga agg gag 1008
ata acc ttc cct ata tat gga agg gag gca aag cag gaa cct cca aga 1056
tcc ttt act ttt aat gga cct gtt ttt agg acg tta tca aat cct acc 1104
cta aga cca tta caa caa cct gca cca gct cct cct ttt aat tta cgt 1152
ggc ttg gaa ggt gta gaa ttt tat aca cct aca aat acc ttt acg tat 1200
cgg gga aga ggc ccg cgt gat tct tta act gaa tta ccg cct gga gat 1248
aca agt gta cta cct cgc gaa gga tat agt cac cgg tta tgt cat gca 1296
aca ttt att caa aga tct ggc aca cct ttt tta aca aca ggc gta gtc 1344
ttt tct tgg aca cat cgt agt gct gat gaa acg aat ata att tat cca 1392
gat aag att act caa att cca tgg gta aag gcg cat acc ctt gaa tcg 1440
ggg gcc act gtt att aag ggt cct gga ttt aca gga ggg gat att ctt 1488
act gtt ctt act agt ctt ggt tcc ttg ggc gct tta cga gta act ttt 1536
acg ggg caa tta cca caa aca tat aat ata cga atc cga tat gcc tcg 1584
gtg cta aat aaa tat ggt aca ctc cat ttt tca cag cca cct gca tat 1632
ggg ctc aca ttt cca aaa act atg gat ata gat gaa cca tta aca tct 1680

```

ES 2 337 952 T3

```

5  cgc tcg ttt gct ttt aca act ctt tgg aca cca ata acc ttt aca cga 1728
   gca caa gag gaa ttt aat cta aca ata caa tca ggt gtt tat ata gat 1776
   aga att gaa ttt gtt ccg gca gaa gta aca ttt gag gca gac tat gac 1824
   ttg gaa aaa gcg caa aag gcg gtg aat gct ttg ttt act tct aaa aac 1872
   caa agg ggg cta aaa aca gag gtg acg gat tat cat att gat caa gta 1920
   tct aat tta gtc gaa tgt tta tcc gat gag ttt tgc cta gat gaa aag 1968
   cga gaa tta cct gag aaa gtc aaa cag gcg aag cga ctc agt gat gag 2016
   cga aac cta ctt caa gat tca aaa ttt agg gga atc aac agg caa cca 2064
   gac agc gga tgg aga gga agt acg gat att acc atc caa gga gga gat 2112
10 gac gta ttc aaa gag aat tac gtc aca cta cca ggt acc ttt gat gag 2160
   tgc tat cct aag cgg ccg caa gcc gaa ttc cag cac act ggc ggc cgt 2208
   tac tag                                     2214

```

```

15 <210>2
   <211> 737
   <212> PRT
   <213> Bacillus thuringiensis
20 <400> 2

```

```

25 Met Glu Asn Asn Ile Gln Asn Gln Cys Val Pro Tyr Asn Cys Leu Ser
   1 5 10 15
   Thr Pro Glu Lys Ile Leu Leu Asp Glu Glu Arg Ile Glu Thr Gly Asn
   20 25 30
   Thr Ser Ile Asp Leu Ser Leu Ser Leu Val Ser Leu Leu Leu Gly Glu
   35 40 45
   Phe Val Pro Gly Ala Ser Phe Val Leu Gly Leu Ile Asp Ile Ile Trp
   50 55 60
30 Gly Phe Ala Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile Glu
   65 70 75 80
   Gln Leu Ile Asp Glu Arg Ile Gly Gln Phe Ala Arg Asn Gln Ala Ile
   85 90 95
   Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu Asp
   100 105 110
35 Phe Thr Gln Trp Glu Ala Asp Pro Asp Asn Pro Ala Leu Arg Glu Glu
   115 120 125
   Met Arg Thr Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala Ile
   130 135 140
   Pro Leu Leu Ala Val Gln Asn Tyr Gln Ile Pro Leu Leu Ser Val Tyr
   145 150 155 160
40 Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser Val
   165 170 175
   Phe Gly Gln Ser Trp Gly Phe Asp Ala Thr Ile Asn Ser Arg Tyr
   180 185 190
   Asn Asp Leu Thr Arg Leu Ile Ser Ser Tyr Thr Asp His Val Val Arg
   195 200 205
45 Trp Tyr Asp Thr Gly Leu Asp Arg Leu Arg Gly Ser Thr Tyr Gln Asp
   210 215 220
   Trp Phe Arg Tyr Asn Arg Phe Arg Arg Glu Leu Thr Leu Thr Ala Leu
   225 230 235 240
50 Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ile Lys Met Tyr Pro Ile
   245 250 255
   Gln Pro Val Ser Gln Leu Thr Arg Glu Val Tyr Thr Asp Pro Leu Ile
   260 265 270
   Asn Phe Asn Pro Gln Leu Gln Ser Val Ala Gln Leu Pro Thr Phe Asn
   275 280 285
55 Val Met Glu Ser Asn Ala Ile Arg Asn Pro His Leu Val Asp Phe Leu
   290 295 300
   Asn Asn Leu Arg Ile Phe Thr Asp Trp Phe Ser Val Gly Arg His Tyr
   305 310 315 320
   Tyr Trp Gly Gly His Arg Val Ile Ser Lys Arg Val Gly Gly Arg Glu
   325 330 335
60 Ile Thr Phe Pro Ile Tyr Gly Arg Glu Ala Lys Gln Glu Pro Pro Arg
   340 345 350

```

65

ES 2 337 952 T3

5 Ser Phe Thr Phe Asn Gly Pro Val Phe Arg Thr Leu Ser Asn Pro Thr
 Leu Arg Pro Leu Gln Gln Pro Ala Pro Ala Pro Pro Phe Asn Leu Arg
 Gly Leu Glu Gly Val Glu Phe Tyr Thr Pro Thr Asn Thr Phe Thr Tyr
 Arg Gly Arg Gly Pro Arg Asp Ser Leu Thr Glu Leu Pro Pro Gly Asp
 Thr Ser Val Leu Pro Arg Glu Gly Tyr Ser His Arg Leu Cys His Ala
 10 Thr Phe Ile Gln Arg Ser Gly Thr Pro Phe Leu Thr Thr Gly Val Val
 Phe Ser Trp Thr His Arg Ser Ala Asp Glu Thr Asn Ile Ile Tyr Pro
 Asp Lys Ile Thr Gln Ile Pro Trp Val Lys Ala His Thr Leu Glu Ser
 Gly Ala Thr Val Ile Lys Gly Pro Gly Phe Thr Gly Gly Asp Ile Leu
 Thr Val Leu Thr Ser Leu Gly Ser Leu Gly Ala Leu Arg Val Thr Phe
 20 Thr Gly Gln Leu Pro Gln Thr Tyr Asn Ile Arg Ile Arg Tyr Ala Ser
 Val Leu Asn Lys Tyr Gly Thr Leu His Phe Ser Gln Pro Pro Ala Tyr
 Gly Leu Thr Phe Pro Lys Thr Met Asp Ile Asp Glu Pro Leu Thr Ser
 Arg Ser Phe Ala Phe Thr Thr Leu Trp Thr Pro Ile Thr Phe Thr Arg
 Ala Gln Glu Glu Phe Asn Leu Thr Ile Gln Ser Gly Val Tyr Ile Asp
 Arg Ile Glu Phe Val Pro Ala Glu Val Thr Phe Glu Ala Asp Tyr Asp
 30 Leu Glu Lys Ala Gln Lys Ala Val Asn Ala Leu Phe Thr Ser Lys Asn
 Gln Arg Gly Leu Lys Thr Glu Val Thr Asp Tyr His Ile Asp Gln Val
 Ser Asn Leu Val Glu Cys Leu Ser Asp Glu Phe Cys Leu Asp Glu Lys
 Arg Glu Leu Pro Glu Lys Val Lys Gln Ala Lys Arg Leu Ser Asp Glu
 Arg Asn Leu Leu Gln Asp Ser Lys Phe Arg Gly Ile Asn Arg Gln Pro
 40 Asp Ser Gly Trp Arg Gly Ser Thr Asp Ile Thr Ile Gln Gly Gly Asp
 Asp Val Phe Lys Glu Asn Tyr Val Thr Leu Pro Gly Thr Phe Asp Glu
 Cys Tyr Pro Lys Arg Pro Gln Ala Glu Phe Gln His Thr Gly Gly Arg
 Tyr

<210> 3

<211> 2013

<212> ADN

<213> *Bacillus thuringiensis*

ES 2 337 952 T3

<400> 3

	atg	aat	tca	tat	aaa	aat	aaa	aat	gaa	tat	gaa	ata	ttg	gat	gct	tca	48
	cga	aac	aac	tct	act	atg	tct	act	cat	tat	cca	agg	tat	cca	cta	gca	96
5	aat	aat	cca	caa	gct	tct	atg	caa	aat	acg	aat	tat	aaa	gac	tgg	cta	144
	aac	atg	tgt	aca	aat	aat	aac	ctt	att	cct	ata	gaa	cct	gta	gat	ttc	192
	acc	tgg	caa	aat	gtt	ctt	gtt	tca	acc	ttc	gct	atc	gct	gca	gct	atc	240
	gga	aca	ttg	tta	acc	gct	cca	att	act	ggt	gga	aca	tct	ctg	gta	gct	288
	gga	tca	gct	ata	ata	gcc	gct	ata	tta	cca	ctg	acc	ttc	ccc	gct	aat	336
10	gat	act	agt	gtt	ccg	gat	aag	ctt	atg	gat	gcc	ata	caa	gat	tta	gtt	384
	agg	cgt	gag	ata	gat	cag	tac	gtt	aga	aat	aga	gca	aat	tcg	gag	cta	432
	ctc	agc	ttg	aga	gca	cag	ttg	gat	tct	ttt	aaa	ggg	cta	ttt	gat	tat	480
	tgg	cgc	gcc	aac	caa	ggc	aat	cca	aat	gca	act	aat	tca	gtt	agt	cag	528
	cgc	ttt	act	gca	gtt	cat	aat	aat	ttc	ata	ggg	gca	atg	gca	ctc	ttt	576
15	aaa	ata	gag	ggg	tat	gaa	gaa	tta	ctg	tta	cca	gta	tat	gtt	cag	gct	624
	gca	cgt	tgg	cat	ttg	ttc	cat	tta	aga	gat	ggt	atc	acg	tac	gcc	gat	672
	caa	tgg	cag	tta	gct	gat	cca	act	cat	gca	act	aat	gca	gga	gag	tac	720
	cac	tat	agt	gaa	ttt	aag	aaa	tat	tct	gcg	caa	tat	gca	gat	cat	tgt	768
	gaa	tta	gta	att	aag	aat	cag	cta	gat	aag	att	aaa	aat	gac	tca	aat	816
	aaa	aca	tgg	aaa	gac	tac	aat	caa	tat	cgt	cga	att	atg	aca	ttt	gct	864
20	gtt	tcg	gat	att	gtt	gct	gaa	ttt	tca	atc	att	gat	cca	att	tta	tat	912
	aaa	gga	ggg	ata	aat	agg	gaa	att	tta	acg	agg	aaa	ata	tat	aca	gac	960
	cct	gtt	aat	ttt	tca	cct	ggt	gat	tca	att	gca	gat	gat	gaa	aat	aga	1008
	tat	aca	gtc	cca	cca	tca	gct	gtt	aga	aaa	cta	gtc	ggc	gca	aca	tta	1056
	ttt	act	tct	cag	acg	cct	gct	gat	cct	gat	gtg	gag	ggg	gag	ttt	att	1104
25	gga	aat	cga	aac	cgt	tat	tta	cgt	tta	gaa	ggt	gga	gaa	cca	ttt	gat	1152
	ggt	cct	caa	atc	gga	aac	tcg	aca	agc	cgt	tcg	ata	cca	gta	gga	atc	1200
	ccg	aca	act	gaa	tcg	gtt	tat	gaa	ggt	ggt	gta	agg	ggg	cgt	agt	ggg	1248
	gct	cca	cgt	att	tta	ggt	ttg	aga	tgg	ggt	tcg	tta	act	gac	ttt	caa	1296
	cag	ttt	agt	gct	gga	gga	gac	gtg	tat	aat	tta	gtt	atg	aat	agg	gtt	1344
30	tct	ttg	cca	cct	gga	gac	aga	ttc	ccc	ata	aat	gct	ttt	aat	ttt	act	1392
	tat	cga	tta	tca	gat	ata	att	cct	cct	gga	aat	aag	gga	tca	tct	ttt	1440
	gca	tgg	act	cat	cgt	gag	gtc	gat	cct	aca	gga	aac	tat	tta	tca	aca	1488
	aat	cag	att	aat	tta	ata	cct	gct	aca	aaa	ttt	tct	gaa	aca	cca	tct	1536
	tca	cta	ggg	ata	ctc	aag	gga	ccg	gga	ttt	ata	ggg	gga	gat	tta	gtc	1584
35	gaa	gtt	tca	tat	acc	gga	att	tct	tat	aag	ttt	aaa	ttg	aga	tcc	gtt	1632
	agc	tca	act	agt	ttt	aga	att	cgt	gta	cgt	tat	gca	ggt	tct	ggt	acg	1680
	ggt	cct	tcg	ctt	agt	ggg	cag	ata	tat	ttt	aaa	tta	ggg	aat	gat	atg	1728
	tct	cca	gca	act	cct	tgg	ttg	aat	act	gga	ttt	aac	tct	tcc	aat	gcg	1776
	atg	tat	aat	cac	ttt	aaa	gta	tta	gag	ctt	tat	gga	act	gca	caa	aat	1824
40	att	aca	gac	aac	aac	ttg	gag	att	ata	gta	agg	tca	gca	agc	tca	ggt	1872
	gct	gag	cgg	ttt	tat	cta	gaa	aga	ttg	gaa	ata	atc	cca	att	ggg	ata	1920
	cca	aca	gaa	tac	gct	gaa	tca	caa	aaa	ttg	gaa	aca	gca	aag	aaa	gca	1968
	gtg	acc	gac	ttg	ttt	cca	att	aaa	aca	aag	tat	tta	ctg	aag	tag		2013

<210> 4

45 <211> 670

<212> PRT

<213> *Bacillus thuringiensis*

50 <400> 4

	Met	Asn	Ser	Tyr	Lys	Asn	Lys	Asn	Glu	Tyr	Glu	Ile	Leu	Asp	Ala	Ser	
	1				5				10						15		
	Arg	Asn	Asn	Ser	Thr	Met	Ser	Thr	His	Tyr	Pro	Arg	Tyr	Pro	Leu	Ala	
55				20					25					30			
	Asn	Asn	Pro	Gln	Ala	Ser	Met	Gln	Asn	Thr	Asn	Tyr	Lys	Asp	Trp	Leu	
			35					40					45				
	Asn	Met	Cys	Thr	Asn	Asn	Asn	Leu	Ile	Pro	Ile	Glu	Pro	Val	Asp	Phe	
		50					55					60					
60	Thr	Trp	Gln	Asn	Val	Leu	Val	Ser	Thr	Phe	Ala	Ile	Ala	Ala	Ala	Ile	
	65				70					75					80		
	Gly	Thr	Leu	Leu	Thr	Ala	Pro	Ile	Thr	Gly	Gly	Thr	Ser	Leu	Val	Ala	
				85					90					95			
	Gly	Ser	Ala	Ile	Ile	Ala	Ala	Ile	Leu	Pro	Leu	Thr	Phe	Pro	Ala	Asn	
				100				105						110			
65	Asp	Thr	Ser	Val	Pro	Asp	Lys	Leu	Met	Asp	Ala	Ile	Gln	Asp	Leu	Val	
			115				120						125				
	Arg	Arg	Glu	Ile	Asp	Gln	Tyr	Val	Arg	Asn	Arg	Ala	Asn	Ser	Glu	Leu	

ES 2 337 952 T3

[illegible]

<210> 5
<211> 3491
<212> ADN

ES 2 337 952 T3

<213> *Bacillus thuringiensis*

<400> 5

5	atgaatagaa	atgaatatata	taaaaaagaa	gtaaagccct	cttcatcgaa	tctttctccc	60
	aaataccac	tgatgaattg	tcttgatag	aaaagtcaca	atgttaatta	taaagatagt	120
	ttgaatgta	taaaaggcga	ttatcaagag	ttagatccat	caacgtcgac	aagagcagct	180
	tctgatgcga	ttactgctgc	ccttagtatt	acctcaacaa	tgtagggggc	attaggtcac	240
	ccaattactg	gtgcgcttct	aggcggtttt	aatacactga	caggattgtt	gtggccaggc	300
10	gggggagatc	ttacatggga	agaattgttg	gcggaagggt	aagaaatcat	taataaaact	360
	ataactgaaa	gtataaaaaa	tgatgctttg	ctgcgattag	aggaagcacg	caaacaggta	420
	gatgagtatt	ctggggcttt	aaatgattgg	atattaaatc	ctgggtgtacg	tattagtgtc	480
	gcaacaatga	ggagtatttt	tataactact	gtacagcaat	tagcagtaca	gatgggtatat	540
	tttgaaaggg	acaattatca	aatccctctg	ttaccagtat	atgcacaagc	tgcaaatcta	600
	cacttagttg	tattgaaaga	tatcacaata	ttcgggggag	aatggggatt	ccctgaaaca	660
15	gacattaatt	tttattataa	atcagaattt	ttagaacata	tagaggaata	cactgattat	720
	gttgtaaat	ggatataaga	agggttaaat	gaattacaga	agtcatcgtc	tacagattgg	780
	gttgcatata	atcggtatcg	aagagaaatg	actatattgg	cactggatat	tattgcactc	840
	ttcccggcat	atgatggatt	tttatatcca	ttggagacaa	acactgagtt	gacaagagaa	900
	gtgtatatgg	accctgaagg	tgggcgacca	actaactggc	gtcaataccc	aatctctttc	960
20	acgcaaatag	aaaatttaag	tcgtccgcct	catttatatta	cttggctaaa	ttccatacaa	1020
	attgaaactg	acaagggaat	aactgcggac	gatttatttg	tatgtatgtg	ggcaggtagt	1080
	ttgctaggct	tgcatatata	aaattcgtca	agtgttttca	cagtgaaaaa	tggcagtggg	1140
	gactataaaa	agcttttatga	tctctcaggg	aaagatgttt	atcaaacggt	attatacgca	1200
	agtttcgagtc	atgaagctgg	ctttagcaaa	gtcactgatt	ttgttggaag	cgaagggtgg	1260
25	gcttttaact	tagtatctga	gagtggatcc	acaagctctt	ttcttctctg	actttgcgga	1320
	atatcaaaaca	cacatcttat	agattcagta	acgggtattac	ctccggatat	atcggaccca	1380
	tctcaaagtc	tttcaaagga	ttatactcat	agactatctt	atgttactaa	cgtgtcctat	1440
	tatagaaatc	agtctgttga	ttttgatcga	gtagagctac	ccatattagg	ttggacccat	1500
	agaagtgtag	gttttgataa	tcgaatttat	ccataataaaa	ttacgcagggt	tccagctgtg	1560
30	aaagctaata	aactggataa	tagtgctatg	acagttgtag	aaggggccaaa	ttttacagggt	1620
	ggagccctag	ttaaaggtaaa	cagcactggg	gggtggtagat	atagtttgaa	attttctgtt	1680
	aaggcagatc	caaataaactt	atcgcaaaaa	tatttatataa	gattacgata	tgccaggtgaa	1740
	tggtcggttca	atcaaagtaa	ttatagaagt	ttactattta	gtattaacgg	caataaaatt	1800
	gagggtaaag	gatttttaag	taataactaga	ttggaaaccg	atgaagttcc	tacgcttttt	1860
	aatgactttg	gggtggaccag	tattaaatct	acctttactt	ttccaaaatc	ggaatgtgaa	1920
35	ataagtttgg	atttttatga	tgagaattca	actattggac	cagttggggc	agtattcatt	1980
	gatagaattg	aggttggtccc	agcggatgat	aattacggag	aaaaagaaaa	tttagaaaaa	2040
	gcacagaaag	ccgtgaatgc	cttggtttaca	gcgggaagac	atgcactcca	aacagggtgtg	2100
	acagattaga	aagttgacca	ggtttcaatt	ttagtggatt	gtgtatcagg	agagttattt	2160
	cccaattaga	aacgcgaact	acaaaatcta	atcaaatcag	caaaacggtt	gagctatttc	2220
40	cgtaattttac	ttctagatcc	aacattcgat	tctatcaatt	catctgagga	gaatggctgg	2280
	tacggcagta	atgggtattgc	aattggaaat	gggaattcttg	tatttaaagg	gaactatata	2340
	attttctcag	gtaccaatga	tacacaatac	ccaacgtatc	tctatcaaaa	aattgatgaa	2400
	tccaagctta	agaatatatac	acgctataaa	ctgagaggat	ttatcgagaa	tagtcaagat	2460
	ttagaagcat	atgtgggttcg	ctatgatgca	aaacatgaaa	cattggatgt	atccaataat	2520
45	ctattcccag	atattttctc	tgtaaatgca	tgcgagagaa	ccaatcggtg	tgccggcacta	2580
	ccatacctgg	atgaaaatcc	gaggttagaa	tgtagttcga	tacaagatgg	cattttatct	2640
	gattcgcatt	catttttctct	caatatagat	acaggttcta	ttgattccaa	tgagaacgta	2700
	ggcattttggg	tggtgttttaa	aatttccaca	ccggaagggt	atgcgaaatt	tggaaccta	2760
	gaagtgtattg	aagatggccc	ggtcatttga	gaagcattag	cccgtgtgaa	acgtcaagaa	2820
	acgaagtggg	gaaacaagtt	gacacaactg	cgaacggaaa	cacaagcgat	ttatacacgc	2880
50	gcaaaacaag	ccattgataa	tttattcaca	aatgcacagg	actctcactt	aaaaataggt	2940
	gctacattcg	cgtaatttgt	ggctgcacga	aagattgtcc	aatccatacg	tgaagcgtat	3000
	atgtcatgggt	tatctatcgt	cccagggtga	aattatcccta	tttttacaga	gttgaatgag	3060
	agagtacagc	gagcattttca	attatatgat	gtacgggaatg	tcgtgcgtaa	tggccgatttc	3120
55	ctgaatggag	tatcgggattg	gattgtgaca	tctaattgtaa	aggtacaaga	agaaaatggg	3180
	aacaatgtat	tagttcttttc	caattgggat	gcgcaagtat	tacaatgtct	gaagctctat	3240
	caagatcgcg	gatataatctt	gcgtgtaacg	gcacgtaaaag	aaggattggg	agaaggatat	3300
	attacaatta	cggatgaaga	agggcataca	gatcaattga	catttggcac	atgtgaggaa	3360
	atagatgcat	ctaacacggt	cgtatccaca	ggttatatta	caaaagaact	agaatttttc	3420
60	ccagatacag	agaaagtgcg	tatagaaatt	ggggagacag	aaggaaacctt	ccaggtagaa	3480
	agtgtagaat	t					3491

<210> 6

<211> 1163

<212> PRT

<213> *Bacillus thuringiensis*

ES 2 337 952 T3

<400> 6

	Met	Asn	Arg	Asn	Glu	Tyr	Asn	Lys	Lys	Glu	Val	Lys	Pro	Ser	Ser	Ser
	1				5					10					15	
5	Asn	Leu	Ser	Pro	Lys	Tyr	Pro	Leu	Met	Asn	Cys	Leu	Asp	Thr	Lys	Ser
				20					25					30		
	Gln	Asn	Val	Asn	Tyr	Lys	Asp	Ser	Leu	Asn	Val	Ile	Lys	Gly	Asp	Tyr
			35					40					45			
	Gln	Glu	Leu	Asp	Pro	Ser	Thr	Ser	Thr	Arg	Ala	Ala	Ser	Asp	Ala	Ile
		50					55					60				
10	Thr	Ala	Ala	Leu	Ser	Ile	Thr	Ser	Thr	Met	Leu	Gly	Ala	Leu	Gly	His
	65					70					75				80	
	Pro	Ile	Thr	Gly	Ala	Leu	Leu	Gly	Val	Phe	Asn	Thr	Leu	Thr	Gly	Leu
					85					90					95	
	Leu	Trp	Pro	Gly	Gly	Gly	Asp	Leu	Thr	Trp	Glu	Glu	Leu	Leu	Ala	Glu
				100					105					110		
15	Gly	Glu	Glu	Ile	Ile	Asn	Lys	Thr	Ile	Thr	Glu	Ser	Ile	Lys	Asn	Asp
			115					120					125			
	Ala	Leu	Leu	Arg	Leu	Glu	Glu	Ala	Arg	Lys	Gln	Val	Asp	Glu	Tyr	Ser
		130					135					140				
20	Gly	Ala	Leu	Asn	Asp	Trp	Ile	Leu	Asn	Pro	Gly	Val	Arg	Ile	Ser	Ala
	145				150						155				160	
	Ala	Thr	Met	Arg	Ser	Ile	Phe	Ile	Thr	Thr	Val	Gln	Gln	Leu	Ala	Val
					165					170					175	
	Gln	Met	Val	Tyr	Phe	Glu	Arg	Asp	Asn	Tyr	Gln	Ile	Pro	Leu	Leu	Pro
				180					185					190		
25	Val	Tyr	Ala	Gln	Ala	Ala	Asn	Leu	His	Leu	Val	Val	Leu	Lys	Asp	Ile
			195					200					205			
	Thr	Ile	Phe	Gly	Glu	Glu	Trp	Gly	Phe	Pro	Glu	Thr	Asp	Ile	Asn	Phe
		210					215					220				
	Tyr	Tyr	Lys	Ser	Glu	Phe	Leu	Glu	His	Ile	Glu	Glu	Tyr	Thr	Asp	Tyr
	225				230						235				240	
30	Val	Val	Lys	Trp	Tyr	Lys	Glu	Gly	Leu	Asn	Glu	Leu	Gln	Lys	Ser	Ser
					245					250				255		
	Ser	Thr	Asp	Trp	Val	Ala	Tyr	Asn	Arg	Tyr	Arg	Arg	Glu	Met	Thr	Ile
				260					265					270		
35	Leu	Ala	Leu	Asp	Ile	Ile	Ala	Leu	Phe	Pro	Ala	Tyr	Asp	Gly	Phe	Leu
			275					280					285			
	Tyr	Pro	Leu	Glu	Thr	Asn	Thr	Glu	Leu	Thr	Arg	Glu	Val	Tyr	Met	Asp
		290					295					300				
	Pro	Glu	Gly	Gly	Gly	Pro	Thr	Asn	Trp	Arg	Gln	Tyr	Pro	Ile	Ser	Phe
		305				310					315				320	
40	Thr	Gln	Ile	Glu	Asn	Leu	Ile	Arg	Pro	Pro	His	Leu	Phe	Thr	Trp	Leu
					325					330					335	
	Asn	Ser	Ile	Gln	Ile	Glu	Thr	Asp	Lys	Glu	Leu	Thr	Ala	Asp	Asp	Tyr
				340					345					350		
	Tyr	Val	Cys	Met	Trp	Ala	Gly	Ser	Leu	Leu	Gly	Leu	His	Tyr	Thr	Asn
			355					360					365			
45	Ser	Ser	Ser	Val	Phe	Thr	Val	Lys	Asn	Gly	Ser	Gly	Asp	Tyr	Lys	Lys
		370					375					380				

ES 2 337 952 T3

5 Leu Tyr Asp Leu Ser Gly Lys Asp Val Tyr Gln Thr Leu Leu Tyr Ala
 385 390 400
 Gly Ser Ser His Glu Ala Gly Phe Ser Lys Val Thr Asp Phe Val Gly
 405 410 415
 Asn Glu Gly Gly Thr Phe Asn Leu Val Ser Glu Ser Gly Thr Thr Ser
 420 425 430
 Ser Phe Ser Ser Ala Leu Cys Gly Ile Ser Asn Thr His Leu Ile Asp
 435 440 445
 Ser Val Thr Val Leu Pro Pro Asp Ile Ser Asp Pro Ser Gln Ser Leu
 450 455 460
 10 Ser Lys Asp Tyr Thr His Arg Leu Ser Tyr Val Thr Asn Val Ser Tyr
 465 470 475 480
 Tyr Arg Asn Gln Ser Val Asp Phe Asp Arg Val Glu Leu Pro Ile Leu
 485 490 495
 15 Gly Trp Thr His Arg Ser Val Gly Phe Asp Asn Arg Ile Tyr Pro Tyr
 500 505 510
 Lys Ile Thr Gln Val Pro Ala Val Lys Ala Asn Lys Leu Asn Ser
 515 520 525
 20 Ala Met Thr Val Val Glu Gly Pro Asn Phe Thr Gly Gly Ala Leu Val
 530 535 540
 Lys Val Asn Ser Thr Gly Gly Arg Tyr Ser Leu Lys Phe Ser Val
 545 550 555 560
 Lys Ala Asp Pro Asn Asn Leu Ser Gln Lys Tyr Tyr Ile Arg Leu Arg
 565 570 575
 25 Tyr Ala Gly Glu Trp Arg Phe Asn Gln Ser Asn Tyr Arg Ser Phe Thr
 580 585 590
 Ile Ser Ile Asn Gly Asn Lys Ile Glu Gly Lys Gly Phe Leu Ser Asn
 595 600 605
 Thr Arg Leu Glu Thr Asp Glu Val Pro Thr Leu Phe Asn Asp Phe Gly
 610 615 620
 30 Trp Thr Ser Ile Lys Ser Thr Phe Thr Phe Pro Lys Ser Glu Cys Glu
 625 630 635 640
 Ile Ser Leu Asp Phe Tyr Asp Glu Asn Ser Thr Ile Gly Pro Val Gly
 645 650 655
 35 Pro Val Phe Ile Asp Arg Ile Glu Val Val Pro Ala Asp Asp Asn Tyr
 660 665 670
 Gly Glu Lys Glu Asn Leu Glu Lys Ala Gln Lys Ala Val Asn Ala Leu
 675 680 685
 Phe Thr Ala Gly Arg His Ala Leu Gln Thr Gly Val Thr Asp Tyr Lys
 690 695 700
 40 Val Asp Gln Val Ser Ile Leu Val Asp Cys Val Ser Gly Glu Leu Tyr
 705 710 715 720
 Pro Asn Glu Lys Arg Glu Leu Gln Asn Leu Ile Lys Tyr Ala Lys Arg
 725 730 735
 Leu Ser Tyr Ser Arg Asn Leu Leu Leu Asp Pro Thr Phe Asp Ser Ile
 740 745 750
 45 Asn Ser Ser Glu Glu Asn Gly Trp Tyr Gly Ser Asn Gly Ile Ala Ile
 755 760 765
 Gly Asn Gly Asn Leu Val Phe Lys Gly Asn Tyr Ile Ile Phe Ser Gly
 770 775 780
 Thr Asn Asp Thr Gln Tyr Pro Thr Tyr Leu Tyr Gln Lys Ile Asp Glu
 785 790 795 800
 Ser Lys Leu Lys Glu Tyr Thr Arg Tyr Lys Leu Arg Gly Phe Ile Glu
 805 810 815
 50 Asn Ser Gln Asp Leu Glu Ala Tyr Val Val Arg Tyr Asp Ala Lys His
 820 825 830
 Glu Thr Leu Asp Val Ser Asn Asn Leu Phe Pro Asp Ile Ser Pro Val
 835 840 845
 55 Asn Ala Cys Gly Glu Pro Asn Arg Cys Ala Ala Leu Pro Tyr Leu Asp
 850 855 860
 Glu Asn Pro Arg Leu Glu Cys Ser Ser Ile Gln Asp Gly Ile Leu Ser
 865 870 875 880

ES 2 337 952 T3

	Asp	Ser	His	Ser	Phe	Ser	Leu	Asn	Ile	Asp	Thr	Gly	Ser	Ile	Asp	Ser
	Asn	Glu	Asn	Val	Gly	Ile	Trp	Val	Leu	Phe	Lys	Ile	Ser	Thr	Pro	Glu
5	Gly	Tyr	Ala	Lys	Phe	Gly	Asn	Leu	Glu	Val	Ile	Glu	Asp	Gly	Pro	Val
	Ile	Gly	Glu	Ala	Leu	Ala	Arg	Val	Lys	Arg	Gln	Glu	Thr	Lys	Trp	Arg
		930					935					940				
10	Asn	Lys	Leu	Thr	Gln	Leu	Arg	Thr	Glu	Thr	Gln	Ala	Ile	Tyr	Thr	Arg
	945					950					955					960
	Ala	Lys	Gln	Ala	Ile	Asp	Asn	Leu	Phe	Thr	Asn	Ala	Gln	Asp	Ser	His
					965					970					975	
	Leu	Lys	Ile	Gly	Ala	Thr	Phe	Ala	Ser	Ile	Val	Ala	Ala	Arg	Lys	Ile
15				980					985					990		
	Val	Gln	Ser	Ile	Arg	Glu	Ala	Tyr	Met	Ser	Trp	Leu	Ser	Ile	Val	Pro
			995					1000					1005			
	Gly	Val	Asn	Tyr	Pro	Ile	Phe	Thr	Glu	Leu	Asn	Glu	Arg	Val	Gln	Arg
		1010					1015				1020					
20	Ala	Phe	Gln	Leu	Tyr	Asp	Val	Arg	Asn	Val	Val	Arg	Asn	Gly	Arg	Phe
	1025					1030					1035				1040	
	Leu	Asn	Gly	Val	Ser	Asp	Trp	Ile	Val	Thr	Ser	Asn	Val	Lys	Val	Gln
					1045					1050					1055	
	Glu	Glu	Asn	Gly	Asn	Asn	Val	Leu	Val	Leu	Ser	Asn	Trp	Asp	Ala	Gln
				1060					1065					1070		
25	Val	Leu	Gln	Cys	Leu	Lys	Leu	Tyr	Gln	Asp	Arg	Gly	Tyr	Ile	Leu	Arg
			1075					1080					1085			
	Val	Thr	Ala	Arg	Lys	Glu	Gly	Leu	Gly	Glu	Gly	Tyr	Ile	Thr	Ile	Thr
		1090					1095					1100				
	Asp	Glu	Glu	Gly	His	Thr	Asp	Gln	Leu	Thr	Phe	Gly	Thr	Cys	Glu	Glu
30	1105					1110					1115				1120	
	Ile	Asp	Ala	Ser	Asn	Thr	Phe	Val	Ser	Thr	Gly	Tyr	Ile	Thr	Lys	Glu
					1125					1130					1135	
	Leu	Glu	Phe	Phe	Pro	Asp	Thr	Glu	Lys	Val	Arg	Ile	Glu	Ile	Gly	Glu
				1140					1145					1150		
35	Thr	Glu	Gly	Thr	Phe	Gln	Val	Glu	Ser	Val	Glu					
			1155					1160								