

NUEVOS GENES DE RESISTENCIA ASOCIADOS CON RESISTENCIA A ENFERMEDADES EN SOJA

CAMPO DE LA INVENCION

La presente invención se refiere a composiciones y métodos para identificar, seleccionar y producir plantas mejoradas, resistentes a enfermedades o patógenos, utilizando nuevos genes de resistencia.

DECLARACIÓN REFERENTE A LA PRESENTACIÓN ELECTRÓNICA DE UN LISTADO DE SECUENCIAS

Se proporciona un listado de secuencias en formato de texto ASCII, presentado según 37 C.F.R. § 1.821, titulado "81819-WO-REG-ORG-P-1 -SequenceList_ST25", generado el 1 de agosto de 2019 y presentado mediante la Web EFS en lugar de una copia en papel. Este listado de secuencias se incorpora por la presente por referencia en la memoria descriptiva para su divulgación.

ANTECEDENTES

Se sabe que los patógenos en vegetales provocan un daño considerable a importantes cultivos, provocando pérdidas en agricultura significativas con amplias consecuencias tanto para el suministro de alimentos como para otras industrias que se basan en materiales vegetales. Como tal, el solicitante desea reducir la incidencia y/o el impacto de patógenos en agricultura en la producción de cultivos.

Varios patógenos se han asociado a daños en la soja, que individual y colectivamente tienen el potencial de provocar pérdidas significativas de rendimiento en los Estados Unidos y en todo el mundo. Patógenos a modo de ejemplo incluyen, pero no se limitan a hongos (*p. ej.*, del género *Phytophthora* y roya de la soja asiática *Phakopsora pachyrhizi*), nematodos (*p. ej.*, del género *Meloidogyne*, particularmente, *Meloidogyne javanica*) y el cancro del tallo de soja. Dada la amenaza

importante a los suministros globales de alimentos que presentan estos patógenos, así como al tiempo y a los costes asociados con el tratamiento de cultivos de soja para prevenir pérdidas de cosecha, se necesitan nuevos métodos para producir cultivares de soja resistentes a patógenos. Lo que se necesita son nuevos genes de resistencia (en el presente documento "Genes R") que se puedan introducir en plantas de soja comerciales para controlar los patógenos de la soja.

SUMARIO DE LA INVENCION

Este sumario enumera varias realizaciones de la materia en cuestión que se describe en el presente documento, incluyendo variaciones de estas realizaciones.

Se proporcionan composiciones y métodos para identificar, seleccionar y producir plantas de *Glycine* (incluyendo *Glycines* silvestres (p. ej. *Glycine canescens*, *Glycine clandestine*, *Glycine tomentella*), líneas de *Glycine* de élite y líneas de *Glycine max*) con resistencia potenciada a las enfermedades. También se proporcionan plantas y germoplasmas de soja resistentes a enfermedades. En algunas realizaciones, se proporcionan métodos de producción de una planta de soja resistente a enfermedades.

En un aspecto de la invención, se proporciona una construcción de ADN que comprende un promotor que actúa en células vegetales, enlazado operativamente a un nuevo gen de resistencia (en el presente documento un Gen R). En otro aspecto de la invención se proporciona una planta transgénica que contiene la construcción de ADN y tiene una expresión aumentada del gen R, en donde la planta transgénica es resistente a patógenos de la soja, particularmente la roya de la soja asiática.

En otro aspecto de la invención se incluye un método para preparar una planta transgénica fértil, que comprende proporcionar un casete de expresión de plantas que comprende un gen R y poner en contacto células vegetales receptoras con el casete de expresión de plantas bajo condiciones que permiten la absorción del casete de expresión de plantas por parte de las células receptoras; seleccionar las células vegetales receptoras que contienen el casete de expresión de plantas; regenerar plantas de las células vegetales receptoras seleccionadas; e identificar una planta transgénica fértil que es resistente a patógenos de la soja, particularmente la roya de la soja asiática (en el presente documento "ASR, *Asian Soybean Rust*").

En otro aspecto de la invención se proporciona una planta transgénica fértil que comprende un casete de expresión de plantas que comprende un gen R, en donde la planta tiene niveles de expresión aumentados del gen R, y en donde la planta es resistente a patógenos de la soja, particularmente la roya de la soja asiática. En un aspecto de la invención, la planta es una planta de soja de élite que es resistente a patógenos de la soja.

En otro aspecto de la invención se proporciona un método para controlar la ASR en un campo, que comprende la etapa de plantar la semilla de una planta que comprende un gen R de la invención.

Por lo tanto, es un objeto de la materia en cuestión que se describe en el presente documento proporcionar métodos para transmitir resistencia a patógenos al germoplasma o líneas de plantas no resistentes de soja.

Además, la materia en cuestión que se describe en el presente documento proporciona nuevas líneas de *Glycine max* que comprenden en su genoma un gen R que deriva de una

especie silvestre de *Glycine*, p. ej., *Glycine canescens* o *Glycine clandestina* o *Glycine tomentella*, y confiere además resistencia a la ASR en la nueva línea de *Glycine max*. También se proporcionan plantas y/o germoplasmas de soja identificados, producidos o seleccionados por los métodos de la presente invención, así como cualquier descendencia y/o semillas derivadas de una planta o de un germoplasma de soja identificado, producido o seleccionado por estos métodos.

Como un aspecto adicional más, la invención abarca plantas transgénicas que comprenden una célula vegetal, parte de planta, secuencia de nucleótidos, casete de expresión, vector y/o genes R de la invención.

Como un aspecto adicional hay semillas que producen las plantas transgénicas de la invención y semillas producidas por las plantas transgénicas de la invención.

También se proporcionan productos cosechados derivados de las plantas transgénicas de la invención, en donde el producto cosechado comprende opcionalmente una secuencia de nucleótidos, casete de expresión, vector y/o gen R de la invención. Se proporcionan adicionalmente productos procesados derivados de los productos cosechados de la invención, en donde el producto cosechado comprende opcionalmente una secuencia de nucleótidos, casete de expresión, vector y/o gen R de la invención.

Aún adicionalmente, la invención proporciona, como un aspecto adicional, un método para producir una planta transgénica con resistencia aumentada a un patógeno de la soja. En realizaciones, el método comprende introducir en una planta un polinucleótido, casete de expresión o vector de la invención, en donde el gen R se expresa en la planta, produciendo con ello una planta transgénica con resistencia aumentada a un patógeno de la soja. Opcionalmente, la etapa

de introducción comprende: (i) transformar una célula vegetal con el polinucleótido, casete de expresión o vector y regenerar una planta transgénica; o (ii) cruzar una primera planta que comprende el polinucleótido, casete de expresión o vector con una segunda planta. En realizaciones, el método comprende, además, producir una semilla a partir de la planta transgénica. En realizaciones, el método comprende, además, obtener una planta descendiente de la planta transgénica, en donde la planta descendiente comprende el polinucleótido, el casete de expresión o el vector, expresa el gen R y tiene resistencia aumentada a un patógeno de la soja.

Como aún otro aspecto, la invención proporciona un método para producir una planta transgénica con resistencia aumentada a un patógeno de la planta de soja (*p. ej.*, roya de la soja asiática), comprendiendo el método: (a) plantar una semilla que comprende un polinucleótido, casete de expresión o vector de la invención; y (b) cultivar una planta transgénica a partir de la semilla, en donde la planta transgénica comprende el polinucleótido, casete de expresión o vector, y produce el gen R y tiene resistencia aumentada a un patógeno de la soja. En realizaciones, el método comprende además: (c) cosechar una semilla a partir de la planta transgénica de (b), en donde la semilla cosechada comprende el polinucleótido, casete de expresión, vector y/o el gen R.

Aún más, como otro aspecto, la invención proporciona un método para producir una semilla. En realizaciones, el método comprende: (a) proporcionar una planta transgénica que comprende un polinucleótido, casete de expresión o vector de la invención; y (b) cosechar una semilla a partir de la planta transgénica de (a), en donde la semilla cosechada comprende el polinucleótido, casete de expresión, vector y/o un gen R de la invención.

La invención contempla además un método para producir una semilla de la planta híbrida. En las realizaciones representativas, el método comprende: (a) cruzar una primera planta endogámica, que es una planta transgénica que comprende un polinucleótido, casete de expresión o vector de la invención con una segunda planta endogámica diferente, que puede comprender o no el polinucleótido, casete de expresión o vector de la invención; y (b) cosechar semillas híbridas de una planta descendiente.

La invención también se refiere a métodos para utilizar los polinucleótidos de la invención, por ejemplo, en construcciones de ADN o casetes de expresión o vectores para la transformación y expresión en organismos, incluyendo plantas. Las secuencias de nucleótidos o aminoácidos pueden ser secuencias naturales o sintéticas, que han sido diseñadas para la expresión en un organismo, tal como una planta.

En realizaciones, la invención proporciona un método para utilizar un polinucleótido, casete de expresión o vector de la invención para producir una semilla transgénica, en que la semilla transgénica hace crecer una planta transgénica con resistencia aumentada a un patógeno de la soja.

Los objetos y aspectos precedentes y otros de la presente invención se explican más detalladamente en los dibujos y la memoria descriptiva expuestos más adelante.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

La figura 1 es una ilustración del vector binario 24217.

La figura 2 muestra fotografías de experimentos de bioensayos de roya realizados en hojas recogidas de eventos primarios de soja generados a partir del vector binario 24217.

La figura 3 es un gráfico que muestra la expresión relativa (eje y) del gen de la β -tubulina de la roya de la

soja en eventos primarios de soja de la construcción 24217 14 días después de la inoculación.

La figura 4 es un gráfico que muestra la expresión relativa (eje y) de los genes TIR-A y TIR-B de *G. canescens* en eventos de soja transgénica que confieren resistencia a la roya de la soja.

La figura 5 es un gráfico que resume el análisis por qRT-PCR de GVG00885565 T1, plantas de generación homocigóticas o heterocigóticas para los genes TIR-A y TIR-B.

La figura 6 muestra fotografías de plantas GVG00885565 nulas y homocigóticas para la generación de T1 14 días después de la inoculación con tres poblaciones diferentes de roya de la soja.

La figura 7 es un gráfico que resume la expresión relativa (eje y) del gen de la β -tubulina de la roya de la soja 8, 14 y 21 días después de la inoculación en las plantas nulas y homocigóticas para TIR-A, TIR-B.

La figura 8A es una ilustración de un mapa de vector binario de TIR-A (24205) utilizado en experimentos de cotransformación.

La figura 8B es una ilustración de un mapa de vector binario de TIR-B (24192) utilizado en experimentos de cotransformación.

Figura 9. Síntomas foliares que se correlacionan con el ensayo qRT.

La figura 10 es una ilustración del vector binario 23891.

La figura 11 es una representación de SNP de *Glycine canescens* PI440935 asociados a la resistencia a ASR (es decir, cada punto es igual a 1 SNP que se asocia a la resistencia a ASR).

La figura 12 es una representación de SNP de *Glycine canescens* PI440935 asociados a la resistencia a ASR con el número de cromosomas (es decir, cada punto es igual a 1 SNP que se asocia a la resistencia a ASR).

La figura 13 muestra gráficos de SNP relacionados con intervalos de cartografiado del armazón 000056F.

La figura 14 muestra SNP de *Glycine clandestina* PI339656 asociados a la resistencia a ASR (es decir, cada punto equivale a 1 SNP que se asocia a la resistencia a ASR; armazón 14851 (Cr14) y armazón 9588 (CR9)).

En la figura 15 se muestran gráficos de SNP relacionados con intervalos de cartografiado del armazón 14851.

BREVE DESCRIPCIÓN DEL LISTADO DE SECUENCIAS

SEQ ID NO: 1 es un intervalo cromosómico derivado de la línea de *Glycine canescens* con número de acceso PI440935 que en el presente documento se denomina "Armazón 000056F". El armazón 000056F se ha cartografiado en el cromosoma 14 de *G. canescens*. SEQ ID NO. 2 es un intervalo cromosómico derivado de la línea de *Glycine canescens* con número de acceso PI595799 que en el presente documento se denomina "Armazón 000090F". El armazón 000090F se ha cartografiado en el cromosoma 14 de *G. canescens*. SEQ ID NO: 3 es un intervalo cromosómico derivado de la línea *Glycine canescens* con número de acceso PI 483193 que en el presente documento se denomina "Armazón 92F". El armazón 132F se ha cartografiado en el cromosoma 14 de *G. canescens*. SEQ ID NO: 4 es un intervalo cromosómico derivado de la línea de *Glycine clandestina* con número de acceso PI339656 que en el presente documento se denomina "Armazón 14851". El armazón 14851 se ha cartografiado en el cromosoma 14 de *G. clandestina* en un intervalo aproximado de 3.18 a 3.50 MB del armazón 14851. Estos intervalos cromosómicos o partes de los mismos, se

pueden introducir (es decir, introgresar a través del uso del rescate de embriones y/o la reproducción asistida por marcadores (MAB, *Marker Assisted Breeding*) en líneas de *Glycine max* para crear líneas de *Glycine max* resistentes a la enfermedad (p. ej., roya de la soja asiática (ASR, *Asian Soybean Rust*)). SEQ ID NO: 5 es un intervalo cromosómico derivado de la línea de *Glycine tomentella* con número de acceso PI509501 que en el presente documento se denomina "Armazón 4269". SEQ ID NO: 6 es un intervalo cromosómico derivado de la línea de *Glycine tomentella* con número de acceso PI509501 que en el presente documento se denomina "Armazón 1509". Los armazones 4269 y 1509 se han cartografiado en el cromosoma 14 de *G. tomentella*.

Investigaciones adicionales en los intervalos dieron lugar al descubrimiento de los genes causantes y se identificaron secuencias promotoras y terminadoras nativas asociadas a partir del intervalo o intervalos anteriores ubicados en o correspondientes al cromosoma 14 de *G. canescens* y *G. clandestina*.

SEQ ID NO: 7 es el ADN genómico para un gen de resistencia a la roya de la soja (también denominado en el presente documento "gen TIR-A") de PI595799 que codifica una proteína que contiene un dominio de receptor de Toll/Interleucina-1 (un motivo de gen R "TIR") sintético para Glyma.05G165800 (Soy_william82_v2). SEQ ID NO: 8 es una secuencia de ADNc para el gen TIR-A. SEQ ID NO: 9 comprende un promotor a modo de ejemplo para cualquiera de las SEQ ID NO: 7 u 8. SEQ ID NO: 9 comprende una secuencia terminadora a modo de ejemplo para cualquiera de las SEQ ID NO: 7 u 8. SEQ ID NO: 7 y 8 codifican la proteína de la SEQ ID NO: 11.

SEQ ID NO: 12 es ADN genómico para otro gen de resistencia a la roya de la soja (también denominado en el

presente documento "gen TIR-B") que tiene un motivo de gen R de TIR. SEQ ID NO: 13 es una secuencia de ADNc para el gen TIR-B. SEQ ID NO: 14 comprende un promotor a modo de ejemplo para cualquiera de las SEQ ID NO: 12 u 13. SEQ ID NO: 15 comprende una secuencia terminadora a modo de ejemplo para cualquiera de las SEQ ID NO: 12 u 13. SEQ ID NO: 12 o 13 codifican la proteína de la SEQ ID NO: 16.

Puede introducirse resistencia a la ASR en plantas de *G. max* usando las SEQ ID NO: 7, 8, 11, 12 o cualquier combinación de las mismas. Se contempla además que puede introducirse resistencia a la ASR en plantas de soja de élite ya sea a través de la expresión transgénica de genes que codifican una proteína que tiene un motivo de gen R de TIR o un gen que tiene entre el 70 y el 100% de homología con cualquiera de las SEQ ID NO: 7, 8, 11, 12.

SEQ ID NO: 17 es una secuencia de ADN para el vector 24217_G. *canescens* PI595799 - cGCaRG3b-02 que es una versión truncada en el extremo N del componente TIR-B.

SEQ ID NO: 18 es una secuencia de ADN para el vector 23819_G. *clandestina* PI339656 - TIR-A y TIR-B genómicos.

SEQ ID NO: 19 es una secuencia de ADN para el promotor AtEF1aA1, que deriva del gen del factor de elongación de *Arabidopsis thaliana* EF-1 alfa A1.

SEQ ID NO: 20 es una secuencia de ADN para prUBQ3, que es una versión modificada del promotor de Ubiquitina 3 de *Arabidopsis*.

SEQ ID NO: 21 es una secuencia de ADN para prGmUbi1, que es un promotor de ubiquitina 1 de la soja.

SEQ ID NO: 22 es una secuencia de ADN de prLuFIS1, que es un fragmento de 660 pb del promotor fis1 del lino (*Linum usitatissimum*).

SEQ ID NO: 23 es la secuencia de ADN para un cebador para TIR-B.

SEQ ID NO: 24 es la secuencia de ADN para un cebador para TIR-B.

SEQ ID NO: 25 es la secuencia de ADN para una sonda para TIR-B.

SEQ ID NO: 26 es la secuencia de ADN para un cebador para TIR-A.

SEQ ID NO: 27 es la secuencia de ADN para un cebador para TIR-A.

SEQ ID NO: 28 es la secuencia de ADN para una sonda para TIR-A.

SEQ ID NO: 29 es la secuencia de ADN para un cebador para TIR-A.

SEQ ID NO: 30 es la secuencia de ADN para un cebador para TIR-A.

SEQ ID NO: 31 es la secuencia de ADN para una sonda para TIR-A.

SEQ ID NO: 32 es la secuencia de ADN para el Vector 24205.

SEQ ID NO: 33 es la secuencia de ADN para el Vector 24192.

DESCRIPCIÓN DETALLADA DE LA INVENCION

La materia en cuestión que se describe en el presente documento se refiere a composiciones y métodos para introducir nuevos genes de resistencia (en el presente documento, "Genes R") en plantas de soja comerciales para controlar patógenos de soja. Los métodos implican transformar organismos con secuencias de nucleótidos que codifican los genes R de la invención. Las secuencias de nucleótidos de la invención son útiles para preparar plantas que muestran resistencia aumentada a patógenos de la soja, particularmente la roya de soja asiática (en el presente documento, "ASR"). Por lo tanto, se proporcionan plantas transformadas, células vegetales, tejidos de plantas y semillas. Las composiciones incluyen ácidos nucleicos y proteínas relacionados con plantas resistentes a patógenos de soja, así como plantas transformadas, tejidos de plantas y semillas. Se describen secuencias de nucleótidos de los genes R y las secuencias de aminoácidos de las proteínas codificadas con ello. Las secuencias encuentran uso en la construcción de vectores de expresión para la subsiguiente transformación en plantas de interés, en calidad de sondas para el aislamiento de otros genes R, y similares.

Es posible producir plantas resistentes a patógenos de soja insertando en el genoma de la planta una molécula de ADN que proporciona resistencia (Cook et al, 2012 para resistencia a SCN; Kawashima et al., 2016 para resistencia a la roya de soja de guandú).

La presente descripción no pretende ser un catálogo detallado de todas las diferentes maneras en las que se puede implementar la invención, o de todas las características que se pueden añadir a la presente invención. Por ejemplo, las características ilustradas con relación a una realización se

pueden incorporar a otras realizaciones y las características ilustradas con relación a una realización particular se pueden eliminar de dicha realización. De este modo, la invención contempla que, en algunas realizaciones de la invención, se pueda excluir u omitir cualquier característica o combinación de características expuestas en el presente documento. Además, para los expertos en la materia, a la vista de la presente divulgación, resultarán evidentes numerosas variaciones y adiciones que se pueden aplicar a las diversas realizaciones sugeridas en el presente documento, las cuales no se alejan de la presente invención. Por lo tanto, se pretende que las siguientes descripciones ilustren algunas realizaciones particulares de la invención pero no se pretende que especifiquen exhaustivamente todas sus permutaciones, combinaciones y variaciones.

Todas las referencias enumeradas a continuación, así como todas las referencias de la presente divulgación incluyendo, pero sin limitación, todas las patentes, solicitudes de patente y sus publicaciones, artículos en revistas científicas y entradas de bases de datos (p. ej., entradas de la base de datos GENBANK® y todas las notas allí disponibles), se incorporan a la presente como referencia en su totalidad en cuanto complementan, explican, proporcionan antecedentes o enseñan metodología, técnicas y/o composiciones empleadas en el presente documento.

Secuencias de nucleótidos proporcionadas en el presente documento en la dirección 5' a 3', de izquierda a derecha, se presentan utilizando el código estándar para representar bases de nucleótidos tal como se recogen en 37 CFR §§1.821 - 1.825 y en la Norma ST.25 de la Organización Mundial de la Propiedad Intelectual (WIPO), por ejemplo: adenina (A), citosina (C), timina (T) y guanina (G).

Del mismo modo, los aminoácidos se indican utilizando la Norma ST.25 de WIPO, por ejemplo: alanina (Ala; A), arginina (Arg; R), asparagina (Asn; N), ácido aspártico (Asp; D), cisteína (Cys; C), glutamina (Gln; Q), ácido glutámico (Glu; E), glicina (Gly; G), histidina (His; H), isoleucina (Ile; I), leucina (Leu; L), lisina (Lys; K), metionina (Met; M), fenilalanina (Phe; F), prolina (Pro; P), serina (Ser; S), treonina (Thr; T), triptófano (Trp; W), tirosina (Tyr; Y) y valina (Val; V).

Definiciones

A menos que se defina lo contrario, todos los términos técnicos y científicos utilizados en el presente documento tienen el mismo significado entendido comúnmente por un experto en la materia a la que pertenece la materia objeto que se describe en el presente documento.

Aunque se considera que los siguientes términos son de uso común para los expertos en la materia, las siguientes definiciones se exponen para facilitar la comprensión de la materia en cuestión que se describe en el presente documento.

Tal como se utilizan en la descripción de la invención y las reivindicaciones adjuntas, las formas en singular "un", "una" y "el" y "la" pretenden incluir igualmente las formas en plural, a menos que el contexto indique claramente lo contrario.

Tal como se utiliza en el presente documento, la expresión "y/o", contempla y se refiere a todas y cada una de las posibles combinaciones de uno o más de los elementos enumerados asociados, así como también a la carencia de combinaciones cuando se interpreta en la forma alternativa ("o").

El término "aproximadamente" tal como se utiliza en el presente documento cuando se refiere a un valor medible tal

como una dosificación o período de tiempo y similares, pretende abarcar variaciones de $\pm$ un 20%, $\pm$ un 10%, $\pm$ un 5%, $\pm$ un 1%, $\pm$ un 0.5% o incluso $\pm$ un 0.1% de la cantidad especificada. Tal como se utilizan en el presente documento, frases tales como "entre aproximadamente X e Y" significan "entre aproximadamente X y aproximadamente Y" y frases tales como "de aproximadamente X a Y" significan "de aproximadamente X a aproximadamente Y".

Tal como se utilizan en el presente documento, frases tales como "entre aproximadamente X e Y", "entre aproximadamente X y aproximadamente Y", "de X a Y" y "de aproximadamente X a aproximadamente Y" (y frases similares) deben interpretarse que incluyen X e Y, a menos que el contexto indique lo contrario.

Una "secuencia codificante" es una secuencia de ácido nucleico que se transcribe en forma de ARN tal como ARNm, ARNr, ARNt, ARNnp, ARN sentido o ARN antisentido. En realizaciones, el ARN se traduce posteriormente para producir una proteína.

Tal como se utiliza en el presente documento, una secuencia de nucleótidos "con codones optimizados" se refiere a una secuencia de nucleótidos de un polinucleótido recombinante, transgénico o sintético en la que los codones se seleccionan para que reflejen la preferencia codónica concreta que pueda tener una célula u organismo hospedador. Esto se realiza normalmente de tal forma que se preserve la secuencia de aminoácidos del polipéptido codificado por la secuencia de nucleótidos con codones optimizados. En ciertas realizaciones, una secuencia de nucleótidos tiene codones optimizados para la célula (p. ej., una célula animal, vegetal, fúngica o bacteriana) en la que la construcción se

va a expresar. Por ejemplo, una construcción que se va a expresar en una célula vegetal puede tener toda su secuencia o partes de esta con codones optimizados para la expresión en una planta. Véase, por ejemplo, la Pat. de EE.UU. n.º 6,121,014. En realizaciones, los polinucleótidos de la invención tienen codones optimizados para la expresión en una célula vegetal (p. ej., una célula dicotiledónea o una célula monocotiledónea) o una célula bacteriana.

Los términos "comprenden", "comprende" o "que comprende", cuando se utilizan en esta memoria descriptiva, indican la presencia de características, números enteros, etapas, operaciones, elementos o componentes mencionados, pero no excluyen la presencia o adición de una o más características, números enteros, pasos, operaciones, elementos, componentes diferentes y/o grupos de estos.

Tal como se utiliza en el presente documento, la frase de transición "constituido esencialmente por" (y variantes gramaticales) significa que se debe considerar que el alcance de una reivindicación contempla los materiales o etapas especificados que se mencionan en la reivindicación "y aquellos que no afectan materialmente a la(s) característica(s) básica(s) y novedosa(s)" de la invención reivindicada. De este modo, no se pretende que la expresión "constituido esencialmente por", cuando se utiliza en una reivindicación de esta invención, se interprete como equivalente a "que comprende".

"Casete de expresión", tal como se utiliza en el presente documento, significa una molécula de ácido nucleico capaz de dirigir la expresión de al menos un polinucleótido de interés, tal como un polinucleótido del gen R que codifica proteína de la invención, en una célula hospedadora apropiada, que comprende un promotor enlazado operativamente

al polinucleótido de interés que está enlazado operativamente a una señal de terminación. Un "casete de expresión" también suele comprender polinucleótidos adicionales para facilitar la traducción correcta del polinucleótido de interés. El casete de expresión también puede comprender otros polinucleótidos no relacionados con la expresión de un polinucleótido de interés pero que están presentes debido a sitios de restricción convenientes para eliminar el casete de un vector de expresión. En realizaciones, al menos uno de los componentes del casete de expresión puede ser heterólogo (es decir, exógeno) con respecto a al menos uno de los otros componentes (p. ej., un promotor heterólogo asociado operativamente a un polinucleótido de interés). El casete de expresión también puede ser de origen natural, pero que se haya obtenido en una forma recombinante útil para la expresión heteróloga. Normalmente, sin embargo, el casete de expresión es heterólogo con respecto al hospedador, es decir, el casete de expresión (o incluso el polinucleótido de interés) no está presente de manera natural en la célula hospedadora y se ha introducido en la célula hospedadora o en una célula predecesora de esta mediante un proceso de transformación o un proceso de cultivo selectivo. La expresión del (de los) polinucleótido(s) de interés en el casete de expresión está controlada normalmente por un promotor. En el caso de un organismo multicelular, tal como una planta, el promotor también puede ser específico o preferencial para un tejido, u órgano o una etapa de desarrollo particular (tal como se describe más detalladamente en el presente documento). Un casete de expresión, o un fragmento de este, también se puede denominar "polinucleótido insertado" o "polinucleótido de inserción" cuando se utiliza para transformar una planta.

En el presente documento, un "gen" se define como una unidad hereditaria que comprende uno o más polinucleótidos que ocupa una ubicación específica en un cromosoma o un plásmido y que contiene las instrucciones genéticas para una característica o rasgo particular en un organismo.

El término "introducido", tal como se utiliza en el presente documento, en relación a una planta, significa realizado de cualquier manera, incluyendo, pero no limitado a; introgresión, transgénico, modificación de Repeticiones Palindrómicas Cortas Agrupadas y Regularmente Interespaciadas (CRISPR, *Clustered Regularly Interspaced Short Palindromic Repeats*), nucleasas efectoras de tipo activador de la transcripción (TALEN, *Transcription activator-like effector nucleases*) (Feng et al. 2013, Joung y Sander 2013), meganucleasas o nucleasas con dedos de zinc (ZFN, *zinc finger nucleases*).

Tal como se utiliza en el presente documento, la expresión "glicina silvestre" se refiere a una planta perenne de *Glycine*, por ejemplo una cualquiera de *G. canescens*, *G. argyrea*, *G. clandestina*, *G. latrobeana*, *G. albicans*, *G. aphyonota*, *G. arenaria*, *G. curvata*, *G. cyrtoloba*, *G. dolichocarpa*, *G. falcata*, *G. gracei*, *G. hirticaulis*, *G. lactovirens*, *G. latifolia*, *G. microphylla*, *G. montis-douglas*, *G. peratosa*, *G. pescadrensis*, *G. pindanica*, *G. pullenii*, *G. rubiginosa*, *G. stenophita*, *G. syndetika* o *G. tomentella*.

Tal como se utiliza en el presente documento, el término "alelo" se refiere a uno de dos o más nucleótidos o secuencias de nucleótidos diferentes que aparecen en un locus específico.

Un marcador está "asociado a" un rasgo, cuando está enlazado al mismo y cuando la presencia del marcador es un indicador de si y/o en qué medida aparecerá el rasgo o forma

de rasgo deseado en una planta/un germoplasma que comprende el marcador. De manera similar, un marcador está "asociado a" un alelo cuando está enlazado al mismo y cuando la presencia del marcador es un indicador de si el alelo está presente en una planta/un germoplasma que comprende el marcador. Por ejemplo, "un marcador asociado a resistencia a patógenos potenciada" se refiere a un marcador, cuya presencia o ausencia puede utilizarse para predecir si y/o en qué medida una planta exhibirá un fenotipo resistente a patógenos.

Tal como se utilizan en el presente documento, los términos "retrocruce" y "retrocruzamiento" se refieren al proceso mediante el cual una planta descendiente se retrocruza repetidamente con una de sus plantas parentales. En un esquema de retrocruzamiento, la planta progenitora "donante" se refiere a la planta progenitora con el gen o locus deseado que va a someterse a introgresión. La planta progenitora "receptora" (utilizada una o más veces) o la planta progenitora "recurrente" (utilizada dos o más veces) se refiere a la planta progenitora en donde se está realizando la introgresión del gen o locus. Por ejemplo, véase Ragot, M. et al. *Marker-assisted Backcrossing: A Practical Example*, en *TECHNIQUES ET UTILISATIONS DES MARQUEURS MOLECULAIRES LES COLLOQUES*, Vol. 72, págs. 45-56 (1995); y Openshaw et al., *Marker-assisted Selection in Backcross Breeding*, en *PROCEEDINGS OF THE SYMPOSIUM "ANALYSIS OF MOLECULAR MARKER DATA"*, págs. 41-53 (1994). El cruce inicial da lugar a la generación F1. El término "BC1" se refiere al segundo uso del progenitor recurrente, "BC2" se refiere al tercer uso del progenitor recurrente, y así sucesivamente.

Un centimorgan ("cM") es una unidad de medida de frecuencia de recombinación. Un cM equivale a una probabilidad de un 1% de que un marcador en un locus genético

se separe de un marcador en un segundo locus debido al cruzamiento en una única generación.

Tal como se utiliza en el presente documento, la expresión "intervalo cromosómico definido por y que incluye", utilizado en referencia a loci y/o alelos particulares, se refiere a un intervalo cromosómico delimitado por y que abarca los loci/alelos establecidos.

Los términos "cruce" o "cruzado", tal como se utilizan en el presente documento, se refieren a la fusión de gametos mediante la polinización para producir la descendencia (p. ej., células, semillas o plantas). El término engloba tanto cruzamientos sexuales (la polinización de una planta por otra) como la autofecundación (autopolinización, p. ej., cuando el polen y el óvulo proceden de la misma planta). El término "cruzamiento" se refiere al acto de fusionar gametos mediante la polinización para producir la descendencia.

Los términos "cultivar" y "variedad", tal como se utilizan en el presente documento, se refieren a un grupo de plantas similares que, mediante características estructurales o genéticas y/o de rendimiento, se pueden distinguir de otras variedades dentro de la misma especie.

Tal como se utiliza en el presente documento, las expresiones "alelo deseado", "alelo favorable" y "alelo de interés", se utilizan de manera indistinta para referirse a un alelo asociado a un rasgo deseado (p. ej., resistencia a la ASR).

Tal como se utiliza en el presente documento, las expresiones "resistencia a patógenos potenciada" o "resistencia a enfermedades potenciada", se refiere a una mejora, potenciamiento o aumento en la capacidad de la planta de perdurar y/o progresar a pesar de estar infectada con una enfermedad (p. ej., roya de soja asiática) en comparación con

una o más plantas control (p. ej., uno o ambos de los progenitores, o una planta que carece de un marcador asociado a resistencia a patógenos potenciada a respectivos patógenos/enfermedades). Resistencia a enfermedades potenciada incluye cualquier mecanismo (distinto de la inmunidad o resistencia de la planta entera) que reduce la expresión de síntomas indicativos de infección para una enfermedad respectiva tal como la roya de la soja asiática, el nematodo del quiste de soja, *Pytophthora*, etc.

Una "línea de élite" o una "variedad de élite" es una línea agronómicamente superior que ha resultado de muchos ciclos de reproducción y selección para un comportamiento agronómico superior. Hay disponibles numerosas líneas de élite y son conocidas por los expertos en la materia del cultivo de soja. Una "población de élite" es un surtido de individuos o líneas de élite que se pueden utilizar para representar el estado de la técnica en términos de genotipos agronómicamente superiores de una especie de cultivo dada, tal como soja. De manera similar, un "germoplasma de élite" o cepa de germoplasma de élite es un germoplasma agronómicamente superior, derivado normalmente de y/o capaz de dar origen a una planta con comportamiento agronómico superior, tal como una línea de élite de soja existente o recientemente desarrollada.

Una planta de "élite" es cualquier planta de una línea de élite tal que una planta de élite es una planta representativa de una variedad de élite. Ejemplos no limitantes de variedades de soja de élite que se encuentran disponibles comercialmente para granjeros o para obtentores de soja incluyen: AG00802, A0868, AG0902, A1923, AG2403, A2824, A3704, A4324, A5404, AG5903, AG6202 AG0934; AG1435; AG2031; AG2035; AG2433; AG2733; AG2933; AG3334; AG3832;

AG4135; AG4632; AG4934; AG5831; AG6534; y AG7231 (Asgrow Seeds, Des Moines, Iowa, EE. UU.); BPR0144RR, BPR 4077NRR y BPR 4390NRR (Bio Plant Research, Camp Point, Ill., EE. UU.); DKB17-51 y DKB37-51 (DeKalb Genetics, DeKalb, Ill., EE. UU.); DP 4546 RR, y DP 7870 RR (Delta & Pine Land Company, Lubbock, Tex., EE. UU.); JG 03R501, JG 32R606C ADD y JG 55R503C (JGL Inc., Greencastle, Ind., EE. UU.); NKS 13-K2 (NK Division de Syngenta Seeds, Golden Valley, Minnesota, EE. UU.); 90M01, 91M30, 92M33, 93M11, 94M30, 95M30, 97B52, P008T22R2; P16T17R2; P22T69R; P25T51R; P34T07R2; P35T58R; P39T67R; P47T36R; P46T21R; y P56T03R2 (Pioneer Hi-Bred International, Johnston, Iowa, EE. UU.); SG4771NRR y SG5161NRR/STS (Soygenetics, LLC, Lafayette, Ind., EE. UU.); S00-K5, S11-L2, S28-Y2, S43-B1, S53-A1, S76-L9, S78-G6, S0009-M2; S007-Y4; S04-D3; S14-A6; S20-T6; S21-M7; S26-P3; S28-N6; S30-V6; S35-C3; S36-Y6; S39-C4; S47-K5; S48-D9; S52-Y2; S58-Z4; S67-R6; S73-S8; y S78-G6 (Syngenta Seeds, Henderson, Ky., EE. UU.); Richer (Northstar Seed Ltd. Alberta, CA); 14RD62 (Stine Seed Co. Ia., EE. UU.); o Armor 4744 (Armor Seed, LLC, Ar., EE. UU.).

Las expresiones "élite agronómica", tal como se utilizan en el presente documento, significan un genotipo que tiene una culminación de muchos rasgos distinguibles, tales como emergencia, vigor, vigor vegetativo, resistencia a enfermedades, siembra, durabilidad, rendimiento y capacidad de desgrane, que permite a un productor cosechar un producto de importancia comercial.

Tal como se utiliza en el presente documento, la expresión "rendimiento comercialmente significativo" o "rendimiento agronómicamente aceptable" se refiere a una producción de granos de al menos 100% de una variedad de control comercial tal como AG2703 o DKB23-51.

Un ácido nucleico, secuencia de nucleótidos, polipéptido o secuencia de aminoácidos "nativo" o "de tipo silvestre" se refiere a un ácido nucleico, secuencia de nucleótidos, polipéptido o secuencia de aminoácidos presente de manera natural o endógeno. De este modo, por ejemplo, un "ARNm de tipo silvestre" es un ARNm que está presente de manera natural en el organismo o que es endógeno respecto a éste.

Las expresiones "ácido nucleico", "molécula de ácido nucleico", "secuencia de nucleótidos", "oligonucleótido", "ácidos polinucleicos" y "polinucleótido" se utilizan de manera indistinta en el presente documento, a menos que el contexto indique lo contrario, y se refieren a un heteropolímero de nucleótidos. Estas expresiones y términos incluyen, sin carácter limitativo, moléculas de ADN y ARN, incluido ADNc, ADN genómico, ADN y ARN sintéticos (*p. ej.*, sintetizados químicamente), ADN plasmídico, ARNm, ARN antisentido e híbridos de ARN/ADN, pudiendo ser cualquiera de los cuales lineal o ramificado, monocatenario o bicatenario o una combinación de los mismos. Cuando se produce ARNbc sintéticamente, se pueden utilizar bases menos comunes, tales como inosina, 5-metilcitosina, 6-metiladenina, hipoxantina y otras, para el apareamiento de ribozimas, ARNbc y antisentido. Por ejemplo, se ha demostrado que los polinucleótidos que contienen análogos de tipo propino en la posición C-5 de la uridina y citidina se unen a ARN con una alta afinidad y que son inhibidores antisentido potentes de la expresión génica. También se pueden realizar otras modificaciones, tales como la modificación del esqueleto fosfodiéster o el 2'-hidroxi en el grupo de azúcar ribosa del ARN. En realizaciones, "ácido nucleico", "molécula de ácido nucleico", "secuencia de nucleótidos", "oligonucleótido" o "polinucleótido" se refieren a ADN.

Con "enlazado operativamente" o "asociado operativamente", tal como se utilizan en el presente documento, se pretende dar a entender que los elementos indicados están relacionados funcionalmente entre sí, y que por lo general también están relacionados físicamente. De esta manera, la expresión "enlazado operativamente" o "asociado operativamente", tal como se utiliza en el presente documento, se refiere a secuencias de nucleótidos en una molécula de ácido nucleico sencillo que están asociadas de forma funcional. De este modo, una primera secuencia de nucleótidos que se encuentra enlazada operativamente a una segunda secuencia de nucleótidos se refiere a una situación en la que la primera secuencia de nucleótidos se encuentra en una relación funcional con la segunda secuencia de nucleótidos. Por ejemplo, un promotor está asociado operativamente a una secuencia de nucleótidos si el promotor efectúa la transcripción o expresión de dicha secuencia de nucleótidos. Los expertos en la materia apreciarán que no es necesario que las secuencias de control (p. ej., el promotor) estén contiguas a la secuencia de nucleótidos a las que están asociadas operativamente, siempre que las secuencias de control sirvan para dirigir su expresión. De este modo, por ejemplo, pueden estar presentes secuencias intermedias sin traducir, pero transcritas, entre un promotor y una secuencia de nucleótidos y el promotor puede aun así considerarse "enlazado operativamente" a la secuencia de nucleótidos o "asociado operativamente" con ésta.

Tal como se utiliza en el presente documento, las expresiones "tolerancia a la enfermedad" y "resistente a la enfermedad" se refieren a la capacidad de una planta de perdurar y/o progresar, a pesar de estar infectada con una enfermedad respectiva. Cuando se utiliza en referencia a

germoplasma, las expresiones se refieren a la capacidad de una planta que surge del germoplasma de perdurar y/o progresar a pesar de estar infectada con una enfermedad respectiva. En algunas realizaciones, plantas de soja resistentes a enfermedades infectadas pueden producir tan bien (o casi tan bien) como plantas de soja no infectadas. En general, una planta o un germoplasma se marca como "resistente a enfermedades", si muestra "una resistencia a los patógenos potenciada".

Tal como se utiliza en el presente documento, el término "endógeno" se refiere a materiales que proceden del interior de un organismo o célula. "Exógeno" se refiere a materiales que proceden del exterior de un organismo o célula. Esto se aplica normalmente a moléculas de ácido nucleico utilizadas en producir células hospedadoras y plantas transformadas o transgénicas.

Tal como se utiliza en el presente documento, el término "exótico" y las expresiones "línea exótica" y "germoplasma exótico" se refieren a cualquier planta, línea o germoplasma que no es élite. En general, las plantas/germoplasmas exóticos no se han derivado de ninguna planta o germoplasma élite conocidos, sino que en su lugar, se han seleccionado para introducir uno o más elementos genéticos deseados en un programa de reproducción (*p. ej.*, para introducir alelos novedosos en un programa de reproducción).

Tal como se utiliza en el presente documento, un "mapa genético" es una descripción de relaciones de enlace genético entre loci en uno o más cromosomas dentro de un especie dada, generalmente representada en una forma diagramática o tabular. Para cada mapa genético, se miden las distancias entre loci mediante las frecuencias de recombinación entre ellos. Se puede detectar la recombinación entre loci

utilizando diversos marcadores. Un mapa genético es un producto de la población mapeada, tipos de marcadores utilizados y el potencial polimórfico de cada marcador entre diferentes poblaciones. El orden y las distancias genéticas entre locus pueden variar entre un mapa genético y otro.

Tal como se utiliza en el presente documento, el término "genoma", tal como se aplica a células vegetales, abarca no solo ADN cromosómico encontrado dentro del núcleo, sino ADN de organelas encontrado dentro de componentes subcelulares de la célula. El término "gen" se refiere a ácidos polinucleicos que comprenden ADN cromosómico, ADN plasmídico, ADNc, un polinucleótido de ADN artificial u otro ADN que es transcrito en una molécula de ARN, en donde el ARN puede codificar un péptido, polipéptido o proteína y los elementos genéticos que flanquean la secuencia codificante que están implicados en la regulación de la expresión del ARNm o polipéptido de la presente invención. Un "fragmento" de un gen es una porción de una molécula de ácido polinucleico de longitud completa que es de al menos una longitud mínima capaz de transcripción en un ARN, traducción en un péptido, o útil como una sonda o cebador en un método de detección de ADN.

El término "genotipo", tal y como se utiliza en el presente documento, se refiere a la constitución genética de un individuo (o un grupo de individuos) en uno o más loci genéticos, a diferencia del rasgo observable y/o detectable y/o manifestado (el fenotipo). El genotipo está definido por el alelo o los alelos de uno o más loci conocidos que el individuo ha heredado de sus progenitores. El término genotipo se puede utilizar para referirse a la constitución genética de un individuo en un único locus, en múltiples loci o, más generalmente, el término genotipo se puede utilizar para referirse a los caracteres genéticos de un individuo

para todos los genes en su genoma. Los genotipos se pueden caracterizar de forma indirecta, p. ej., empleando marcadores y/o se pueden caracterizar de forma directa mediante la determinación de la secuencia de los ácidos nucleicos.

Tal como se utiliza en el presente documento, el término "germoplasma" se refiere a material genético de o procedente de un individuo (p. ej., una planta), un grupo de individuos (p. ej., una línea, variedad o familia de planta) o un clon derivado de una línea, variedad, especie o cultivo. El germoplasma puede ser parte de un organismo o célula, o puede estar separado del organismo o célula. En general, el germoplasma proporciona material genético con una composición molecular específica que proporciona una base física para algunas o todas las cualidades hereditarias de un organismo o cultivo celular. Tal como se utiliza en el presente documento, germoplasma se puede referir a semillas, células (incluyendo protoplastos y callos) o tejidos a partir de los cuales pueden crecer nuevas plantas, así como partes de plantas que se pueden cultivar en una planta entera (p. ej., tallos, brotes, raíces, hojas, etc.).

Tal como se utiliza en el presente documento, una secuencia de "ADN heterólogo" se refiere a una secuencia de polinucleótidos que procede de una fuente o especie extraña o, si es de la misma fuente, está modificada de su forma original.

Tal como se utiliza en el presente documento, un "ADN homólogo" se refiere a ADN de la misma fuente que la de la célula receptora.

Tal como se utiliza en el presente documento, el término "híbrido" se refiere a una semilla y/o planta producida cuando se cruzan al menos dos parentales genéticamente distintos.

El término "endogámico", tal como se utiliza en el presente documento, se refiere a una planta o variedad sustancialmente homocigótica. El término se puede referir a una planta o variedad que es sustancialmente homocigótica a lo largo de todo su genoma o que es sustancialmente homocigótica con respecto a una porción del genoma que es de un interés particular.

El término "indel", tal como se utiliza en el presente documento, se refiere a una inserción o delección en un par de secuencias de nucleótidos, donde de la primera secuencia se dice que tiene una inserción respecto a una segunda secuencia o de la segunda secuencia se dice que tiene una delección respecto a la primera secuencia.

Los términos y expresiones "introgresión", "introgresar" y "sometido a introgresión", tal como se utilizan en el presente documento, se refieren tanto a la transmisión natural como artificial de un alelo deseado o una combinación de alelos deseados de un locus genético o loci genéticos desde un acervo genético a otro. Por ejemplo, un alelo deseado en un locus específico se puede transmitir a al menos una descendencia mediante un cruce sexual entre dos progenitores de la misma especie, donde al menos uno de los progenitores tenga el alelo deseado en su genoma. Como alternativa, por ejemplo, la transmisión de un alelo se puede producir mediante la recombinación entre dos genomas donantes, p. ej., en un protoplasto fusionado, donde al menos uno de los protoplastos donantes posee el alelo deseado en su genoma. El alelo deseado puede ser un alelo seleccionado de un marcador, un QTL, un transgén o similares. La descendencia que comprende el alelo deseado se puede retrocruzar repetidamente hasta obtener una línea que tiene un acervo genético deseado y seleccionar en función del alelo deseado,

para que el resultado consista en fijar el alelo deseado en el acervo genético deseado. Por ejemplo, un marcador asociado a la tolerancia potenciada a la ASR se puede introgresar a partir de un donante en un progenitor recurrente que no sea resistente a la enfermedad. La descendencia resultante se podría entonces retrocruzar y seleccionar repetidamente hasta que la descendencia poseyera el o los alelos de la tolerancia a la ASR en el acervo del progenitor recurrente.

Tal como se utiliza en el presente documento, una molécula de ácido nucleico "aislada" está sustancialmente separada de otras secuencias de ácidos nucleicos con las que el ácido nucleico está normalmente asociado, tal como del ADN cromosómico o extracromosómico de una célula en la que se produce de forma natural el ácido nucleico. Una molécula de ácido nucleico es una molécula de ácido nucleico aislada cuando comprende un transgén o parte de un transgén presente en el genoma de otro organismo. El término también abarca ácidos nucleicos que están bioquímicamente purificados, con el fin de separar sustancialmente ácidos nucleicos contaminantes y otros componentes celulares. El término "transgén" se refiere a cualquier molécula de ácido polinucleico normativa a una célula u organismo transformado en la célula u organismo. "Transgén" abarca también las partes componentes de un gen de planta nativa, modificadas por la inserción de una molécula de ácido polinucleico normativa mediante recombinación dirigida o mutación específica del sitio.

Un polipéptido está "aislado", si ha sido separado de los componentes celulares (ácidos nucleicos, lípidos, hidratos de carbono, y otros polipéptidos) que lo acompañan de forma natural o que ha sido sintetizado químicamente o es recombinante. Una molécula polipeptídica es una molécula

polipeptídica aislada cuando se expresa a partir de un transgén en otro organismo. Un polipéptido monomérico está aislado, cuando al menos 60% en peso de una muestra está compuesta del polipéptido, preferiblemente 90% o más, más preferiblemente 95% o más, o lo preferiblemente más de 99%. La pureza u homogeneidad de la proteína se indica, por ejemplo, mediante electroforesis en gel de poliacrilamida de una muestra de proteína, seguida de visualización de una banda única de polipéptido tras la tinción del gel de poliacrilamida; cromatografía líquida a alta presión; u otros métodos convencionales. Las proteínas se pueden purificar por cualquiera de los medios conocidos en la técnica, por ejemplo tal como se describe en *Guide to Protein Purification*, ed. Deutscher, Meth. Enzymol. 185, Academic Press, San Diego, 1990; y *Scopes, Protein Purification: Principles and Practice*, Springer Verlag, Nueva York, 1982.

Utilizando métodos bien conocidos, el experto puede producir fácilmente variantes de secuencias de nucleótidos y aminoácidos de genes y proteínas que proporcionan un producto génico modificado. Síntesis químicas de ácidos nucleicos se pueden realizar, por ejemplo, en sintetizadores de oligonucleótidos automatizados. Variantes de este tipo no cambian preferiblemente el marco de lectura de la región codificante de proteínas del ácido nucleico. La presente invención también abarca fragmentos de una proteína que carece de al menos un residuo de una proteína de longitud completa, pero que sustancialmente mantiene la actividad de la proteína.

Un "locus" es una posición en un cromosoma en el que está ubicado un gen o marcador o alelo. En algunas realizaciones, un locus puede abarcar uno o más nucleótidos.

Una "variedad de soja que se produce de forma no natural" es cualquier variedad de soja que no existe de forma natural en la naturaleza. Una "variedad de soja que se produce de forma no natural" se puede producir por cualquier método conocido en el técnica, incluyendo, pero no limitado a, transformar una planta o un germoplasma de soja, transfectar una planta o un germoplasma de soja y cruzar una variedad de soja que se produce de forma natural con una variedad de soja que no se produce de forma natural. En algunas realizaciones, una "variedad de soja que no se produce de forma natural" puede comprender una o más secuencias de nucleótidos heterólogas. En algunas realizaciones, una "variedad de soja que no se produce de forma natural" puede comprender una o más copias que no se produzcan de forma natural de una secuencia de nucleótidos que se produzca de forma natural (es decir, copias extrañas de un gen que se produce de forma natural en la soja). En algunas realizaciones, una "variedad de soja que no se produce de forma natural" puede comprender una combinación no natural de dos o más secuencias de nucleótidos de origen natural (es decir, dos o más genes que se produzcan de forma natural que no existan de forma natural en la misma soja, por ejemplo, genes no encontrados en líneas de *Glycine max*).

Tal como se utilizan en el presente documento el término "fenotipo" y las expresiones "rasgo fenotípico" o "rasgo" se refieren a uno o más rasgos y/o manifestaciones de un organismo. El fenotipo puede ser una manifestación que es observable a simple vista o por cualesquiera otros medios de evaluación conocidos en la técnica, p. ej., microscopía, análisis bioquímico o un ensayo electromecánico. En algunos casos, un fenotipo o rasgo está directamente controlado por un solo gen o locus genético, es decir, un "rasgo de un solo

gen". En otros casos, un fenotipo o rasgo es el resultado de varios genes. Se señala que, tal como se utiliza en el presente documento, la expresión "fenotipo resistente a enfermedades" tiene en cuenta condiciones medioambientales que pudieran afectar a la enfermedad respectiva de modo que el efecto es real y reproducible.

Tal como se utiliza en el presente documento, el término "planta" puede referirse a una planta entera, a cualquier parte de la misma o a un cultivo celular o tisular derivado de una planta. Por lo tanto, el término "planta" puede referirse a cualquiera de: plantas enteras, componentes u órganos de plantas (p. ej., raíces, tallos, hojas, brotes, flores, vainas, etc.), tejidos vegetales, semillas y/o células vegetales. Una célula vegetal es una célula de una planta, extraída de una planta u obtenida mediante el cultivo de una célula extraída de una planta. Por lo tanto, la expresión "planta de soja" puede referirse a planta de soja entera, a una o más partes de una planta de soja (p. ej., raíces, puntas de las raíces, tallos, hojas, brotes, flores, vainas, semillas, cotiledones, etc.), células de la planta de soja, protoplastos de la planta de soja y/o callos de la planta de soja.

Una "célula vegetal" es una unidad estructural y fisiológica de una planta, que comprende un protoplasto y una pared celular. La célula vegetal puede estar en forma de una única célula aislada o de una célula cultivada, o puede formar parte de una unidad con una organización superior tal como, por ejemplo, un tejido vegetal, un órgano vegetal o una planta entera. En realizaciones, la célula vegetal no es propagante y/o no puede regenerar una planta completa.

Un "cultivo de células vegetales" se refiere a un cultivo de unidades vegetales tal como, por ejemplo,

protoplastos, células de cultivo celular, células en tejidos vegetales, polen, tubos polínicos, óvulos, sacos embrionarios, cigotos y embriones en diferentes etapas de desarrollo.

La expresión "material vegetal" se refiere a las hojas, vástagos, raíces, flores o partes de las flores, frutas, polen, óvulos, cigotos, semillas, esquejes, cultivos de tejidos o celulares, o cualquier otra parte o producto de una planta.

Un "órgano vegetal" es una parte distintiva visiblemente diferenciada y estructurada de una planta tal como una raíz, vástago, hoja, botón floral o embrión.

Tal como se utiliza en el presente documento, la expresión "parte de la planta" incluye, sin carácter limitativo, embriones, polen, óvulos, semillas, hojas, flores, ramas, fruta, tallos, raíces, ápices radiculares, anteras y/o células vegetales, incluidas las células vegetales que están intactas en plantas y/o partes de las plantas, protoplastos vegetales, tejidos vegetales, cultivos de tejidos de células vegetales, callos de plantas, agregados vegetales y similares.

La expresión "tejido vegetal", tal como se utiliza en el presente documento, se refiere a un grupo de células vegetales organizadas en una unidad estructural y funcional. Se incluye cualquier tejido de una planta *in planta* o en cultivo. Esta expresión incluye, sin limitación, plantas enteras, órganos vegetales, semillas de plantas, el cultivo tisular y grupos cualesquiera de células vegetales organizadas en unidades estructurales o funcionales. El uso de esta expresión junto con, o en ausencia de, cualquier tipo específico de tejido vegetal, tal como el que se indica anteriormente o que esté contemplado de otra manera por esta

definición, no pretende ser exclusivo de cualquier otro tipo de tejido vegetal.

"Señal de poliadenilación" o "señal poliA" se refiere a una secuencia de ácidos nucleicos localizada en 3' con respecto a una región codificante que provoca la adición de nucleótidos de adenilato en el extremo 3' del ARNm transcrito a partir de la región codificante.

"Reacción en cadena de la polimerasa (PCR)" se refiere a un método de amplificación de ADN que utiliza una técnica enzimática para crear múltiples copias de una secuencia de ácidos nucleicos (amplicón). Copias de una molécula de ADN se preparan lanzando una ADN polimerasa entre dos amplímeros. La base de este método de amplificación son múltiples ciclos de cambios de temperatura para desnaturalizar, después re-asociar amplímeros (moléculas de cebador de ADN), seguido de extensión para sintetizar nuevas hebras de ADN en la región localizada entre los amplímeros flanqueantes. La amplificación de ácidos nucleicos se puede conseguir por cualquiera de los diversos métodos de amplificación de ácidos nucleicos conocidos en la técnica, incluyendo la reacción en cadena de la polimerasa (PCR). En la técnica se conoce una diversidad de métodos de amplificación y se describe, entre otros, en las Pat. de EE. UU. N° 4,683,195 y 4,683,202 y en Protocolos de PCR: A Guide to Methods and Applications, ed. Innis et al., Academic Press, San Diego, 1990. Se han desarrollado métodos de amplificación por PCR para amplificar hasta 22 kb de ADN genómico y hasta 42 kb de ADN bacteriófago (Cheng et al., Proc. Natl. Acad. Sci. USA 91:5695-5699, 1994). Estos métodos, así como otros métodos conocidos en la técnica de la amplificación de ADN se pueden utilizar en la práctica de la presente invención.

Tal como se utiliza en el presente documento, el término "cebador" se refiere a un oligonucleótido que es capaz de asociarse a una diana de ácido nucleico y servir como punto de inicio de la síntesis de ADN cuando se coloca bajo condiciones en las que se induce la síntesis de un producto de extensión de un cebador (p. ej., en presencia de nucleótidos y un agente para la polimerización, tal como ADN polimerasa, y una temperatura y un pH adecuados). Un cebador (en algunas realizaciones un cebador de extensión y en otras realizaciones un cebador de amplificación) en algunas realizaciones es de una sola hebra para obtener la mayor eficiencia de extensión y/o amplificación. En algunas realizaciones, el cebador es un oligodesoxirribonucleótido. Un cebador normalmente es lo suficientemente largo como para cebar la síntesis de productos de extensión y/o amplificación en presencia de un agente para polimerización. Las longitudes mínimas de los cebadores pueden depender de varios factores, incluyendo, pero sin limitación, de la temperatura y composición (contenido A/T vs G/C) del cebador. En el contexto de cebadores de amplificación, estos se proporcionan normalmente como un par de cebadores bidireccionales que consisten en un cebador directo y un cebador inverso o se proporcionan como un par de cebadores directos como se utiliza comúnmente en la técnica de amplificación de ADN, por ejemplo, en la amplificación por PCR. Como tal, se entenderá que el término "cebador" tal como se utiliza en el presente documento, se puede referir a más de un cebador, particularmente en el caso en el que haya algo de ambigüedad en la información que se refiere a la(s) secuencia(s) terminal(es) de la región diana que se quiere amplificar. Por lo tanto, un "cebador" puede incluir un conjunto de oligonucleótidos cebadores que contienen secuencias que

representan las posibles variaciones en la secuencia o incluye nucleótidos que permiten un apareamiento de bases típico. Los cebadores pueden prepararse por cualquier método adecuado conocido en la técnica. Los métodos para preparar oligonucleótidos de secuencia específica son conocidos en la técnica e incluyen, por ejemplo, la clonación y restricción de secuencias apropiadas y la síntesis química directa. Los métodos de síntesis química pueden incluir, por ejemplo, el método de fosfo di- o triéster, el método de dietilfosforamidato y el método en soporte sólido divulgado en la Patente de EE. UU. N.º 4,458,066. Se pueden marcar los cebadores, si se desea, incorporando porciones detectables, por ejemplo, mediante porciones espectroscópicas, fluorescentes, fotoquímicas, bioquímicas, inmunoquímicas o químicas. Pueden crearse cebadores de diagnóstico (es decir, que pueden identificar o seleccionar basándose en la presencia de alelos resistentes a ASR) para resistencia a ASR para cualquiera SNP favorable como se describe en una cualquiera de las tablas 1-5. El método de PCR está bien descrito en los manuales y es conocido para los expertos en la materia. Después de la amplificación mediante PCR, los polinucleótidos diana se pueden detectar mediante hibridación con un polinucleótido sonda que forma un híbrido estable con la secuencia diana, en condiciones de hibridación y de lavado rigurosas a moderadamente rigurosas. Si se espera que las sondas sean esencialmente complementarias por completo (es decir, un 99% o más) con respecto a la secuencia diana, se pueden utilizar condiciones rigurosas. Si se prevé algún desacierto, por ejemplo si se esperan hebras variantes con el resultado de que la sonda no sea completamente complementaria, se puede reducir la rigurosidad de la hibridación. En algunas realizaciones, las condiciones se

eligen para descartar las uniones no específicas/fortuitas. Las condiciones que afectan la hibridación, y que seleccionan para que no se produzcan las uniones no específicas se conocen en la técnica y se describen, por ejemplo, en Sambrook & Russell (2001). Molecular Cloning: A Laboratory Manual, tercera edición, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York, Estados Unidos de América. Generalmente la concentración de sal más baja y la temperatura más elevada aumentan la rigurosidad de las condiciones de hibridación.

Tal como se utiliza en el presente documento, el término "descendencia" y la expresión "planta descendiente" se refieren a una planta generada a partir de una reproducción vegetativa o sexual de una o más plantas progenitoras. Puede obtenerse una planta descendiente mediante la clonación o autofecundación de una sola planta progenitora, o mediante el cruzamiento de dos plantas progenitoras.

El término "promotor" o la expresión "región promotora" se refiere a una molécula de ácido nucleico que funciona como un elemento regulador, que habitualmente se encuentra aguas arriba (5') de una secuencia codificante, que controla la expresión de la secuencia codificante controlando la producción de ARN mensajero (ARNm) proporcionando el sitio de reconocimiento para la ARN polimerasa y/u otros factores necesarios para comenzar la transcripción en el sitio correcto. Tal como se contempla en el presente documento, un promotor o una región promotora incluye variaciones de promotores derivadas por medio de ligadura a diversas secuencias reguladoras, mutagénesis aleatoria o controlada y adición o duplicación de secuencias de potenciador. La región de promotor descrita en el presente documento, y equivalentes biológicamente funcionales de la misma, son los responsables

de impulsar la transcripción de secuencias codificantes bajo su control cuando se introducen en un hospedador como parte de una construcción de ADN recombinante adecuada, tal como se demuestra por su capacidad de producir ARNm.

Un ácido nucleico "recombinante" está hecho de una combinación de dos segmentos separados de otro modo de secuencia de ácidos nucleicos, por ejemplo, por síntesis química o mediante la manipulación de segmentos aislados de ácidos polinucleicos mediante técnicas de ingeniería genética. La expresión "construcción de ADN recombinante" se refiere a cualquier agente, tal como un plásmido, cósmido, virus, secuencia de replicación autónoma, fago o secuencia de nucleótidos de ADN o ARN monocatenario o bicatenario, lineal o circular, derivado de cualquier fuente, capaz de integración genómica o replicación autónoma, que comprende una molécula de ADN a la que se han enlazado una o más secuencias de ADN de una manera funcionalmente operativa. Construcciones de ADN recombinante de este tipo son capaces de introducir una secuencia reguladora o región promotora 5' y una secuencia de ADN para un producto génico seleccionado en una célula de tal manera que la secuencia de ADN se transcribe en un ARNm funcional que se traduce y, por lo tanto, se expresa. Construcciones de ADN recombinante se pueden construir de modo que sean capaces de expresar ARNs antisentido, o ARNs antisentido bicatenarias estabilizadas.

La frase "sustancialmente idénticas", en el contexto de dos secuencias de ácidos nucleicos o dos aminoácidos, se refiere a dos o más secuencias o subsecuencias que tienen al menos aproximadamente 50% de identidad de residuos nucleótidos o aminoácidos cuando se compara y alinea para una máxima correspondencia, tal como se mide utilizando un algoritmo de comparación de secuencias o mediante inspección

visual. En ciertas realizaciones, las secuencias sustancialmente idénticas tienen al menos una identidad entre residuos de nucleótidos o aminoácidos de aproximadamente un 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o más. En determinadas realizaciones, una sustancial identidad existe a lo largo de una región de las secuencias que es al menos aproximadamente 50 residuos de aminoácidos, 100 residuos de aminoácidos, 150 residuos de aminoácidos, 200 residuos de aminoácidos, 250 residuos de aminoácidos, 300 residuos de aminoácidos, 350 residuos de aminoácidos, 400 residuos de aminoácidos, 450 residuos de aminoácidos, 500 residuos de aminoácidos, 525 residuos de aminoácidos, 526 residuos de aminoácidos, 527 residuos de aminoácidos, 528 residuos de aminoácidos, 529 residuos de aminoácidos, 530 residuos de aminoácidos, 531 residuos de aminoácidos, 532 residuos de aminoácidos, 533 residuos de aminoácidos, 534 residuos de aminoácidos, 535 residuos de aminoácidos, 536 residuos de aminoácidos o más con respecto a la secuencia de proteínas o la secuencia de nucleótidos que codifican las mismas. En las realizaciones adicionales, las secuencias son sustancialmente idénticas cuando son idénticas a lo largo de toda la longitud de las regiones codificantes.

"Identidad" o "porcentaje de identidad" se refiere al grado de identidad entre dos secuencias de ácidos nucleicos o aminoácidos. A la hora de comparar secuencias, normalmente una secuencia actúa como secuencia de referencia con la que se comparan las secuencias de ensayo. Cuando se utiliza un algoritmo para la comparación de secuencias, se introducen las secuencias de ensayo y de referencia en una computadora, se designan las coordenadas de las subsecuencias, en caso necesario, y se designan los parámetros del programa

algorítmico de las secuencias. El algoritmo para la comparación de las secuencias calcula entonces el porcentaje de identidad secuencial para la secuencia o las secuencias de ensayo con relación a la secuencia de referencia, en función de los parámetros del programa designados.

A la hora de comparar secuencias, normalmente una secuencia actúa como secuencia de referencia con la que se comparan las secuencias de ensayo. Cuando se utiliza un algoritmo para la comparación de secuencias, se introducen las secuencias de ensayo y de referencia en una computadora, se designan las coordenadas de las subsecuencias, en caso necesario, y se designan los parámetros del programa algorítmico de las secuencias. El algoritmo para la comparación de las secuencias calcula entonces el porcentaje de identidad secuencial para la secuencia o las secuencias de ensayo con relación a la secuencia de referencia, en función de los parámetros del programa designados.

El alineamiento óptimo de las secuencias para la comparación se puede llevar a cabo, *p. ej.*, mediante el algoritmo de homología local de Smith y Waterman, *Adv. Appl. Math.* 2: 482 (1981), mediante el algoritmo de alineación por homología de Needleman y Wunsch, *J. Mol. Biol.* 48: 443 (1970), mediante el método de búsqueda de similitud de Pearson y Lipman, *Proc. Nat'l. Acad. Sci. USA* 85: 2444 (1988), mediante implementaciones computarizadas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), o mediante inspección visual (véase, en general,).

Un ejemplo de un algoritmo que es adecuado para determinar el porcentaje de identidad de secuencia y la similitud secuencial es el algoritmo BLAST, que se describe

en Altschul et al., J. Mol. Biol. 215: 403-410 (1990). Software para realizar análisis BLAST está disponible al público a través del National Center for Biotechnology Information (en la web en ncbi.nlm.nih.gov/). Este algoritmo incluye identificar en primer lugar pares de secuencias de puntuación elevada (HSP, por sus siglas en inglés) mediante la identificación de palabras cortas de longitud W en la secuencia problema que igualen o satisfagan cierta puntuación umbral T evaluada positivamente cuando se alinean con una palabra de la misma longitud en una secuencia de la base de datos. A T se la alude como el umbral de puntuación de palabras de vecindad (Altschul et al., J. Mol. Biol. 215: 403-410 (1990)). Estas coincidencias iniciales de las palabras vecinas actúan como semillas para iniciar las búsquedas con el fin de encontrar los HSP más largos que las contengan. Las coincidencias de las palabras se extienden posteriormente en ambas direcciones a lo largo de cada secuencia en la medida en que se pueda aumentar la puntuación de alineación acumulada. Las puntuaciones acumuladas para las secuencias de nucleótidos se calculan utilizando los parámetros M (puntuación de recompensa para un par de residuos coincidentes; siempre > 0) y N (puntuación de penalización para residuos no coincidentes; siempre < 0). Para las secuencias de aminoácidos, se utiliza una matriz de puntuación para calcular la puntuación acumulada. La extensión de las coincidencias de las palabras en cada dirección se ve interrumpida cuando la puntuación de alineación acumulada disminuye una cantidad X respecto a su valor máximo alcanzado, la puntuación acumulada tiende a cero o por debajo de cero debido a la acumulación de una o más alineaciones de residuos con puntuación negativa, o se alcanza el final de cualquiera de las secuencias. Los

parámetros W, T y X del algoritmo BLAST determinan la sensibilidad y la velocidad de la alineación. El programa BLASTN (para secuencias de nucleótidos) utiliza por defecto una longitud de palabra (W) de 11, una esperanza (E) de 10, un límite de corte de 100, $M = 5$, $N = -4$ y una comparación de ambas hebras. Para secuencias de aminoácidos, el programa BLASTP utiliza como defectos una longitud de palabras (W) de 3, una expectativa (E) de 10 y la matriz de puntuación BLOSUM62 (véase, Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89: 10915 (1989)).

Además de calcular el porcentaje de identidad de secuencia, el algoritmo BLAST realiza también un análisis estadístico de la similitud entre dos secuencias (véase, p. ej., Karlin y Altschul, Proc. Nat'l. Acad. Sci. USA 90: 5873-5877 (1993)). Una medida de la similitud proporcionada por el algoritmo BLAST es la probabilidad de la suma más pequeña ($P(N)$), que proporciona una indicación de la probabilidad de que una coincidencia entre dos secuencias de nucleótidos o de aminoácidos se produzca por casualidad. Por ejemplo, una secuencia de ácido nucleico de ensayo se considera similar a una secuencia de referencia si la probabilidad de la suma más pequeña en una comparación de la secuencia de ácido nucleico de ensayo respecto a la secuencia de ácido nucleico de referencia es inferior a aproximadamente 0.1, más preferentemente inferior a aproximadamente 0.01 y de la manera más preferida inferior a aproximadamente 0.001.

Otro programa informático de uso extendido y aceptado para llevar a cabo las alineaciones de las secuencias es CLUSTALW v1.6 (Thompson, et al., Nuc. Acids Res., 22: 4673-4680, 1994). El número de bases o de aminoácidos coincidentes se divide por el número total de bases o de aminoácidos y se multiplica por 100 para obtener un porcentaje de identidad.

Por ejemplo, si dos secuencias de 580 pares de base tenían 145 bases coincidentes, serían idénticas en un 25 por ciento. Si las dos secuencias comparadas tienen longitudes diferentes, el número de coincidencias se divide por la menor de las dos longitudes. Por ejemplo, si había 100 aminoácidos coincidentes entre proteínas de 200 y 400 aminoácidos, estas son idénticas en un 50 por ciento con respecto a la secuencia más corta. Si la secuencia más corta tiene una longitud menor de 150 bases o 50 aminoácidos, el número de coincidencias se divide por 150 (para las bases de ácidos nucleicos) o 50 (para los aminoácidos) y se multiplica por 100 para obtener un porcentaje de identidad. Además de las posiciones de identidad, también se puntúan comúnmente posiciones consenso. Aminoácidos consenso son aquellos que se sabe que tienen propiedades de aminoácidos similares, tales como carga, tamaño, polaridad y aromaticidad.

También se puede considerar que dos secuencias de nucleótidos son sustancialmente idénticas cuando las dos secuencias se hibridan entre sí en condiciones rigurosas. En las realizaciones representativas, dos secuencias de nucleótidos que se consideran sustancialmente idénticas se hibridan entre sí en condiciones muy rigurosas.

Las expresiones "condiciones rigurosas" o "condiciones de hibridación rigurosas" incluyen la referencia a condiciones en las cuales un ácido nucleico se hibridará selectivamente con una secuencia diana, en un grado detectablemente mayor que con otras secuencias (p. ej., al menos 2 veces con respecto a una secuencia no diana) y de manera opcional puede excluir sustancialmente la unión a las secuencias que no son diana. Las condiciones rigurosas dependen de la secuencia y variarán en circunstancias diferentes. Controlando la rigurosidad de las condiciones de

hibridación y/o de lavado, se pueden identificar secuencias diana que pueden presentar hasta un 100% de complementariedad respecto a la secuencia de nucleótidos de referencia. Como alternativa, se pueden emplear condiciones de rigurosidad moderada o incluso baja para permitir ciertas discrepancias en las secuencias de forma que se detecten grados de similitud secuencial menores. Por ejemplo, los expertos en la materia apreciarán que para actuar como un cebador o sonda, una secuencia de ácido nucleico únicamente necesita ser lo suficientemente complementaria con la secuencia diana para unirse sustancialmente a esta de modo que forme una estructura bicatenaria estable en las condiciones empleadas. Por lo tanto, se pueden utilizar cebadores o sondas en condiciones de rigurosidad elevada, moderada o incluso baja. Asimismo, las condiciones de rigurosidad baja o moderada pueden ser convenientes para detectar secuencias homólogas, ortólogas y/o parálogas que tienen grados menores de identidad secuencial de los que se hubieran identificado en condiciones muy rigurosas.

Los términos "complementario" o "complementariedad" (y términos similares), tal como se utilizan en el presente documento, se refieren a la unión natural de los polinucleótidos mediante el apareamiento de bases en condiciones permisivas de temperatura y sales. Por ejemplo, la secuencia "A-G-T" se une a la secuencia complementaria "T-C-A". La complementariedad entre dos moléculas monocatenarias puede ser parcial, donde únicamente algunos de los nucleótidos se unen, o puede ser completa cuando existe una complementariedad total entre las moléculas monocatenarias. El grado de complementariedad entre las hebras de ácidos nucleicos tiene efectos significativos sobre la eficacia y fuerza de hibridación entre las moléculas. Como se utiliza en

el presente documento, la expresión "sustancialmente complementario" (y términos similares) significa que dos secuencias de ácidos nucleicos son complementarias en al menos aproximadamente un 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% o más. Como alternativa, la expresión "sustancialmente complementario" (y expresiones similares) puede significar que dos secuencias de ácidos nucleicos se pueden hibridar entre sí en condiciones de rigurosidad elevada (tal como se describe en el presente documento).

Tal como se utiliza en el presente documento, se hibrida "específicamente" o "selectivamente" (y términos similares) se refiere a la unión, duplexación o hibridación de una molécula a una secuencia objetivo de ácidos nucleicos particular bajo condiciones rigurosas cuando la secuencia está presente en una mezcla compleja (p. ej., ADN o ARN celular total) a la exclusión sustancial de ácidos nucleicos no objetivo, o incluso con unión, duplexación o hibridación no detectable a secuencias no objetivo. Las secuencias que se hibridan específica o selectivamente que son, de manera habitual, complementarias en al menos aproximadamente un 40% y son, de manera opcional, sustancialmente complementarias o incluso completamente complementarias (es decir, idénticas en un 100%).

Para híbridos ADN-ADN, se puede aproximar la T_m a partir de la ecuación de Meinkoth y Wahl, Anal. Biochem., 138:267-84 (1984): $T_m = 81.5\text{ }^{\circ}\text{C} + 16.6 (\log M) + 0.41 (\% \text{ de GC}) - 0.61 (\% \text{ de formamida}) - 500/L$; donde M es la molaridad de cationes monovalentes, % de GC es el porcentaje de nucleótidos de guanósina y citosina en el ADN, % de formamida 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 (a un pH y una fuerza iónica definidos) a la que

un 50% de una secuencia diana complementaria se hibrida con una sonda perfectamente coincidente. La T_m se reduce en aproximadamente 1 °C por cada 1% de discrepancias; por lo tanto, la T_m y las condiciones de hibridación y/o lavado se pueden ajustar para que se hibriden con secuencias con el grado de identidad deseado. Por ejemplo, si se buscan secuencias con >90% de identidad, la T_m se puede reducir 10 °C. En general, se seleccionan condiciones rigurosas para que sean aproximadamente 5 °C inferiores que el punto de fusión térmico (T_m) para la secuencia específica y su complemento a una fuerza iónica y pH definidos. Sin embargo, las condiciones muy rigurosas pueden emplear una hibridación y/o un lavado en el punto de fusión térmica (T_m) o 1, 2, 3 o 4 °C por debajo del punto de fusión térmica (T_m); las condiciones de rigurosidad moderada pueden utilizar una hibridación y/o un lavado 6, 7, 8, 9 o 10 °C por debajo del punto de fusión térmica (T_m); las condiciones de rigurosidad baja pueden emplear una hibridación y/o un lavado 11, 12, 13, 14, 15 o 20 °C por debajo del punto de fusión térmica (T_m). Si el grado deseado de emparejamiento erróneo resulta en una T_m menor que 45 °C (solución acuosa) o 32 °C (solución de formamida), opcionalmente se puede incrementar la concentración de SSC, de modo que se puede utilizar una temperatura más elevada. Se puede consultar una guía exhaustiva sobre la hibridación de ácidos nucleicos en Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, parte I, capítulo 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, Nueva York (1993); y en Current Protocols in Molecular Biology, capítulo 2, Ausubel et al., eds., Greene Publishing y Wiley-Interscience, Nueva York (1995); y Green y Sambrook, en:

Molecular Cloning, A Laboratory Manual, 4.^a edición, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (2012).

Normalmente, las condiciones rigurosas son aquellas en las que la concentración salina es inferior a aproximadamente 1.5 M para el ion Na, normalmente una concentración de aproximadamente 0.01 a 1.0 M para el ion Na (u otras sales) de aproximadamente un pH de 7.0 a un pH de 8.3, y la temperatura es de al menos aproximadamente 30 °C para sondas cortas (p. ej., de 10 a 50 nucleótidos) y de al menos aproximadamente 60 °C para sondas largas (p. ej., de más de 50 nucleótidos). Las condiciones rigurosas también se pueden lograr mediante la adición de agentes desestabilizantes tales como formamida o solución de Denhardt (5 g de Ficoll, 5 g de polivinilpirrolidona, 5 g de albúmina sérica bovina en 500 ml de agua). Las condiciones ilustrativas de rigurosidad baja incluyen la hibridación con una solución de tampón de un 30% a un 35% de formamida, NaCl 1 M, 1% de SDS (dodecilsulfato sódico) a 37 °C y un lavado en 1X-2X SSC (20X SSC = NaCl 3.0 M/citrato trisódico 0.3 M) a 50 °C-55 °C. Las condiciones ilustrativas de rigurosidad moderada incluyen la hibridación en un 40% a 45% de formamida, NaCl 1 M, 1% de SDS a 37 °C y un lavado en 0.5X-1X SSC a 55 °C-60 °C. Las condiciones ilustrativas de rigurosidad elevada incluyen la hibridación en un 50% de formamida, NaCl 1 M, 1% de SDS a 37 °C y un lavado en 0.1X SSC a 60 °C-65 °C. Un ejemplo adicional no limitativo de condiciones de rigurosidad elevada incluye la hibridación en 4X SSC, 5X solución de Denhardt, 0.1 mg/mL de ADN de esperma de salmón hervido, y fosfato de Na 25 mM a 65 °C y un lavado en 0.1X SSC, 0.1% de SDS a 65 °C. Otro ejemplo de condiciones de hibridación de rigurosidad elevada incluye la hibridación en 7% de SDS, NaPO₄ 0.5 M, EDTA 1 mM a 50 °C con un lavado en 2X SSC, 0.1% de SDS a 50 °C, de manera

alternativa con un lavado en 1X SSC, 0.1% de SDS a 50 °C, de manera alternativa con un lavado en 0.5X SSC, 0.1% de SDS a 50 °C, o de manera alternativa con un lavado en 0.1X SSC, 0.1% de SDS a 50 °C, o incluso con un lavado en 0.1X SSC, 0.1% de SDS a 65°C. Los expertos en la materia apreciarán que la especificidad normalmente es una función de los lavados post-hibridación, siendo factores relevantes la fuerza iónica y la temperatura de la solución de lavado final.

Los ácidos nucleicos que no se hibridan entre sí en condiciones rigurosas siguen siendo sustancialmente idénticos si las proteínas que estos codifican son sustancialmente idénticas (p. ej., debido a la degeneración del código genético).

Una indicación adicional de que dos ácidos nucleicos o proteínas son sustancialmente idénticos es que la proteína codificada por el primer ácido nucleico presente reactividad cruzada inmunológicamente con la proteína codificada por el segundo ácido nucleico. Por lo tanto, normalmente una proteína es sustancialmente idéntica a una segunda proteína, por ejemplo, si las dos proteínas se diferencian únicamente por sustituciones conservadoras.

El término "vector" se refiere a una composición para transferir, suministrar o introducir un ácido nucleico (o ácidos nucleicos) en una célula. Un vector comprende una molécula de ácido nucleico que comprende la o las secuencias de nucleótidos que se han de transferir, suministrar o introducir.

Expresión de una Secuencia Codificante de un gen R en Plantas

Se preparan construcciones de ADN que contienen diversos elementos genéticos, necesarios para la expresión de una secuencia codificante de un gen R en plantas, transmitiendo de este modo resistencia a patógenos a las plantas en las que

se expresan. "Construcción de ADN" se refiere a los elementos genéticos heterólogos enlazados operativamente entre sí que constituyen una molécula de ADN recombinante y pueden comprender elementos que proporcionan expresión de una molécula de polinucleótido de ADN en una célula hospedadora y elementos que proporcionan el mantenimiento de la construcción en la célula hospedadora. Un casete de expresión en plantas comprende el enlace operativo de elementos genéticos que, cuando se transfieren a una célula vegetal, proporciona la expresión de un producto génico deseable. "Casete de expresión en plantas" se refiere a segmentos de ADN quimérico que comprenden los elementos reguladores que están enlazados operativamente para proporcionar la expresión de un producto transgénico en plantas. Promotores, conductores, intrones, ácidos polinucleicos codificantes de péptidos de tránsito, regiones de terminación de la transcripción 3' son todos elementos genéticos que pueden ser enlazados operativamente por los expertos en la materia de la biología molecular vegetal para proporcionar un nivel de expresión o funcionalidad deseable a un gen R de la presente invención. Una construcción de ADN puede contener uno o más casetes de expresión en plantas que expresan las moléculas de ADN de la presente invención u otras moléculas de ADN útiles en la ingeniería genética de plantas de cultivo.

Se puede utilizar una diversidad de promotores específicamente activos en tejidos vegetativos, tales como hojas, tallos, raíces y tubérculos, para expresar las moléculas de ácidos polinucleicos del gen R de la presente invención.

La secuencia conductora de la traducción significa una molécula de ADN situada entre el promotor de un gen y la secuencia codificante. La secuencia conductora de la

traducción está presente en el ARNm totalmente procesado aguas arriba de la secuencia de inicio de la traducción. La secuencia conductora de la traducción puede afectar al procesamiento del transcrito primario a ARNm, la estabilidad de ARNm o la eficacia de la traducción. Ejemplos de secuencias conductoras de la traducción incluyen conductores de proteínas de choque térmico de maíz y petunia, conductores de proteínas de recubrimiento de virus de plantas, conductores de genes rubisco, entre otros (Turner y Foster, *Molecular Biotechnology* 3:225, 1995).

Las "secuencias no traducidas 3'" significan secuencias de ADN situadas aguas abajo de una secuencia polinucleotídica estructural e incluyen secuencias que codifican señales de poliadenilación y otras señales reguladoras capaces de afectar al procesamiento del ARNm o la expresión génica. La señal de poliadenilación actúa en plantas provocando la adición de nucleótidos adenilato múltiples en el extremo 3' del precursor de ARNm. La secuencia de poliadenilación se puede derivar del gen natural, de una diversidad de genes de plantas o de T-ADN. Un ejemplo de la secuencia de poliadenilación es la secuencia nopalina sintasa 3' (nos 3'; Fraley et al., *Proc. Natl. Acad. Sci. USA* 80: 4803-4807, 1983). El uso de diferentes secuencias no traducidas 3' se ejemplifica por parte de Ingelbrecht et al., *Plant Cell* 1:671-680, 1989.

Los procedimientos de laboratorio en tecnología de ADN recombinante utilizados en el presente documento son aquellos bien conocidos y comúnmente empleados en la técnica. Se utilizan técnicas convencionales para la clonación, el aislamiento, la amplificación y purificación de ADN y ARN. Generalmente, reacciones enzimáticas que implican ADN ligasa, ADN polimerasa, endonucleasas de restricción y similares se

realizan de acuerdo con las especificaciones del fabricante. Estas técnicas y otras diversas técnicas se realizan generalmente de acuerdo con Sambrook et al. (1989).

La construcción de ADN de la presente invención se puede introducir en el genoma de un hospedador de la planta deseada mediante una diversidad de técnicas de transformación convencionales que son bien conocidas por los expertos en la materia. "Transformación" se refiere a un proceso de introducir una molécula de ácido polinucleico exógena (por ejemplo, una construcción de ADN, una molécula de ácido polinucleico recombinante) en una célula o protoplasto y a que esa molécula de ácido polinucleico exógena se incorpora en un genoma de la célula hospedadora o en un genoma de la organela (por ejemplo, cloroplasto o mitocondria) o es capaz de replicación autónoma. "Transformado" o "transgénico" se refiere a una célula, tejido, órgano u organismo, en el que un ácido polinucleico extraño, tal como un vector de ADN o una molécula de ácido polinucleico recombinante. Una célula o un organismo "transgénico" o "transformado" incluye también la descendencia de la célula u organismo producida a partir de un programa de reproducción que emplea una planta "transgénica" como un progenitor en un cruce y que exhibe un fenotipo alterado que resulta de la presencia de la molécula de ácido polinucleico extraña.

Métodos de transformación de células o tejidos vegetales incluyen, pero no se limitan a un método de transformación mediada por *Agrobacterium* y a la biolística o método de transformación mediada por pistola de partículas. Vectores para transformación de plantas adecuados para el fin de la transformación mediada por *Agrobacterium* incluyen los elementos derivados de un plásmido inductor de tumores (Ti) de *Agrobacterium tumefaciens*, por ejemplo, regiones del borde

derecho (RB) y regiones del borde izquierdo (LB) y otras descritas por Herrera-Estrella et al., Nature 303:209 (1983); Bevan, Nucleic Acids Res. 12:8711-8721 (1984); Klee et al., Bio-Technology 3(7):637-642 (1985). Además de vectores para transformación de plantas derivados de los plásmidos Ti o inductores de raíces (Ri) de *Agrobacterium*, se pueden utilizar métodos alternativos para insertar construcciones de ADN de esta invención en células vegetales. Métodos de este tipo pueden implicar, pero no se limitan, por ejemplo, al uso de liposomas, electroporación, productos químicos que aumentan la absorción de ADN, el suministro de ADN libre a través de bombardeo de microproyectiles y transformación utilizando virus o polen.

Se pueden preparar construcciones de ADN que incorporan las secuencias codificantes del gen R de la presente invención para uso en dirigir la expresión de las secuencias directamente a partir del plástido de la célula vegetal hospedadora. Ejemplos de construcciones de este tipo adecuadas para este propósito y métodos que son conocidos en la técnica se describen generalmente, por ejemplo, en Svab et al., Proc. Natl. Acad. Sci. USA 87:8526-8530, (1990) y Svab et al., Proc. Natl. Acad. Sci. USA 90:913-917 (1993) y en la Pat. de EE. UU. n.º 5,693,507.

Cuando se obtienen números adecuados de células que contienen los polipéptidos que codifican moléculas de ácido polinucleico exógeno de la presente invención, las células se pueden cultivar y luego regenerar en plantas enteras. "Regeneración" se refiere al proceso de cultivar una planta a partir de una célula de la planta (por ejemplo, protoplasto vegetal o explante). Técnicas de regeneración de este tipo se basan en la manipulación de determinadas fitohormonas en un medio de crecimiento de cultivo tisular, que se basa

normalmente en un marcador biocida y/o herbicida que ha sido introducido junto con las secuencias de nucleótidos deseadas. La elección de la metodología para la etapa de regeneración no es crítica. Véase, por ejemplo, Ammirato et al., Handbook of Plant Cell Culture—Crop Species. Macmillan Publ. Co. (1984); Shimamoto et al., Nature 338:274-276 (1989); Fromm, UCLA Simposio sobre Estrategias Moleculares para la Mejora de Cultivos, 16-22 de abril de 1990. Keystone, Colo. (1990); Vasil et al., Bio/Technology 8:429-434 (1990); Vasil et al., Bio/Technology 10:667-674 (1992); Hayashimoto, Plant Physiol. 93:857-863 (1990); y Datta et al., Bio-technology 8:736-740 (1990). Técnicas de regeneración de este tipo se describen, en general, en Klee et al., Ann. Rev. Plant Phys. 38:467-486 (1987).

El desarrollo o la regeneración de plantas transgénicas que contienen la molécula de ácido polinucleico exógeno que codifica un polipéptido de interés es bien conocido en la técnica. Preferiblemente, las plantas regeneradas se autopolinizan para proporcionar plantas transgénicas homocigóticas, tal como se ha comentado anteriormente. De lo contrario, el polen obtenido a partir de las plantas regeneradas se cruza con plantas cultivadas a partir de semillas de líneas importantes desde un punto de vista agronómico. Y a la inversa, el polen de plantas de estas líneas importantes se utiliza para polinizar plantas regeneradas.

Plantas y germoplasmas de soja resistentes a enfermedades

La presente invención proporciona plantas y germoplasmas de soja resistentes a enfermedades. Una planta o un germoplasma de soja resistente a enfermedades se puede producir por cualquier método, en el que un gen R se introduce en la planta o el germoplasma de soja, que incluye,

pero no se limita a transformación, transformación o fusión de protoplastos, una técnica haploide doble, rescate de embriones, edición de genes y/o por cualquier otro sistema de transferencia de ácidos nucleicos.

En algunas realizaciones, la planta o el germoplasma de soja comprende una variedad de soja que se produce de forma no natural. En algunas realizaciones, la planta o germoplasma de soja es al menos aproximadamente un 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 99% o 100% idéntico al de una variedad de élite de soja.

La planta o el germoplasma de soja resistente a enfermedades puede ser la descendencia de un cruce entre una variedad de élite de soja y una variedad de soja que comprende un gen R para tolerancia reforzada a las enfermedades (p. ej., ASR), en donde el gen R se selecciona del grupo que consiste en genes que codifican una proteína que tiene un motivo de gen R de TIR o un gen que tiene entre el 70 y el 100% de homología con cualquiera de las SEQ ID NO: 7, 8, 11 o 12 mientras que proporciona resistencia a la ASR en la planta. En muchos ejemplos, realizaciones alternativas tendrán al menos uno de al menos un 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, o 99% de homología con cualquiera de las SEQ ID NO: 7, 8, 11 o 12.

La planta o el germoplasma de soja resistente a enfermedades puede ser la descendencia de una introgresión, en donde el progenitor recurrente es una variedad de élite de soja y el donante comprende un gen R asociado a tolerancia y/o resistencia reforzada a las enfermedades, en donde el donante porta un motivo de gen R de TIR o un gen que tiene entre el 70 y el 100% de homología con cualquiera de las SEQ ID NO: 7, 8, 11 o 12.

La planta o el germoplasma de soja resistente a enfermedades puede ser la descendencia de un cruce entre una primera variedad de élite de soja (p. ej., una línea de ensayo) y la descendencia de un cruce entre una segunda variedad de élite de soja (p. ej., un progenitor recurrente) y una variedad de soja que comprende un gen R.

Una planta o un germoplasma de soja resistente a enfermedades de la presente invención puede comprender uno o genes R de la presente invención (p. ej., cualquiera de las SEQ ID NO: 7, 8, 11 o 12).

Semillas de soja resistentes a enfermedades

La presente invención proporciona semillas de soja resistentes a enfermedades. Tal como se ha comentado anteriormente, los métodos de la presente invención se pueden utilizar para identificar, producir y/o seleccionar una semilla de soja resistente a enfermedades. Además de los métodos descritos anteriormente, una semilla de soja resistente a enfermedades se puede producir por cualquier método, en el que un gen R se introduce en la semilla de soja, que incluye, pero no se limita a transformación, transformación o fusión de protoplastos, una técnica haploide doble, rescate de embriones, edición de genes (p. ej., CRISPR o TALEN o MegaNucleasas) y/o por cualquier otro sistema de transferencia de ácidos nucleicos.

En algunas realizaciones, la semilla de soja resistente a enfermedades comprende una variedad de soja que se produce de forma no natural. En algunas realizaciones, la semilla de soja es al menos aproximadamente un 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 99% o 100% idéntica a la de una variedad de élite de soja.

La semilla de soja resistente a enfermedades se puede producir por una planta de soja resistente a enfermedades

identificada, producida o seleccionada por los métodos de la presente invención.

Una semilla de soja resistente a enfermedades de la presente invención puede comprender, seleccionarse o producirse mediante el uso de uno o más genes R de la presente invención.

EJEMPLOS

No se pretende que los siguientes ejemplos sean un catálogo detallado de todas las diferentes formas en que se puede implementar la presente invención ni de todas las características que se pueden añadir a la presente invención. Los expertos en la materia se darán cuenta de que se pueden realizar numerosas variaciones y adiciones a las diferentes realizaciones sin alejarse de la presente invención. Por lo tanto, se pretende que las siguientes descripciones ilustren algunas realizaciones particulares de la invención pero no se pretende que especifiquen exhaustivamente todas sus permutaciones, combinaciones y variaciones.

Ejemplo 1 Identificación de líneas de *Glycine silvestre* resistentes a ASR

Se evaluaron múltiples líneas de glicina silvestre (*Glycine canescens*, *Glycine clandestine*, *Glycine tomentella*) en cuanto a la resistencia a la ASR frente a dieciséis cepas de roya recogidas en una gama diversa de entornos. Los datos de las royas se generaron utilizando aislados derivados de pústula única de USDA-ARS (FL Q09, FL Q12, LABR13, FLQ11) y poblaciones de campo (FL Q15, NC06, Vero, GLC15, UBL, BR sur y BR central). El cribado se realizó en instalaciones de contención.

Cada línea de glicina silvestre se evaluó durante un curso de múltiples días de infección y se clasificó en diversos puntos temporales. Cada acceso se cribó >2 veces con ~4 plantas cada vez en América del Norte y del Sur utilizando un gran panel diverso de aislados de roya.

De las líneas de silvestre cribadas para determinar la resistencia a ASR, las siguientes líneas mostraron una amplia resistencia contras todas las cepas de ASR sometidas a ensayo: PI440935, PI483193, PI595799, PI339656 y PI509501.

Ejemplo 2 Minería de Alelos y Asociaciones a Loci de ASR **PI440935, PI483193, PI595799, PI339656 o PI509501**

Se cruzó una línea parental resistente con una línea susceptible de la misma especie (es decir, *G. canescens* x *G. canescens* y *G. clandestina* x *G. clandestina*) y se generó una planta F1 (véase la Tabla 5). La planta F1 se autofertilizó y se cosechó la semilla F2 de la planta F1 autofertilizada. Se sembraron alrededor de 200 semillas F2 y se recogió tejido foliar de cada planta para preparaciones de ADN y luego se inocularon plantas con *Phakopsora pachyrhizi* para determinar la resistencia/el fenotipo susceptible de cada individuo F2. Se combinó tejido de 50 F2 resistentes y 50 F2 susceptibles

en agrupaciones separadas y el ADN genómico se preparó de cada una de las agrupaciones. Se prepararon bibliotecas de secuenciación Illumina a partir de ADN de cada una de las agrupaciones y cada biblioteca se secuenció en dos carriles de extremos apareados (PE) de 2 x 100 pb de Illumina HiSeq2000. El rendimiento medio por muestra fue de 383 millones de pares leídos, lo que equivale a 77 gigabases de secuencia por biblioteca. Las lecturas de secuenciación se recortaron para retirar bases con puntuaciones de calidad PHRED de <15.

Las lecturas recortadas en calidad se alinearon a la secuencia del genoma de referencia PI440935 y PI595799 para *G. canescens* y al genoma de referencia PI339656 para *G. clandestina* utilizando GSNAP (WU y NACU 2010) como fragmentos de extremos apareados. En caso de no poderse alinear entre sí un par de lecturas, estas se trataron como singulones para la alineación. Las lecturas se utilizaron en posteriores análisis si estas se cartografiaron únicamente a la referencia (≤ 2 apareamientos erróneos cada 36 pb y menos de 5 bases cada 75 pb como colas).

Los SNP se filtraron antes del análisis por segregación a granel (BSA) en base a la profundidad de lectura, reteniéndose los SNP que tenían entre 40 y 200x de profundidad de lectura. Se utilizó un ensayo de ji al cuadrado para seleccionar los SNP con recuentos de lectura significativamente diferentes entre los dos alelos en las dos agrupaciones. Se utilizó un enfoque Bayesiano empírico (LIU et al. 2012) para estimar la probabilidad condicionada de que no exista recombinación entre cada uno de los marcadores de SNP y el locus causal tanto en la agrupación resistente como en la agrupación susceptible. La probabilidad de enlace entre el SNP y el gen causal es la media geométrica de estas dos

probabilidades condicionadas. Se encontró que alrededor de 1000 SNP tenían un posible enlace al locus diana. Se utilizó un subconjunto de estos SNP supuestamente enlazados para cartografiar con precisión el locus utilizando individuos F2 fenotipados. Véanse las referencias: LIU, S., C.-T. YEH, H. M. TANG, D. NETTLETON Y P. S. SCHNABLE, 2012 Gene Mapping via Bulk Segregant RNA-Seq (BSR-Seq). PLoS ONE 7: e36406 & Wu, T. D. y S. Nacu, 2010 Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26: 873-881.

TABLA 5: Cruces de plantas y tipo de estudio

Especie	N.º PI (masculino)	N.º PI (femenino)	Tipo de Estudio	F2: Proporci ón de resisten tes con respecto a suscepti bles
G. canescens	PI 440935	PI 505154	BSA	3:1
G. canescens	PI 440935	PI 505154	Indexación	3:1
G. canescens	PI 483193	PI 505154	BSR + DEG	3:1
G. canescens	PI 595799	PI 505154	Indexación	3:1
G. <i>clandestina</i>	PI 339656	PI 446943	BSA	3:1

1. Metodología de laboratorio Data2Bio LLC (Ames, Iowa) de PI440935 para el análisis de la secuencia de gBSA para soja tetraploide.

El descubrimiento de cromosomas para loci causales en la población de soja tetraploide, PI440935 se llevó a cabo utilizando la tecnología de análisis segregante genómico a granel (gBSA) de Data2Bio. Se elaboró la teoría de que la resistencia está controlada por un único gen dominante. Data2Bio generó varias bibliotecas a partir de muestras de ADN extraídas de dos grupos de tejidos susceptibles y dos grupos de tejidos resistentes, y los secuenciaron en ocho (8) carriles de extremos pareados (PE) de 2 x 100 pb de Illumina HiSeq2000. Un resumen de los genomas de referencia utilizados para los análisis posteriores, el procesamiento de la lectura a partir de datos sin procesar para el recorte de calidad, el alineamiento, el descubrimiento de SNP y el impacto de SNP se demuestran en las figuras 11-13. Tras distintas etapas de filtrado, se identificaron 1,099,754 y 324,931 SNP informativos en el genoma PI440935. Se utilizó después un enfoque Bayesiano para calcular las probabilidades asociadas al rasgo. A continuación, se crearon mapas de SNP asociados al rasgo (corte de probabilidad en 0.01) para los contigios superiores (figuras 13-15). Se identificó un armazón, armazón 000056F, y se cartografió en el cromosoma 14 de PI440953 (SEQ ID NO: 1). Los SNP de este armazón enriquecido se cartografiaron en el genoma de *G. max* público. También se observó en el cromosoma 6 un grupo adicional más pequeño de SNP de otros armazones (véanse las figuras 11-12).

2. Metodología de laboratorio Data2Bio LLC (Ames, Iowa) de PI595799 para el análisis de la secuencia de gBSA para soja tetraploide

El descubrimiento de cromosomas para loci causales en la población de soja tetraploide, PI595799 se realizó. Se creó un mapa físico de SNP asociados al rasgo en cóntigos. La agrupación de estos SNP sugiere que el gen de resistencia está ubicado en o cerca del armazón 000090F (SEQ ID NO: 2). Las secuencias contexto asociadas a estos SNP se alinearon también al genoma público de *G. max* para crear una comprensión al nivel de cromosoma del intervalo de cartografiado. Las posiciones cromosómicas de los SNP asociados a los rasgos se mostraron a continuación gráficamente.

3. Metodología de laboratorio Data2Bio LLC (Ames, Iowa) de PI339656 para el análisis de la secuencia de gBSA para soja tetraploide.

El descubrimiento de cromosomas para loci causales en la población de soja tetraploide, PI339656 se llevó a cabo utilizando la tecnología de análisis segregante genómico a granel (gBSA) de Data2Bio. Se teorizó que la resistencia está controlada por un solo alelo dominante. Data2Bio generó varias bibliotecas a partir de muestras de ADN extraídas de dos grupos de tejidos susceptibles y dos grupos de tejidos resistentes, y los secuenciaron en ocho (8) carriles de extremos pareados (PE) de 2 x 100 pb de Illumina HiSeq2000. Un sumario de los genomas de referencia utilizados para los análisis posteriores, el procesamiento de la lectura a partir de datos sin procesar para el recorte de calidad, el alineamiento, el descubrimiento de SNP y el impacto de SNP se demuestran en las figuras 4-6. Tras distintas etapas de filtrado, se identificaron 4,610,234 SNP informativos en el genoma PI339656. Se utilizó después un enfoque Bayesiano para calcular las probabilidades asociadas al rasgo. A

continuación, se crearon mapas de SNP asociados al rasgo (corte de probabilidad en 0.01) para los cóntigos superiores (figuras 13-15). Se identificó un armazón, armazón 14851, y se cartografió en el cromosoma 14 de PI339656 (SEQ ID NO: 1). Los SNP de este armazón enriquecido se cartografiaron en el genoma de *G. max* público. Los datos sugieren que el gen R causal puede cartografiarse dentro o cerca del intervalo de 3.18 a 3.5 MB del cromosoma 14.

Ejemplo 3 Rescate del embrión e introgresión de intervalos del gen R en líneas de *Glycine max*

Se realiza el rescate de embriones y se aplica un tratamiento químico con el fin de generar brotes anfidiploides. Si las plantas anfidiploides son fértiles, se utilizarán para el retrocruzamiento con *G. max*. Será necesario realizar el retrocruzamiento con *G. max* y el posterior rescate del embrión para varias generaciones a fin de eliminar gradualmente los cromosomas de *Glycine* silvestre perenne.

Cruces amplios: Se utilizaron las líneas de soja de élite Syngenta (RM 3.7 a 4.8) como líneas femeninas (receptoras de polen) y se utilizaron *Glycine canescens*, *G. clandestine* y *G. tomentella* con múltiples números de acceso como líneas masculinas o donantes de polen. La selección de flores de la planta de glicina que contengan anteras en la etapa de desarrollo adecuada es importante. Las flores nuevas, totalmente abiertas, y de colores brillantes conservan anteras con polen maduro. El polen debería aparecer como un polvo amarillo suelto. Se retiran estas flores de la planta de glicina y se llevan a la planta de soja para la polinización. El polen de las plantas de *Glycine* debe utilizarse en los 30 minutos posteriores a la retirada de la flor. También es importante identificar y seleccionar

capullos de flores de soja que estén listos para la polinización. Un capullo de una flor está generalmente listo cuando tiene un mayor tamaño cuando se compara con un capullo inmaduro. Los sépalos de flores de soja son de un color más claro y los pétalos están justo comenzando a aparecer. Primeramente, se utiliza un par de pinzas de punta fina para retirar con cuidado los sépalos del capullo de flor para dejar al descubierto el conjunto externo de pétalos. Después, se cogen suavemente y se separan los pétalos (5 en total) de la flor, dejando al descubierto el anillo de estambres que rodean al pistilo. Dado que el estigma es receptivo al polen 1 día antes de que las anteras comiencen a desprenderse, es importante reconocer el desarrollo de fases de "flor femenina lista, masculina no lista". Cuando se polinizan flores de soja en esta fase del desarrollo, no es necesario castrar a la flor femenina. Se localiza el estigma en la flor de soja. Después, utilizando 1 flor masculina, se separan con cuidado los pétalos para dejar al descubierto las anteras y se espolvorean suavemente los granos de polen sobre el estigma de la flor de soja. Debe tenerse cuidado de no dañar al estigma en ningún momento durante este proceso. Partiendo del día después de la polinización, se pulveriza una mezcla de hormonas sobre la flor polinizada y la vaina F1 en desarrollo eventual 1 vez al día hasta la cosecha. La flor polinizada o la vaina se satura con una nube ligera de la mezcla de hormonas, teniendo cuidado de no provocar que la flor/vaina se separe prematuramente de la planta. La mezcla contiene 100 mg de GA3, 25 mg de NAA y 5 mg de quinetina/l de agua destilada. Estas hormonas ayudan a la retención de la vaina en desarrollo y a un mayor crecimiento de la vaina.

Cosecha: Se cosechan vainas de cruces amplios aproximadamente de 14 a 16 días después de la polinización.

(Las fechas de cosecha del material publicado sugieren de 19 a 21 días). Antes de seleccionar una vaina individual para la cosecha, se comprueba la retirada de los sépalos (esto indica un intento de cruce amplio) y que el tamaño de la semilla sea el esperado para un cruce amplio. Las vainas se recolectan y recuentan de acuerdo con una combinación de cruce amplio para determinar el éxito del cruce. El éxito medio del cruce a través de múltiples sojas femeninas y 5 accesos diferentes de *Glycine canescens* es de aproximadamente el 40%. Las vainas de cruce amplio pueden contener 1 a 3 semillas, pero generalmente en cada una de las vainas de F1 se encuentran 2 semillas.

Rescate de embriones: Se recogen las vainas cosechadas y se devuelven al laboratorio para ser esterilizadas. Las vainas se aclaran primero con EtOH al 70% durante 2 a 3 minutos y luego se colocan en blanqueador Clorox al 10% durante 30 minutos adicionales en un agitador de plataforma a aproximadamente 130 rpm. Finalmente, las vainas se aclaran múltiples veces con agua estéril para separar cualquier blanqueador residual. El aislamiento de los embriones puede comenzar inmediatamente después de la esterilización de las vainas, o las vainas se pueden almacenar a 4 °C durante hasta 24 horas antes del aislamiento de los embriones. Las vainas esterilizadas se llevan luego a una campana de flujo laminar en donde se pueden rescatar los embriones. Las vainas individuales se colocan en una placa de Petri y se abren utilizando un bisturí y forceps. Se practica una incisión a lo largo de la vaina de cruce amplio lejos de la semilla. La vaina se puede entonces abrir fácilmente para dejar la semilla al descubierto. Alternativamente, se pueden utilizar dos pares de forceps para separar la envuelta de la vaina. Se separa con cuidado la semilla de la vaina y se coloca en una

placa de Petri bajo el microscopio de disección. Se necesitan forceps muy finos para aislar el embrión de la semilla. Con el forceps en una mano, se mantiene suavemente el lado de la semilla alejado del embrión, con el hilio hacia arriba. Se utiliza otro par de forceps en la otra mano para separar la cubierta de la semilla del lado de la semilla que contiene el embrión. Se pela la membrana que rodea al embrión y se empuja el embrión hacia arriba desde su lado inferior. Los embriones deberían pasar la fase de desarrollo globular y preferiblemente pasar la fase temprana de desarrollo del corazón (se desean embriones con fase del corazón media a tardía, fase de cotiledón y fase de maduración temprana). Los embriones aislados se transfieren al medio de rescate de embriones, tal como Soy ER1-1. Los embriones pueden ser tratados para inducir la duplicación del cromosoma en este momento. (Véase más adelante para detalles de duplicación del cromosoma). Los embriones aislados permanecen en medio de rescate de embriones durante de 21 a 30 días a 24 °C. Los embriones pueden permanecer en la oscuridad durante toda la incubación en ER1-1, pueden comenzar la incubación en la oscuridad y completarla a la luz, o pueden pasar toda la incubación a la luz. No hay una etapa de inducción de callos en este protocolo. Los brotes se desarrollan directamente de los embriones.

Tratamientos de duplicación de los cromosomas: Para inducir la duplicación de los cromosomas se puede utilizar colchicina o trifluralina. De manera ideal, se tratan embriones de cruce amplio de fase de corazón tardía (o mayores) químicamente para inducir la duplicación de cromosomas en cualquier momento desde inmediatamente después del aislamiento hasta 1 semana después del aislamiento. El agente de duplicación se puede mezclar en medio sólido o

líquido y aplicar durante varias horas o hasta unos pocos días. Se utiliza trifluralina a una concentración de 10 - 40 uM en medios sólidos o líquidos. Además, se utiliza colchicina a una concentración de 0.4 - 1 mg/ml en medios sólidos o líquidos. Después del tratamiento químico, los embriones se transfieren a medio de rescate de embriones reciente.

Regeneración de los brotes: Los embriones en desarrollo se transfieren del medio de rescate al medio de germinación, tal como Soy ER GSMv2, durante aproximadamente 3 a 5 semanas a la luz a 24 °C. Alternativamente, los embriones en desarrollo pueden transferirse del medio de rescate a medio de elongación, tal como Soy E1 0 No TCV, durante aproximadamente 3 a 5 semanas a la luz a 24 °C. Los brotes en desarrollo pueden transferirse de placas de medios a recipientes de cultivo PhytoCon que contienen medio de germinación o de alargamiento para un mayor desarrollo de brotes. Los brotes establecidos se mueven al suelo. El cuidado inicial de la planta es crítico para la supervivencia de estos brotes.

Análisis de la ploidía: El análisis de la ploidía se realiza utilizando un citómetro de flujo. Se extrae tejido foliar para el análisis de la ploidía de pequeños brotes en cultivo o tras establecerse en el suelo. Se recoge tejido en hielo seco y se almacena a -80 °C hasta el análisis, o se recoge en hielo seco y se analiza el mismo día. Es suficiente un tamaño de muestra de 0.5 cm². Se preparan muestras de acuerdo con las instrucciones del kit Sysmex. Cada uno de los conjuntos de muestras contiene una planta F1 no tratada (no tratada para inducir la duplicación de los cromosomas) como control. A continuación, se presentan notas del método: Cruces amplios: la tasa de éxito de cruces amplios de los

presentes inventores es significativamente mayor que la presentada en el material publicado. No se realiza la castración de las flores femeninas, lo que ahorra tiempo y reduce el riesgo de dañar el estigma.

Cosecha: cosecha de vainas a los 14 a 16 días de la polinización (~ 5 días antes que la del material publicado, lo que reduce la línea de tiempo)

Rescate de embriones: el protocolo de rescate de embriones de los presentes inventores implica la regeneración directa de brotes a partir de embriones, en lugar de la regeneración a través de la embriogénesis, lo que hace que la recuperación de la planta sea más rápida (recuperación de brotes en de aproximadamente 2 a 3 meses, en comparación con la línea de tiempo informada de hasta 1 año). El presente protocolo no incluye el cultivo en la oscuridad después de la transferencia al medio de germinación. El presente protocolo no requiere una transferencia al medio de enraizamiento.

Duplicación de los cromosomas: se ha utilizado con éxito el tratamiento con trifluralina para producir una planta F1 duplicada: el tratamiento con trifluralina se puede realizar inmediatamente después del rescate del embrión.

Ejemplo 4 Introgresión del rasgo de resistencia a ASR

Líneas anfidiploides generadas a partir del cruce amplio (es decir, *Glycine canescens* o *Glycine clandestina* cruzada con *Glycine max*), seguido del rescate de embriones tal como se describe en el Ejemplo 3, se retrocruzaron múltiples veces un línea de élite recurrente de *Glycine max*. Se sabe en la técnica que se necesitan múltiples retrocruzamientos para generar líneas híbridas fértiles, en particular, la bibliografía sugiere que es necesaria una generación BC3. En este caso, se determinó que son necesarios retrocruzamientos adicionales, p. ej. BC4 en el caso de *G. tomentella* x *G. max*

para generar plantas híbridas fértiles. Plantas híbridas F1, producidas por los métodos según se describe anteriormente, se crearon a partir de cruces amplios que comprenden PI440935, PI483193, PI595799 y PI339656. Las plantas F1 a continuación se cruzaron como femenina con una planta de *G. max* recurrente masculina para realizar un primer retrocruzamiento (descendencia BC1). La descendencia BC1 se retrocruzó adicionalmente para múltiples generaciones (p. ej., BC2). Las plantas BC se evalúan para la resistencia a ASR, los números de cromosomas y, en algunos casos, las líneas se genotipan mediante el uso de marcadores moleculares como se describe en este documento para detectar la presencia de intervalos cromosómicos correspondientes a las SEQ ID NO: 1-4 o cualquier marcador identificado en las tablas 1-4.

Tabla 1. SNPS a modo de ejemplo del Armazón 56F

Nombre del SNP	Posición (SEQ ID NO: 1)	Alelo favorable (resistente a la roya)	Alelo desfavorable (susceptible a la roya)
56F	3046682	T	C
56F	3065127	T	C
56F	3131652	C	T
56F	3159094	T	C
56F	3239763	T	A
56F	3265198	T	A
56F	3383844	C	A
56F	3390809	T	A
56F	3404306	G	C
56F	3410070	C	T
56F	3527033	A	C

Tabla 2. SNPS a modo de
ejemplo del Armazón 90F

Nombre del SNP	Posición (SEQ ID NO: 1)	Alelo favorable (resistente a la roya)	Alelo desfavorable (susceptible a la roya)
90F	33235	A	T
90F	33551	T	G
90F	33556	T	G
90F	33586	G	T
90F	33738	C	T
90F	34016	T	A
90F	34054	A	G
90F	34299	C	A
90F	34696	C	A
90F	35273	G	T
90F	37176	C	T
90F	37222	A	T
90F	37838	G	C
90F	37841	C	T
90F	37989	G	T
90F	38358	C	T

Tabla 3. SNPS a modo de
ejemplo del Armazón 132FF

Nombre del SNP	Posición (SEQ ID NO: 1)	Alelo favorable (resistente a la roya)	Alelo desfavorable (susceptible a la roya)
000132F	1036303	T	G
000132F	1066085	T	C

000132F	1085690	G	A
000132F	1088855	G	A
000132F	1090029	C	T
000132F	1095842	G	A
000132F	1104225	G	A
000132F	1105287	A	G
000132F	1105896	C	T
000132F	1130936	C	T
000132F	1131612	G	T
000132F	1135444	A	G
000132F	1147457	T	G
000132F	1148939	G	A
000132F	1180730	G	T
000132F	1233671	C	T
000132F	1253344	T	C
000132F	1266772	G	A
000132F	1276594	G	C
000132F	1282394	G	A
000132F	1293582	G	A
000132F	1340471	T	C
000132F	1420388	T	C
000132F	1483043	C	A
000132F	1501651	T	G
000132F	1504883	C	A
000132F	1508972	C	A
000132F	1564562	G	T
000132F	1589089	A	G
000132F	1597852	T	C
000132F	1607501	T	A
000132F	1615924	A	G
000132F	1679827	A	G

000132F	1694841	C	T
000132F	1731209	T	G
000132F	1747932	G	C
000132F	1749670	A	G
000132F	1775700	T	C
000132F	1788283	A	T
000132F	1792363	A	G
000132F	1801200	A	G
000132F	1814495	T	G
000132F	1817365	C	A
000132F	1820064	T	C

Tabla 4. SNPS a modo de ejemplo del Armazón 14851

Nombre del SNP	Posición (SEQ ID NO: 1)	Alelo favorable (resistente a la roya)	Alelo desfavorable (susceptible a la roya)
14851	61341	T	C
14851	61461	G	T
14851	91811	A	G
14851	92509	A	G
14851	92510	T	A
14851	116144	C	A
14851	116271	G	C
14851	116676	G	T
14851	118844	G	A
14851	119667	A	G
14851	119964	C	A
14851	120680	C	G
14851	167877	T	A
14851	168709	C	T

14851	169754	T	G
14851	186158	T	G
14851	188346	A	G
14851	189085	A	T
14851	192255	G	A
14851	195663	A	C
14851	209847	C	T
14851	211195	G	A
14851	212035	G	A
14851	239552	G	A
14851	240480	G	T
14851	240526	C	T
14851	259531	C	A
14851	260142	C	G
14851	261948	C	T
14851	263314	G	T
14851	298445	T	C

Los SNPS de la tabla 1-4 se proporcionan solamente a modo de ilustración. Se pueden encontrar SNPS adicionales para las líneas de PI mencionadas anteriormente en las Solicitudes de Patente de EE. UU. N.º 62/881035 y 62/881008, que se incorporan en el presente documento por referencia.

Ejemplo 5 Identificación de dos genes causantes ubicados dentro de intervalos del cromosoma 14.

El genotipado adicional de los diversos intervalos de *G. canescens* dio lugar al descubrimiento de dos genes causantes para resistencia a ASR ubicados en el cromosoma 14 dentro de los intervalos divulgados. El tipo de gen codifica un motivo de gen R de TIR y se representa en las SEQ ID NO: 7 y 11 como ejemplo. Se contempla que cualquiera de estos genes u homólogos de los mismos se pueda emplear en un método

transgénico, editor de genes o de reproducción utilizando el rescate de embriones tal como se describe anteriormente para generar plantas que tengan resistencia aumentada a la ASR. Además, u opcionalmente, las secuencias de polinucleótidos que tienen una identidad de secuencia del 70-100% con estos genes (o las SEQ ID NO 7 y 11) pueden emplearse en el método de reproducción para generar las plantas descritas anteriormente.

Ejemplo 6 Construcción de vectores que comprenden genes R y transformación

Los componentes genéticos cGcaRG3a-01 (TIR-A) y cGcaRG3b-02 (TIR-B) se apilaron juntos en la forma nativa que incluye los exones e intrones. Sin embargo, TIR-B se expresaba como una forma truncada N-terminal en la que se retiraron los primeros 90 nucleótidos. La transcripciones de TIR-A y TIR-B fueron impulsadas por las secuencias promotoras y terminadoras de *Medicago truncatula* prMt51186-03 y tMt51186-01, y prMt12344-02 y tMt12344-01, respectivamente. Estos casetes génicos individuales se apilaron juntos en el vector binario 24217 (Figura 1).

A continuación se describe brevemente las características del vector 24217.

cGcaRG3b-02 (TIR-B). Inicio: 2226 Final: 3955. Un gen candidato de resistencia a la roya de la soja, PI595799 de *Glycine canescens*, que codifica una proteína que contiene dos dominios de receptor Toll/Interleucina-1 (TIR). La versión 2 se basa en un modelo génico alternativo que es 90 pb más corto que la versión 1. El gen es sintético a Glyma.14G024500 (Soja_william82_v2)

cGcaRG3a-01 (TIR-A). Inicio: 7499 Final: 10058. Un gen candidato de resistencia a la roya de la soja, PI595799 de *Glycine canescens*, que codifica una proteína que contiene dos

dominios de receptor Toll/Interleucina-1 (TIR). El gen es sintético a Glyma.14G024400 (Soja_william82_v2).

cmEPSPSpt-29. Inicio: 13162 Final: 14619. EPSPS de *A. tumefaciens* de clase 2 modificado con codón de soja dirigido a plástido con una modificación T273S que tiene una tolerancia al glifosato Ki superior a dmEPSPS *in vitro*.

cSpec-03. Inicio: 15398 Final: 16186. Gen que codifica la enzima aminoglicósido 3'adeniltransferasa que confiere resistencia a espectinomicina y estreptomicina para el mantenimiento del vector en *E. coli* y *Agrobacterium*.

cVirG-01. Inicio: 16486 Final: 17211.

cRepA-03. Inicio: 17241 Final: 18314.

bNRB-04. Inicio: 4 Final: 143 (Complementario). Región del borde derecho de ADN-T de plásmido ti de nopalina de *Agrobacterium tumefaciens*.

bNRB-01-01. Inicio: 101 Final: 125 (Complementario). Repetición del borde derecho

bNLB-05. Inicio: 14989 Final: 15118 (Complementario). Región del borde izquierdo de ADN-T de plásmido ti de nopalina de *Agrobacterium tumefaciens*. Difiere de bNLB-03 en un par de bases para retirar el codón de inicio ATG para ORF apuntando a la región flanqueante.

bNLB-01-01. Inicio: 15024 Final: 15048 (Complementario). Región de repetición del borde izquierdo de 25 pb de ADN-T de plásmido ti de nopalina de *Agrobacterium tumefaciens*.

intrón1. Inicio: 2706 Final: 3054.

intrón2. Inicio: 3566 Final: 3843.

iMt51186-04. Inicio: 7217 Final: 7467. El primer intrón del gen de *Medicago truncatula* identificado.

iGmEF-02. Inicio: 12233 Final: 13141. El primer intrón del gen del factor de elongación (EF) de la soja. iGmEF-01

con un sitio BamHI interno y un ORF retirado no intencional de extremo 3'.

xTAG-06. Inicio: 144 Final: 183. Un sitio de 40 pb para ensayos para determinar si están intactos los insertos de plantas y para detener la lectura de los ORF.

xPARADAS-01. Inicio: 184 Final: 195. Parada de 6 marcos para minimizar la lectura no intencionada del ORF.

Inicio. Inicio: 2041 Final: 2041. El sitio de inicio de la transcripción basado en la alineación de ADNc/ADNg.

PARADA. Inicio: 3853 Final: 3855.

Inicio. Inicio: 7080 Final: 7080. Inicio de la transcripción basado en la alineación de ADNc/ADNg.

Parada. Inicio: 10364 Final: 10364. Parada de la transcripción basada en la alineación de ADNc/ADNg.

xPARADAS-01. Inicio: 11075 Final: 11086. Parada de 6 marcos para minimizar la lectura no intencionada del ORF.

Inicio. Inicio: 12170 Final: 12170. Sitio de inicio de la transcripción.

xPARADAS-01. Inicio: 14620 Final: 14631. Parada de 6 marcos para minimizar la lectura no intencionada del ORF.

xPARADAS-01. Inicio: 14929 Final: 14940. Parada de 6 marcos para minimizar la lectura no intencionada del ORF.

xTAG-02. Inicio: 14941 Final: 14980. Un sitio de 40 pb para ensayos para determinar si están intactos los insertos de plantas y para detener la lectura de los ORF.

prMt12344-02. Inicio: 215 Final: 2216. El promotor del gen de *Medicago truncatula*.

prMt51186-03. Inicio: 4984 Final: 7488. Un promotor del gen de *Medicago truncatula*; diferente de prMt51186-01 en una sustitución de un pb (pos2374) para retirar una mutación de alérgeno de 8 monómeros.

prGmEF-05. Inicio: 11087 Final: 13152. Promotor del factor de elongación de traducción EF-1 alfa/Tu.

prVirG-01. Inicio: 16281 Final: 16411. Promotor virG compuesto por dos elementos promotores, uno que responde a la acetosiringona y a la inanición de fosfatos (pb 45 a 83) y otro a la acidificación del medio (86 a 128).

oVS1-02. Inicio: 18357 Final: 18761. Origen de replicación y región de reparto del plásmido pVS1 de *Pseudomonas*; sirve como origen de replicación en el hospedador *Agrobacterium tumefaciens*.

oCOLE-06. Inicio: 19439 Final: 20245 (Complementario). El origen de replicación ColE1 funcional en *E. coli* derivado de pUC19.

tMt12344-01. Inicio: 3962 Final: 4968. El terminador basado en el gen de *Medicago truncatula*.

tMt51186-01. Inicio: 10068 Final: 11067. El terminador basado en el gen de *Medicago truncatula*.

tNOS-05-01. Inicio: 14654 Final: 14906. Terminador de la sintetasa de nopalina sintética.

u5GmEF-01. Inicio: 12170 Final: 12232. Primera UTR 5' del gen del factor de elongación (EF) de la soja.

u5GmEF-02. Inicio: 13142 Final: 13152. Segunda UTR 5' del gen del factor de elongación (EF) de la soja.

xPSIp28-03. Inicio: 13162 Final: 13251. Péptido de tránsito de plástido optimizado con codones de soja derivado del gen PSI P28 de *Chlamydomonas reinhardtii*.

La transformación y la selección de los eventos de soja primarios se consiguió usando *Agrobacterium* con glifosato como agente de selección (prGmEF-05: cmEPSPSpt-29: tNOS-05-01).

El análisis Taqman de estos eventos de soja primarios identificó tres eventos que albergan los casetes génicos *TIR*-

A y TIR-B, sobreviviendo solamente dos eventos (GVG00885565; GVG00885566) hasta la madurez (Tabla 2A).

Tabla 2A: Sumario del análisis Taqman (datos en bruto y llamada de copia) realizado en eventos primarios de soja de la construcción 24217.

ID de la planta	Pedigrí	cm EPSP ^{spt-29}	cm EPSP ^{spt-29}	cGcaRG3b-02+Mt12344-01	cGcaRG3b-02+Mt12344-01	cGcaRG3a-01	cGcaRG3a-01
GVG00885565	SYDC181509A001A	1.25	1	0.92	1	6.88	>2
GVG00885566	SYDC181501A001A	0.95	1	0.84	1	1.09	1
GVG00885564	SYDC181509A002A	0.52	1	0.58	1	1.04	1

Las tablas 2B, 2C y 2D describen sondas y cebadores utilizables para determinar la presencia de los genes TIR-A y TIR-B, p. ej., en una planta transformada o en su descendencia. En algunos ejemplos pueden ser útiles además para evaluar la introgresión de los genes TIR-A o TIR-B, p. ej., mediante reproducción. Por consiguiente, la presente divulgación también se refiere a la detección de los genes nuevos divulgados en el presente documento en plantas, por ejemplo, soja.

Tabla 2B: Ensayo de TIR-B

Nombre	SEQ ID NO: 23	SEQ ID NO: 24	SEQ ID NO: 25
Tipo de Oligo	CEBADOR	CEBADOR	SONDA
Tipo de tecnología	TAQ	TAQ	TAQ
Secuencia de nucleótidos	ACCCTCTCATTGGAAT CTCTAAATATT	GAGTGTATCCGTGG CAAGTAAGG	AGACATAGATGCTCCATA GTTACCGCCCTG

Tabla 2C: Ensayo de TIR-A

Nombre	SEQ ID NO: 26	SEQ ID NO: 27	SEQ ID NO: 28
Tipo de	CEBADOR	CEBADOR	SONDA

Oligo			
Tipo de tecnología	TAQ	TAQ	TAQ
Secuencia de nucleótidos	CATTGGAAGCATCCA GGATTTTC	TTCAAGGCACCAC GAGGAA	ATCGTTGTTCTCTCCAAA AGCTTCGCG

Tabla 2D: Ensayo de TIR-A

Nombre	SEQ ID NO: 29	SEQ ID NO: 30	SEQ ID NO: 31
Tipo de Oligo	CEBADOR	CEBADOR	SONDA
Tipo de tecnología	TAQ	TAQ	TAQ
Secuencia de nucleótidos	TGTGTGAGAGAGTTTCT TGCTTGAA	TGCACTGGTTACGTTTT TCTTTATCT	CCGTCTTAAAGT AATCAAT

Ejemplo 7 Validación de los genes R contra ASR

Se colocaron hojas de los eventos primarios GVG00885565 y GVG00885566 en una placa de Petri sobre una toalla de papel húmeda y después se inocularon con una suspensión de esporas de roya de la soja RTP1 o FLQ12. Hojas de las plantas de soja de tipo silvestre sirvieron como controles negativos. Después de 14 días, se evaluó la resistencia de estas hojas frente a la roya de la soja. Como se muestra en la figura 2, las hojas de GVG00885565 muestran pruebas claras de resistencia a la roya de la soja RTP1 y FLQ12 en comparación con el control de tipo silvestre. Las hojas de GVG00885566 muestran una resistencia parcial.

Con el fin de cuantificar el nivel de resistencia a la enfermedad, se tomaron muestras de estas hojas para un análisis por qRT-PCR para cuantificar la expresión relativa del gen de la β -tubulina de la roya de la soja (Figura 3).

Los resultados del análisis por qRT-PCR muestran que el nivel de expresión de β -tubulina de la roya tanto RTP1 como FLQQ12 se reduce drásticamente en GVG00885565 en comparación con el control de tipo silvestre. El evento GVG00885566 solo tiene una resistencia parcial. En el mismo experimento se determinó la expresión relativa de los genes TIR-A y TIR-B (Figura 4). Estos datos muestran que los genes tanto TIR-A como TIR-B se expresan en los eventos GVG00885565 y GVG00885566.

Ejemplo 8 Transformación estable con vectores de genes candidatos de ASR

Se sembraron semillas de la primera generación (generación T1) del evento GVG00885565 en macetas de 4 pulgadas y se cultivaron en condiciones de invernadero. Estas plántulas se muestrearon para ensayos por análisis de Taqman para determinar la cigosidad de segregación (homocigótica, heterocigótica o nula) de cada plántula T1. El análisis por qRT-PCR de plantas homocigóticas y heterocigóticas confirmó que la expresión tanto de TIR-A como de TIR-B se ha heredado de manera estable del evento primario a la generación posterior (Figura 5).

Las plántulas homocigóticas y nulas fueron inoculadas con una suspensión de esporas de roya de la soja de tres poblaciones diferentes (RTP1; BRO1-Brasil; BRO3-Brasil). A los 14 días de la inoculación, estas plantas se clasificaron en cuanto al desarrollo de la enfermedad de la roya de la soja. Las plantas nulas son completamente susceptibles a las tres poblaciones de roya, mientras que las plantas

homocigóticas GVG00885565 *TIR-A* y *TIR-B* tienen un tipo de lesión típica resistente rojo/marrón (RB) con muy poca esporulación (figura 6).

En este mismo experimento, las hojas de las plántulas GVG00885565 nulas y homocigóticas fueron las muestras para el análisis por qRT-PCR para determinar la expresión relativa del gen de la β -tubulina de la roya de la soja. Estos datos muestran un nivel significativamente superior de expresión de β -tubulina de la roya de la soja en las plantas nulas 8, 14 y 21 días después de la inoculación en comparación con las plantas homocigóticas *TIR-A* y *TIR-B* (figura 7).

Ejemplo 9. Cotransformación

Se crearon eventos primarios de soja mediante la cotransformación con dos cepas diferentes de *Agrobacterium* que albergaban vectores binarios independientes para *TIR-A* (construcción 24205; SEQ ID NO: 32) y *TIR-B* (construcción 24192; SEQ ID NO: 33) Figuras 8A y 8B. Estas construcciones independientes expresan los casetes génicos idénticos utilizados en la pila molecular de la construcción 24217. El análisis Taqman de los eventos primarios de soja de los experimentos de cotransformación identificó cuatro eventos con *TIR-A* y *TIR-B* (Tabla 3A). La expresión tanto de *TIR-A* como de *TIR-B* se confirmó mediante qRT-PCR.

Tabla 3A: Sumario del análisis Taqman y por qRT-PCR de los eventos primarios de soja cotransformados con los vectores binarios independientes 24205 y 24192. Nótese que el análisis por qRT-PCR confirma que los genes *TIR-A* y *TIR-B* se expresan en estos eventos primarios.

ID de la planta	cGcaRG 3a-01	cGcaRG 3a-01	cGcaRG 3b-02	cGcaRG 3b-02	qRT-PCR <i>TIR-A</i>	qRT-PCR <i>TIR-B</i>
GVG011738	0.90	1	2.02	2	2.34	89.4

20						
GVG011738						
23	0.87	1	3.30	>2	8.74	112.17
GVG011743						
13	0.80	1	0.68	1	2.41	12.49
GVG011755						
68	1.65	2	1.83	2	2.79	24.87
GVG011758						
42	0.95	1	11.67	>2	0.27	127.15

Dos de estos eventos primarios GVG01173820 y GVG01173823 se sometieron a ensayo para determinar la resistencia a diferentes poblaciones de roya de la soja RTP1, BRO1-Brasil y BRO3-Brasil. Se recogieron hojas y se colocaron en una placa de Petri con una toalla de papel húmeda. Las hojas fueron inoculadas con una suspensión de esporas de roya de la soja. A los 14 días de la inoculación, se muestrearon las hojas para el análisis por qRT-PCR de la expresión del gen de la β -tubulina de la roya de la soja. Los resultados muestran que el evento GVG01173820 tiene una expresión del gen de la β -tubulina significativamente menor frente a las tres poblaciones de roya, mientras que GVG01173823 es más resistente a las poblaciones de roya de Brasil BRO1 y BRO3 (figura 9).

Ejemplos adicionales de construcciones de vectores

A modo de ejemplo, se proporcionan también otras construcciones de vectores.

a) Construcción de Vector 23891 para el Gen R que comprende TIR-A y TIR-B derivados de PI339656 de *G. clandestina*

SEQ ID NO: 1 corresponde al Vector 23891. La figura 10 proporciona una ilustración del vector 23891. Se describen características a continuación.

1° TIR-. Inicio: 1868 Final: 3892.

2° TIR-. Inicio: 5957 Final: 8167.

cNtALS-01. Inicio: 12147 Final: 14141. El fragmento de ADN NtALS codifica un mutante doble de Acetolactato sintasa (P191A, W568L) de *Nicotiana tabacum*.

cSpec-03. Inicio: 15491 Final: 16279. Un gen que codifica la enzima aminoglicósido 3'adeniltransferasa que confiere resistencia a espectinomicina y estreptomycin para el mantenimiento del vector en *E. coli* y *Agrobacterium*.

cVirG-01. Inicio: 16579 Final: 17304.

cRepA-03. Inicio: 17334 Final: 18407.

bNRB-04. Inicio: 4 Final: 143 (Complementario). Región del borde derecho de ADN-T de plásmido ti de nopalina de *Agrobacterium tumefaciens*.

bNRB-01-01. Inicio: 101 Final: 125 (Complementario).

bNLB-05. Inicio: 15082 Final: 15211 (Complementario). Región del borde izquierdo de ADN-T de plásmido ti de nopalina de *Agrobacterium tumefaciens*. Difiere de bNLB-03 en un par de bases para retirar el codón de inicio ATG para ORF apuntando a la región flanqueante.

bNLB-01-01. Inicio: 15117 Final: 15141 (Complementario). Región de repetición del borde izquierdo de 25 pb de ADN-T de plásmido ti de nopalina de *Agrobacterium tumefaciens*.

iGMEF-02. Inicio: 11216 Final: 12124. El primer intrón del gen del factor de elongación (EF) de la soja. iGMEF-01 con un sitio BamHI interno y un ORF retirado no intencional de extremo 3'.

gGclRG6b-01. Inicio: 217 Final: 4834. Un fragmento genómico derivado de la secuencia PI339656 de *Glycine clandestina* que alberga un gen de resistencia a la roya de la soja que codifica una proteína que contiene dos dominios de

receptor similar a Toll Interleucina (TIR) sintéticos a Glyma.14G024500 (Soy_william82_v2).

gGclRG6a-01. Inicio: 4842 Final: 10050. Un fragmento genómico derivado de la secuencia PI339656 de *Glycine clandestina* que alberga un gen de resistencia a la roya de la soja que codifica una proteína que contiene dos dominios de receptor similar a Toll Interleucina (TIR) sintéticos a Glyma.14G024400 (Soy_william82_v2).

xTAG-06. Inicio: 144 Final: 183.

xPARADAS-01. Inicio: 184 Final: 195.

INICIO. Inicio: 1868 Final: 1870.

PARADAS. Inicio: 3890 Final: 3892.

INICIO. Inicio: 5957 Final: 5959.

PARADA. Inicio: 8165 Final: 8167.

xPARADAS-01. Inicio: 10058 Final: 10069.

Inicio. Inicio: 11153 Final: 11153.

xPARADAS-01. Inicio: 14946 Final: 14957.

xPARADAS-01. Inicio: 15022 Final: 15033.

xTAG-02. Inicio: 15034 Final: 15073

promotor. Inicio: 217 Final: 1867.

Promotor. Inicio: 4842 Final: 5956.

prGmEF-02. Inicio: 10058 Final: 12135. Promotor del factor de elongación de la traducción EF-1 alfa/Tu, incluyendo el primer intrón y la utr adyacente, de soja (williams 82). Como prGmEF-01, pero retirando los sitios NcoI, BamHI y SanDI internos, codones de parada de múltiples marcos añadidos en el extremo 5' y se retiró un ORF no intencionado en el extremo 3'.

prVirG-01. Inicio: 16374 Final: 16504.

oVS1-02. Inicio: 18450 Final: 18854.

oCOLE-06. Inicio: 19532 Final: 20338 (Complementario).
El origen de replicación ColE1 funcional en *E. coli* derivado de pUC19.

Terminador. Inicio: 3893 Final: 4834.

Terminador. Inicio: 8168 Final: 10050.

tGmEPSPS-02. Inicio: 14148 Final: 14957. Versión modificada de tGMEPS-01; un terminador de EPSPS de *Glycine max*.

u5GmEF-01. Inicio: 11153 Final: 11215. UTR 5' del gen del factor de elongación (EF) de la soja, antes del intrón.

u5GmEF-02. Inicio: 12125 Final: 12135. UTR 5' del gen del factor de elongación (EF) de la soja, después del intrón.

Los ejemplos anteriores ilustran claramente las ventajas de varias realizaciones de la invención. Aunque la presente invención se ha descrito con referencia a detalles específicos de determinadas realizaciones de la misma, no se pretende que estos detalles deban considerarse como limitaciones del alcance de la invención, excepto en la medida y en el grado en el que estén incluidos en las reivindicaciones que se acompañan.

A lo largo de esta solicitud, se hace referencia a diversas patentes, publicaciones de patente y publicaciones que no son patente. Las divulgaciones de estas patentes, publicaciones de patente y publicaciones que no son patente en su totalidad, se incorporan como referencia en el presente documento en esta solicitud con el fin de describir más completamente el estado de la técnica al que pertenece la invención.

Se hace constar que con relación a esta fecha, el mejor método conocido por la solicitante para llevar a la práctica la citada invención, es el que resulta claro de la presente descripción de la invención.