

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 989 717**

51 Int. Cl.:

C07K 14/415 (2006.01)

C12N 15/82 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **27.02.2013** **E 20156547 (0)**

97 Fecha y número de publicación de la concesión europea: **17.07.2024** **EP 3689895**

54 Título: **Modulación del vigor de las semillas**

30 Prioridad:

29.02.2012 EP 12157514

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
27.11.2024

73 Titular/es:

SYNGENTA CROP PROTECTION AG (100.0%)
Rosentalstrasse 67
4058 Basel, CH

72 Inventor/es:

FINCH-SAVAGE, WILLIAM EDWARD;
BARKER, GUY CAMERON;
BRUGGINK, GERHARD TONKO;
VAN DEN WIJNGAARD, PAUL WILLEM JAN y
MORRIS, KARL

74 Agente/Representante:

LEHMANN NOVO, María Isabel

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 989 717 T3

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

DESCRIPCIÓN

Modulación del vigor de las semillas

5 **CAMPO DE LA INVENCIÓN**

La presente invención se refiere en general a un polinucleótido que posibilita la modulación del vigor de las semillas, particularmente potencia el vigor de las semillas, y más particularmente posibilita la modificación de la velocidad de germinación. La invención también se refiere a una semilla de una planta que comprende dicho polinucleótido. Además, también se proporcionan dentro de la presente invención un método de producción de la semilla de la planta, un método para mejorar la germinación y el vigor de la semilla de la planta, la planta transgénica y el uso del polinucleótido de la invención para producir plantas que generan semillas con germinación y características de vigor mejoradas. Las plantas de la presente invención se refieren particularmente a *Brassica*, más particularmente a *Brassica oleracea*.

15 **ANTECEDENTES DE LA INVENCIÓN**

La calidad de las semillas, que se define por el número de plantas útiles uniformes que se obtienen de un lote de semillas, está llegando a ser un rasgo cada vez más importante en los mercados de horticultura desarrollados. El desarrollo de plantas jóvenes es una actividad altamente tecnológica en estos mercados y las demandas con respecto a la calidad de las semillas, por lo tanto, son altas. Se requiere una calidad de las semillas fiable y constantemente alta para esto. Además, la germinación en condiciones adversas es un rasgo de calidad importante de las semillas. Esto significa que un requisito para el éxito comercial de variedades de semillas es una calidad constante y robusta de las semillas. Actualmente, sin embargo, la calidad de las semillas de variedades comerciales no siempre es estable y predecible.

Los parámetros de calidad de las semillas se ven muy influidos por las condiciones ambientales maternas durante el desarrollo de las semillas. Dados los volúmenes de semillas necesarios y la viabilidad comercial, son posibles únicamente controles limitados de estas condiciones. Por lo tanto, la constancia de la calidad de las semillas está limitada por la susceptibilidad a las condiciones maternas. Las investigaciones han mostrado que el entorno materno puede afectar potencialmente a todos los parámetros de calidad de las semillas, incluyendo la uniformidad y la germinación en condiciones adversas. Por lo tanto, disminuir la influencia de las condiciones maternas daría lugar a una calidad más constante y robusta de las semillas, lo que proporciona ventajas en los mercados tanto desarrollados como en desarrollo.

El establecimiento de plántulas predecibles y uniformes es esencial para la producción de cultivos que sean tanto sostenibles como provechosos. Un contribuyente clave a esta previsibilidad es el rendimiento de germinación de las semillas, que se ve influido directamente por el letargo y el vigor de las semillas. El letargo *per se* (ausencia de germinación en condiciones en general permisibles) no se considera un problema práctico con muchas especies de cultivo, pero el bajo vigor de las semillas (poco rendimiento de las semillas en la práctica) influye enormemente no solo en el número de plántulas que brotan, sino también en la cronología y uniformidad del brote de las plántulas en todos los cultivos. Los efectos de esto tienen un impacto principal sobre muchos aspectos de producción de cultivos que determinan la rentabilidad y las inversiones requeridas, y también hay influencias directas específicas de cultivo sobre la producción comercializable (Finch-Savage, 1995). El bajo vigor de las semillas puede ser el resultado de deterioro o daño de muchos tipos en las semillas y esto tiene gran importancia comercial. Sin embargo, también hay diferencias inherentes en el vigor inicial de las semillas antes de que empiecen a deteriorarse, pero la base genética, molecular y fisiológica de esto sigue estando poco comprendido.

Se han identificado mutaciones en muchos genes que muestran fenotipos con rendimiento alterado de la germinación de las semillas y estas han sido decisivas en el desarrollo de nuestra comprensión actual del control de la germinación (revisado por Finch-Savage y Leubner-Metzger, 2006; Holdsworth *et al.*, 2008a y 2008b). Sin embargo, el impacto relativo de estos genes en semillas de tipo silvestre o de cultivo está poco comprendido y no se han revelado candidatos claros que formen la base de un ensayo discriminador para el vigor de las semillas. Una fuente alternativa de variación genética para mutaciones inducidas en laboratorio está disponible en poblaciones naturales y genotipos de cultivo. Usar esta variación para identificar QTL asociados con el vigor de las semillas y después genes candidatos que influyan en estos rasgos puede proporcionar una ruta para identificar genes esencialmente importantes.

Se ha explotado la variación tanto de plantas naturales como de plantas de cultivo en análisis genéticos cuantitativos de una serie de rasgos de vigor de las semillas en tomate (Foolad *et al.*, 1999), *Brassica oleracea* (Bettey *et al.*, 2000, Finch-Savage *et al.*, 2005) y *Arabidopsis* (Groot *et al.*, 2000, Clerkx *et al.*, 2004). La velocidad de QTL de germinación de las semillas se ha identificado en las tres especies.

La distinción entre el letargo y el bajo vigor de las semillas en semillas sanas no envejecidas en términos de velocidad de germinación, si existe alguna, no se comprende y puede tener la misma base (Hilhorst y Toorop, 1997). En la mayoría de situaciones, por ejemplo, en *Arabidopsis*, el letargo fisiológico no es absoluto, pero las semillas son condicionalmente letárgicas, es decir, la germinación tiene a ser lenta y es posible únicamente en una serie limitada

de entornos. Como el letargo se pierde progresivamente, la germinación tiende a acelerarse y llega a ser posible en una serie más amplia de entornos y, por lo tanto, puede aparecer como un aumento en el vigor.

Entre los factores que explican el establecimiento de la germinación de las semillas y la regulación del letargo de las semillas, el ácido abscísico (ABA), una hormona vegetal bien conocida, desempeña una función importante. El ABA es en particular esencial para la germinación de las semillas y los procesos de maduración de las semillas (para una revisión, véase Finkelstein *et al.* 2002), ya que es responsable del establecimiento de un periodo de letargo de las semillas. En cuanto a las yemas, es importante que las semillas no germinan prematuramente, por ejemplo, durante condiciones inoportunamente suaves antes del comienzo del invierno o una estación seca. El ABA en la semilla impone este letargo. El letargo se suprime únicamente si la semilla se ha expuesto a un episodio de frío prolongado y/u otra señal ambiental apropiada y si hay suficiente agua para mantener la germinación. Además de su función en el vigor de las semillas, el ABA también regula muchos aspectos importantes de la vida de la planta, incluyendo las respuestas fisiológicas a amenazas bióticas y agresiones abióticas como sequía y desecación.

Por tanto, hay una necesidad permanente de semillas con un vigor más fiable y constante de las semillas; especialmente con una velocidad oportuna y uniforme de germinación, para proporcionar semillas que germinen a una tasa más constante, independientemente de las condiciones maternas y sean cuales sean las condiciones ambientales externas. Dicho vigor aumentado de las semillas sería de particular interés en casos donde la semilla se recubre con una preparación dada (química, biológica), ya que habitualmente se observa un retardo en la germinación de las semillas. En consecuencia, las semillas que comprenden secuencias que potencian el vigor de las semillas, más particularmente potencian la velocidad y la uniformidad de la germinación de las semillas, serían de importancia fundamental para contrarrestar el efecto del tratamiento de recubrimiento, mientras aún se aplican insecticidas y fungicidas.

Además, aunque en muchos aspectos, aumentar el vigor de las semillas sería un rasgo muy útil y deseado, parece que, en algunos casos, disminuir el vigor de las semillas sería de gran interés. En particular en semillas vivíparas, disminuir el vigor de las semillas podría ser beneficioso. La viviparidad se define como la germinación de la semilla mientras aún está en la planta madre o antes de secarse y puede producirse tanto en semillas inmaduras como en semillas completamente maduras. La viviparidad se ha observado en muchas especies de cultivo diferentes incluyendo cultivos de *Brassica* (Ruan *et al.* 2000). Las semillas que comprenden secuencias que pueden disminuir el vigor de las semillas, por tanto, pueden retardar, si no eliminar, el fenotipo de viviparidad no deseado.

SUMARIO DE LA INVENCIÓN

Por lo tanto, es mérito de los autores de la invención, en dicho estado de la técnica, haber identificado una planta de una población que contenía una pequeña región de introgresión que abarca *SOG1* (*Speed Of Germination 1* - velocidad de germinación 1), un QTL de velocidad de germinación de las semillas identificado en *B. oleracea* por Bettey *et al.* (2000) y particularmente haber demostrado e identificado los genes correspondientes que están implicados en la modulación del vigor de las semillas, en particular implicados en la regulación de la velocidad de germinación de las semillas. En particular se demuestra que estos genes, y sus secuencias correspondientes, pueden usarse como herramientas posibilitadoras para obtener semillas (o plantas que suministran semillas) que muestran un fenotipo de vigor modificado de las semillas. En particular, pueden introducirse secuencias génicas en un nuevo fondo para modular el vigor de las semillas. Más particularmente, pueden usarse secuencias génicas para genomanipular una planta novedosa, cuyas semillas brotarán más pronto, más uniformemente, independientemente de las condiciones ambientales externas, y con independencia de las condiciones maternas. Además, dichas secuencias génicas pueden usarse como herramientas para modificar el contenido de ABA en la semilla y/o la respuesta de la semilla a ABA, influyendo de ese modo al comportamiento de la semilla con respecto al letargo de la semilla y la síntesis de proteínas y lípidos de almacenamiento en la semilla.

La invención es como se define en el conjunto de reivindicaciones adjuntas.

REALIZACIONES

Por lo tanto, en un 1.^{er} aspecto, se proporciona un polinucleótido, particularmente un polinucleótido aislado, que comprende una molécula de ácido nucleico que comprende una secuencia de nucleótidos como se representa en la SEQ ID NO: 3 (GD33 versión de BolC.VG1.a); en el que dicha molécula de ácido nucleico, tras la expresión en una planta o parte de la planta, da lugar a un vigor modificado de las semillas.

En otro aspecto, la presente divulgación proporciona un polinucleótido, particularmente un polinucleótido aislado, que comprende una molécula de ácido nucleico seleccionada del grupo que consiste en

- a. molécula de ácido nucleico que comprende una secuencia de nucleótidos que tiene al menos un 95 % de identidad de secuencia con cualquiera de las secuencias representadas en el grupo que comprende SEQ ID NO:3;
- b. molécula de ácido nucleico que comprende una secuencia de nucleótidos que tiene al menos un 98 % de identidad de secuencia con cualquiera de las secuencias representadas en el grupo que comprende: SEQ ID NO:3;

en el que dicha molécula de ácido nucleico como se define en cualquiera de a) – b), tras la expresión en una planta o parte de la planta, da lugar a un vigor modificado de las semillas.

En otro aspecto, la presente divulgación se refiere a un polinucleótido de acuerdo con el primer, segundo o tercer aspecto, en el que el fenotipo de vigor modificado de las semillas se caracteriza por un fenotipo adicional seleccionado del grupo que comprende: velocidad modificada de germinación, velocidad modificada de brote de plántulas, uniformidad modificada de germinación de las semillas, uniformidad modificada de brote de plántulas, porcentaje modificado de germinación de las semillas, tolerancia modificada de la semilla frente al entorno externo y/o condiciones maternas, sensibilidad modificada a ABA o contenido modificado de ABA.

En otro aspecto, la presente divulgación se refiere a un polinucleótido, particularmente un polinucleótido aislado, que comprende una molécula de ácido nucleico seleccionada del grupo que comprende una secuencia de nucleótidos como se representa en la SEQ ID NO: 3; en el que dicha molécula de ácido nucleico, tras la expresión en una planta o parte de la planta, da lugar a un vigor aumentado de las semillas.

En el contexto de la presente invención, la expresión "vigor aumentado de las semillas" significa que la velocidad de germinación de la semilla puede modificarse en el sentido de que la semilla germina más rápido y, por tanto, es más vigorosa. En una realización particular, la velocidad de germinación se aumenta permitiendo que se obtengan semillas que germinan más rápido gracias a la presencia de un polinucleótido de acuerdo con la presente invención en su genoma en comparación con una semilla que no comprende dicho polinucleótido. Dicha velocidad aumentada de germinación provoca un brote de plántulas anterior significativo y, por tanto, proporciona a las semillas una flexibilidad potenciada y una mejor adaptación a diversos entornos. Se descubrió que los 2 genes que se identificaron tenían un fenotipo de germinación significativo, lo que indica que estos genes son reguladores de la germinación en diferente grado. El alelo correspondiente al polinucleótido de la SEQ ID NO: 3 y SEQ ID NO: 4 (alelos GD33 de los genes) de acuerdo con la presente invención se ha demostrado que corresponde al fenotipo de vigor aumentado de las semillas y, por tanto, permite modificar la velocidad de germinación de las semillas de la planta en que se introgresan hacia una velocidad aumentada de germinación.

En otro aspecto, la presente divulgación se refiere a un polinucleótido de acuerdo con el aspecto anterior, en el que el fenotipo de vigor aumentado de las semillas se caracteriza por un fenotipo adicional seleccionado del grupo que comprende: velocidad aumentada de germinación, velocidad aumentada de brote de plántulas, uniformidad aumentada de germinación de las semillas, uniformidad aumentada de brote de plántulas, porcentaje aumentado de germinación de las semillas, tolerancia aumentada de la semillas frente al entorno externo y/o condiciones maternas, sensibilidad disminuida a ABA o contenido disminuido de ABA.

En otro aspecto, la presente divulgación proporciona un casete de expresión que comprende un polinucleótido de cualquiera de los aspectos precedentes.

En otro aspecto, la presente divulgación proporciona una molécula de vector que comprende el casete de expresión de acuerdo con el aspecto anterior.

En un otro aspecto, la presente divulgación se refiere al uso de un polinucleótido de acuerdo con realizaciones anteriores para modificar el vigor de las semillas.

En una realización, la presente invención se refiere a una planta híbrida o parte de la planta que contiene dentro de su genoma una introgresión que comprende el polinucleótido de cualquiera de los aspectos anteriores y muestra una modificación del vigor de las semillas en comparación con una planta o parte de la planta que no comprende dicho polinucleótido, casete de expresión o molécula de vector.

En otra realización, la presente invención se refiere a una planta o parte de la planta que contiene dentro de su genoma una introgresión que comprende un polinucleótido de acuerdo con cualquiera de las realizaciones anteriores y muestra un vigor aumentado de las semillas en comparación con una semilla suministrada por una planta o parte de la planta que no comprende dicho polinucleótido.

En otro aspecto, la divulgación proporciona un método para seleccionar una planta o parte de la planta con vigor modificado de las semillas, que comprende la detección en la planta o parte de la planta a ensayar de la presencia o ausencia de un polinucleótido de acuerdo con cualquiera de los aspectos anteriores.

En otro aspecto, la divulgación proporciona un método para seleccionar una planta o parte de la planta con vigor modificado de las semillas, que comprende poner en contacto una planta candidata o parte de la planta con una herramienta de selección seleccionada del grupo que comprende los polinucleótidos de cualquiera de los aspectos anteriores.

En otra realización, la presente invención se refiere una planta o parte de la planta de acuerdo con cualquiera de las realizaciones precedentes que es una planta cultivada o parte de la planta cultivada y se seleccionada del grupo que comprende *Brassica oleracea*, *Brassica napus*, *Brassica rapa*, *Brassica campestris*, *Brassica juncea*, *Brassica nigra*,

Brassica pekinensis, Brassica chinensis, Brassica rosularis, Eruca vesicaria, Eruca sativa, Raphanus sativus, Lepidium sativum, Nasturtium officinale, Wasabia japonica.

En otro aspecto, la divulgación proporciona un método no biológico para obtener una planta o parte de la planta con vigor modificado de las semillas, que comprende introducir un polinucleótido de acuerdo con cualquiera de los aspectos anteriores en el genoma de dicha planta o parte de la planta.

En otro aspecto, la divulgación se refiere al método de acuerdo con el aspecto anterior, en el que la presencia del polinucleótido se verifica mediante el uso de un marcador molecular, particularmente mediante un marcador molecular físicamente localizado en una posición que está dentro o fuera del locus genético que contiene el polinucleótido.

En otro aspecto, la divulgación se refiere al método de acuerdo con aspectos anteriores, en el que la presencia del polinucleótido se verifica mediante el uso de al menos dos marcadores moleculares, particularmente mediante al menos dos marcadores moleculares físicamente localizados en una posición que está flanqueando el locus genético que contiene el polinucleótido.

En otra realización, la invención se refiere a una semilla que comprende un polinucleótido de acuerdo con cualquiera de los aspectos anteriores, en la que dicha semilla está recubierta con cualquier tipo de recubrimiento.

En otra realización, la invención se refiere a una planta o parte de la planta de acuerdo con cualquiera de los aspectos anteriores, en la que la planta es una planta híbrida.

En otra realización, la invención se refiere a una planta o parte de la planta que se puede obtener de la semilla depositada en el NCIMB con el número de depósito NCIMB 41951.

DESCRIPCIÓN DETALLADA DE LA INVENCION

Las secuencias polinucleotídicas de acuerdo con la realización previa se han identificado en y aislado de una planta de *Brassica oleracea* con diferentes fenotipos de vigor de las semillas. Los alelos con las secuencias polinucleotídicas de acuerdo con la presente invención se sometieron a introgresión en una planta con diferente calidad de las semillas y diferente vigor de las semillas. Gracias a la introgresión de cualquiera de las secuencias polinucleotídicas de acuerdo con la presente invención, el vigor de las semillas se modifica significativamente y, por consiguiente, la calidad de la planta o parte de la planta, particularmente la calidad de las semillas de la planta se modificó significativamente, particularmente el vigor de las semillas, que destaca por la velocidad de germinación. En una realización, la semilla podía germinar mejor y más rápido, particularmente en condiciones frías, lo que muestra una calidad de las semillas más robusta. La introgresión de una secuencia polinucleotídica de acuerdo con la presente invención permite obtener una calidad de las semillas de la planta menos susceptible al entorno de producción de las semillas. Esto significa que la introgresión de las secuencias polinucleotídicas de acuerdo con la presente invención posibilita el suministro de una calidad constante de las semillas a partir de producción convencional de semillas.

La introducción de las secuencias polinucleotídicas de acuerdo con la presente invención en plantas comerciales cultivadas, particularmente plantas de *Brassica* cultivadas, más particularmente plantas de *Brassica oleracea* cultivadas, garantizará una producción más fiable de semillas con suficiente calidad de las semillas. Esto aumenta significativamente la flexibilidad operativa.

Las semillas híbridas que comprenden cualquiera de las secuencias polinucleotídicas de acuerdo con la presente invención muestran un establecimiento de la población más uniforme, especialmente en condiciones adversas. Esto definitivamente aumentará el valor del producto comercial de semillas.

Detalles del depósito de semillas

Las siguientes muestras de semillas se depositaron en NCIMB, Ferguson Building, Craibstone Estate, Bucksburn, Aberdeen AB21 9YA, Escocia, R.U., el 4 de abril de 2012, según las disposiciones del Tratado de Budapest en nombre de Syngenta Participations AG:

NCIMB 41950 *Brassica oleracea* A12
NCIMB 41951 *Brassica oleracea* SL101

Definiciones

A las expresiones y términos técnicos usados dentro del alcance de esta solicitud en general se les da el significado habitualmente aplicado a los mismos en la técnica relevante de ingeniería genética en plantas, mejoramiento y cultivo de plantas si no se indica lo contrario en este documento a continuación.

Como se usa en esta memoria descriptiva y las reivindicaciones adjuntas, las formas en singular "un/o", "una" y "el/la" incluyen sus plurales salvo que el contexto indique claramente lo contrario. Por tanto, por ejemplo, la referencia a "una planta" incluye una o más plantas y la referencia a "una célula" incluye mezclas de células, tejidos y similares.

Se entiende, dentro del alcance de la invención, que una "planta cultivada", particularmente "planta de *Brassica* cultivada", más particularmente "planta de *Brassica oleracea* cultivada" se refiere a una planta que ya no está en estado natural, sino que se ha desarrollado mediante cuidados humanos y para uso en seres humanos y/o con propósitos de cultivos y/o para consumo. Se entiende, además, que "planta cultivada", particularmente "planta de *Brassica* cultivada", más particularmente "planta de *Brassica oleracea* cultivada" excluyen aquellas especies de tipo silvestre que comprenden el rasgo que es objeto de esta invención como un rasgo natural y/o parte de su genética natural.

Se entiende, dentro del alcance de la invención, que un "alelo" se refiere a formas alternativas o variantes de diversas unidades genéticas idénticas o asociadas con diferentes formas de un gen o de cualquier tipo de elemento genético identificable, que son alternativas en herencia porque están situadas en el mismo locus en cromosomas homólogos. Dichas formas alternativas o variantes pueden ser el resultado de polimorfismos, inserciones, inversiones, translocaciones o eliminaciones de un solo nucleótido, o la consecuencia de la regulación génica provocada, por ejemplo, por modificación química o estructural, regulación de la transcripción o modificación/regulación postraduccional. En una célula u organismo diploide, los dos alelos de un gen o elemento genético dado típicamente ocupan los locus correspondientes en un par de cromosomas homólogos.

Se entiende, dentro la presente invención, que un alelo "truncado" de un gen representa un alelo de un gen que ha perdido una sola secuencia de nucleótidos parcial o múltiples, en comparación con su alelo equivalente del gen de longitud completa.

Un alelo asociado con un rasgo cualitativo puede comprender formas alternativas o variantes de diversas unidades genéticas, incluyendo las que son idénticas o están asociadas con un solo gen o múltiples genes o sus productos, o incluso un gen alterado o controlado por un factor genético que contribuye al fenotipo representado por el locus.

Como se usa en este documento, la expresión "alelo marcador" se refiere a una forma alternativa o variante de una unidad genética como se define en este documento anteriormente, cuando se usa como un marcador para ubicar locus genéticos que contienen alelos en un cromosoma que contribuyen a la variabilidad de los rasgos fenotípicos.

Como se usa en este documento, el término "reproducción", y variantes gramaticales del mismo, se refieren a cualquier proceso que genere un individuo de la descendencia. Las reproducciones pueden ser sexuales o asexuales, o cualquier combinación de las mismas. Tipos no limitantes ejemplares de reproducción incluyen cruces, autopolinización, generación de derivados haploides duplicados y combinaciones de los mismos, que son todas técnicas conocidas por los expertos en la materia.

Se entiende, dentro del alcance de la invención, que "retrocruzamiento" se refiere a un proceso en que una descendencia híbrida se retrocruza repetidamente de nuevo con uno de los progenitores. Se pueden usar diferentes progenitores recurrentes en retrocruzamientos posteriores. Pueden producirse líneas recombinantes autopolinizando los descendientes resultantes del retrocruzamiento.

Se entiende, dentro del alcance de la invención, que "locus" se refiere a una región en un cromosoma, que comprende un gen o cualquier otro elemento genético o factor que contribuye a un rasgo.

Se entiende, dentro del alcance de la presente invención, que "introgresión" (o sometido a introgresión) se refiere a un movimiento de uno o más genes o segmentos de ácido nucleico desde una especie a la combinación de genes de otra especie, o desde una línea a la combinación de genes de otra línea dentro de la misma especie. La introgresión puede conseguirse por cruce sexual, hibridación sexual o por transformación genética.

Como se usa en este documento, "locus marcador" se refiere a una región en un cromosoma, que comprende un nucleótido o una secuencia polinucleotídica que está presente en el genoma de un individuo y que está asociada con uno o más locus de interés, que pueden comprender un gen o cualquier otro elemento o factor genético que contribuya a un rasgo. "Locus marcador" también se refiere a una región en un cromosoma, que comprende una secuencia polinucleotídica complementaria a una secuencia genómica, tal como una secuencia de un ácido nucleico usado como sondas.

Para los fines de la presente invención, el término "segregación" o "cosegregación" se refiere al hecho de que el alelo para el rasgo y el alelo o los alelos para el marcador o los marcadores tienden a transmitirse juntos porque están físicamente muy juntos en el mismo cromosoma (recombinación reducida entre ellos a causa de su proximidad física), lo que provoca una asociación no aleatoria de sus alelos como resultado de su proximidad en el mismo cromosoma. La "cosegregación" también se refiere a la presencia de dos o más rasgos dentro de una única planta, de los que se sabe que al menos uno es genético y que no puede explicarse fácilmente por casualidad.

Como se usa en este documento, la expresión "marcador genético" o "marcador" se refiere a una característica del genoma de un individuo (por ejemplo, un nucleótido o una secuencia polinucleotídica que está presente en el genoma de un individuo) que está asociada a uno o más locus de interés. En algunos aspectos, un marcador genético es

polimórfico en una población de interés, o el locus ocupado por el polimorfismo, dependiendo del contexto. Los marcadores genéticos incluyen, por ejemplo, polimorfismos mononucleotídicos (SNP), indels (es decir, inserciones/eliminaciones), repeticiones de secuencia simple (SSR), polimorfismos de longitud de fragmentos de restricción (RFLP), ADN polimórficos amplificados aleatoriamente (RAPD), marcadores de secuencia polimórfica amplificada y escindida (CAPS), marcadores de tecnología de matrices de diversidad (DArT) y polimorfismos de longitud de fragmentos amplificados (AFLP), entre muchos otros ejemplos. Los marcadores genéticos pueden usarse, por ejemplo, para localizar locus genéticos que contienen alelos en un cromosoma que contribuye a variabilidad de rasgos fenotípicos. La expresión "marcador genético" o "marcador" también puede referirse a una secuencia polinucleotídica complementaria a una secuencia genómica, tal como una secuencia de un ácido nucleico usado como sondas.

Un "marcador genético" o "marcador" puede estar físicamente localizado en una posición en un cromosoma que está dentro o fuera del locus genético con el que está asociado (es decir, es intragénico o extragénico, respectivamente). Indicado de otra manera, mientras los "marcadores genéticos" o "marcadores" típicamente se emplean cuando la ubicación en un cromosoma del gen o de una mutación funcional, por ejemplo, dentro de un elemento de control fuera de un gen que corresponde al locus de interés, no se ha identificado y hay una tasa de recombinación que no es cero entre el "marcador genético" o "marcador" y el locus de interés, la materia en cuestión divulgada en la presente también puede emplear "marcadores genéticos" o "marcadores" que están físicamente dentro de los límites de un locus genético (por ejemplo, dentro de una secuencia genómica que corresponde a un gen tal como, aunque sin limitación, un polimorfismo dentro de un intrón o un exón de un gen). En algunos aspectos de la materia en cuestión divulgada en la presente, el uno o más "marcadores genéticos" o "marcadores" comprenden entre uno y diez marcadores, y en algunas realizaciones el uno o más marcadores genéticos comprenden más de diez marcadores genéticos.

Como se usa en este documento, el término "genotipo" se refiere a la constitución genética de una célula u organismo. El "genotipo para un conjunto de marcadores genéticos" de un individuo incluye los alelos específicos para uno o más locus marcadores genéticos, presentes en el haplotipo de un individuo. Como se sabe en la técnica, un genotipo puede estar relacionado con un único locus o con múltiples locus, pudiendo estar los locus relacionados o no relacionados y/o ligados o no ligados. En algunas realizaciones, el genotipo de un individuo se refiere a uno o más genes que están relacionados en el sentido de que uno o más de los genes están implicados en la expresión de un fenotipo de interés. Por tanto, en algunas realizaciones, un genotipo comprende un sumario de uno o más alelos presentes dentro de un individuo en uno o más locus genéticos de un rasgo cuantitativo. En algunas realizaciones, un genotipo se expresa en función de un haplotipo.

Como se usa en este documento, el término "ligamiento" y variantes gramaticales del mismo, se refiere a la tendencia de alelos en diferentes locus en el mismo cromosoma de segregar juntos más a menudo que lo que se esperaría por probabilidad si su transmisión fuera independiente, en algunas realizaciones como una consecuencia de su proximidad física.

Por ejemplo, "secuencia polinucleotídica", como se usa en este documento, se refiere a todas las formas de tipos de ácidos nucleicos y/o secuencias de nucleótidos de origen natural o generados de forma recombinante, así como a ácidos nucleicos/secuencias de nucleótidos sintetizados químicamente. Esta expresión también abarca análisis de ácido nucleico y derivados de ácido nucleico, tales como, por ejemplo, ADN bloqueado, ARN, ADNc, APN, oligonucleótidos tiofosfato y ribooligonucleótidos sustituidos. Además, la expresión "secuencia polinucleotídica" también se refiere a cualquier molécula que comprende nucleótidos o análogos nucleotídicos. La expresión "ácido nucleico" o "polinucleótido" se refiere a cualquier cadena física de unidades monoméricas que puede corresponder a una cadena de nucleótidos, incluyendo un polímero de nucleótidos (por ejemplo, un ADN típico, ADNc o polímero de ARN), que opcionalmente contiene bases nucleotídicas sintéticas, no naturales o alteradas que tienen capacidad de incorporación en polímeros de ADN o ARN, oligonucleótidos modificados (por ejemplo, oligonucleótidos que comprenden bases que no son típicas de ARN o ADN biológico, tales como oligonucleótidos 2'-O-metilados), y similares. Se entiende en este documento que el término "polinucleótido" se refiere a una molécula polimérica de alto peso molecular, que puede ser monocatenaria o bicatenaria, multicatenaria o combinaciones de las mismas, compuesta de monómeros (nucleótidos) que contienen un glúcido, fosfato y una base que es una purina o una pirimidina. A menos que se indique lo contrario, una secuencia particular de ácido nucleico de la materia en cuestión divulgada en la presente comprende o codifica opcionalmente secuencias complementarias, además de cualquier secuencia explícitamente indicada.

Preferiblemente, la expresión "secuencia polinucleotídica" se refiere a una molécula de ácido nucleico, es decir, ácido desoxirribonucleico (ADN) y/o ácido ribonucleico (ARN). La "secuencia polinucleotídica" en el contexto de la presente invención puede prepararse por metodología química sintética conocida por los expertos en la materia, o mediante el uso de tecnología recombinante, o puede aislarse de fuentes naturales, o mediante una combinación de los mismos. El ADN y el ARN pueden comprender opcionalmente nucleótidos no naturales y pueden ser monocatenarios o bicatenarios. "Secuencia polinucleotídica" también se refiere a ADN y ARN de sentido y de antisentido, es decir, una secuencia polinucleotídica que es complementaria a una secuencia específica de nucleótidos en ADN y/o ARN. Además, la expresión "secuencia polinucleotídica" puede referirse a ADN o ARN o híbridos de los mismos o cualquier modificación de los mismos, que son conocidos en el estado de la técnica (véanse, por ejemplo, los documentos US 5525711, US 4711955, US 5792608 o EP 302175 para ejemplos de modificaciones). La secuencia polinucleotídica

puede ser monocatenaria o bicatenaria, lineal o circular, natural o sintética, y sin ninguna limitación de tamaño. Por ejemplo, la secuencia polinucleotídica puede ser ADN genómico, ADNc, ARNm, ARN de antisentido, ribozima o un ADN que codifica dichos ARN o quimeroplastos (Gamper, *Nucleic Acids Research*, 2000, 28, 4332 – 4339). Dicha secuencia polinucleotídica puede estar en forma de un plásmido o de ADN o ARN vírico. "Secuencia polinucleotídica" también puede referirse a uno o más oligonucleótidos, en los que se incluyen cualquiera de las modificaciones del estado de la técnica tales como fosfotioatos o ácidos peptidonucleicos (APN).

Un "fragmento polinucleotídico" es una fracción de una molécula polinucleotídica dada o de una "secuencia polinucleotídica". En plantas superiores, el ácido desoxirribonucleico (ADN) es el material genético, mientras que el ácido ribonucleico (ARN) está implicado en la transferencia de la información contenida dentro del ADN en proteínas. Un "genoma" es el cuerpo completo de material genético contenido en cada célula de un organismo.

A menos que se indique lo contrario, una secuencia de ácido nucleico particular de esta invención también abarca de forma implícita variantes modificadas de forma conservadora de la misma (por ejemplo, sustituciones de codones degenerados) y secuencias complementarias y también la secuencia indicada explícitamente. Específicamente, las sustituciones de codones degenerados pueden conseguirse generando secuencias en que la tercera posición de uno o más codones seleccionados (o todos) se sustituye con residuos de base mixta y/o de desoxiinosina (Batzer *et al.*, 1991; Ohtsuka *et al.*, 1985; Rossolini *et al.*, 1994). La expresión polinucleótido se usa indistintamente con ácido nucleico, secuencia de nucleótidos y pueden incluir genes, ADNc y ARNm codificados por un gen, etc.

Se entiende que el polinucleótido de la invención se proporciona en forma aislada. El término "aislado" significa que el polinucleótido divulgado y reivindicado en este documento no es un polinucleótido que se produce en su contexto natural, si acaso es que tiene un equivalente de origen natural. Por consiguiente, los otros compuestos de la invención descritos adicionalmente a continuación se entiende que están aislados. Si se reivindica en el contexto de un genoma vegetal, el polinucleótido de la invención se distingue sobre los equivalentes de origen natural por el sitio de inserción en el genoma y las secuencias flanqueantes en el sitio de inserción.

Como se usa en este documento, el término "gen" se refiere a cualquier segmento de ácido nucleico asociado con una función biológica. Por tanto, los genes incluyen secuencias codificantes y/o las secuencias reguladoras requeridas para su expresión. Por ejemplo, gen se refiere a un fragmento de ácido nucleico que expresa ARNm o ARN funcional, o codifica una proteína específica, y que incluye secuencias reguladoras. Los genes también incluyen segmentos de ADN no expresados que, por ejemplo, forman secuencias de reconocimiento para otras proteínas. Los genes pueden obtenerse de una diversidad de fuentes, incluyendo clonación desde una fuente de interés o síntesis a partir de información de secuencia conocida o predicha, y pueden incluir secuencias diseñadas para que tengan parámetros deseados.

Se entiende, dentro del alcance de la invención, que "selección basada en marcadores" se refiere, por ejemplo, al uso de marcadores genéticos para detectar uno o más ácidos nucleicos de la planta, donde el ácido nucleico está asociado con un rasgo deseado para identificar plantas que portan genes para rasgos deseables (o indeseables), de modo que esas plantas puedan usarse (o evitarse) en un programa de mejoramiento selectivo. Un "gen marcador" codifica un rasgo de selección o cribado.

Los marcadores adecuados usados dentro de la invención pueden seleccionarse, por ejemplo, del grupo que consiste en marcadores de polimorfismo mononucleotídico (SNP), marcadores de indel (es decir, inserciones/eliminaciones), marcadores de repetición de secuencia (SSR), marcadores de polimorfismo de longitud de fragmentos de restricción (RFLP), marcadores de ADN polimórfico amplificado aleatoriamente (RAPD), marcadores de secuencia polimórfica ampliada y escindida (CAPS), marcadores de tecnología de matrices de diversidad (DART) y marcadores de polimorfismos de longitud de fragmentos amplificados (AFLP).

Por ejemplo, el RFLP implica el uso de enzimas de restricción para cortar el ADN cromosómico en sitios de restricción cortos y específicos, los polimorfismos se producen por duplicaciones o eliminaciones entre los sitios o mutaciones en los sitios de restricción.

La RAPD utiliza la amplificación por reacción en cadena de la polimerasa (PCR) de baja rigurosidad con cebadores únicos de secuencia arbitraria para generar matrices específicas de cepa de fragmentos de ADN anónimos. El método requiere solamente pequeñas muestras de ADN y analiza un gran número de locus polimórficos.

El AFLP requiere la digestión de ADN celular con una o más enzimas de restricción antes de usar la PCR y los nucleótidos selectivos en los cebadores para amplificar fragmentos específicos. Con este método, usando técnicas de electroforesis para visualizar los fragmentos obtenidos, se pueden medir hasta 100 locus polimórficos por combinación de cebador y solamente se requiere una pequeña muestra de ADN para cada uno de los ensayos.

El análisis por SSR se basa en secuencias de microsatélites de ADN (repetición corta) que están muy dispersas por todo el genoma de los eucariotas, que se amplifican de forma selectiva para detectar variaciones en repeticiones de secuencias simples. Solamente se requieren pequeñas muestras de ADN para un análisis por SSR. Los SNP usan ensayos de extensión por PCR que detectan de manera eficaz las mutaciones puntuales. El procedimiento requiere

poco ADN por muestra. Uno o dos de los métodos anteriores se pueden usar en un programa típico de mejoramiento por selección basado en marcadores.

El método más preferido de conseguir amplificación de fragmentos de nucleótidos que abarcan una región polimórfica del genoma vegetal emplea la reacción en cadena de la polimerasa ("PCR") (Mullis *et al.*, Cold Spring Harbor Symp. Quant. Biol. 51: 263 273 (1986)), usando pares de cebadores que implican un cebador directo y un cebador inverso que pueden hibridar con las secuencias proximales que definen un polimorfismo en su forma bicatenaria. Como se divulga en este documento, dichos cebadores pueden usarse para un cartografiado preciso, clonación basada en cartografiado y para análisis de expresión.

Se entiende, dentro del alcance de la invención, que "microsatélite o marcador de SSR (repeticiones de secuencia simples)" se refiere a un tipo de marcador genético que consiste en numerosas repeticiones de secuencias cortas de bases de ADN, que se encuentran en locus por todo el genoma de la planta y tiene una probabilidad de ser altamente polimórfico.

Se entiende, dentro del alcance de la invención, que "PCR (reacción en cadena de la polimerasa)" se refiere a un método de producción de cantidades relativamente grandes de regiones específicas de ADN o uno o más subconjuntos del genoma, haciendo posible de ese modo diversos análisis que se basan en esas regiones.

Se entiende, dentro del alcance de la invención, que "cebador de PCR" se refiere a fragmentos relativamente cortos de ADN monocatenario usados en la amplificación por PCR de regiones específicas de ADN.

Se entiende, dentro del alcance de la invención, que "fenotipo" se refiere a una o más características distinguibles de un rasgo genéticamente controlado.

Como se usa en este documento, la frase "rasgo fenotípico" se refiere a la apariencia u otra característica detectable de un individuo, que resulta de la interacción de su genoma, proteoma y/o metaboloma con el entorno.

Se entiende, dentro del alcance de la invención, que "polimorfismo" se refiere a la presencia en una población de dos o más formas diferentes de un gen, marcador genético o rasgo hereditario o un producto génico obtenible, por ejemplo, a través de corte y empalme alternativo, metilación del ADN, etc.

"Sonda", como se usa en este documento, se refiere a un grupo de átomos o moléculas que puede reconocer y unirse a una molécula o estructura celular diana específica y, por tanto, permitir la detección de la molécula o estructura diana. Particularmente, "sonda" se refiere a una secuencia de ADN o ARN marcada que puede usarse para detectar la presencia de y para cuantificar una secuencia complementaria por hibridación molecular.

Dichas secuencias polinucleotídicas que pueden hibridar pueden identificarse y aislarse usando las secuencias polinucleotídicas descritas en este documento o partes o complementos inversos de las mismas, por ejemplo, por hibridación de acuerdo con métodos convencionales (véase, por ejemplo, Sambrook y Russell (2001), Molecular Cloning: A Laboratory Manual, CSH Press, Cold Spring Harbor, NY, EE. UU.). Secuencias de nucleótidos que comprenden las mismas o sustancialmente las mismas secuencias de nucleótidos que las indicadas en las SEQ ID NO: 1, 2 o 3 enumeradas, o partes/fragmentos de las mismas, pueden usarse, por ejemplo, como sondas de hibridación. Los fragmentos usados como sondas de hibridación también pueden ser fragmentos sintéticos que se preparan por técnicas habituales de síntesis, cuya secuencia es sustancialmente idéntica a la de una secuencia de nucleótidos de acuerdo con la invención.

El término "hibridar", como se usa en este documento, se refiere a condiciones convencionales de hibridación, preferiblemente a condiciones de hibridación en que se usa SSPE 5x, SDS al 1 %, solución de Denhardt 1x como una solución y/o las temperaturas de hibridación están entre 35 °C y 70 °C, preferiblemente 65 °C. Después de la hibridación, el lavado se realiza preferiblemente, en primer lugar, con SSC 2x, SDS al 1 % y posteriormente con SSC 0,2x a temperaturas entre 35 °C y 75 °C, particularmente entre 45 °C y 65 °C, pero especialmente a 59 °C (con respecto a la definición de SSPE, SSC y solución de Denhardt, véase Sambrook *et al. loc. cit.*). Las condiciones de hibridación de alta rigurosidad, como se describen, por ejemplo, en Sambrook *et al., supra*, son particularmente preferidas. Condiciones de hibridación rigurosas particularmente preferidas están presentes, por ejemplo, si la hibridación y el lavado se producen a 65 °C, como se indica anteriormente. Condiciones de hibridación no rigurosas, por ejemplo, con hibridación y lavado realizados a 45 °C, son menos preferidas y a 35 °C incluso menos.

"Homología de secuencia" o "identidad de secuencia" se usa en este documento indistintamente. Los términos "idéntica" o porcentaje de "identidad", en el contexto de dos o más secuencias de ácido nucleico o proteínicas, se refieren a dos o más secuencias o subsecuencias que son iguales o tienen un porcentaje especificado de residuos aminoácidos o nucleótidos que son iguales, cuando se comparan y alinean para una máxima correspondencia, medida usando uno de los siguientes algoritmos de comparación de secuencias o por inspección visual. Por ejemplo, este término se usa en este documento en el contexto de una secuencia de nucleótidos que tiene una homología, es decir, una identidad de secuencia, de al menos un 50 %, 55 %, 60 %, preferiblemente de al menos un 70 %, 75 %, más preferiblemente de al menos un 80 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 %, 100 %.

96 %, 97 %, 98 %, e incluso mucho más preferiblemente de al menos un 99 % con otra secuencia de nucleótidos, preferiblemente completa.

Si las dos secuencias que deben compararse entre sí difieren en longitud, la identidad de secuencia se refiere preferiblemente al porcentaje de los residuos nucleotídicos de la secuencia más corta que son idénticos a los residuos de nucleotídicos de la secuencia más larga. Como se usa en este documento, el porcentaje de identidad/homología entre dos secuencias es una función del número de posiciones idénticas compartidas por las secuencias (es decir, % de identidad = $n.^{\circ}$ de posiciones idénticas/número total de posiciones $\times 100$), teniendo en cuenta el número de huecos y la longitud de cada uno de los huecos, que deben introducirse para una alineación óptima de las dos secuencias. La comparación de secuencias y la determinación del porcentaje de identidad entre dos secuencias se puede lograr usando un algoritmo matemático, como se describe a continuación en este documento. Por ejemplo, la identidad de secuencia se puede determinar convencionalmente con el uso de programas informáticos, tal como el programa Bestfit (Wisconsin Sequence Analysis Package, Versión 8 para Unix, Genetics Computer Group, University Research Park, 575 Science Drive Madison, WI 53711). Bestfit utiliza el algoritmo de homología local de Smith y Waterman, *Advances in Applied Mathematics* 2 (1981), 482-489, para encontrar el segmento con la mayor identidad de secuencia entre dos secuencias. Cuando se usa Bestfit u otro programa de alineación de secuencias para determinar si una secuencia particular tiene, por ejemplo, una identidad de un 95 % con una secuencia de referencia de la presente invención, los parámetros se ajustan preferiblemente de modo que el porcentaje de identidad se calcule en toda la longitud de la secuencia de referencia y que se permitan brechas de homología de hasta un 5 % del número total de nucleótidos en la secuencia de referencia. Cuando se usa Bestfit, los denominados parámetros opcionales se dejan preferiblemente en sus valores predeterminados ("por defecto"). Las desviaciones que aparecen en la comparación entre una secuencia dada y las secuencias de la invención descritas anteriormente pueden estar provocadas, por ejemplo, por adición, eliminación, sustitución, inserción o recombinación. Dicha comparación de secuencias también se puede realizar preferiblemente con el programa "fasta20u66" (versión 2.0u66, septiembre de 1998 por William R. Pearson y la Universidad de Virginia; véase también W.R. Pearson (1990), *Methods in Enzymology* 183, 63-98, ejemplos adjuntos y <http://workbench.sdsc.edu/>). Para este fin, pueden usarse los ajustes de parámetros "por defecto".

Otra indicación de que dos secuencias de ácido nucleico son sustancialmente idénticas es que las dos moléculas hibridan entre sí en condiciones rigurosas. La frase: "hibridación específicamente con" se refiere a la unión, duplicación o hibridación de una molécula solamente con una secuencia de nucleótidos particular en condiciones rigurosas cuando esa secuencia está presente en una mezcla compleja de ADN o ARN (por ejemplo, celular total). "Se une(n) sustancialmente" se refiere a la hibridación complementaria entre un ácido nucleico sonda y un ácido nucleico diana y abarca desajustes menores que pueden acomodarse reduciendo la rigurosidad de los medios de hibridación para lograr la detección deseada de la secuencia de ácido nucleico diana.

Las "condiciones de hibridación rigurosas" y "condiciones de lavado de hibridación rigurosas", en el contexto de los experimentos de hibridación de ácidos nucleicos, tales como las hibridaciones de Southern y Northern, dependen de la secuencia y son diferentes para parámetros ambientales distintos. Las secuencias más largas hibridan específicamente a temperaturas más elevadas. Una guía extensa para la hibridación de ácidos nucleicos se encuentra en Tijssen (1993) *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. En general, las condiciones de hibridación y lavado altamente rigurosas se seleccionan para que sean aproximadamente 5 °C más bajas que el punto de fusión térmico para la secuencia específica a una fuerza iónica y pH definidos. Típicamente, en "condiciones rigurosas" hibridará una sonda con su subsecuencia diana, pero no con otras secuencias.

El punto de fusión térmico es la temperatura (a una fuerza iónica y pH definidos) a la que el 50 % de la secuencia diana hibrida con una sonda perfectamente emparejada. Las condiciones muy rigurosas se seleccionan para que sean iguales a la temperatura de fusión (T_m) para una sonda particular. Un ejemplo de condiciones de hibridación rigurosas para la hibridación de ácidos nucleicos complementarios que tienen más de 100 residuos complementarios en un filtro en una transferencia de Southern o Northern son formamida al 50 % con 1 mg de heparina a 42 °C, realizándose la hibridación durante la noche. Un ejemplo de condiciones de lavado altamente rigurosas es NaCl 0,15 M a 72 °C durante aproximadamente 15 minutos. Un ejemplo de condiciones de lavado rigurosas es un lavado con SSC 0,2x a 65 °C durante 15 minutos (véase, Sambrook, *infra*, para una descripción del tampón SSC). A menudo, un lavado de rigurosidad alta va precedido por un lavado de rigurosidad baja para eliminar la señal de fondo de la sonda. Un ejemplo de lavado de rigurosidad media para una estructura bicatenaria de, por ejemplo, más de 100 nucleótidos, es SSC 1x a 45 °C durante 15 minutos. Un ejemplo de lavado de rigurosidad baja para una estructura bicatenaria de, por ejemplo, más de 100 nucleótidos, es SSC 4-6x a 40 °C durante 15 minutos. Para sondas cortas (por ejemplo, de aproximadamente 10 a 50 nucleótidos), las condiciones rigurosas implican típicamente concentraciones de sal de menos de aproximadamente 1,0 M de ion Na, típicamente de aproximadamente 0,01 a 1,0 M de ion Na (u otras sales) a pH 7,0 a 8,3, y la temperatura es típicamente de al menos aproximadamente 30 °C. También se pueden lograr condiciones rigurosas con la adición de agentes desestabilizantes tales como formamida. En general, una relación de señal a ruido de 2 veces (o mayor) que la observada para una sonda no relacionada en el ensayo de hibridación particular indica detección de una hibridación específica. Los ácidos nucleicos que no hibridan entre sí en condiciones rigurosas siguen siendo sustancialmente idénticos si las proteínas que codifican son sustancialmente idénticas. Esto

ocurre, por ejemplo, cuando se crea una copia de un ácido nucleico usando la máxima degeneración de codones permitida por el código genético.

Una "planta" es cualquier planta en cualquier estado de desarrollo, particularmente una planta de semilla.

Una "parte de la planta", como se usa en este documento, se refiere a subunidades estructurales y/o funcionales de una planta incluyendo, aunque sin limitación, células vegetal, tejidos vegetales, material vegetal, órganos vegetales, partes recolectables de la planta, etc., como se define a continuación en este documento.

Una "parte recolectable de la planta" es una parte de una planta que se refiere a esas partes de la planta que se recolectan en cualquier momento adecuado y pueden procesarse además para uso industrial o consumo, incluyendo flores, frutos, hojas, semillas, fibras, etc.

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 como parte de una unidad con una organización superior tal como, por ejemplo, un tejido vegetal, un órgano vegetal o una planta entera.

El "cultivo de células vegetales" significa cultivos de unidades vegetales tales como, por ejemplo, protoplastos, células de cultivo celular, células en tejidos vegetales, polen, tubos polínicos, óvulos, sacos embrionarios, cigotos y embriones en diferentes estados de desarrollo.

"Material vegetal" o "material vegetal que se obtiene de una planta" se refiere a hojas, tallos, raíces, flores o partes de flores, frutos, polen, óvulos, cigotos, semillas, esquejes, cultivos celulares o de tejidos, o cualquier otra parte o producto de una planta.

Un "órgano de una planta" es una parte distinta y visiblemente estructurada y diferenciada de una planta, tal como una raíz, un tallo, una hoja, una yema floral o un embrión.

"Tejido vegetal", como se usa en este documento, se refiere a un grupo de células vegetales organizadas en una unidad estructural y funcional. Se incluye cualquier tejido de una planta en la planta o en cultivo. Esta expresión incluye, aunque sin limitación, plantas enteras, órganos vegetales, semillas de plantas, cultivo tisular y cualquier grupo de células vegetales organizadas en unidades estructurales y/o funcionales. El uso de esta expresión junto con, o en ausencia de, cualquier tipo específico de tejido vegetal, como los enumerados anteriormente o que esté contemplado de otra manera por esta definición, no pretende ser exclusivo de ningún otro tipo de tejido vegetal.

Como se usa en este documento, el término "población" significa un conjunto genéticamente heterogéneo de plantas que comparten una derivación genética común.

Como se usa en este documento, el término "variedad" o "cultivar" significa un grupo de plantas similares que, por sus características estructurales y su rendimiento, pueden identificarse a partir de otras variedades dentro de la misma especie.

Como se usa en este documento, la expresión "técnicas de transformación de plantas" se refiere a la introducción de un transgén, que confiere un rasgo específico, en la planta hospedadora. El transgén se incorpora en el genoma de la planta hospedadora y se hereda de forma estable a través de futuras generaciones. Se añaden las secuencias reguladoras correctas al gen de interés, es decir, promotores y terminadores, después el ADN se transfiere al cultivo de células vegetales usando un vector apropiado. En algunas realizaciones, el gen se fija a un marcador de selección que permite la selección por la presencia del transgén como se describe en este documento anteriormente. Una vez se ha transformado el tejido vegetal, las células que contienen el transgén se seleccionan y se realiza la regeneración de nuevo en plantas completas.

La transformación de las plantas puede realizarse de varias maneras diferentes para generar hospedadores vegetales con funciones delegadas requeridas que hacen que el hospedador vegetal sea competente para la expresión de la secuencia de interés, dependiendo de la especie de la planta en cuestión. Por ejemplo, puede usarse transformación mediada por *Agrobacterium* para transformar plantas de acuerdo con la invención. Dentro de este método de transformación, se corta la planta o el tejido vegetal (por ejemplo, hojas) en pequeños trozos, por ejemplo, de 10 x 10 mm, y se remojan durante 10 minutos en un líquido que contiene *Agrobacterium* suspendida que contiene un vector de plásmido Ti portado por *Agrobacterium* (documentos US 5591616; US 4940838; US 5464763). Colocadas en un medio de enraizamiento y despunte de selección, las plantas volverán a crecer.

Las técnicas que no se basan en *Agrobacterium* implican la captación de material genético exógeno directamente por los protoplastos o las células. Estas técnicas incluyen, aunque sin limitación, captación mediada por PEG o electroporación, suministro mediado por bombardeo de partículas y microinyección. Ejemplos de estas técnicas se describen en Paszkowski *et al.*, EMBO J 3, 2717-2722 (1984), Potrykus *et al.*, Mol. Gen. Genet. 199,169-177 (1985), Reich *et al.*, Biotechnology 4:1001-1004 (1986) y Klein *et al.*, Nature 327,70-73 (1987). En cada caso, las células

transformadas se regeneran en plantas completas usando técnicas convencionales. Por ejemplo, en el bombardeo de partículas, se recubren partículas de oro o tungsteno con ADN y después se disparan en células vegetales jóvenes o embriones vegetales (documentos US 05100792; EP 00444882B1; EP 00434616B1). Algo de material genético quedará en las células y las transformará. Este método también puede usarse para transformar plástidos vegetales. Además, puede usarse la técnica de electroporación para transformar plantas de acuerdo con la invención. Durante la electroporación, se preparan orificios transitorios en las membranas celulares usando choque eléctrico que permite que el ADN entre en la célula. Otra técnica de transformación dentro de la invención puede ser transformación vírica (transducción). Aquí, el material genético deseado se empaqueta en un virus de plantas adecuado y se permite que este virus modificado infecte la planta. Si el material genético es ADN, puede recombinarse con los cromosomas para producir células transformadas. Sin embargo, los genomas de la mayoría de virus de plantas consisten en ARN monocatenario que se replica en el citoplasma de la célula infectada.

La transformación o genomanipulación de la planta o célula vegetal con una secuencia de nucleótidos o el vector de acuerdo con la invención puede realizarse por métodos convencionales, como se describe, por ejemplo, en Sambrook y Russell (2001), *Molecular Cloning: A Laboratory Manual*, CSH Press, Cold Spring Harbor, NY, EE. UU. Además, la célula vegetal transgénica de la presente invención se cultiva en medio nutritivo que cumple los requisitos de la célula particular usada, en particular con respecto al valor de pH, la temperatura, la concentración salina, la aireación, los antibióticos, las vitaminas, los oligoelementos, etc.

La expresión "vector" o "molécula de vector", como se usa en este documento, puede comprender un casete de expresión que puede comprender secuencias de control de la expresión unidas de forma funcional a dicho polinucleótido. El vector o vectores pueden estar en forma de un plásmido, y pueden usarse en solitario o en combinación con otros plásmidos, para proporcionar plantas transformadas, usando métodos de transformación como se describe a continuación para incorporar transgenes en el material genético de la planta o las plantas. Vectores adicionales pueden comprender cósmidos, virus, bacteriófagos y otros vectores habitualmente usados en ingeniería genética.

Un "transgén" se refiere a un gen que se ha introducido en el genoma por transformación y se mantiene de forma estable. Los transgenes pueden incluir, por ejemplo, genes que son heterólogos u homólogos a los genes de una planta particular a transformar. Además, los transgenes pueden comprender genes naturales insertados en un organismo no natural, o genes quiméricos.

"Secuencia codificante" se refiere a una secuencia de ADN o ARN que codifica una secuencia de aminoácidos específica y excluye las secuencias no codificantes. Puede constituir una "secuencia codificante ininterrumpida", es decir, que carece de un intrón, tal como en un ADNc o puede incluir uno o más intrones unidos mediante uniones de corte y empalme apropiadas. Un "intrón" es una secuencia de ARN que está contenida en el transcrito primario, pero que se retira a través de escisión y religamiento del ARN dentro de la célula para crear el ARNm maduro que puede traducirse en una proteína.

La expresión "casete de expresión", como se usa en este documento, puede estar compuesto de una o más secuencias de nucleótidos de la presente invención en unión funcional con secuencias de nucleótidos reguladoras que controlan su expresión. Como se sabe en la técnica, un casete de expresión puede consistir en una secuencia promotora (promotor), un marco abierto de lectura o una parte funcional del mismo, una región no traducida 3' y una secuencia terminadora (terminador). El casete puede ser parte de una molécula de vector como se describe en este documento anteriormente. Diferentes casetes de expresión pueden transformarse en plantas o células vegetales siempre que se usen las secuencias reguladoras correctas.

Como se usa en este documento, el término "promotor" se refiere a una secuencia de nucleótidos, habitualmente anterior (5') a su secuencia codificante, que controla la expresión de la secuencia codificante proporcionando el reconocimiento para la ARN-polimerasa y otros factores necesarios para una transcripción apropiada. Un "promotor" incluye un promotor mínimo que es una secuencia corta de ADN compuesta de un dominio TATA y otras secuencias que sirven para especificar el sitio de inicio de la transcripción, al que se añaden elementos reguladores para el control de la expresión. Además, un "promotor" también se refiere a una secuencia de nucleótidos que incluye un promotor mínimo más elementos reguladores, que puede controlar la expresión de una secuencia codificante o ARN funcional. Este tipo de secuencia promotora consiste en elementos anteriores proximales y más distales, denominándose los últimos elementos a menudo potenciadores. Por consiguiente, un "potenciador" es una secuencia de ADN que puede estimular la actividad promotora y puede ser un elemento innato del promotor o un elemento heterólogo insertado para potenciar el nivel o especificidad de tejido de un promotor. Puede funcionar en ambas orientaciones (normal o invertido), y puede funcionar incluso cuando se mueve en dirección 5' o 3' desde el promotor. Tanto los potenciadores como otros elementos promotores en dirección 5' se unen a proteínas de unión a ADN específicas de secuencia que median sus efectos. Los promotores pueden derivar en su totalidad de un gen natural, o pueden estar compuestos de diferentes elementos derivados de diferentes promotores encontrados en la naturaleza, o incluso pueden comprender segmentos de ADN sintético. Un promotor también puede contener secuencias de ADN que están implicadas en la unión de factores proteínicos que controlan la eficacia del inicio de la transcripción en respuesta a condiciones fisiológicas o del desarrollo.

El "sitio de inicio" es la posición que rodea el primer nucleótido que es parte de la secuencia transcrita, que también se define como la posición +1. Con respecto a este sitio, se numeran todas las demás secuencias del gen y sus regiones de control. Las secuencias posteriores (es decir, secuencias adicionales que codifican proteína en la dirección 3') se denominan positivas, mientras que las secuencias anteriores (principalmente las regiones de control en la dirección 5') se denominan negativas.

Los elementos promotores, particularmente un elemento TATA, que son inactivos o que tienen actividad promotora enormemente reducida en ausencia de activación en dirección 5' se denominan "promotores mínimos o centrales". En presencia de un factor de transcripción adecuado, el promotor mínimo funciona permitiendo la transcripción. Un "promotor mínimo o central", por tanto, consiste únicamente en todos los elementos básicos necesarios para el inicio de la transcripción, por ejemplo, un dominio TATA y/o un iniciador.

De acuerdo con la invención, el término "promotor" es una región reguladora de ácido nucleico (por ejemplo, ADN) que dirige la transcripción de un gen. Los promotores están ubicados típicamente cerca de los genes que regulan, en la misma hebra y anteriores (hacia la región 5' de la hebra de sentido). Varios tipos de promotores son bien conocidos en las técnicas de transformación, así como otros elementos reguladores que pueden usarse en solitario o en combinación con otros promotores. Un "promotor vegetal" es un promotor que puede iniciar la transcripción en células vegetales. Ejemplos de promotores bajo control por el desarrollo incluyen "promotores específicos de tejidos" que se refieren a promotores regulados que no se expresan en todas las células vegetales, sino solamente en uno o más tipos celulares en órganos específicos (tal como hojas o semillas), tejidos específicos (tal como embrión o cotiledón), o tipos celulares específicos (tal como parénquima de las hojas o células de almacenamiento de las semillas). Estos también incluyen promotores que se regulan temporalmente, tal como en embriogénesis temprana o tardía, durante la maduración de los frutos en semillas o frutos en desarrollo, en hojas completamente diferenciadas, o en el inicio de la senescencia.

Un promotor específico de "tipo celular" o también llamado "promotor inducible" principalmente dirige la expresión en determinados tipos celulares en uno o más órganos, por ejemplo, células vasculares en raíces u hojas mediante un estímulo externo, tal como un agente químico, luz, hormona, agresión o un patógeno.

"Expresión constitutiva" se refiere a la expresión usando un promotor constitutivo o regulado. "Expresión regulada" o "condicionada" se refiere a la expresión controlada por un promotor regulado.

"Promotor constitutivo" se refiere a un promotor que puede expresar el marco abierto de lectura (ORF) que predomina en todos o casi todos los tejidos vegetales durante todos o casi todos los estados de desarrollo de la planta. Cada uno de los elementos activadores de la transcripción no muestra una especificidad de tejido absoluta, sino que media la activación transcripcional en la mayoría de partes de la planta a un nivel de ≥ 1 % del nivel alcanzado en la parte de la planta en que la transcripción es más activa.

"Promotor regulado" se refiere a promotores que no dirigen la expresión génica de forma constitutiva, sino de un modo regulado temporal y/o espacialmente, e incluyen tanto promotores específicos de tejido como inducibles. Incluye secuencias naturales y sintéticas, así como secuencias que pueden ser una combinación de secuencias sintéticas y naturales. Diferentes promotores pueden dirigir la expresión de un gen en diferentes tejidos o tipos celulares, o en diferentes estados de desarrollo, o en respuesta a diferentes condiciones ambientales. Constantemente se están descubriendo nuevos promotores de diversos tipos útiles en células vegetales, numerosos ejemplos pueden encontrarse en la compilación de Okamoto *et al.* (1989). Los promotores regulados típicos útiles en plantas incluyen, aunque sin limitación, promotores inducibles por protectores, promotores derivados del sistema inducible por tetraciclina, promotores derivados de sistemas inducibles por salicilato, promotores derivados de sistemas inducibles por alcohol, promotores derivados del sistema inducible por glucocorticoesteroides, promotores derivados de sistemas inducibles por patógenos y promotores derivados de sistemas inducibles por ecdisona.

"Unido de forma funcional" se refiere a la asociación de secuencias de ácido nucleico en un solo fragmento de ácido nucleico de modo que la función de uno se vea afectada por el otro. Por ejemplo, una secuencia de ADN reguladora se dice que está "unida de forma funcional a" o "asociada con" una secuencia de ADN que codifica un ARN o un polipéptido si las dos secuencias están situadas de modo que la secuencia de ADN reguladora afecta a la expresión de la secuencia de ADN codificante (es decir, que la secuencia codificante o ARN funcional está bajo el control transcripcional del promotor). Las secuencias codificantes pueden unirse de forma funcional a secuencias reguladoras en orientación de sentido o de antisentido.

En el contexto anterior, el término "terminador" se refiere a una región reguladora de una secuencia de ADN que marca el final de un gen en ADN genómico para la transcripción y, por tanto, inicia la terminación de la transcripción del gen respectivo.

"Expresión", como se usa en este documento, se refiere a la transcripción y/o traducción de un gen endógeno, ORF o parte del mismo, o un transgén en plantas. Por ejemplo, en el caso de construcciones de antisentido, la expresión puede referirse a la transcripción del ADN de antisentido solamente. Además, expresión se refiere a la transcripción y

acumulación estable de ARN de sentido (ARNm) o funcional. Expresión también puede referirse a la producción de proteína.

5 "Sobreexpresión" se refiere al nivel de expresión en células u organismos transgénicos que excede los niveles de expresión en células u organismos normales o no transformados (no transgénicos).

"Inhibición antisentido" se refiere a la producción de transcritos de ARN de antisentido que pueden suprimir la expresión de la proteína desde un gen endógeno o un transgén.

10 "Silenciamiento génico" se refiere a supresión dependiente de homología de genes víricos, transgenes o genes nucleares endógenos. El silenciamiento génico puede ser transcripcional, cuando la supresión se debe a transcripción disminuida de los genes afectados, o postranscripcional, cuando la supresión se debe a recambio (degradación) aumentado de especies de ARN homólogas a los genes afectados (English *et al.*, 1996). El silenciamiento génico incluye silenciamiento génico inducido por virus (Ruiz *et al.*, 1998).

15 Como se usa en este documento, el término "BAC" significa cromosoma artificial bacteriano y define una construcción de ADN, basada en un plásmido de fertilidad funcional usado para transformación y clonación en bacterias, tales como *E. coli*. Los BAC pueden usarse para secuenciar el genoma de organismos tales como plantas. Un corto trozo del ADN del organismo se amplifica como un inserto en BAC, y después se secuencía. Finalmente, las partes secuenciadas se reordenan *in silico*, produciendo la secuencia genómica del organismo.

20 Como se usa en este documento, "vigor aumentado de las semillas" se refiere a la capacidad de una semilla de germinar rápidamente, de una manera uniforme, y de conseguir un alto porcentaje de germinación; de producir plántulas que brotan rápidamente del suelo y tienen un alto porcentaje de brote; de producir plántulas que crecen rápidamente y demuestran tolerancia superior a diversas agresiones incluyendo, aunque sin limitación, frío. El vigor de las semillas es la cuantificación de cualquiera de los parámetros mencionados anteriormente y cualquier combinación de estos.

25 "Germinación de las semillas" se define como el brote de la radícula desde el tegumento.

30 "Velocidad de germinación", a partir de ahora en este documento, se refiere al promedio de tiempo observado entre la imbibición de la semilla y el brote de la radícula desde el tegumento.

35 A lo largo de la aplicación, la velocidad de germinación puede medirse finalmente calculando el periodo de tiempo que se necesita para observar un 50 % de germinación de las semillas (resultados expresados como una medición T50).

40 Debe entenderse "velocidad aumentada de germinación" como una diferencia observable significativa entre la germinación de la semilla que comprende un polinucleótido de acuerdo con la presente invención frente a una semilla que no comprende uno de dichos polinucleótidos. Típicamente, una velocidad aumentada de germinación significa que la semilla que comprende un polinucleótido de acuerdo con la presente invención germina antes que una semilla que no comprende la SEQ ID NO 3.

45 Debe entenderse "velocidad modificada de germinación" como una diferencia observable significativa entre la germinación de la semilla que comprende un polinucleótido de acuerdo con la presente invención frente a una semilla que no comprende uno de dichos polinucleótidos. Típicamente, una velocidad modificada de germinación significa que la semilla que comprende un polinucleótido de acuerdo con la presente invención germina de forma diferente a la semilla que no comprende la SEQ ID NO 3.

50 Como se usa en este documento, la uniformidad de la germinación de las semillas puede expresarse como el tiempo entre un 10 % de germinación y un 90 % de germinación. Cuanto más corto sea este periodo, mejor será la uniformidad de la germinación de las semillas.

55 Las "condiciones ambientales externas adversas" son condiciones que inhiben o posponen la germinación de la semilla o el brote de la plántula. En el contexto de la presente invención, el frío es un factor, entre otros, que puede considerarse adverso respecto a las condiciones de germinación normales.

Se entiende que "brote (de una plántula)" se refiere al crecimiento de la planta que es observable.

60 Debe entenderse "brote aumentado de plántulas" como una diferencia observable significativa entre el brote de una plántula desde una semilla que comprende un polinucleótido de acuerdo con la presente invención frente a una semilla que no comprende la SEQ ID NO 3.

Breve descripción de las figuras

La figura 1 es una representación gráfica de las curvas de germinación acumulativa de semillas de la línea de sustitución SL101 (○) y líneas progenitoras (A12DHd (■) y GDDH33 (●)) a 15 °C en agua. Las líneas verticales son los errores típicos.

5 La figura 2 es una representación gráfica de las curvas de germinación acumulativa de SL101 (Δ) y el progenitor recurrente A12DHd (■) y líneas de retrocruzamiento F₁ recíprocas (A12DHd x SL101 (Δ) y SL101 x A12DHd (○)). Las líneas verticales son los errores típicos.

10 La figura 3 es una representación gráfica del brote de plántulas acumulativo en el terreno a partir de semillas de la línea de sustitución SL101 (▲) y líneas progenitoras A12DHd (●) y GDD33H (■).

La figura 4 es un diagrama que ilustra las concentraciones endógenas de ABA durante la germinación en semillas de la línea de sustitución SL101 (columna blanca) y líneas progenitoras, GDDH33 (columna negra) y A12DHd (columna gris).

15 La figura 5 es un diagrama que ilustra la velocidad de germinación de las semillas en 3, para el gen *At3g01060* (columnas negras), y 2, para el gen *At3g01150* (columnas grises), líneas KO. La línea de control de tipo silvestre (columna blanca) es ColO.

20 La figura 6 es un diagrama que ilustra la velocidad de germinación de las semillas en líneas KO 102 (línea gris claro) y 18 (línea gris medio bajo) (para el gen *At3g01060*) y en líneas KO 15 (líneas gris medio alto) y 3 (línea gris oscuro) (para el gen *Atg01150*) en comparación con la línea de control de tipo silvestre ColO (línea negra), después de la producción de semillas en el invernadero (condiciones más agresivas).

25 Ejemplo 1

Material y métodos

Producción de semillas y comparación de líneas

30 Se obtuvieron muestras de semillas de la Universidad de Birmingham, R.U., para una serie de líneas de sustitución cromosómica de *Brassica oleracea* derivadas de las líneas progenitoras doble haploides A12DHd (var. *alboglabra*) y GDDH33 (var. *italica*; Rae *et al.*, 1999). Entonces se produjeron semillas en masa y se recogieron de 10 plantas replicadas individuales de las líneas de sustitución y el progenitor GDDH33 y 20 plantas del progenitor recurrente A12DHd como líneas de sustitución se comparan con la última en experimentos de germinación. Las plantas se disponen en un bloque aleatorizado con 10 réplica en un invernadero a 16-18 °C durante las 16 h del día y a 10-15 °C en la noche, como se describe por Bettey *et al.* (2000). Se aportaba iluminación complementaria (lámparas de sodio de alta presión de 400 W; Osram Ltd, St Helens, R.U.) cuando la intensidad de la luz caía por debajo de 300 w m² durante las 16 h del día. Las plantas se autopolinizaron encerrando las inflorescencias en bolsas de polietileno perforadas que contenían moscardones antes de que las flores se abrieran. Se permitió que las vainas de la semilla se secaran completamente sobre la planta dentro de las bolsas circundantes antes de la recolección. Las semillas se limpiaron, se equilibraron a un 15 % de HR y 15 °C, y después se almacenaron a -20 °C antes de que se realizaran los experimentos de germinación. La germinación acumulativa en papel de filtro húmedo se registró a 15 °C en 4 réplicas de 15 semillas recogidas de cada una de las 10 plantas replicadas (o 20, A12DHd) descritas anteriormente. Trabajos previos habían demostrado que estas son suficientes semillas (Bettey *et al.*, 2000). Se hicieron recuentos frecuentes para permitir un cálculo preciso del tiempo hasta un 50 % de germinación a partir de estas mediciones. El porcentaje de germinación fue elevado en todos los lotes de semillas.

50 En los últimos experimentos, se produjeron semillas F₁ de retrocruzamientos recíprocos de la misma manera que se describe anteriormente. La polinización de yemas se realizó para hacer el cruce que produce F₁. Las semillas de las líneas progenitoras se produjeron al mismo tiempo para hacer comparaciones para minimizar la influencia de las diferencias ambientales durante la producción de semillas. Además, en varias ocasiones en diferentes momentos del año, se produjeron semillas de plantas replicadas tanto del progenitor A12DHd como de la línea de sustitución SL101 en invernaderos como se describe anteriormente. Aunque los ajustes de calentamiento, ventilación e iluminación del invernadero permanecieron igual que los descritos anteriormente, la temperatura ambiente difería y se registró.

Ensayos de germinación

60 Tres réplicas biológicas de 50 semillas de líneas de *Brassica oleracea* de sustitución y progenitoras o de 3 a 15 réplicas biológicas de 50 semillas de líneas de tipo silvestre y mutantes de *Arabidopsis* se colocaron para que germinaran en 2 capas de papel de filtro (Whatman International Ltd., R.U.) que mantuvieron húmedas durante todo el tratamiento con agua.

65 En todos los experimentos de germinación, las semillas en el papel de filtro se mantuvieron en cajas de poliestireno transparentes dispuestas en bloques aleatorizados y se mantuvieron en la oscuridad. No se observaron evidencias de

infección fúngica y, por tanto, las semillas no se esterilizaron para evitar influir en su germinación. La germinación (brote de la radícula) se registró a intervalos para construir curvas de germinación acumulativa.

Brote en el terreno

Como parte de una comparación no publicada más grande del brote de plántulas a partir de genotipos de *B. oleracea*, se sembraron 100 semillas de las líneas progenitoras GDDH33, A12DHd y la línea de sustitución SL101 el 31 de mayo, en 4 filas de 1 m replicadas dispuestas en un bloque aleatorizado. Las semillas se sembraron a mano en un surco de 15 mm de profundidad, se cubrieron con suelo tamizado (tamaño de orificio del tamiz < 4 mm) y a la superficie se le pasó rodillo una vez con una rueda de prensa taladradora de semillas Stanhay. El suelo era un suelo franco arenoso y se aplicó irrigación para mantener la humedad del suelo durante toda la germinación de las semillas y el brote de plántulas. Lo último se registró a intervalos regulares hasta que no brotaron más plántulas.

Análisis de hormonas

Se tomaron muestras de semillas de líneas de sustitución y progenitoras como semillas secas no embebidas y semillas embebidas durante 24 y 48 horas a 15 °C en papel de filtro húmedo. Cada muestra contenía 1 g de semillas secas medido antes de la imbibición. Las muestras se colocaron inmediatamente en nitrógeno líquido, se liofilizaron y después se colocaron en un congelador doméstico a -20 °C hasta la extracción. Las muestras estaban en 10 ml de metanol al 80 % (que contenía 20 mg l-l de BHT) frío (4 °C), después se añadieron 500 ng de ABA y 100 ng de cada uno de los patrones de GA. Las muestras se agitaron durante una noche en una vitrina fría y después se centrifugaron. El sobrenadante se decantó y el sedimento se resuspendió en 10 ml adicionales de metanol al 80 %. Las muestras se agitaron durante 4 h, se centrifugaron y los sobrenadantes se combinaron. Los sobrenadantes se evaporaron a acuosidad (aprox. 3-4 ml) y se añadieron 10 ml de tampón fosfato 0,5 M pH 8,2, las muestras se repartieron con 2 x 15 ml de diclorometano, y el diclorometano se descartó. La fracción acuosa se ajustó hasta pH 3 con ácido fosfórico 1 M repartido con 3 x 15 ml de acetato de etilo.

Las fracciones de acetato de etilo combinadas se lavaron con 2 x 3,5 ml de agua pH 3,0 y se evaporaron a sequedad, se volvieron a disolver en 5 ml de agua y el pH se ajustó hasta 8. La solución entonces se cargó en cartuchos QMA que se lavaron posteriormente con 5 ml de metanol al 15 % pH 8,0. Los GA y ABA se eluyeron de los cartuchos QMA directamente en cartuchos C18 21 con ácido fórmico 0,2 M en metanol al 5 %. Los cartuchos C18 entonces se lavaron con 5 ml de metanol al 20 % y las muestras se recuperaron en 5 ml de metanol al 80 % y se evaporaron a sequedad. Las muestras entonces se disolvieron en metanol y se metilaron con exceso de diazometano etéreo. Después de la evaporación a sequedad, las muestras se volvieron a disolver en acetato de etilo seco y se pasaron a través de cartuchos de amino. Las muestras resultantes se analizaron directamente para el contenido de ABA mediante CG-EM, después se evaporaron a sequedad y se volvieron a disolver en BSTFA antes del análisis para el contenido de GA por CG-EM.

Análisis de marcadores

Se diseñaron pares de cebadores para 30 modelos de genes de *Arabidopsis* que se propagaron a intervalos por la región *SOG1* usando el programa informático Primer 3 (http://gene.pbi.nrc.ca/cgibin/primer/primer3_www.cgi) y los datos de genes de Tair (<http://www.arabidopsis.org/>) para dar productos de PCR de 200 a 700 pb. La mezcla de PCR usada era convencional, pero se usó un programa de rampa descendente. Esta consistía en los parámetros de ciclado siguientes: 94 °C durante 5 min; después hibridación a 65 °C hasta 55 °C durante 10 ciclos bajando un grado en cada ciclo con 30 s de extensión a 72 °C y 30 s de desnaturalización a 94 °C durante los 10 ciclos; seguidos de 30 ciclos de 94 °C de 30 s, 55 °C de 30 s, 72 °C de 45 s; y una extensión final a 72 °C durante 15 min. Las secuencias de los cebadores para los modelos de genes que dieron resultados polimórficos se seleccionaron como marcadores (tabla 1).

Análisis de los datos

Todos los análisis se realizaron usando el paquete estadístico Genstat 5 (Payne *et al.* 1993), y cuando era apropiado, los datos se sometieron a análisis de la varianza.

Resultados

El QTL para la velocidad de germinación (SOG1) en el grupo de ligamiento C1 se confirmó y se cartografió de forma precisa

El análisis de la varianza de los datos de germinación que compara las líneas de sustitución y progenitoras mostró que el progenitor GDDH33 germinaba significativamente ($P < 0,001$) más rápido que el progenitor A12DHd, lo que confirma que los alelos positivos de la velocidad de germinación se proporcionan por GDDH33, como se muestra por Bettey *et al.* (2000) (figura 1). Había 4 líneas de sustitución que abarcaban el QTL *SOG1* (SL101, SL111, SL118, SL119) y todas estas tenían una germinación significativamente ($P < 0,005$) más rápida que el progenitor A12DHd. La línea de sustitución SL101 tenía la región de introgresión más pequeña (1-9 cM; Rae *et al.*, 1999) que potenciaba la

velocidad de germinación en comparación con A12DHd y explicaba gran parte de la diferencia en la velocidad de germinación entre las líneas progenitoras (figura 1) y, por lo tanto, se seleccionó para estudio adicional de *SOG1*.

El fenotipo de germinación rápida SOG1 no se ve influido por el genotipo materno

La velocidad de germinación se determina por el embrión, pero también puede verse influido significativamente por los tejidos que lo rodean que son de origen materno. Se realizaron retrocruzamientos recíprocos entre SL101 y A12DHd y entre GDDH33 y A12DHd para determinar los componentes genéticos maternos y cigóticos en el locus *SOG1*. La germinación se registró a partir de las semillas F1 de cada cruce y a partir de las semillas de las líneas progenitoras autopolinizadas producidas al mismo tiempo. No hubo diferencias significativas en la velocidad de germinación de SL101 y la F1 de los retrocruzamientos recíprocos con A12DHd (SL101 como planta madre y polen de A12DHd y viceversa), pero la germinación de las semillas de las tres fue significativamente ($P < 0,01$) más rápida que la de semillas de A12DHd (figura 2). Esto muestra que el alelo GDDH33 de germinación más rápida es dominante sin influencia materna genética sobre la herencia del rasgo, que confirma que se basa en el embrión.

Diferencias en la velocidad de germinación da lugar a diferencias en la cronología del brote de plántulas en el terreno

Los datos anteriores muestran que la velocidad de germinación de semillas GDDH33 y SL101 era significativamente mayor que la de A12DHd en condiciones de temperatura constante. En el terreno esto provocó brote de plántulas significativamente más pronto de GDDH33 y SL101 que de A12DHd (figura 3).

La concentración de ABA endógeno difiere entre genotipos en la madurez

La concentración endógena de ABA en semillas secas y de imbibición de los tres genotipos se midió usando CGEM. No hubo diferencias significativas en la concentración endógena de ABA en la semilla seca de SL101 y GD33 o durante la imbibición hasta 48 horas justo antes del brote de la radícula. La concentración de ABA en estos dos lotes de semillas permaneció igual durante las primeras 24 horas y después disminuyó hasta la mitad en 48 horas. Por el contrario, la concentración endógena de ABA en semillas de A12DHd inicialmente fue tres veces mayor que en SL101 y GDDH33 y después disminuyó progresivamente durante el periodo de 48 horas de imbibición, pero permaneció significativamente por encima de la de los otros dos lotes de semillas (figura 4). De forma interesante, si ABA sigue disminuyendo a la misma velocidad en A12DHd alcanzaría el mismo nivel después de 72 horas, el punto de germinación, que se observa en las otras dos líneas inmediatamente antes de su germinación.

Los resultados para las líneas GDDH33, SL101 y A12DHd presentados aquí muestran una clara relación determinada genéticamente entre el mayor contenido de ABA endógeno y la menor velocidad de germinación y viceversa.

Se ha realizado un análisis genético cuantitativo del rasgo de velocidad de germinación en Brassica oleracea.

Mediante cartografiado preciso del QTL con la línea de sustitución descrita previamente SL101, finalmente se identificaron dos genes subyacentes del QTL (*Speed Of Germination (SOG1)* - velocidad de germinación). Este línea SL101 tiene una corta introgresión en el extremo telomérico de C1 del progenitor de terminación rápida (GD33, brécol) en el fondo del progenitor de germinación lenta (A12, col rizada) en *Brassica oleracea*. Usando marcadores en toda esta región de introgresión de SL101, se identificaron 30 recombinaciones a lo largo de esta zona dentro de 1300 líneas, gracias a una estrategia de solapamiento de BAC. Esta estrategia permite reunir la líneas en 5 grupos distintos y las 2 líneas progenitoras. Entonces se evaluó la germinación de las semillas en la totalidad de estos 7 grupos y el análisis estadístico reveló que la germinación más rápida estaba asociada con 2 marcadores en el extremo telomérico de C1. La colocalización de estos marcadores con un solo BAC en el extremo telomérico de C1 se evaluó adicionalmente por hibridación de fluorescencia *in situ* (Fluorescence In-Situ Hybridization - FISH). El BAC se secuenció y se encontraron 12 genes de longitud completa presentes dentro del mismo.

Se han identificado supuestos ortólogos de estos genes de *Brassica* en *Arabidopsis*, en el brazo superior del cromosoma 3, donde se ha localizado un QTL *SOG1* (Clerkx *et al.* 2004). Basándonos en la buena colinealidad genética entre el extremo telomérico de C1 en *Brassica oleracea* y el brazo superior del cromosoma 3 en *Arabidopsis* (véase anteriormente y a continuación), es razonable creer que los genes que se identificaron en *Brassica* y en *Arabidopsis* comparten una función común en la germinación de las semillas. Para reforzar nuestra hipótesis, hemos identificado líneas mutantes de inactivación (KO) de *Arabidopsis* en supuestos genes ortólogos de los que se han descubierto en *Brassica*. Se descubrió que dos de estas líneas KO tenían un fenotipo de germinación significativo (germinación más rápida en comparación con la línea de control ColO), lo que sugiere que estos genes actúan como reguladores negativos de la germinación (figura 5). Es interesante observar que al menos dos líneas KO diferentes se usaron para evaluar el fenotipo de germinación y que estas distintas líneas KO mostraron resultados similares con respecto a la velocidad de germinación. Cuando se evalúan después de la producción de las semillas en condiciones más agresivas (figura 6), los resultados son muy comparables.

Este estudio funcional, por tanto, ha confirmado la función de los genes *Atg01060* y *Atg01150* en la regulación del vigor de las semillas, en particular en la regulación de la velocidad de germinación. Los ortólogos de *Brassica* de estos genes, *BoIC.VG1.a* y *BoIC.VG1.b*, que se han identificado dentro del QTL *SOG1* en *Brassica oleracea*, por lo tanto,

pueden considerarse, de hecho, herramientas que permiten la modulación del vigor de las semillas, más particularmente la modulación de la velocidad de la germinación de las semillas en *Brassicaceae*.

El ligamiento entre el grupo de ligamiento C1 de B. oleracea y el brazo superior del cromosoma tres de Arabidopsis se confirma

Estudios sobre SL101 anteriores han mostrado que una sola región sometida a introgresión en el extremo telomérico del grupo de ligamiento C1 (LGC1) contiene el QTL para *SOG1*. En el trabajo actual, tenemos como objetivo establecer la colinealidad entre esta región en *B. oleracea* y el genoma de *Arabidopsis* para posibilitar la comparación con el análisis exhaustivo de QTL realizado en esta especie modelo. Previamente se han mostrado varios ligamientos entre el genoma de *Brassica* y *Arabidopsis* (Cogan *et al.*, 2004; Parkin *et al.*, 2005).

En particular, el ligamiento entre el LGC1 y *Arabidopsis* (Cogan *et al.*, 2004) se utilizó para ayudar en el desarrollo de marcadores informativos adicionales. Usando esta estrategia, se diseñaron pares de cebadores para 30 modelos de genes de *Arabidopsis*, que se propagaron por esta región. Estos cebadores se ensayaron para determinar si amplificaban un producto de *B. oleracea* y después si había algún polimorfismo entre SL101 y las líneas progenitoras A12DHD y GD33DHD (tabla 1). Un patrón de bandas que es igual en SL101 y GD33DHD, pero diferente del de A12DHD indica su presencia en este locus y, por lo tanto, su utilidad como marcador para *SOG1*. Se identificaron cebadores para tres modelos de genes como marcadores informativos (*At3g01190*, *At3g07130*, *At3g02420*, tabla 1) que anclaron el ligamiento entre el brazo superior del cromosoma tres de *Arabidopsis* y la región *SOG1* de LGC1 de *B. oleracea*. Esta confirmación de colinealidad justifica la comparación de *SOG1* con QTL para el rendimiento de las semillas ubicado en esta región del genoma de *Arabidopsis*.

Tabla 1 Cebadores para marcadores seleccionados usados para anclar la introgresión SL101 en *Arabidopsis*

Modelo de gen	Cebadores
<i>At3g01190</i>	F: TTCTTCCACGACTGCTTCG
	R: CTAACAAAAGTATCCGTCAC
<i>At3g02420</i>	F: GTTGCGTTGCCATCTGCAG
	R: CAGGCTGAGATAGCCATTGG
<i>At3g07130</i>	F: CTACTAACCATGGAGTTACC
	R: AACGCTGGTGGGATTCAC

Conjuntamente, estos resultados abren la posibilidad de usar los alelos *Bo/C.VG1.a* A12/GD33 o los alelos *Bo/C.VG2.a* A12/GD33 o los alelos *Bo/C.VG2.b* A12/GD33 de acuerdo con cualquiera de las realizaciones 1 a 23, en particular para genomanipular plantas en que se ha modulado el vigor de las semillas, más particularmente plantas en que se ha modulado la velocidad de germinación.

Ejemplo 2. Experimentos que resaltan fenotipos adicionales ligados al vigor de las semillas.

Estos experimentos subrayan las diferencias en las características de germinación entre semillas con (SL101) o sin (A12) el alelo GD33.

Las semillas de las líneas A12 y SL101 se originaron en R.U. y se replicaron en 2009 en Sudáfrica mediante un productor de semillas comerciales. La determinación de las características de germinación de las semillas tuvo lugar a primeros de 2010 en Enkhuizen, Países Bajos.

Rendimiento de las semillas en condiciones comerciales convencionales

Se sembraron dos réplicas de 100 semillas en bandejas convencionales rellenas de suelo como en una práctica comercial normal. Las bandejas se colocaron en una cámara de germinación a 18 °C en la oscuridad durante tres días. Las bandejas entonces se transfirieron a un invernadero con un promedio de temperatura de 20 °C. A los 10 días después de la siembra, se contó el número de plántulas desarrolladas normalmente y el número de semillas que no habían brotado.

El rendimiento de SL101 en estas condiciones fue considerablemente mejor que para A12 (tabla 2).

Tabla 2 Porcentaje de plántulas normales y porcentaje de semillas que no han germinado, medidos en condiciones prácticas de producción de plántulas

	plántulas normales	semillas no germinadas
A12	19	80
SL101	91	7

Sensibilidad a la temperatura de la germinación sobre papel

5 Se sembraron dos réplicas de 100 semillas de cada uno de A12 y SL101 para cada combinación de condiciones sobre papel de filtro húmedo en cajas de germinación de plástico transparentes a temperaturas de 10, 20 o 30 °C. Las cajas se mantuvieron en la oscuridad a las temperaturas mencionadas o se colocaron bajo luz fluorescente.

10 La germinación, medida como prominencia radical, se contó diariamente. La germinación se contó hasta 10 días después del inicio de la medición cuando no se observó germinación adicional.

Los datos muestran que, en cada condición, el porcentaje de germinación era mayor para SL101 en comparación con A12 (tabla 3).

15 Tabla 3 Porcentaje de germinación final de A12 y SL101 a diferentes temperaturas en luz o en oscuridad

		temperatura		
		10	20	30
luz	A12	47	94	59
	SL101	94	100	99
oscuridad	A12	65	24	2
	SL 101	95	78	50

Velocidad de germinación sobre papel

20 La velocidad de germinación se determinó en condiciones de 20 °C en luz, una condición donde ambas líneas tenían su germinación máxima. Se colocaron dos réplicas de 50 semillas por línea sobre papel en cajas de germinación de plástico en las condiciones mencionadas. Se determinó el t50 (tiempo hasta que el 50 % de las semillas en germinación han germinado) (tabla 4).

25 Como se muestra en la tabla, tiempo hasta el 50 % de germinación fue considerablemente más corto para SL101 en comparación con A12.

Tabla 4 Tiempo hasta el 50 % de germinación a 20 °C en luz para A12 y SL101

línea	t50 (h)
A12	59,0
SL101	47,5

Sensibilidad a bajas temperaturas durante la germinación sobre papel

30 Se incubaron dos réplicas de 50 semillas por tratamiento a 5 °C o 10 °C en luz como se describe anteriormente para un ensayo de germinación sobre papel. La incubación fue en agua o en una solución 1 mM de GA3. Después de 10 días, se retiraron las semillas que habían germinado. Las cajas con las semillas no germinadas restantes se colocaron entonces a 20 °C en la luz. Después de 5 días más, se contó el número de semillas germinadas. Como comparación, se germinaron semillas de A12 y SL101 a 20 °C en luz sin un pretratamiento. El porcentaje de germinación de estas semillas se contó después de 10 días.

40 Especialmente a 5 °C, había una gran diferencia entre las dos líneas (tabla 5). El pretratamiento en agua evitó en gran medida la germinación de A12 a 20 °C en luz más tarde. Este no fue el caso para el pretratamiento en GA3 1 mM. SL101 mostró solamente una reducción mínima en la germinación después del pretratamiento a 5 °C en agua.

45 Tabla 5 Porcentaje de germinación de semillas de A12 y SL101 con o sin un pretratamiento a 5 o 10 °C en luz con un medio de incubación de agua o una solución acuosa 1 mM de GA3

	A12	SL101
sin pretratamiento	99	100
10 días 5 °C en agua	15	84
10 días 5 °C en GA3 1 mM	100	100
10 días 10 °C en agua	82	99
10 días 10 °C en GA3 1 mM	100	100

Ejemplo 3 Efecto del genotipo sobre el rendimiento de semillas híbridas

50 Se produjeron semillas híbridas en condiciones de producción de semillas comerciales en Sudáfrica polinizando flores de dos líneas estériles macho con polen de otras 7 líneas.

La línea 1 hembra contenía el alelo GD33, la línea 2 hembra contenía el alelo A12.

Se ensayo la sensibilidad a baja temperatura durante la germinación para estas semillas. Las semillas se incubaron sobre papel de filtro humedecido a 5 °C en luz fluorescente blanca durante 10 días. El porcentaje de germinación se registró después de 10 días y las semillas no germinadas se transfirieron a 20 °C en luz. Después de cinco días, se registró el porcentaje de semillas germinadas.

Las semillas producidas en la línea hembra que contenían el alelo GD33 mostraron germinación considerablemente mayor a 5 °C y germinaron más de un 90 % después de transferencia a 20 °C. Las semillas producidas en la hembra sin el alelo GD33 mostraron solamente porcentajes de germinación bajos y no se recuperaron después de transferencia a 20 °C.

Tabla 6 Porcentaje de germinación de semillas híbridas producidas usando una línea hembra GD33 o A12 con un pretratamiento durante 10 días a 5 °C en luz seguido de incubación de semillas no germinadas durante 5 días a 20 °C.

	hembra 1 (GD33)			hembra 2 (A12)	
	10 días 5 °C	+5 días 20 °C		10 días 5 °C	+5 días 20 °C
macho 1	35	96		0	6
macho 2	28	96		0	10
macho 3	72	94		19	44
macho 4	9	96		0	4
macho 5	43	100		0	8
macho 6	28	99		0	4
macho 7	69	97		35	58
promedio	41	97		8	19

Tabla 6 Porcentaje de germinación de semillas híbridas producidas usando una línea hembra GD33 o A12 con un pretratamiento durante 10 días a 5 °C en luz seguido de incubación de semillas no germinadas durante 5 días a 20 °C.

Se ensayaron semillas de los mismos lotes de semillas híbridas en un ensayo de germinación sobre papel a 25 °C en luz fluorescente blanca. De los recuentos diarios de germinación, se calculó el tiempo hasta el 50 % de germinación (t50). El porcentaje de germinación final se determinó después de 5 días. Todos los híbridos tenían al menos un 85 % de germinación, la mayoría más de un 95 %. Las semillas producidas en la línea hembra con el alelo GD33 mostraron una germinación de promedio aproximadamente un 25 % más rápida, mostrada por el mejor t50 de 1,33 días frente a 1,80 días para semillas producidas en la hembra con el alelo A12.

Tabla 7 Porcentaje de germinación y tiempo que se tarda hasta el 50 % de germinación (t50) de las semillas híbridas producidas usando una línea hembra GD33 o A12 en luz a 25 °C.

	hembra 1 (GD33)			hembra 2 (A12)	
	% de germ.	t50 (d)		% de germ.	t50 (d)
macho 1	100	1,08		99	1,73
macho 2	97	1,95		98	1,85
macho 3	98	1,85		100	1,90
macho 4	100	1,15		95	1,84
macho 5	99	1,07		98	1,68
macho 6	99	1,14		85	1,84
macho 7	95	1,05		98	1,79
promedio	98	1,33		96	1,80

Referencias

Betty M., Finch-Savage W.E., King G.J., Lynn J.R. 2000. Quantitative genetic analysis of seed vigour and pre-emergence seedling growth traits in *Brassica oleracea* L. New Phytol 148: 277-286.

Clercks E.J.M., El-Lithy M.E., Vierling E., Ruys G.J., Blankestijn-De Vries H., Groot S.P.C., Vreugdenhil D., Koornneef M. 2004. Analysis of natural allelic variation of *Arabidopsis* seed germination and seed longevity traits between the accessions Landsberg erecta and Shakdara, using a new recombinant inbred line population. Plant Physiol 135: 432-443.

Finch-Savage W.E. 1995. Influence of seed quality on crop establishment, growth and yield. En AS Basra, ed. Seed quality: Basic mechanisms and agricultural implications. Haworth Press, Inc, Nueva York, pág. 361-384.

Finch-Savage W.E., Côme D., Ly J.R., Corbineau F. 2005. Sensitivity of *Brassica oleracea* seed germination to hypoxia: a QTL analysis. *Plant Sci* 169: 753-759.

- 5 Finch-Savage W.E. y Leubner-Metzger G. 2006. Seed dormancy and the control of germination. *New Phytol.* 171: 501-523.

Finkelstein R.R., Gampala S.S., y Rock C.D. 2002. Absciscic acid signaling in seeds and seedlings. *Plant Cell* 14, S15-45.

- 10 Foolad M.R., Lin G.Y., Chen F.Q. 1999. Comparison of QTLs for seed germination under non-stress, cold stress and salt stress in tomato. *Plant Breed* 118: 167-173.

- 15 Groot S.P.C., van der Geest A.H.M., Tesnier K.J.Y., Alonso-Blanco C., Bentsink L., Donkers H., Koornneef M., Vreugdenhil D., Bino R.J. 2000. Molecular genetic analysis of *Arabidopsis* seed quality. En M Black, J Vasquez-Ramos, eds, *Seed Biology: Advances and applications*, CAB International., Londres, pág. 123-132.

- 20 Hilhorst H.W.M. y Toorop P.E. 1997. Review on dormancy, germinability and germination in crop and weed seeds. *Advances Agron* 61: 115-165.

Holdsworth M.J., Bentsink L., Soppe W.J.J. 2008a. Molecular networks regulating *Arabidopsis* seed maturation, after-ripening, dormancy and germination. *New Phytol* 179: 33-54.

- 25 Holdsworth M.J., Finch-Savage W.E., Grappin P., Job J. 2008b. Post-genomics dissection of seed dormancy and germination. *Trends Plant Sci* 13: 7-13.

Ruan S.L., Duan X.M. y Hu W.M. 2000. Occurrence of seed vivipary in hybrid rape (*Brassica napus* L.) and its effect on seed quality. *Journal of Zhejiang University (Agriculture and Life Sciences)* 2000 Vol. 26 n.º 5 pág. 573-578.

30 Alineaciones de secuencias

Alineación de secuencias de ADN para los genes VG2 de *Brassica oleracea*. Las alineaciones se llevaron a cabo utilizando el programa localizado en la red ClustalW (<http://www.genome.jp/tools/clustalw/>) y la anotación se representa utilizando el programa localizado en la red BOXSHADE (http://www.ch.embnet.org/software/BOX_form.html). A12_BOLC.VG2.A es una copia de longitud completa de A12 del gen ubicada en la región SOG1 del grupo de ligamiento C1. A12_BOLC.VG2.B es una copia truncada de un gen similar localizada dentro de 50 Kb del gen de longitud completa. Se ha utilizado una anotación similar para describir la misma región del contexto genómico de GD33.

```
A12_BOLC.VG2.A      1 GTAGGCCAACCCAGCCATATGGATCACTTAGTTGAGAGAGTTATGGCAAGAAATAAGCC
GD33_BOLC.VG2.A     1 GTAGGCCAACCCAGCCATATGGATCACTTAGTTGAGAGAGTTATGGCAAGAAATAAGCC
A12_BOLC.VG2.B      1 -ACCTCTATCACCAGCA--TA--GTCTCCTTCGTGAGACCGGTCACTCAAACCAATCAT
GD33_BOLC.VG2.B     1 -ACCTCTATCACCAGCA--TA--GTCTCCTTCGTGAGACCGGTCACTCAAACCAATCAT
```

```
A12_BOLC.VG2.A      61 ATGCCGATTAAAGAGAGTCCAAACAACTAAGTTCAGTCTAAAGCTAACTTTAT
GD33_BOLC.VG2.A     61 ATGCCGATTAAAGAGAGTCCAAACAACTAAGTTCAGTCTAAAGCTAACTTTAT
A12_BOLC.VG2.B      58 GTTCCCATATCTCTCCATTCTTTAAGGTATTTAACGTTTCTGGTTTATTGA-
GD33_BOLC.VG2.B     58 GTTCCCATATCTCTCCATTCTTTAAGGTATTTAACGTTTCTGGTTTATTGA-
```

```
A12_BOLC.VG2.A      120 ATCAATTAAACCAATATTATGGATGATACCTTCACAAAACATTTCGTGAATTTTCAG
GD33_BOLC.VG2.A     120 ATCAATTAAACCAATATTATGGATGATACCTTCACAAAACATTTCGTGAATTTTCAG
A12_BOLC.VG2.B      117 CCATCTTAAGCACTCGATGGATCTATCAGCTGACAATCCGGGACACGATTAG
GD33_BOLC.VG2.B     117 CCATCTTAAGCACTCGATGGATCTATCAGCTGACAATCCGGGACACGATTAG
```

```
A12_BOLC.VG2.A      179 ACCCTCTGGACCAATCAAGACTAAGATGTAAACCTTAACGACAGTTGTTTCATGT
GD33_BOLC.VG2.A     179 ACCCTCTGGACCAATCAAGACTAAGATGTAAACCTTAACGACAGTTGTTTCATGT
A12_BOLC.VG2.B      176 ATCATGCGG-----TACAGATTCTAGA-GGAACCTTAATGCCCAATCCGCCAGA
GD33_BOLC.VG2.B     176 ATCATGCGG-----TACAGATTCTAGA-GGAACCTTAATGCCCAATCCGCCAGA
```

```
A12_BOLC.VG2.A      239 GGCCAATTAACAGCTCAATTCGTTACTTTCAAATTCATTTTCTGGGTTGTGTTTTG
GD33_BOLC.VG2.A     239 GGCCAATTAACAGCTCAATTCGTTACTTTCAAATTCATTTTCTGGGTTGTGTTTTG
A12_BOLC.VG2.B      229 GTCTCATTTACAGTGTCA-----TTCTATGCAAATCTCTGGGATTTCGTTGTTTTTT
GD33_BOLC.VG2.B     229 GTCTCATTTACAGTGTCA-----TTCTATGCAAATCTCTGGGATTTCGTTGTTTTTT
```

40

ES 2 989 717 T3

A12_BOLC.VG2.A	299	CTTTTTCAGGGTTGATTTGGAATTTCCTCAAAGAGGTCACATAAATTAAAAACAATTTCT
GD33_BOLC.VG2.A	299	CTTTTTCAGGGTTGATTTGGAATTTCCTCAAAGAGGTCACATAAATTAAAAACAATTTCT
A12_BOLC.VG2.B	283	CTTTTTCAGGGTTGATTTGGAATTTCCTCAAAGAGGTCACATAAATTAAAAACAATTTCT
GD33_BOLC.VG2.B	283	CTTTTTCAGGGTTGATTTGGAATTTCCTCAAAGAGGTCACATAAATTAAAAACAATTTCT
A12_BOLC.VG2.A	359	CTTAATTCTTTTGACAAG--GCAGCAAAATCCATTAAAGAAGATCACACACAAAGAAAA
GD33_BOLC.VG2.A	359	CTTAATTCTTTTGACAAG--GCAGCAAAATCCATTAAAGAAGATCACACACAAAGAAAA
A12_BOLC.VG2.B	343	CCCTTATCCGCGCGCCAC--GCAGTATTAATTTTAGTTTTTAAATTATTTATTTTAT
GD33_BOLC.VG2.B	343	CCCTTATCCGCGCGCCAC--GCAGTATTAATTTTAGTTTTTAAATTATTTATTTTAT
A12_BOLC.VG2.A	417	TAAAAACCAACAAAGAATTTCCCATTTACAAAGTGATGAACAGATCACTACGTA
GD33_BOLC.VG2.A	417	TAAAAACCAACAAAGAATTTCCCATTTACAAAGTGATGAACAGATCACTACGTA
A12_BOLC.VG2.B	401	CAATATGTATTTTATATTTCGTTTCATTTATTTTCGTTAAGTA--AATATTTTTTGT
GD33_BOLC.VG2.B	401	CAATATGTATTTTATATTTCGTTTCATTTATTTTCGTTAAGTA--AATATTTTTTGT
A12_BOLC.VG2.A	477	TCTCCACCAATGCATGACACTACTGAAACATTA--CAAACAA--TCCTATAATT
GD33_BOLC.VG2.A	477	TCTCCACCAATGCATGACACTACTGAAACATTA--CAAACAA--TCCTATAATT
A12_BOLC.VG2.B	460	TCTTAAACTATCTATTTTTTTTAC--GAAT--TGT--TATCATATAAAATATTA--AAAT
GD33_BOLC.VG2.B	460	TCTTAAACTATCTATTTTTTTTAC--GAAT--TGT--TATCATATAAAATATTA--AAAT
A12_BOLC.VG2.A	535	GAGCACCTCACATCAAAAAACTGCCATTAATTTCAA--ATTTGACAGGATTTGTAAACA
GD33_BOLC.VG2.A	535	GAGCACCTCACATCAAAAAACTGCCATTAATTTCAA--ATTTGACAGGATTTGTAAACA
A12_BOLC.VG2.B	517	GAGTATA--ATTAATTAATAATTTTAAACA--AAATTTTTCTTTTGTGCGATATCA
GD33_BOLC.VG2.B	517	GAGTATA--ATTAATTAATAATTTTAAACA--AAATTTTTCTTTTGTGCGATATCA
A12_BOLC.VG2.A	595	CTTAATCACTTATTTCTTTAAGGATAATGTTGCACAGAATATTATCATTTTTCTATG
GD33_BOLC.VG2.A	595	CTTAATCACTTATTTCTTTAAGGATAATGTTGCACAGAATATTATCATTTTTCTATG
A12_BOLC.VG2.B	577	TATAATTAATGATCCGCCA-----TTACA--AA--ATCATGATTATTTTATG
GD33_BOLC.VG2.B	577	TATAATTAATGATCCGCCA-----TTACA--AA--ATCATGATTATTTTATG
A12_BOLC.VG2.A	655	TAAAGCATAGCAACATTTGTAAGAAAG--ATCATATCATCTCTATCTTATAGGAGTTTA
GD33_BOLC.VG2.A	655	TAAAGCATAGCAACATTTGTAAGAAAG--ATCATATCATCTCTATCTTATAGGAGTTTA
A12_BOLC.VG2.B	626	TAATTTTCTTTAT---TTTACTATTTTCCTTAAACTATATAAATTTTACAT--ATTTTA
GD33_BOLC.VG2.B	626	TAATTTTCTTTAT---TTTACTATTTTCCTTAAACTATATAAATTTTACAT--ATTTTA
A12_BOLC.VG2.A	715	ATTTATTTCCAGTTTCACATTTAGTCATCCTCTATCTTCAA--CAATAAAACATAACCAA
GD33_BOLC.VG2.A	715	ATTTATTTCCAGTTTCACATTTAGTCATCCTCTATCTTCAA--CAATAAAACATAACCAA
A12_BOLC.VG2.B	682	ATTACAATTTTTTATAATCT--TTACCTTTTATTTTAAAT--CAACTCAATATTTTTTT
GD33_BOLC.VG2.B	682	ATTACAATTTTTTATAATCT--TTACCTTTTATTTTAAAT--CAACTCAATATTTTTTT
A12_BOLC.VG2.A	773	CTTATAATCAAA--TAACCAATTTGAATGCCAGTTTTCAAAA--AAAT--ATACATATCTAA
GD33_BOLC.VG2.A	773	CTTATAATCAAA--TAACCAATTTGAATGCCAGTTTTCAAAA--AAAT--ATACATATCTAA
A12_BOLC.VG2.B	740	TTAACAATTTATA--CAAAATATTTTAA--AAATTTTTTAAATTTT--TTTAAATATATGAA
GD33_BOLC.VG2.B	740	TTAACAATTTATA--CAAAATATTTTAA--AAATTTTTTAAATTTT--TTTAAATATATGAA
A12_BOLC.VG2.A	831	CTGATAGTCAGTCAAATCAACCAATAAGTAGAAA--AACCAACAA--CACA--CTACCT
GD33_BOLC.VG2.A	831	CTGATAGTCAGTCAAATCAACCAATAAGTAGAAA--AACCAACAA--CACA--CTACCT
A12_BOLC.VG2.B	800	TTTACATTT---TTAAATTTAA---AATTATCCTTAAATAT--ATAAGTTTGTATTAAC
GD33_BOLC.VG2.B	799	TTTACATTT---TTAAATTTAA---AATTATCCTTAAATAT--ATAAGTTTGTATTAAC
A12_BOLC.VG2.A	891	GACACTTCTATTATTTT--TAGTTTCGTTTGGATTCTTCTCTTTCTTATTTTATTC
GD33_BOLC.VG2.A	890	GACACTTCTATTATTTT--TAGTTTCGTTTGGATTCTTCTCTTTCTTATTTTATTC
A12_BOLC.VG2.B	852	GAAACTCAAAATATCAAAAT---AT---TATTTATCAAT--ATAATA--CAATTTTAAAT
GD33_BOLC.VG2.B	852	GAAACTCAAAATATCAAAAT---AT---TATTTATCAAT--ATAATA--CAATTTTAAAT

GD33_BOLC.VG2.B 851 GAAACTCAAATTATTCAAAAT---ATTATATTCAATATAATCAATTTTAATT

A12_BOLC.VG2.A 951 TAGAATTTTTTAAACTGAAGTAAACCATAAATTTCAGTCTTTTGATTCCGATCATTT

GD33_BOLC.VG2.A 950 TAGAATTTTTTAAACTGAAGTAAACCATAAATTTCAGTCTTTTGATTCCGATCATTT

A12_BOLC.VG2.B 908 GATTATTTTTTAAA---AATA---CATTTCAAAAATATTTTTGAAATAATTTCATT

GD33_BOLC.VG2.B 907 GATTATTTTTTAAA---AATA---CATTTCAAAAATATTTTTGAAATAATTTCATT

A12_BOLC.VG2.A 1011 TGTGTTTCATTCTTTTAATCCATTCCTTAATTTAATTCCTTTTAAATAATCATTT

GD33_BOLC.VG2.A 1010 TGTGTTTCATTCTTTTAATCCATTCCTTAATTTAATTCCTTTTAAATAATCATTT

A12_BOLC.VG2.B 962 TATCTTTCAAT---ATAAATGAAAATTTATAATTAAATAATAAAAAATATAATAAA

GD33_BOLC.VG2.B 961 TATCTTTCAAT---ATAAATGAAAATTTATAATTAAATAATAAAAAATATAATAAA

A12_BOLC.VG2.A 1071 TTTAAACACGGTCATATATCATCTGTTTGTGACTATGAAAAATAAATTATATAATT

GD33_BOLC.VG2.A 1070 TTTAAACATGTCATATATCATCTGTTTGTGACTATGAAAAATAAATTATATAATT

A12_BOLC.VG2.B 1020 ACCATAAATTTAGCAAATACAACTGAGTTA---TATTATGCTAAAAAAAATTTCTAAC

GD33_BOLC.VG2.B 1019 ACCATAAATTTAGCAAATACAACTGAGTTA---TATTATGCTAAAAAAAATTTCTAAC

A12_BOLC.VG2.A 1131 AAATCTAATAT-ATAACTGAAACAAATATCCTAAAAAACACAAAAAATTTATAAGAAC

GD33_BOLC.VG2.A 1130 AAATCTAATAT-ATAACTGAAACAAATATCCTAAAAAACACAAAAAATTTATAAGAAC

A12_BOLC.VG2.B 1078 AAATATCAATGACAAATGAGTTTATCACT---AAATATACCATATAATTTTAAAAC

GD33_BOLC.VG2.B 1077 AAATATCAATGACAAATGAGTTTATCACT---AAATATACCATATAATTTTAAAAC

A12_BOLC.VG2.A 1190 ACAATCAACCCAAAT-TTACTAAATAAATAGAAATTAATAAAATAA-CAATTCAT

GD33_BOLC.VG2.A 1189 ACAATCAACCCAAAT-TTACTAAATAAATAGAAATTAATAAAATAA-CAATTCAT

A12_BOLC.VG2.B 1134 ATATCCATTCTTAATTTTTTTAAATAATAATTATTTAAATTTAATCAACTAAAA

GD33_BOLC.VG2.B 1133 ATATCCATTCTTAATTTTTTTAAATAATAATTATTTAAATTTAATCAACTAAAA

A12_BOLC.VG2.A 1248 AAAATTATTAAGTACTTTTGTGTTAGATTAGTTTAAAATGTTTATTCAAATAATA

GD33_BOLC.VG2.A 1248 AAAATTATTAAGTACTTTTGTGTTAGATTAGTTTAAAATGTTTATTCAAATAATA

A12_BOLC.VG2.B 1194 AATATCTTACATTTGTGTGAGTCATTTCTAGTTTAA-----TTATGCAATTTATAT

GD33_BOLC.VG2.B 1193 AATATCTTACATTTGTGTGAGTCATTTCTAGTTTAA-----TTATGCAATTTATAT

A12_BOLC.VG2.A 1308 CAAACACCTTGGTTTTTCGGTGCAATTTGATTCGGTTCCGATTCCTTCGGTCTGATT

GD33_BOLC.VG2.A 1308 CAAACACCTTGGTTTTTCGGTGCAATTTGATTCGGTTCCGATTCCTTCGGTCTGATT

A12_BOLC.VG2.B 1248 TTAGTGCTCATGTTTGTAGCTCATTTTATTTGCACGTTTTCATTTCTCCATTA

GD33_BOLC.VG2.B 1247 TTAGTGCTCATGTTTGTAGCTCATTTTATTTGCACGTTTTCATTTCTCCATTA

A12_BOLC.VG2.A 1368 TTTTCTTCTCTGAACACAACGACACCGACATACTTTTTTAAACATTACGTGAATA

GD33_BOLC.VG2.A 1367 TTTTCTTCTCTGAACACAACGACACCGACATACTTTTTTAAACATTACGTGAATA

A12_BOLC.VG2.B 1308 TTAATAATTT---TGACCTAATGATTAATCATTAATTTT---TTTATTAATGACATT

GD33_BOLC.VG2.B 1307 TTAATAATTT---TGACCTAATGATTAATCATTAATTTT---TTTATTAATGACATT

A12_BOLC.VG2.A 1428 TTTTAACTATTGGCCTTAGTAAGGCTAATGACAGATTAAATAGCTTATAACAC

GD33_BOLC.VG2.A 1427 TTTTAACTATTGGCCTTAGTAAGGCTAATGACAGATTAAATAGCTTATAACAC

A12_BOLC.VG2.B 1365 ATGTCCAAATTT---TTTAAGGATATTTCAATTAAGTAATTAATAAGACAACATTA

GD33_BOLC.VG2.B 1364 ATGTCCAAATTT---TTTAAGGATATTTCAATTAAGTAATTAATAAGACAACATTA

A12_BOLC.VG2.A 1488 AATTGCATGCTATCTATCG---AAGCATAATCTCAATAATTCATTTAAGACTCA

GD33_BOLC.VG2.A 1487 AATTGCATGCTATCTATCG---AAGCATAATCTCAATAATTCATTTAAGACTCA

A12_BOLC.VG2.B 1422 AATAGGAAGTTTAAATCATTTAAGCATATTCAT---TAATGTAACCTAAAGACA-A

GD33_BOLC.VG2.B 1421 AATAGGAAGTTTAAATCATTTAAGCATATTCAT---TAATGTAACCTAAAGACA-A

A12_BOLC.VG2.A 1547 TTTGCAATTTTGCAAAGATTTTAGCTGTACAGATAGGCTCATATAAATCCGTAAC

GD33_BOLC.VG2.A 1546 TTTGCAATTTTGCAAAGATTTTAGCTGTACAGATAGGCTCATATAAATCCGTAAC

A12_BOLC.VG2.B 1477 GTATAAATTTGGCA---TTTATTCGTTT---TATGATATTTCATAAA---TGCAAC

GD33_BOLC.VG2.B 1476 GTATAAATTTGGCA---TTTATTCGTTT---TATGATATTTCATAAA---TGCAAC

A12_BOLC.VG2.A 1607 CCGAAATCCGAACCAACCAAAAAACCGTCCCTACCCATCCAAATGTAAAAATAC

GD33_BOLC.VG2.A 1606 CCGAAATCCGAACCAACCAAAAAACCGTCCCTACCCATCCAAATGTAAAAATAC

ES 2 989 717 T3

A12_BOLC.VG2.B	1527	TGAAAAGACAAACAA--AAAAAATA--GACTTTAATTAAT--AAAT--AAACATT
GD33_BOLC.VG2.B	1526	TGAAAAGACAAACAA--AAAAAATA--GACTTTAATTAAT--AAAT--AAACATT
A12_BOLC.VG2.A	1667	CTCAATGAATATTGTAG--TTTATAAAATATATCTGAACCCGAA--TGTTATTAAACCGAA
GD33_BOLC.VG2.A	1666	CTCAATGAATATTGTAG--TTTATAAAATATATCTGAACCCGAA--TGTTATTAAACCGAA
A12_BOLC.VG2.B	1582	TCAAATGATACTTC-----TCTTTTAATA-ATAT-----ACATTTCATTAAT-GTA
GD33_BOLC.VG2.B	1581	TCAAATGATACTTC-----TCTTTTAATA-ATAT-----ACATTTCATTAAT-GTA
A12_BOLC.VG2.A	1727	CCCGAACGGATAATCC--AAAACCCG--AAAACCAAAAATTC--AA--AAATATCCGAAAAAA
GD33_BOLC.VG2.A	1726	CCCGAACGGATAATCC--AAAACCCG--AAAACCAAAAATTC--AA--AAATATCCGAAAAAA
A12_BOLC.VG2.B	1627	ACTGAA--GACAA--TA--TAAATTAG--CA--TTTATTTC--TTTAA--TATATTTC---ATAAA
GD33_BOLC.VG2.B	1626	ACTGAA--GACAA--TA--TAAATTAG--CA--TTTATTTC--TTTAA--TATATTTC---ATAAA
A12_BOLC.VG2.A	1787	CT--ATCCGAAT--CGTAAATTAATATAAAATATA--TATTT--AA--CAT--CA--TTG--TAT
GD33_BOLC.VG2.A	1786	CT--ATCCGAAT--CGTAAATTAATATAAAATATA--TATTT--AA--CAT--CA--TTG--TAT
A12_BOLC.VG2.B	1684	TGCAACTGAA--AAG--CAAGAAAA--AAAATA--AA--AATTTAATT--ATG--A--GT--AG--ATA
GD33_BOLC.VG2.B	1683	TGCAACTGAA--AAG--CAAGAAAA--AAAATA--AA--AATTTAATT--ATG--A--GT--AG--ATA
A12_BOLC.VG2.A	1846	ATATCTAACAA--TAACTATTT--AATTTAATA--AT--CTTTAAATACTCCATT--TATACA
GD33_BOLC.VG2.A	1845	ATATCTAACAA--TAACTATTT--AATTTAATA--AT--CTTTAAATACTCCATT--TATACA
A12_BOLC.VG2.B	1741	TTTTC--AA--AGTATTTCTCTTTTAAATATA--CAT--GAT--CAAAAATTATATTGGATA--A
GD33_BOLC.VG2.B	1740	TTTTC--AA--AGTATTTCTCTTTTAAATATA--CAT--GAT--CAAAAATTATATTGGATA--A
A12_BOLC.VG2.A	1905	AA--GAAGTATATATTTT--TGTTTACTTTTGAATTTTATAGATTTTACTAT--GATATAT
GD33_BOLC.VG2.A	1904	AA--GAAGTATATATTTT--TGTTTACTTTTGAATTTTATAGATTTTACTAT--GATATAT
A12_BOLC.VG2.B	1800	CC--CAAGCA--AAATCCATTGAGAAATCA--CAACAACAAAAGA--AGGAACA--CACACAA
GD33_BOLC.VG2.B	1799	CC--CAAGCA--AAATCCATTGAGAAATCA--CAACAACAAAAGA--AGGAACA--CACACAA
A12_BOLC.VG2.A	1963	CC--AGCCGATCCGATA--CAATCCGAATCCGAAT--TT--T--TGGCTACTTT--GATATATCT
GD33_BOLC.VG2.A	1962	CC--AGCCGATCCGATA--CAATCCGAATCCGAAT--TT--T--TGGCTACTTT--GATATATCT
A12_BOLC.VG2.B	1857	CC--CG--CGA--CAGATAACTAA--TGTA--CTGGACT--TT--CCCTTAATATAG--CTAAT--
GD33_BOLC.VG2.B	1856	CC--CG--CGA--CAGATAACTAA--TGTA--CTGGACT--TT--CCCTTAATATAG--CTAAT--
A12_BOLC.VG2.A	2021	GAACCGATCCGAAACCGA--GTGTTTATATCCGAATCTCATCCGTACTT--TAAATTTACTA
GD33_BOLC.VG2.A	2020	GAACCGATCCGAAACCGA--GTGTTTATATCCGAATCTCATCCGTACTT--TAAATTTACTA
A12_BOLC.VG2.B	1914	GAACAGATT---AAATG--GCTA--TTATAGACGTC--TTGCAT--GTATCCATCTATCT--GAA
GD33_BOLC.VG2.B	1913	GAACAGATT---AAATG--GCTA--TTATAGACGTC--TTGCAT--GTATCCATCTATCT--GAA
A12_BOLC.VG2.A	2081	GA--AT--AA--AC--T--GGGGAT--ATAC--AAATAG--ACCGAAATCCGAAAA--ACCGATCCGAA
GD33_BOLC.VG2.A	2080	GA--AT--AA--AC--C--GGGGAT--ATAC--AAATAG--ACCGAAATCCGAAAA--ACCGATCCGAA
A12_BOLC.VG2.B	1971	GC--ATA--TAATTTG-----TTT--TAG--G--ACTG--TTTGCAATTTTACAGAT--ATTTTAGC
GD33_BOLC.VG2.B	1970	GC--ATA--TAATTTG-----TTT--TAG--G--ACTG--TTTGCAATTTTACAGAT--ATTTTAGC
A12_BOLC.VG2.A	2140	CCGAACGCCGAGG--GCTAGTACAGAGTTCCGATCAGAA--AT--GAGAGAATCCTTTGAGCGTC
GD33_BOLC.VG2.A	2139	C--GAACGCCGAGGCTAGTACAGAGTTCCGATCAGAA--AT--GAGAGAATCCTTTGAGCGTC
A12_BOLC.VG2.B	2024	T-----GTACAGTACAGAG-----ACAGAAATCCTTTGAGCGTC
GD33_BOLC.VG2.B	2023	T-----GTACAGTACAGAG-----ACAGAAATCCTTTGAGCGTC
A12_BOLC.VG2.A	2200	G---GCCTTGCTTTGTGCGTAGGGATCTCT--CCGACTAATTGTTGTAATTATATAA--GGC
GD33_BOLC.VG2.A	2198	G---GCCTTGCTTTGTGCGTAGGGATCTCT--CCGACTAATTGTTGTAATTATATAA--GGC
A12_BOLC.VG2.B	2057	GTCGGCCTTGCTTTGTGCTAGGGATCTCCG--CTGACTAATTGTTGTAACCTACAAAAGGG
GD33_BOLC.VG2.B	2056	GTCGGCCTTGCTTTGTGCTAGGGATCTCCG--CTGACTAATTGTTGTAACCTACAAAAGGG
A12_BOLC.VG2.A	2256	GCC-----GCGCCTTTT--TAGCGGAGAGATGGCGAGCTCA-----CAGTTTCAGGTAT
GD33_BOLC.VG2.A	2254	GCC-----GCGCCTTTT--TAGCGGAGAGATGGCGAGCTCA-----CAGTTTCAGGTAT
A12_BOLC.VG2.B	2117	GCCCCCCCCCGCCTTTTACT---ACATGGCGAGCTCAT-----CACAGTTTAGGTAT
GD33_BOLC.VG2.B	2116	GCCCCCCCCCGCCTTTTACT---AGATGGCGAGCTCAT-----CACAGTTTAGGTAT
A12_BOLC.VG2.A	2303	AC--CAGACGCCGTCGAAGGTGGTGCACCTGAGGAATCTGCCGTGGGAATGCGTGGAAGAG
GD33_BOLC.VG2.A	2301	AC--CAGACGCCGTCGAAGGTGGTGCACCTGAGGAATCTGCCGTGGGAATGCGTGGAAGAG
A12_BOLC.VG2.B	2167	ACC--CAGACGCCGTCGAAGGTGGTGCACCTGAGGAATCTGCCGTGGGAATGCGTGGAAGAG

ES 2 989 717 T3

GD33_BOLC.VG2.B	2166	ACCCAGACGCGCTCGAAGTGGTGCACCTGAGGAATCTGCCGTGGGAATGCGTGGAAGAG
A12_BOLC.VG2.A	2363	GAGCTCATCGACCTATGCAAACGATTCCGGCAAGATCGTCAATACGAAGACCAATGTCCGG
GD33_BOLC.VG2.A	2361	GAGCTCATCGACCTATGCAAACGATTCCGGCAAGATCGTCAATACGAAGACCAATGTCCGG
A12_BOLC.VG2.B	2227	GAGCTCATCGACCTATGCAAACGATTCCGGCAAGATCGTCAATACGAAGACCAATGTCCGG
GD33_BOLC.VG2.B	2226	GAGCTCATCGACCTATGCAAACGATTCCGGCAAGATCGTCAATACGAAGACCAATGTCCGG
A12_BOLC.VG2.A	2423	GCCAAATCGCAACCAAGCCTTTGTGCGAATTCGTAA---CAACT-----TTTATTTCTC
GD33_BOLC.VG2.A	2421	GCCAAATCGCAACCAAGCCTTTGTGCGAATTCGTAA---CAACT-----TTTATTTCTC
A12_BOLC.VG2.B	2287	GCCAAATCGCAACCAAGCCTTTGTGCGAATTCGTAA---CAACT-----TTTATTTCTC
GD33_BOLC.VG2.B	2286	GCCAAATCGCAACCAAGCCTTTGTGCGAATTCGTAA---CAACT-----TTTATTTCTC
A12_BOLC.VG2.A	2472	TTGGATCATCAGATTGTTTCT-----
GD33_BOLC.VG2.A	2470	TTGGATCATCAGATTGTTTCTGAAG-----CTTT---TG---TTTCT---TGAT---AACT---CTT
A12_BOLC.VG2.B	2336	TTGGATCATCAGATTGTTGCTTCCC-----CCCC---TG---TTTCT---TTAGAAATTGAA
GD33_BOLC.VG2.B	2335	TTGGATCATCAGATTGTTGCTTCCC-----CCCC---TG---TTTCT---TTAGAAATTGAA
A12_BOLC.VG2.A	2493	-----TTTTTTGGTATCTCGTTGTTGT-----TTTGGTAT-----T
GD33_BOLC.VG2.A	2521	-----TTTTTTTGGTATCTCGTTGTTGT-----TTTGGTAT-----T
A12_BOLC.VG2.B	2388	-----TAAATAAAAGGACTCGA---CTCTTTT-----TTTGGTATCTCTTT
GD33_BOLC.VG2.B	2387	-----TAAATAAAAGGACTCGA---CTCTTTT-----TTTGGTATCTCTTT
A12_BOLC.VG2.A	2524	GTT-----GTTTTGAAGATGAAACTGTATACTTT-----GATTCA
GD33_BOLC.VG2.A	2554	GTT-----GTTTTGAAGATGAAACTGTATACTTT-----GATTCA
A12_BOLC.VG2.B	2426	GTT-----GTTTTGAAGATGAAACTGTATACTTT-----GATTCA
GD33_BOLC.VG2.B	2425	GTT-----GTTTTGAAGATGAAACTGTATACTTT-----GATTCA
A12_BOLC.VG2.A	2559	TATT-----CCAGGGTGACGTGAATCAGGCAATATCAATGGTT
GD33_BOLC.VG2.A	2589	TATT-----CCAGGGTGACGTGAATCAGGCAATATCAATGGTT
A12_BOLC.VG2.B	2461	TATT-----CCAGGGTGACGTGAATCAGGCAATATCAATGGTT
GD33_BOLC.VG2.B	2460	TATT-----CCAGGGTGACGTGAATCAGGCAATATCAATGGTT

ES 2 989 717 T3

A12_BOLC.VG2.A	2598	TCTTACTATGCTTCGTCTTCAGAGCCGGCTCAGATTCGAGGGAAGACTGTTTATATTTCAG
GD33_BOLC.VG2.A	2628	TCTTAATATGCTTCGTCTTCAGAGCCGGCTCAGATTCGAGGGAAGACTGTTTATATTTCAG
A12_BOLC.VG2.B	2500	TCTTACTATGCTTCGTCTTCAGAGCCGGCTCAGATTCGAGGGAAGACTGTTTATATTTCAG
GD33_BOLC.VG2.B	2499	TCTTACTATGCTTCGTCTTCAGAGCCGGCTCAGATTCGAGGGAAGACTGTTTATATTTCAG
A12_BOLC.VG2.A	2658	TACTCTAATCGCCATGAGATTGTCAACAATCAGAGTCCTGGAGAGGTCCTGGCAATGTC
GD33_BOLC.VG2.A	2688	TACTCTAATCGCCATGAGATTGTCAACAATCAGAGTCCTGGAGAGGTCCTGGCAATGTC
A12_BOLC.VG2.B	2560	TACTCTAATCGCCATGAGATTGTCAACAATCAGAGTCCTGGAGAGGTCCTGGCAATGTC
GD33_BOLC.VG2.B	2559	TACTCTAATCGCCATGAGATTGTCAACAATCAGAGTCCTGGAGAGGTCCTGGCAATGTC
A12_BOLC.VG2.A	2718	CTCTTGGTCACCTTTGAAGGAGTCCAAATCTCACCATGTCTGCATCGATGTCATCCATCTG
GD33_BOLC.VG2.A	2748	CTCTTGGTCACCTTTGAAGGAGTCCAAATCTCACCATGTCTGCATCGATGTCATCCATCTG
A12_BOLC.VG2.B	2620	CTATTGGTCACCTTTGAAGGAGTCCAAATCTCACCATGTCTGCATCGATGTCATCCATCTG
GD33_BOLC.VG2.B	2619	CTATTGGTCACCTTTGAAGGAGTCCAAATCTCACCATGTCTGCATCGATGTCATCCATCTG
A12_BOLC.VG2.A	2778	GTATGTGAATATTCAGCTTACCTTCCACTATTGTTTCTT-GTTATGTTAAGTGATTTT
GD33_BOLC.VG2.A	2808	GTATGTGAATATTCAGCTTACCTTCCACTATTGTTTCTT-GTTATGTTAAGTGATTTT
A12_BOLC.VG2.B	2680	GTATGTGAATATTCAGCTTACCTTCCACTATTGTTTCTT-GTTATGTTAAGTGATTTT
GD33_BOLC.VG2.B	2679	GTATGTGAATATTCAGCTTACCTTCCACTATTGTTTCTT-GTTATGTTAAGTGATTTT
A12_BOLC.VG2.A	2837	TCG-----TTTCTTCG-----AGTAGATTCTAATCTATGAAAATATTTCAACTTG
GD33_BOLC.VG2.A	2867	TCG-----TTTCTTCG-----AGTAGATTCTAATCTATGAAAATATTTCAACTTG
A12_BOLC.VG2.B	2735	TCG-----TTTCTTCG-----AGTAGATTCTAATCTATGAAAATATTTCAACTTG
GD33_BOLC.VG2.B	2734	TCG-----TTTCTTCG-----AGTAGATTCTAATCTATGAAAATATTTCAACTTG
A12_BOLC.VG2.A	2882	TTGTTATTAGGCAAACCTT-----CTTTTG-----AGTGATTTTTTTCC-----
GD33_BOLC.VG2.A	2912	TTGTTATTAGGCAAACCTT-----CTTTTG-----AGTGATTTTTTTCC-----
A12_BOLC.VG2.B	2780	TTGTTATTAGGCAAACCTT-----CTTTTG-----AGTGATTTTTTTCC-----
GD33_BOLC.VG2.B	2779	TTGTTATTAGGCAAACCTT-----CTTTTG-----AGTGATTTTTTTCC-----
A12_BOLC.VG2.A	2921	-----ACTTATTGTTAGACATAC-----AGTATGTC-----ACATACTATTG
GD33_BOLC.VG2.A	2951	-----AGTTATTGTTAGACATAC-----AGTATGTC-----ACATACTATTG
A12_BOLC.VG2.B	2819	-----AGTTATTGTTAGACATAC-----AGTATGTC-----ACATACTATTG
GD33_BOLC.VG2.B	2818	-----AGTTATTGTTAGACATAC-----AGTATGTC-----ACATACTATTG
A12_BOLC.VG2.A	2958	TAAATTACAGTATATCTGACGTTAATGAAAATGCTCGAATCACAGATGTTGATGCCCTCTT
GD33_BOLC.VG2.A	2988	TAAATTACAGTATATCTGACGTTAATGAAAATGCTCGAATCACAGATGTTGATGCCCTCTT
A12_BOLC.VG2.B	2856	TAAATTACAGTATATCTGACGTTAATGAAAATGCTCGAATCACAGATGTTGATGCCCTCTT
GD33_BOLC.VG2.B	2855	TAAATTACAGTATATCTGACGTTAATGAAAATGCTCGAATCACAGATGTTGATGCCCTCTT
A12_BOLC.VG2.A	3018	TATTATAATCTTTCTGGAGAGAGTTTGGAAAATAGTTTCATGTTCCGTCATTCTCAT-----
GD33_BOLC.VG2.A	3048	TATTATAATCTTTCTGGAGAGAGTTTGGAAAATAGTTTCATGTTCCGTCATTCTCAT-----
A12_BOLC.VG2.B	2916	TATTATAATCTTTCTGGAGAGAGTTTGGAAAATAGTTTCATGTTCCGTCATTCTCAT-----
GD33_BOLC.VG2.B	2915	TATTATAATCTTTCTGGAGAGAGTTTGGAAAATAGTTTCATGTTCCGTCATTCTCAT-----
A12_BOLC.VG2.A	3074	-----GGACAGGTGTCACCTCTGCACCTTATCTAGTCATTCTCTT
GD33_BOLC.VG2.A	3104	-----GGACAGGTGTCACCTCTGCACCTTATCTAGTCATTCTCTT
A12_BOLC.VG2.B	2972	-----GGACAGGTGTCACCTCTGCACCTTATCTAGTCATTCTCTT
GD33_BOLC.VG2.B	2971	-----GGACAGGTGTCACCTCTGCACCTTATCTAGTCATTCTCTT
A12_BOLC.VG2.A	3112	TTTLAGTCTCCTAA-----TTTGAGTTTATTTGATTGATTGCTTCCTTAGTT
GD33_BOLC.VG2.A	3142	TTTLAGTCTCCTAAACCGGCAACA-----TTTGAGTTTATTTGATTGATTGCTTCCTTAGTT
A12_BOLC.VG2.B	3010	TTTLAGTCTCCTAA-----TTTGAGTTTATTTGATTGATTGCTTCCTTAGTT
GD33_BOLC.VG2.B	3009	TTTLAGTCTCCTAA-----TTTGAGTTTATTTGATTGATTGCTTCCTTAGTT
A12_BOLC.VG2.A	3162	ATTATCAATTTACTCCACTGTATTATATGGACATGACTCATATCTAGTCCAAACTTTTGT
GD33_BOLC.VG2.A	3202	ATTATCAATTTACTCCACTGTATTATATGGACATGACTCATATCTAGTCCAAACTTTTGT
A12_BOLC.VG2.B	3060	ATTATCAATTTACTCCACTGTATTATATGGACATGACTCATATCTAGTCCAAACTTTTGT
GD33_BOLC.VG2.B	3059	ATTATCAATTTACTCCACTGTATTATATGGACATGACTCATATCTAGTCCAAACTTTTGT

A12_BOLC.VG2.A	3222	TTTACGCTCTTTTCTGCTTATGGCTTCGTGCACAAAATTGCCACTTTTGAGAAAAGCTGCTG
GD33_BOLC.VG2.A	3262	TTCAGGTGTTTTCTGCTTATGGCTTCGTGCACAAAATTGCCACTTTTGAGAAAAGCTGCTG
A12_BOLC.VG2.B	3120	TTCAGGTGTTTTCTGCTTATGGCTTCGTGCACAAAATTGCCACTTTTGAGAAAAGCTGCTG
GD33_BOLC.VG2.B	3119	TTCAGGTGTTTTCTGCTTATGGCTTCGTGCACAAAATTGCCACTTTTGAGAAAAGCTGCTG
A12_BOLC.VG2.A	3282	CTTTCCAGGTTAGAGATATCGAGTTTGTITTTCAAGGCTTGCCT-----A--AATC
GD33_BOLC.VG2.A	3322	GTTTCCAGGTTAGAGATATCGAGTTTGTITTTCAAGGCTTGCCT-----A--AATC
A12_BOLC.VG2.B	3180	GTTTCCAGGTTAGAGATATCGAGTTTGTITTTCAAGGCTTGCCTTTTTCCTTA--AATC
GD33_BOLC.VG2.B	3179	GTTTCCAGGTTAGAGATATCGAGTTTGTITTTCAAGGCTTGCCTTTTTCCTTA--AATC
A12_BOLC.VG2.A	3332	TAATGATGTTTAAACTCATCTCTGCGGTCATAACCTGCTCAGGCACCTTGTTTCAGTTTAC
GD33_BOLC.VG2.A	3372	TAATGATGTTTAAACTCATCTCTGCGGTCATAACCTGCTCAGGCACCTTGTTTCAGTTTAC
A12_BOLC.VG2.B	3238	TAATGATGTTTAAACTCATCTCTGCGGTCATAACCTGCTCAGGCACCTTGTTTCAGTTTAC
GD33_BOLC.VG2.B	3237	TAATGATGTTTAAACTCATCTCTGCGGTCATAACCTGCTCAGGCACCTTGTTTCAGTTTAC
A12_BOLC.VG2.A	3392	TGATGTGGACACTGCCTTAGCGGCAAGGACTGCGCTGGATGGTAGAAGTATACCCAGTA
GD33_BOLC.VG2.A	3432	TGATGTGGACACTGCCTTAGCGGCAAGGACTGCGCTGGATGGTAGAAGTATACCCAGTA
A12_BOLC.VG2.B	3298	TGATGTGGACACTGCCTTAGCGGCAAGGACTGCGCTGGATGGTAGAAGTATACCCAGTA
GD33_BOLC.VG2.B	3297	TGATGTGGACACTGCCTTAGCGGCAAGGACTGCGCTGGATGGTAGAAGTATACCCAGTA
A12_BOLC.VG2.A	3452	TGCTCAAATCCTTCATTTCATGCTTTTGACCATAAGATAAAGCTCTGTTGATGGTTTCTTC
GD33_BOLC.VG2.A	3492	TGCTCAAATCCTTCATTTCATGCTTTTGACCATAAGATAAAGCTCTGTTGATGGTTTCTTC
A12_BOLC.VG2.B	3358	TGCTCAAATCCTTCATTTCATGCTTTTGACCATAAGATAAAGCTCTGTTGATGGTTTCTTC
GD33_BOLC.VG2.B	3357	TGCTCAAATCCTTCATTTCATGCTTTTGACCATAAGATAAAGCTCTGTTGATGGTTTCTTC
A12_BOLC.VG2.A	3512	CTTTTCTTTTGGTAAATTCAGATATCTGCTTCCAGAACATGTAGGCTCATGCAATTTGCG
GD33_BOLC.VG2.A	3552	CTTTTCTTTTGGTAAATTCAGATATCTGCTTCCAGAACATGTAGGCTCATGCAATTTGCG
A12_BOLC.VG2.B	3418	CTTTTCTTTTGGTAAATTCAGATATCTGCTTCCAGAACATGTAGGCTCATGCAATTTGCG
GD33_BOLC.VG2.B	3417	CTTTTCTTTTGGTAAATTCAGATATCTGCTTCCAGAACATGTAGGCTCATGCAATTTGCG
A12_BOLC.VG2.A	3572	AATGCTTTACTCAGCTCATACTGATCTAAATATCAAATTTTCAGTCCCACCGCAGCAGGTA
GD33_BOLC.VG2.A	3612	AATGCTTTACTCAGCTCATACTGATCTAAATATCAAATTTTCAGTCCCACCGCAGCAGGTA
A12_BOLC.VG2.B	3478	AATGCTTTACTCAGCTCATACTGATCTAAATATCAAATTTTCAGTCCCACCGCAGCAGGTA
GD33_BOLC.VG2.B	3477	AATGCTTTACTCAGCTCATACTGATCTAAATATCAAATTTTCAGTCCCACCGCAGCAGGTA
A12_BOLC.VG2.A	3632	GAGTTTGTAGTCCTCGCAAAATGTGCCTCTCCATTGTTTATACTGCTCTTACTGTTGGAAGC
GD33_BOLC.VG2.A	3672	GAGTTTGTAGTCCTCGCAAAATGTGCCTCTCCATTGTTTATACTGCTCTTACTGTTGGAAGC
A12_BOLC.VG2.B	3538	GAGTTTGTAGTCCTCGCAAAATGTGCCTCTCCATTGTTTATACTGCTCTTACTGTTGGAAGC
GD33_BOLC.VG2.B	3537	GAGTTTGTAGTCCTCGCAAAATGTGCCTCTCCATTGTTTATACTGCTCTTACTGTTGGAAGC
A12_BOLC.VG2.A	3692	CTTATGCTTGAATAGTTCTTCATGTGTTT--AATCTCTTACCGACTCAGGCACTACACA
GD33_BOLC.VG2.A	3732	CTTATGCTTGAATAGTTCTTCATGTGTTT--AATCTCTTACCGACTCAGGCACTACACA
A12_BOLC.VG2.B	3598	CTTATGCTTGAATAGTTGATTCATGTGTTTGAATTTCTGACGTACTCAGGCACTACACA
GD33_BOLC.VG2.B	3597	CTTATGCTTGAATAGTTGATTCATGTGTTTGAATTTCTGACGTACTCAGGCACTACACA
A12_BOLC.VG2.A	3751	AATCCATACCTTCCGGTGAATCAAACCGCTATGGACGGTTCTATGCAGGTATATTTTCC
GD33_BOLC.VG2.A	3791	AATCCATACCTTCCGGTGAATCAAACCGCTATGGACGGTTCTATGCAGGTATATTTTCC
A12_BOLC.VG2.B	3658	AATCCATACCTTCCGGTGAATCAAACCTGCTATGGATGGTTCTATGCAGGTATATTTTCC
GD33_BOLC.VG2.B	3657	AATCCATACCTTCCGGTGAATCAAACCTGCTATGGATGGTTCTATGCAGGTATATTTTCC
A12_BOLC.VG2.A	3811	TTTGATCTATCTTATCTCCAACAGCTTATATTCTTACACTGAATGCCGTTTGTATTA
GD33_BOLC.VG2.A	3851	TTTGATCTATCTTATCTCCAACAGCTTATATTCTTACACTGAATGCCGTTTGTATTA
A12_BOLC.VG2.B	3718	TCTT-TCTATCT-ATCTCCAACAGCTTAGACAATTCTTACATTGAATGCCGTT-----
GD33_BOLC.VG2.B	3717	TCTT-TCTATCT-ATCTCCAACAGCTTAGACAATTCTTACATTGAATGCCGTT-----
A12_BOLC.VG2.A	3871	CGTTACTGATCGTCTTTGTCCTCACTGCTTCAGGCTGCTTTGGGTGCTGATGGAAGA
GD33_BOLC.VG2.A	3911	CGTTACTGATCGTCTTTGTCCTCACTGCTTCAGGCTGCTTTGGGTGCTGATGGAAGA
A12_BOLC.VG2.B	3769	----ACTGATTCGCTTTGAC--TCATTGCAATTCAGCCTGCTTTGGGTGCTGATGGAAGA
GD33_BOLC.VG2.B	3768	----ACTGATTCGCTTTGAC--TCATTGCAATTCAGCCTGCTTTGGGTGCTGATGGAAGA
A12_BOLC.VG2.A	3931	GGGTTGAACCTCAGAGCAAGCTCCTCTTGCTTTGATTGAGAATATGCAGTACGCTGTCA

ES 2 989 717 T3

GD33_BOLC.VG2.A	3971	CGCTTCGAAACTCAGAGCAACGTCCTCTTGCTTTGATTGAGAATATGCAGTAGGCTGTCA
A12_BOLC.VG2.B	3824	GGGTAGAAACTCAGAGCAACGTCCTCTTGCTTTGATTGAGAATATGCAATACGCTGTCA
GD33_BOLC.VG2.B	3823	GGGTAGAAACTCAGAGCAACGTCCTCTTGCTTTGATTGAGAATATGCAATACGCTGTCA
A12_BOLC.VG2.A	3991	CGGTGGATGTTCTTCACACGGTGAGGAACAACATATGCGCTGTTTCTTCATCTCATCAT
GD33_BOLC.VG2.A	4031	CGGTGGATGTTCTTCACACGGTGAGGAACAACATATGCGCTGTTTCTTCATCTCATCAT
A12_BOLC.VG2.B	3884	C-----ACGGTGAGGAACAACATATGCGCTGTTTCTTCATCTCATCAT
GD33_BOLC.VG2.B	3883	C-----ACGGTGAGGAACAACATATGCGCTGTTTCTTCATCTCATCAT
A12_BOLC.VG2.A	4050	GCTTTCATATCTCAAAATTAACATATCCTTTGCTTCTCATAGGTGTTTTCCGCTTAAGG
GD33_BOLC.VG2.A	4090	GCTTTCATATCTCAAAATTAACATATCCTTTGCTTCTCATAGGTGTTTTCCGCTTAAGG
A12_BOLC.VG2.B	3927	CCTTTCATTTCTCAAAATTAATTAATTCCTTTGCTTCTCATAGGTGTTTTCCGCTTAAGG
GD33_BOLC.VG2.B	3926	CCTTTCATTTCTCAAAATTAATTAATTCCTTTGCTTCTCATAGGTGTTTTCCGCTTAAGG
A12_BOLC.VG2.A	4109	AACTGTGCAGAAGATTGCAATATTTGAGAAAAATGGTTCAACGCAAGCCTTAATTCATA
GD33_BOLC.VG2.A	4149	AACTGTGCAGAAGATTGCAATATTTGAGAAAAATGGTTCAACGCAAGCCTTAATTCATA
A12_BOLC.VG2.B	3987	AACTGTGCAGAAGATTGCAATATTTGAGAAAAATGGTTCAACGCAAGCCTTAATTCATA
GD33_BOLC.VG2.B	3986	AACTGTGCAGAAGATTGCAATATTTGAGAAAAATGGTTCAACGCAAGCCTTAATTCATA
A12_BOLC.VG2.A	4169	CTCTGGTACATGACCTTGATGATCTGAATACATATATGATTCTACTATACTTCTTTTGGT
GD33_BOLC.VG2.A	4209	CTCTGGTACATGACCTTGATGATCTGAATACATATATGATTCTACTATACTTCTTTTGGT
A12_BOLC.VG2.B	4047	CTCTGGTACATGACCTTGATGATCTGAATACATATATGATTCTACTATACTTCTTTTGGT
GD33_BOLC.VG2.B	4046	CTCTGGTACATGACCTTGATGATCTGAATACATATATGATTCTACTATACTTCTTTTGGT
A12_BOLC.VG2.A	4229	TATATGGTGAATCTTCTTAATGTTGATGCAAAAGAGAACAGCCTTAAGAATTTGAATATAGTC
GD33_BOLC.VG2.A	4269	TATATGGTGAATCTTCTTAATGTTGATGCAAAAGAGAACAGCCTTAAGAATTTGAATATAGTC
A12_BOLC.VG2.B	4107	TATATGGCGAATCATCTTATCTGTATGC-----TAAGAATTTGAAT--AGTC
GD33_BOLC.VG2.B	4106	TATATGGCGAATCATCTTATCTGTATGC-----TAAGAATTTGAAT--AGTC
A12_BOLC.VG2.A	4289	ATCCTTTGCTTCATGTTCCGTTTTATTGATTCTTCTGTAGACATTTCAACGGCGACATG
GD33_BOLC.VG2.A	4329	ATCCTTTGCTTCATGTTCCGTTTTATTGATTCTTCTGTAGACATTTCAACGGCGACATG
A12_BOLC.VG2.B	4152	ATCCTTTGCTTCATGTTCCGTTTTATTGATTCTTCTGTAGACATTTCAACGGCGACATG
GD33_BOLC.VG2.B	4151	ATCCTTTGCTTCATGTTCCGTTTTATTGATTCTTCTGTAGACATTTCAACGGCGACATG
A12_BOLC.VG2.A	4349	GCGAAAGAAGCACTGGAGGGGACACTGCATATATGACGG--AGGCTACTGTAAGCTTCGACT
GD33_BOLC.VG2.A	4389	GCGAAAGAAGCACTGGAGGGGACACTGCATATATGACGG--AGGCTACTGTAAGCTTCGACT
A12_BOLC.VG2.B	4212	GCGAAAGAAGCACTGGAGGGGACACTGCATATATGACGGGAGGCTACTGTAAGCTTCGACT
GD33_BOLC.VG2.B	4211	GCGAAAGAAGCACTGGAGGGGACACTGCATATATGACGGGAGGCTACTGTAAGCTTCGACT
A12_BOLC.VG2.A	4408	AACCTACTCTCGGCATACTGATCTCAATGTAAAGGTACATAAGATCAGTTGCTTT-----
GD33_BOLC.VG2.A	4448	AACCTACTCTCGGCATACTGATCTCAATGTAAAGGTACATAAGATCAGTTGCTTT-----
A12_BOLC.VG2.B	4272	AACATACTC--GTCATACTGATCTCAATGTAAAGGTACATAAGATCAGTTGCTTTCTTGA
GD33_BOLC.VG2.B	4271	AACATACTC--GTCATACTGATCTCAATGTAAAGGTACATAAGATCAGTTGCTTTCTTGA
A12_BOLC.VG2.A	4463	---ATGTACCCAAACAAAAA--GTCCTTAAAGAGAATAACAAATATTTCTGCAAGAAC
GD33_BOLC.VG2.A	4503	---ATGTACCCAAACAAAAA--GTCCTTAAAGAGAATAACAAATATTTCTGCAAGAAC
A12_BOLC.VG2.B	4330	TTTGTGCACCCAAACAAAAAGCTTGGGAAAGAGAATAATAAATATATGTCTGCAAGAAC
GD33_BOLC.VG2.B	4329	TTTGTGCACCCAAACAAAAAGCTTGGGAAAGAGAATAATAAATATATGTCTGCAAGAAC
A12_BOLC.VG2.A	4519	CGCATGCT----TTTTTCTCATCAACAATTG--AGTGAATCTGAATTCGACTTTCTAGT
GD33_BOLC.VG2.A	4559	CGCATGCT----TTTTTCTCATCAACAATTG--AGTGAATCTGAATTCGACTTTCTAGT
A12_BOLC.VG2.B	4390	CATATGCTATTTT-----TTTTTCTCTTCAACAATTG--TATGAATCTGAATTCGACTTTCTAGT
GD33_BOLC.VG2.B	4389	CATATGCTATTTT-----TTTTTCTCTTCAACAATTG--TATGAATCTGAATTCGACTTTCTAGT
A12_BOLC.VG2.A	4573	TATTCAACTAATGATTTT--CGTCTTCTTTCTTCAGGCATTTAGCGACAAAAGCAGAGAC
GD33_BOLC.VG2.A	4613	TATTCAACTAATGATTTT--CGTCTTCTTTCTTCAGGCATTTAGCGACAAAAGCAGAGAC
A12_BOLC.VG2.B	4448	TTATTCAACTATGATTTTTCGTCTTCTTTCTTCAGGCATTTAGCGACAAAAGCAGAGAC
GD33_BOLC.VG2.B	4447	TTATTCAACTATGATTTTTCGTCTTCTTTCTTCAGGCATTTAGCGACAAAAGCAGAGAC

A12_BOLC.VG2.A	4632	TACACACTGCCTGATCTAAGCCAACTGGTGGGCCAAAAGGTTCCAGGAGTGGCTGCAGCT
GD33_BOLC.VG2.A	4672	TACACACTGCCTGATCTAAGCCAACTGGTGGGCCAAAAGGTTCCAGGAGTGGCTGCAGCT
A12_BOLC.VG2.B	4508	TACACACTGCCTGATCTAAGCCAACTGGTGGGCCAAAAGGTTCCAGGAGTGGCTGCAGCT
GD33_BOLC.VG2.B	4507	TACACACTGCCTGATCTAAGCCAACTGGTGGGCCAAAAGGTTCCAGGAGTGGCTGCAGCT
A12_BOLC.VG2.A	4692	AGTGGGCCAACAGATGGTTGGCCCAATGGGCAGGTGCAGACTCAATACATGG-----
GD33_BOLC.VG2.A	4732	AGTGGGCCAACAGATGGTTGGCCCAATGGGCAGGTGCAGACTCAATACATGG-----
A12_BOLC.VG2.B	4568	AGTGGGCCAACAGATGGTTGGCACAAATGGGCAGGTGCAGACTCAATACACAG-----
GD33_BOLC.VG2.B	4567	AGTGGGCCAACAGATGGTTGGCACAAATGGGCAGGTGCAGACTCAATACACAG-----
A12_BOLC.VG2.A	4745	--AAGTTCCTATATGTACCCACC--AAGTTCCTATATGTACCCACCAGGAGCTTCACCTTCTTCTGGTC
GD33_BOLC.VG2.A	4785	--AAGTTCCTATATGTACCCACC--AAGTTCCTATATGTACCCACCAGGAGCTTCACCTTCTTCTGGTC
A12_BOLC.VG2.B	4621	--AAGTTCATATATGTACCCACTGTAGC--AAGGTACTAGATTTTGATAAAGGCGCGTT
GD33_BOLC.VG2.B	4620	--AAGTTCATATATGTACCCACTGTAGC--AAGGTACTAGATTTTGATAAAGGCGCGTT
A12_BOLC.VG2.A	4801	-ATCCTC---CTTATTATGTTGATCCAT--TACTGT---TTTCTCTTTATATATA--ATTT
CD33_BOLC.VG2.A	4841	-ATCCTC---CTTATTATGTTGATCCAT--TACTGT---TTTCTCTTTATATATA--ATTT
A12_BOLC.VG2.B	4679	TATTTTAAATTTTTCATTGACAAATTTTAGTAAATGTCATATTTCAIAT--ATTT
GD33_BOLC.VG2.B	4678	TATTTTAAATTTTTCATTGACAAATTTTAGTAAATGTCATATTTCAIAT--ATTT
A12_BOLC.VG2.A	4853	CGGATTATTTTGAAGACTAGTTTCCATGTTATAGGTTTCACTAAAGTGTGTTGTTCTT
GD33_BOLC.VG2.A	4893	CGGATTATTTTGAAGACTAGTTTCCATGTTATAGGTTTCACTAAAGTGTGTTGTTCTT
A12_BOLC.VG2.B	4738	-GTCTTTTATTTTATAAAGACTTAACTTTTTCCTTTACTTAT---CGTATT--TTATT
GD33_BOLC.VG2.B	4737	-GTCTTTTATTTTATAAAGACTTAACTTTTTCCTTTACTTAT---CGTATT--TTATT
A12_BOLC.VG2.A	4913	TTATTGTTTAAATAACA---TGGGAAAACCAATAATATTATAGTATTTATTTTAA
GD33_BOLC.VG2.A	4953	TTATTGTTTAAATAACA---TGGGAAAACCAATAATATTATTTATTTTAA
A12_BOLC.VG2.B	4793	TTAATGCTATTTATTTT---TAAATAATTAACCTTATTTCTTTAATGAATTA--TTG
GD33_BOLC.VG2.B	4792	TTAATGCTATTTATTTT---TAAATAATTAACCTTATTTCTTTAATGAATTA--TTG
A12_BOLC.VG2.A	4970	ACATTGCTATTTTTCTCTATTTGACCTTGTTTGAGCTTCCAAAGTTTAACTT---
GD33_BOLC.VG2.A		
A12_BOLC.VG2.B	4850	ATATAACTCTGATAAATTAATTTTATTATGTGGTTAATATTTTATTAAATAAATTATA
GD33_BOLC.VG2.B	4849	ATATAACTCTGATAAATTAATTTTATTATGTGGTTAATATTTTATTAAATAAATTATA
A12_BOLC.VG2.A	5028	--CTTTTCTCTCTCTTTAACTGTTTCATTCCTGAATTTTGAAGATCTCGAGCTT
GD33_BOLC.VG2.A		
A12_BOLC.VG2.B	4910	TACTTTTATAAGATTTTATTTTTCATTAATAATTCAATTTTATGAATCTT
GD33_BOLC.VG2.B	4909	TACTTTTATAAGATTTTATTTTTCATTAATAATTCAATTTTATGAATCTT
A12_BOLC.VG2.A	5086	ACACTTTAGTAACTTTTGTATCC
GD33_BOLC.VG2.A		
A12_BOLC.VG2.B	4970	AAATTATATTAGAAAAAGAAAGA
GD33_BOLC.VG2.B	4969	AAATTATATTAGAAAAAGAAAGA

Alineación de secuencias de proteínas para proteínas ortólogas de Brassica oleracea VG2 a partir de GD33 y A12. Las secuencias de las proteínas se predijeron utilizando el programa bioinformático localizado en la red FGENESH (<http://mendel.cs.rhul.ac.uk/mendel.php?topic=faen-file>).

A12_BOLC.VG2.A	1	---MSSQFRYTQTPSKVVHLRNLFWECVEEELIDLCKRFGKIVNTKTNVGANRNQAFVE
GD33_BOLC.VG2.A	1	---MSSQFRYTQTPSKVVHLRNLFWECVEEELIDLCKRFGKIVNTKTNVGANRNQAFVE
A12_BOLC.VG2.B	1	---MSSQFRYTQTPSKVVHLRNLFWECVEEELIDLCKRFGKIVNTKTNVGANRNQAFVE
GD33_BOLC.VG2.B	1	---MSSQFRYTQTPSKVVHLRNLFWECVEEELIDLCKRFGKIVNTKTNVGANRNQAFVE
A12_BOLC.VG2.A	58	FGDVNQAI SMVSY YASSSEPAQIRGKT VYIQYSNRHEIVNNQSPGEVPGNVLLVTFEGVQ
GD33_BOLC.VG2.A	58	FGDVNQAI SMVSY YASSSEPAQIRGKT VYIQYSNRHEIVNNQSPGEVPGNVLLVTFEGVQ
A12_BOLC.VG2.B	59	FGDVNQAI SMVSY YASSSEPAQIRGKT VYIQYSNRHEIVNNQSPGEVPGNVLLVTFEGVQ
GD33_BOLC.VG2.B	59	FGDVNQAI SMVSY YASSSEPAQIRGKT VYIQYSNRHEIVNNQSPGEVPGNVLLVTFEGVQ

A12_BOLC.VG2.A	118	SHHVCIDVIHLVFSAYGFVHKIATFEKAAGTQALVQFTD VDTALAARTALDGRSIPK YLL
GD33_BOLC.VG2.A	118	SHHVCIDVIHLVFSAYGFVHKIATFEKAAGTQALVQFTD VDTALAARTALDGRSIPK YLL
A12_BOLC.VG2.B	119	SHHVCIDVIHLVFSAYGFVHKIATFEKAAGTQALVQFTD VDTALAARTALDGRSIPK YLL
GD33_BOLC.VG2.B	119	SHHVCIDVIHLVFSAYGFVHKIATFEKAAGTQALVQFTD VDTALAARTALDGRSIPK YLL
A12_BOLC.VG2.A	178	PEHVGSCNLRMSYSAHTDLN IKFQSHRSRDYTNPYLPVNQTAMDGSMQPALGADGK VE
GD33_BOLC.VG2.A	178	PEHVGSCNLRMSYSAHTDLN IKFQSHRSRDYTNPYLPVNQTAMDGSMQPALGADGK VE
A12_BOLC.VG2.B	179	PEHVGSCNLRMSYSAHTDLN IKFQSHRSRDYTNPYLPVNQTAMDGSMQ-----
GD33_BOLC.VG2.B	179	PEHVGSCNLRMSYSAHTDLN IKFQSHRSRDYTNPYLPVNQTAMDGSMQ-----
A12_BOLC.VG2.A	238	QSNVLL LIENMQYAVTV DVLHTVFSAYGT VQKIAIFEKNGSTQALIQYSDI STATMAKE
GD33_BOLC.VG2.A	238	QSNVLL LIENMQYAVTV DVLHTVFSAYGT VQKIAIFEKNGSTQALIQYSDI STATMAKE
A12_BOLC.VG2.B	227	-----VFSAYGT VQKIAIFEKNGSTQALIQYSDI PTATAKE
GD33_BOLC.VG2.B	227	-----VFSAYGT VQKIAIFEKNGSTQALIQYSDI PTATAKE
A12_BOLC.VG2.A	298	ALEGHCIYDGGYCKLRL YSRHTDLNVKAFSDKSRDYTL PDLSQLVGQKVPGVAAASGPT
G GD33_BOLC.VG2.A	298	ALEGHCIYDGGYCKLRL YSRHTDLNVKAFSDKSRDYTL PDLSQLVGQKVPGVAAASGPT
A12_BOLC.VG2.B	264	ALEGHCIYDC-----RLLDKSGDYTL PDLSQLVGQKVPGVAAASGPT
GD33_BOLC.VG2.B	264	ALEGHCIYDC-----RLLDKSGDYTL PDLSQLVGQKVPGVAAASGPT
A12_BOLC.VG2.A	358	DGWP--NGQVQTQYMGSSSYMYPPADPTGASPS SGHP PYYG
GD33_BOLC.VG2.A	358	DGWP--NGQVQTQYMGSSSYMYPPADPTGASPS SGHP PYYG
A12_BOLC.VG2.B	306	DCWH--NGQVQTQYTRSSSYMYPL-----
GD33_BOLC.VG2.B	306	DCWH--NGQVQTQYTRSSSYMYPL-----

Alineación de secuencias de ADN para genes de Brassica oleracea VG1

ES 2 989 717 T3

A12 1 AAAAAGCAGGCTAAGAGTGTAAAGCGACCCATTTACATCGGATCTAAAAATCAATAATC
GD33 1 AAAAAGCAGGCTAAGAGTGTAAAGCGACCCATTTACATCGGATCTAAAAATCAATAATC

A12 61 TAGAGTCTGACAAGAATCAACGTTGATAGAGAGTCCGAGAGAATAGTACAGAGCCAGTCC
GD33 61 TAGAGTGTGACAAGAATCAACGTTGATAGAGAGTGAAGAGAGAATAGTACAGAGCCAGTCC

A12 121 TCGAGGTAGAAATGTTTGTCTGAGCCCAAATGGTATAGTTGATATTGGGATTCCAGAAC
GD33 121 TCGAGGTAGAAATGTTTGTCTGAGCCCAAATGGTATAGTTGATATTGGGATTCCAGAAC

A12 181 TTGTTATCACCCACTATCCATCTCTTAGCGGCCACCTCAGTCACCGGAGCTGCCGCAAGA
GD33 181 TTGTTATCACCCACTATCCATCTCTTAGCGGCCACCTCAGTCACCGGAGCTGCCGCAAGA

A12 241 AAAGCCAGAACCATTGCCGCAGCCATCAACGCAAGTCTAGCCATAGATCACTCTAGAAAG
GD33 241 AAAGCCAGAACCATTGCCGCAGCCATCAACGCAAGTCTAGCCATAGATCACTCTAGAAAG

A12 301 AAAAAGAAGCTCAGTGCTTATGACTCTTAGTCTTATCTCTCAAGAAATAGCAGTTTAT
GD33 301 AAAAAGAAGCTCAGTGCTTATGACTCTTAGTCTTATCTCTCAAGAAATAGCAGTTTAT

A12 361 GATGAATGTATATGAGCGAGTGGGGCTGTTTCCAAGTCAAATCAGACATACAAATTTCT
GD33 361 GATGA-----GTGGGGCTGTTTCCAAGTCAAATCAGACATACAAATTTCT

A12 421 CGTTTTACGGGGAAAAGATCTCTCAAGATCCGTAATTTCTTCAACTAACGTTTGGAAATT
GD33 407 CGTTTTACGGGGAAAAGATCTCTCAAGATCCGTAATTTCTTCAACTAACGTTTGGAAATT

A12 481 TATTACTACTTTAATCAACCTCGATGTTGACCATGAAGTACACGTTAATTAAGGTACGC
GD33 467 TATTACTACTTTAATCAACCTCGATGTTGACCATGACCTACACGTTAATTAAGGTACGC

A12 541 GTTATATAGTTAAAGGAGTAATTAGATTTTAATTGCTGGCATTATTATTATTTTCTCTG
GD33 527 GTTATATAGTTAAAGGAGTAATTAGATTTTAACCTGCTGGCATTATTATTATTTTCTCTG

A12 601 GT----ATATT-----TTTAAAGCAATATTTTCTCACCTTTAGGCGGGAATTTATTT
GD33 587 GTTATTACATTATTCTGTTTTTAGTAAATATTTTCTCACCTTTAGGCGGGAATTTATTT

A12 650 CCATTTTCTTCTTTCAAGGGTTCAGAGATTCAATTCCTTTATTAACATACGAAATGAGTA
GD33 647 CCATTTTCTTCTTTCAAGGGTTCAGAGATTCAATTCCTTTATTAACATACGAAATGAGTA

A12 710 TATAGTATCTTTTAAATATTTAAATAATTAAAAATAAAATAAATTACATAACAGTGATA
GD33 707 TATAGTATCTTTTAAATATTTAAATAATTAAAAATAAAATAAATTACATAACAGTGATA

A12 770 CATATCAAAATTTCCAAAATAATGTGAGTACTTATATTAGTTGTACTTCTTAGTAAAT
GD33 767 CATATCAAAATTTCCAAAATAATGTGAGTACTTATATTAGTTGTACTTCTTAGTAAAT

A12 830 AACGAGATATGTTATGGCTCGTCGCATACA CAAAGCAGAGCATATGTTCCCGCAATAAC
GD33 827 AACGAGATATGTTATGGCTCGTCGCATACA CAAAGCAGAGCATATGTTCCCGCAATAAC

ES 2 989 717 T3

A12 890 ATGCACACGCAACTGATAGCACGTCGTAACCATTTCCGCGAACCACCAATGGATATTGT
GD33 887 ATGCACACGCAACTGATAGCACGTCGTAACCATTTCCGCGAACCACCAATGGATATTGT

A12 950 AATTCGATAAGACGACGTCGTGTTGTTGACCACGAATATTTCTTAAATTTCCCTTTATCT
GD33 947 AATTCGATAAGACGACGTCGTGTTGTTGACCACGAATATTTCTTAAATTTCCCTTTATCT

A12 1010 TTTGTTTTTGACGTCATCACATTGGTTTTATGTCCACACGATATCAGTCATAAAATCCACG
GD33 1007 TTTGTTTTTGACGTCATGACATTGGTTTTATGTCCACACGATATCAGTCATAAAATCCACG

A12 1070 TAGGGAACGACTATCCATTACACCCATCCACGTGTTCTCATAACAGCTTCTAGTTCCATT
GD33 1067 TAGGGAACGACTATCCATTACACCCATCCACGTGTTCTCATAACAGCTTCTAGTTCCATT

A12 1130 AGCTGAACGAACAATAGAAACAGAGCATCGAACTGAGAAAGAAGAAGAAGGGAAGCA
GD33 1126 AGCTGAACGAACAATAGAAACAGAGCATCGAACTGAGAAAGAAGAAGAAGGGAAGCA

A12 1190 AGAGAAATGGCGGCAGCGGCCATGGCCGTTTCATCTCCCGAAACATTCATCGTTTCTTCTCT
GD33 1186 AGAGAAATGGCGGCAGCGGCCATGGCCGTTTCATCTCCCGAAACATTCATCGTTTCTTCTCT

A12 1250 AATCCCAAGCTTCCATTGATCAAAACTCTAACTTCTCGGTCTGTCCTTAAAGATCGGT
GD33 1246 AATCCCAAGCTTCCATTGATCAAAACTCTAACTTCTCGGTCTGTCCTTAAAGATCGGT

A12 1310 CGAACCATGTCTCTTAACCGGAAAAATGAAAGGTCCAGTTACGTTTCAGCTGCTTCGACC
GD33 1306 CGAACCATGTCTCTTAACCGGAAAAATGAAAGGTCCAGTTACGTTTCAGCTGCTTCGACC

A12 1370 GTCGAAGGCGATCGAAGCAAAACAGTTTTACATAAACTTCACTGGATTCCCATTTCTCTT
GD33 1366 GTCGAAGGCGATCGAAGCAAAACAGTTTTACATAAACTTCACTGGATTCCCATTTCTCTT

A12 1430 GGTCTTTCTCTTAACCGGCGCACCATCAGAACCAGGTTAGGCTTTCTCCATCCCTCTTG
GD33 1426 GGTCTTTCTCTTAACCGGCGCACCATCAGAACCAGGTTAGGCTTTCTCCATCCCTCTTG

A12 1490 AGTTTTGATTTGAACTCGTTGAGAATCCCATATGGAT---GTTATGCAGGCGGTTAAAGG
GD33 1486 AGTTTTGATTTGAACTCGTTGAGAATCCCATATGGATAATGTTATGCAGGCGGTTAAAGG

A12 1547 AAGCATATGGATGTTTGAACAAGAACAAGCTTTAGGTTTCAGCAGTGTCTCAACCAATAT
GD33 1546 AAGCATATGGATGTTTGAACAAGAACAAGCTTTAGGTTTCAGCAGTGTCTCAACCAATAT

A12 1607 AAGAATGACTGTCATCAAACTCAAATCCGGTGGCTTATGGGTTTCATGCCCTATTGCTCC
GD33 1606 AAGAATGACTGTCATCAAACTCAAATCCGGTGGCTTATGGGTTTCATGCCCTATTGCTCC

A12 1667 CACCAAAGAGTGATTCAGGTTCCCTCTTTTATTTCATTATCCTTGATAAAGTATCATCC
GD33 1666 CACCAAAGAGTGATTCAGGTTCCCTCTTTTATTTCATTATCCTTGATAAAGTATCATCC

A12 1727 TTTCTTATCATTTGATAA--TAACCAATGTCTAAACTCTTTTGGGTTGTTTGGCAAC
GD33 1717 TTTCTTATCATTTGATAAATTAACCAATGTCTAAACTCTTTTGGGTTGTTTGGCTTAC

ES 2 989 717 T3

A12 1785 TTTTCTTTGAGGAGCTTATTGAGGAGTTGGGAGCTCCGGTTGAGTACATTCTCCTCC
GD33 1777 TTTTCTTTGAGGAGCTTATTGAGGAGTTGGGAGCTCCGGTTGAGTACATTCTCCTCC

A12 1845 AACCTTTTGCTTACGAGCACAAGATCTTCGTCCGTCCCTTCTCTAGAAAGTTCCCCAAGGC
GD33 1837 AACCTTCGCTTACGAGCACAAGATCTTCGTCCGTCCCTTCTCTAGAAAGTTCCCCAAGGC

A12 1905 TCAAGTATGGGTGGCGCCAAGACAATGGAGCTGGCCACTGAACTTACCACTCGAGTTTT
GD33 1897 TCAAGTATGGGTGGCGCCAAGACAATGGAGCTGGCCACTGAACTTACCACTCGAGTTTT

A12 1965 CGGTATCTTTCCGCGCTAAAACCATTAAGACGGAGACTTCTACCCCGTGGGCTGATGA
GD33 1957 CGGTATCTTTCCGCGCTAAAACCATTAAGACGGAGACTTCTACCCCGTGGGCTGATGA

A12 2025 GATCGAGCAGAAAGTCTTAAGCTCTCCTGAAGTCGGTACGTTCCCTTCACTTTCTCAATCT
GD33 2017 GATCGAGCAGAAAGTCTTAAGCTCTCCTGAAGTCGGTACGTTCCCTTCACTTTCTCAATCT

A12 2085 TGTGGCCCTCCTTAGCCCAAGCCTCAGTAAACATTCGCTTAACCCGCACTCTCCTCCT
GD33 2077 TGTGGCCCTCCTTAGCCCAAGCCTCA-TAAACATTCGCTTCAACCC---CTCACC---

A12 2145 CTTCCTTCAAAATTTTGAAGAGTTATATATACATAGCCTCCCAAGTTATTAAGAA
GD33 2129 ---CCACCAAAATTTTGAAGAGTTATATATACATAGCC-CCAAATTTATTAAGAA

A12 2205 AACGTTTGTCTCCAAACTATGAAATCTTTATCAAATTTCTATCGCTCCAAATCC
GD33 2185 ATC-----TATTGAAATCTTTATCAAATTTCTATCTCCCTAGAAATCT

A12 2265 GAGGGTCGGTCCGTGACTTATGATAA-TAATATACAGGAATTGGACCGTATGTGGAAG
GD33 2229 GAGGGTCGGTCCGTGACTTATGATAAATAATATACAGGAATTGGACCGTATGTGGAAG

A12 2324 TAGCGTTCTACCAATAAGCGTTCAAGAACTCTATTAGTCACTGATGCTGTGATCTTCGTCC
GD33 2289 TAGCGTTCTACCAATAAGCGTTCAAGAACTCTATTAGTCACTGACGCTGTGATCTTCGTCC

A12 2384 CACGAAGCCGCCATCGAGTATCAGCAGCGAGTCTTTGCTGGCTCTGCTAAGAACGGAC
GD33 2349 CAAAGGAAGCCGCCATCGAGTATCAGCAGCGAGTCTTTGCTGGCTCTGCTAAGAACGGAC

A12 2444 TCGCTGTGAAGATACTTAGCAAAGGCAAACAAGTACCTATGACCCGTGTCGTTGATACCC
GD33 2409 TCGCTGTGAAGATACTTAGCAAAGGCAAACAAGTACCTATGACCCGTGTCGTTGATACCC

A12 2504 CAAACACCCGCCAAAAGGTGAGCCATCAGCTTCTTGAATTCAAATTCATCAACAAA
GD33 2469 CAAACACCCGCCAAAAGGTGAGCCATCAGCTTCTTGAATTCAAATTCATCAACAAA

A12 2564 GATCTCATGGGTGTCTTTAACGTGCAGGATGGGAAAGAATGGTGTGCAAATCCTGTT
GD33 2528 GATCTCATGGGTGTCTTTAACGTGCAGGATGGGAAAGAATGGTGTGCAAATCCTGTT

A12 2624 TCTTGGTCCCTCTAATCTCTTGAGCCAAACGCGAGCTTCGCAAAATGTACAGAAGCT
GD33 2588 TCTTGGTCCCTCTAATCTCTTGAGCCAAACGCGAGCTTCGCAAAATGTACAGAAGCT

A12 2684 GATCGTTTCTCCATTGTCAAGACTCTGTCTTTAGCAAAGTCCCTGAGAAGGTGAGGGA

GD33 2648 GATTGTTTCTCCCATTTGTCAAGACTCTGGTCTTTAGCAAAGTCCCTGAGAAGGTGAGGGA

A12 2744 CTGGATCGATGAGATAGCGAGTGACTGGAGTTCAAGAGGATAATCCCAGCTCATTTCGA
GD33 2708 CTGGATCGATGAGATAGCGAGTGACTGGAGTTCAAGAGGATAATCCCAGCTCATTTCGA

A12 2804 GGCTCCGTAAACGCGGGAGGTCAGATTTCCTAGCTGCGTTTGGGTTTCTTGATGATCT
GD33 2768 GGCTCCGTAAACGCGGGAGGTCAGATTTCCTAGCTGCGTTTGGGTTTCTTGATGATCT

A12 2864 TCTAGGGGAAAGATATGTGAACCGTCTCCTTCGCTCTCTGTTCTCTTCACTTCGCTGAT
GD33 2828 TCTAGGGGAAAGATATGTGAACCGTCTCCTTCGCTCTCTGTTCTCTTCACTTCGCTGAT

A12 2924 GGGTAAAGCAGCGAGCTATTTTCCTCCGGATGATATGAGACTCTCTCTTCTCTTGATCA
GD33 2888 GGGTAAAGCAGCGAGCTATTTTCCTCCGGATGATATGAGACTCTCTCTTCTCTTGATCA

A12 2984 GTTCTTAGTCTCTGTTGGTGCTGTTAAGAAGACCGTCTCTGGTAGAAAACGAAGATGACG
GD33 2948 GTTCTTAGTCTCTGTTGGTGCTGTTAAGAAGACCGTCTCTGGTAGAAAACGAAGATGACG

A12 3044 GAACAAACCCAGCTTTCT
GD33 3008 GAACAAACCCAGCTTTCT

Alineación de secuencias de proteínas para proteínas ortólogas de Brassica oleracea VG1 a partir de A12 y GD33. Las secuencias de las proteínas se predijeron utilizando el programa bioinformático localizado en la red FGENESH.

5

A12 1 MAAAMAVHLPKHSSFLNPKLELNQNSNFLGVSLKGRPMSVNRKMKGPVTVSAAST--
GD33 1 MAAAMAVHLPKHSSFLTKPKHESDQNSNFLGGSLEGRSMSNRKTKGPVMVSAAST--

A12 59 -----VEGDRSKQFYINFTGFFPLGPFLNRRTIRTEAVKGSIWMFEQEALGFSSVST
GD33 59 -----VEGDRSKQFYINFTGFFPLGPFLNRRTIRTEAVKGSIWMFEQEALGFSSVST

A12 113 NIRMTVILKSGGLWVHAPIAPTKECIQLIEELGAPVEYIVLPTFAYEHKIFVGPFSRKF
GD33 113 NIRMTVILKSGGLWVHAPIAPTKECIQLIEELGAPVEYIVLPTFAYEHKIFVGPFSRKF

A12 173 PKAQVWVAPRQWSWPLNLPLEFFGIFRAKTIKDGLSTPWADEIEQKVLSSPEVGIGPYV
GD33 173 PKAQVWVAPRQWSWPLNLPLEFFGIFRAKTIKDGLSTPWADEIEQKVLSSPEVGIGPYV

A12 233 EVAFYHKRSRTLLVTDVAVIFVPRKPPSSISSESLASAKNGLAVKILSKGKQVPDPVVD
GD33 233 EVAFYHKRSRTLLVTDVAVIFVPRKPPSSISSESLASAKNGLAVKILSKGKQVPNDPVVD

A12 293 TPNTRQKGWERMVLQILFLGPSNLLEPNASFAQMSQKLIVSPIVKTLVFSKVPEKVRDWI
GD33 293 TPNTRQKGWERMVLQILFLGPSNLLEPNASFAQMSQKLIVSPIVKTLVFSKVPEKVRDWI

A12 353 DEIASDWRFKRIIPAHFEAPINAGRSEFLAAFGLDLDLGERYVNRPPSLSVLFTSLMGK
GD33 353 DEIASDWRFKRIIPAHFEAPINAGRSEFLAAFGLDLDLGERYVNRPPSLSVLFTSLMGK

A12 413 AASYFPDDMRTLSSLDQFLVSVGAVKKTVSGRKRR
GD33 413 AASYFPDDMRTLSSLDQFLVSVGAVKKTVSGRKRR

ES 2 989 717 T3

GD33_BOLC.VG2.A	1845	ATATCTAACAA--TAAGTATTTA--AATTTAATA--TTCTTTAAATCTCTATT--TATACA
A12_BOLC.VG2.B	1741	TTTTC--AAT--AGIATTTCTCTTTTAATAATATAC--ATTG--TCAAAAATTATATTGGATA--A
GD33_BOLC.VG2.B	1740	TTTTC--AAT--AGIATTTCTCTTTTAATAATATAC--ATTG--TCAAAAATTATATTGGATA--A
A12_BOLC.VG2.A	1905	AA--GAAGTATAIATTTT--TGTTTACTTTTGAATTTT--AGATTTTACTAT--GATATAT
GD33_BOLC.VG2.A	1904	AA--GAAGTATAIATTTT--TGTTTACTTTTGAATTTT--AGATTTTACTAT--TATATAT
A12_BOLC.VG2.B	1800	CC--CAAGCA--AAATCCATTGAGAAATCACAACAACAAAAGA--AGGAACA--CACACAA
GD33_BOLC.VG2.B	1799	CC--CAAGCA--AAATCCATTGAGAAATCACAACAACAAAAGA--AGGAACA--CACACAA
A12_BOLC.VG2.A	1963	CC--AGCCGATCCGATA--CAATCCGAATCCGAATTTT--TTGGCTACTTT--GTATATCT
GD33_BOLC.VG2.A	1962	CC--AGCCGATCCGATA--CAATCCGAATCCGAATTTT--TTGGCTACTTT--GTATATCT
A12_BOLC.VG2.B	1857	CC--CGCCGA--CAGATAACTAAATGTACTGGACTTTT--CCCTTAATATAG--CTAAT--
GD33_BOLC.VG2.B	1856	CC--CGCCGA--CAGATAACTAAATGTACTGGACTTTT--CCCTTAATATAG--CTAAT--
A12_BOLC.VG2.A	2021	GAACCGATCCGAAACCGA--GTGTGTATATCCGAATCTCATTCCGTACTTT--TAAATTTTACTA
GD33_BOLC.VG2.A	2020	GAACCGATCCGAAACCGA--GTGTGTATATCCGAATCTCATTCCGTACTTT--TAAATTTTACTA
A12_BOLC.VG2.B	1914	GAACAGATT---AAATGGCTA--TTATAGACGTC--TTGCAT--GTATCCATCTATCT--GAA
GD33_BOLC.VG2.B	1913	GAACAGATT---AAATGGCTA--TTATAGACGTC--TTGCAT--GTATCCATCTATCT--GAA
A12_BOLC.VG2.A	2081	GA--AT--AAAC--TTGGGGAT--ATAAC--AAATAG--ACCGAAATCCGAAAA--ACCGATCCGAAAC
GD33_BOLC.VG2.A	2080	GA--AT--AAAC--TTGGGGAT--ATAAC--AAATAG--ACCGAAATCCGAAAA--ACCGATCCGAAAC
A12_BOLC.VG2.B	1971	GC--ATA--TAATTTC---TTT--TAG--TACTGTTTGC--AATTTTACAGAT--ATTTTAGC
GD33_BOLC.VG2.B	1970	GC--ATA--TAATTTC---TTT--TAG--TACTGTTTGC--AATTTTACAGAT--ATTTTAGC
A12_BOLC.VG2.A	2140	CCGAACGCCGAG--CCCTAGTACAGAGTTCCGATCAGAA--AT--AGAGAATCCCTTTGAGCGTC
GD33_BOLC.VG2.A	2139	C--GAACGCCCAG--CCCTAGTACAGAGTTCCGATCAGAA--AT--AGAGAATCCCTTTGAGCGTC
A12_BOLC.VG2.B	2024	T-----GTACAGTACAGAG-----AGAGAATCCCTTTGAGCGTC
GD33_BOLC.VG2.B	2023	T-----GTACAGTACAGAG-----AGAGAATCCCTTTGAGCGTC
A12_BOLC.VG2.A	2200	G---GCCTTGCTTTGTGCGTAGGGATCTCT--CCGACTAATTGTTGTAATTATATAA--GGC
GD33_BOLC.VG2.A	2198	G---GCCTTGCTTTGTGCGTAGGGATCTCT--CCGACTAATTGTTGTAATTATATAA--GGC
A12_BOLC.VG2.B	2057	CTCGCCCTTCTTTCTGCTAGCGATGTCCT--CTGACTAATTGTTGTAAC--TACAAAACGG
GD33_BOLC.VG2.B	2056	GTGCGCCCTTGCTTTGTGCTAGGGATCTCT--CTGACTAATTGTTGTAAC--TACAAAACGG
A12_BOLC.VG2.A	2256	GGC-----GGCCCTTTT--TAG--CCGAG--AGATGGCCAGCTCA-----CACTTCAGGTAT
GD33_BOLC.VG2.A	2254	GGC-----GGCCCTTTT--TAG--CCGAG--AGATGGCCAGCTCA-----CACTTCAGGTAT
A12_BOLC.VG2.B	2117	CGCCCCCGCCGCTTTT--TACT---AGATGGCCAGCTCAT-----CACAGTTTAGGTAT
GD33_BOLC.VG2.B	2116	CGCCCCCGCCGCTTTT--TACT---AGATGGCCAGCTCAT-----CACAGTTTAGGTAT
A12_BOLC.VG2.A	2303	AC--CAGACGCCGTGGAAGGTGGTGCACCTGAGGAATCTGCCGTGGGAATGCGTGGAAGAG
GD33_BOLC.VG2.A	2301	AC--CAGACGCCGTGGAAGGTGGTGCACCTGAGGAATCTGCCGTGGGAATGCGTGGAAGAG
A12_BOLC.VG2.B	2167	ACC--CAGACGCCGTGGAAGGTGGTGCACCTGAGGAATCTGCCGTGGGAATGCGTGGAAGAG
GD33_BOLC.VG2.B	2166	ACC--CAGACGCCGTGGAAGGTGGTGCACCTGAGGAATCTGCCGTGGGAATGCGTGGAAGAG
A12_BOLC.VG2.A	2363	GAGCTCATCGACCTATGCAAACGATTCCGGCAAGATCGTCAATACGAAGACCAATGTCCGC
GD33_BOLC.VG2.A	2361	GAGCTCATCGACCTATGCAAACGATTCCGGCAAGATCGTCAATACGAAGACCAATGTCCGC
A12_BOLC.VG2.B	2227	GAGCTCATCGACCTATGCAAACGATTCCGGCAAGATCGTCAATACGAAGACCAATGTCCGC
GD33_BOLC.VG2.B	2226	GAGCTCATCGACCTATGCAAACGATTCCGGCAAGATCGTCAATACGAAGACCAATGTCCGC
A12_BOLC.VG2.A	2423	GCCAATCGCAACCAAGCCTTTGTGCGAATTCGTAA---CAACT-----TTTTATTCTC
GD33_BOLC.VG2.A	2421	GCCAATCGCAACCAAGCCTTTGTGCGAATTCGTAA---CAACT-----TTTTATTCTC
A12_BOLC.VG2.B	2287	GCCAATCGCAACCAAGCCTTTGTGCGAATTCGTAA---GAACT-----TTTTATTCTC
GD33_BOLC.VG2.B	2286	GCCAATCGCAACCAAGCCTTTGTGCGAATTCGTAA---CAACT-----TTTTATTCTC
A12_BOLC.VG2.A	2472	TTGGATCATCAGATTGTTTCT-----TG-----TTTCTTTGAT--AACTCTT
GD33_BOLC.VG2.A	2470	TTGGATCATCAGATTGTTTCTGAAG-----CTTT--TG-----TTTCTTTGAT--AACTCTT
A12_BOLC.VG2.B	2336	TTGGATCATCAGATTGTTGCTTCC-----CCCC--TG-----TTTCTTTGAT--AACTCTT

GD33_BOLC.VG2.B 2335 TTGGATCATCAGATGTGTGCTTCCC-----CCCCATGATTTCTTAGAAATGCA

A12_BOLC.VG2.A 2493 -----TTTTTGGTATCTCGTTGTTGT-----TTTGGTAT-----T

GD33_BOLC.VG2.A 2521 -----TTTTTTTGGTATCTCGTTGTTGT-----TTTGGTAT-----T

A12_BOLC.VG2.B 2388 -----TAAAAAAGGACTCGAATCTCTTTT-----TTTGGTATCTCTTT

GD33_BOLC.VG2.B 2387 -----TAAAAAAGGACTCGAATCTCTTTT-----TTTGGTATCTCTTT

A12_BOLC.VG2.A 2524 GTT-----GTTTTGAAGATGAAACTGTATACTTT-----GATTCA

GD33_BOLC.VG2.A 2554 GTT-----GTTTTGAAGATGAAACTGTATACTTT-----GATTCA

A12_BOLC.VG2.B 2426 GTT-----GTTTTGAAGATGAAACTGTATACTTT-----GATTCA

GD33_BOLC.VG2.B 2425 GTT-----GTTTTGAAGATGAAACTGTATACTTT-----GATTCA

A12_BOLC.VG2.A 2559 TATT-----CGCAGGGTGACCTGAATCAGGCAATATCAATGGTT

GD33_BOLC.VG2.A 2589 TATT-----CGCAGGGTGACCTGAATCAGGCAATATCAATGGTT

A12_BOLC.VG2.B 2461 TATT-----CGCAGGGTGACCTGAATCAGGCAATATCAATGGTT

GD33_BOLC.VG2.B 2460 TATT-----CGCAGGGTGACCTGAATCAGGCAATATCAATGGTT

A12_BOLC.VG2.A 2593 TCTTACTATGCTTCCTCTTCAGAGCCGECTCAGATTCGAGGGAAGACTGTTTATATTTCAG

GD33_BOLC.VG2.A 2628 TCTTATATGCTTCCTCTTCAGAGCCGECTCAGATTCGAGGGAAGACTGTTTATATTTCAG

A12_BOLC.VG2.B 2500 TCTTACTATGCTTCCTCTTCAGAGCCGECTCAGATTCGAGGGAAGACTGTTTATATTTCAG

GD33_BOLC.VG2.B 2499 TCTTACTATGCTTCCTCTTCAGAGCCGECTCAGATTCGAGGGAAGACTGTTTATATTTCAG

A12_BOLC.VG2.A 2653 TACTCTAATCGCCATGAGATTGTCAACAATCAGASTCCTGGAGAGGTCCCTGGCAATGTC

GD33_BOLC.VG2.A 2688 TACTCTAATCGCATGAGATTGTCAACAATCAGASTCCTGGAGAGGTCCCTGGCAATGTC

A12_BOLC.VG2.B 2560 TACTCTAATCGCCATGAGATTGTCAACAATCAGASTCCTGGAGAGGTCCCTGGCAATGTC

GD33_BOLC.VG2.B 2559 TACTCTAATCGCCATGAGATTGTCAACAATCAGASTCCTGGAGAGGTCCCTGGCAATGTC

A12_BOLC.VG2.A 2713 CTCTTGGTCACCTTTGAAGGAGTCCAATCTCACCATGCTGTCATCGATGTCATCCATCTG

GD33_BOLC.VG2.A 2743 CTCTTGGTCACCTTTGAAGGAGTCCAATCTCACCATGCTGTCATCGATGTCATCCATCTG

A12_BOLC.VG2.B 2620 CTATTGGTCACCTTTGAAGGAGTCCAATCTCACCATGCTGTCATCGATGTCATCCATCTG

GD33_BOLC.VG2.B 2619 CTATTGGTCACCTTTGAAGGAGTCCAATCTCACCATGCTGTCATCGATGTCATCCATCTG

A12_BOLC.VG2.A 2773 GTATGTGAATATTTCAGCTTACCTTCCACTATTGTTTCTT-GTTATGTTAAGTGATTTTTT

GD33_BOLC.VG2.A 2803 GTATGTGAATATTTCAGCTTACCTTCCACTATTGTTTCTT-GTTATGTTAAGTGATTTTTT

A12_BOLC.VG2.B 2680 GTATGTGAATATTTCAGCTTACCTTCCACTATTGTTTCTT-GTTAT-----AGTGATTTTTT

GD33_BOLC.VG2.B 2679 GTATGTGAATATTTCAGCTTACCTTCCACTATTGTTTCTT-GTTAT-----AGTGATTTTTT

A12_BOLC.VG2.A 2837 TCG-----TTTCTTCG-----AGTAGATTCTAATCTATGAAAATATTTCAACTTG

GD33_BOLC.VG2.A 2867 TCG-----TTTCTTCG-----AGTAGATTCTAATCTATGAAAATATTTCAACTTG

A12_BOLC.VG2.B 2735 TCG-----TTTCTTCG-----AGTAGATTCTAATCTATGAAAATATTTCAACTTG

GD33_BOLC.VG2.B 2734 TCG-----TTTCTTCG-----AGTAGATTCTAATCTATGAAAATATTTCAACTTG

A12_BOLC.VG2.A 2882 TTGTTATTAGGCAAACCTT-----CTTTTG-AGTGATTTTTTTTCC-----

GD33_BOLC.VG2.A 2912 TTGTTATTAGGCAAACCTT-----CTTTTG-AGTGATTTTTTTTCC-----

A12_BOLC.VG2.B 2780 TTGTTATTAGGCAAACCTT-----CTTTTG-AGTGATTTTTTTTCC-----

GD33_BOLC.VG2.B 2779 TTGTTATTAGGCAAACCTT-----CTTTTG-AGTGATTTTTTTTCC-----

A12_BOLC.VG2.A 2921 -----AGTTATTGTTAGACATAC-----AGTATGTC-----ACATACTATTG

GD33_BOLC.VG2.A 2951 -----AGTTATTGTTAGACATAC-----AGTATGTC-----ACATACTATTG

A12_BOLC.VG2.B 2819 -----AGTTATTGTTAGACATAC-----AGTATGTC-----ACATACTATTG

GD33_BOLC.VG2.B 2818 -----AGTTATTGTTAGACATAC-----AGTATGTC-----ACATACTATTG

A12_BOLC.VG2.A 2953 TAAATTACAGTATATCTGACGTTAATGAAAATGCTCGAATCACAGATGTTGATGCCTCTT

GD33_BOLC.VG2.A 2988 TAAATTACAGTATATCTGACGTTAATGAAAATGCTCGAATCACAGATGTTGATGCCTCTT

A12_BOLC.VG2.B 2856 TAAATTACAGTATATCTGACGTTAATGAAAATGCTCGAATCACAGATGTTGATGCCTCTT

GD33_BOLC.VG2.B 2855 TAAATTACAGTATATCTGACGTTAATGAAAATGCTCGAATCACAGATGTTGATGCCTCTT

A12_BOLC.VG2.A 3013 TATTATAATCTTTCIGGAGAGAGTTTGGAAAAATAGTTTCATGTTTCGTCATTCTCAT----

GD33_BOLC.VG2.A 3043 TATTATAATCTTTCIGGAGAGAGTTTGGAAAAATAGTTTCATGTTTCGTCATTCTCAT----

A12_BOLC.VG2.B 2916 TATTATAATCTTTCIGGAGAGAGTTTGGAAAAATAGTTTCATGTTTCGTCATTCTCAT----

GD33_BOLC.VG2.B 2915 TATTATAATCTTTCIGGAGAGAGTTTGGAAAAATAGTTTCATGTTTCGTCATTCTCAT----

A12_BOLC.VG2.A 3074 -----GGACAGGTGTCACCTCTGCACTTATCTAGTCATTCTCTT

GD33_BOLC.VG2.A 3104 -----GGACAGGTGTCACCTCTGCACTTATCTAGTCATTCTCTT

A12_BOLC.VG2.B 2972 -----GGACAGGTGTCACCTCTGCACTTATCTAGTCATTCTCTT

GD33_BOLC.VG2.B 2971 -----GGACAGGTGTCACCTCTGCACTTATCTAGTCATTCTCTT

A12_BOLC.VG2.A 3112 TTTTAGTCTCCTAA-----CTTGAGTTTATTTCGATTGATTGCTTCCCTTAGTT

GD33_BOLC.VG2.A 3142 TTTTAGTCTCCTAA-----CTTGAGTTTATTTCGATTGATTGCTTCCCTTAGTT

A12_BOLC.VG2.B 3010 TTTTAGTCTCCTAA-----CTTGAGTTTATTTCGATTGATTGCTTCCCTTAGTT

GD33_BOLC.VG2.B 3009 TTTTAGTCTCCTAA-----CTTGAGTTTATTTCGATTGATTGCTTCCCTTAGTT

A12_BOLC.VG2.A 3162 ATTATCAATTACTCCACTGTATTATATGGACATGACTCATATCTAGTCCAAACTTTTGT

GD33_BOLC.VG2.A	3202	ATTATCAATTTACTCCACTGTATTATATGGACATGCTCATATCTAGTCCAAACTTTTGT
A12_BOLC.VG2.B	3060	ATTATCAATTTACTCCACTGTATTATATGGACATGACTCATATCTAGTCCAAACTTTTGT
GD33_BOLC.VG2.B	3059	ATTATCAATTTACTCCACTGTATTATATGGACATGACTCATATCTAGTCCAAACTTTTGT
A12_BOLC.VG2.A	3222	TTTCAGGTGTTTTCTGCTTATGGCTTCGTGCACAAAATTGCCACTTTTGAGAAAGCTGCTG
GD33_BOLC.VG2.A	3262	TTTCAGGTGTTTTCTGCTTATGGCTTCGTGCACAAAATTGCCACTTTTGAGAAAGCTGCTG
A12_BOLC.VG2.B	3120	TTTCAGGTGTTTTCTGCTTATGGCTTCGTGCACAAAATTGCCACTTTTGAGAAAGCTGCTG
GD33_BOLC.VG2.B	3119	TTTCAGGTGTTTTCTGCTTATGGCTTCGTGCACAAAATTGCCACTTTTGAGAAAGCTGCTG
A12_BOLC.VG2.A	3282	GTTTCCAGGTTAGAGATATCGAGTTTGTGTTTCAAGGTCTTGCAAT-----A--AATC
GD33_BOLC.VG2.A	3322	GTTTCCAGGTTAGAGATATCGAGTTTGTGTTTCAAGGTCTTGCAAT-----A--AATC
A12_BOLC.VG2.B	3180	GTTTCCAGGTTAGAGATATCGAGTTTGTGTTTCAAGGTCTTGCAAT-----A--AATC
GD33_BOLC.VG2.B	3179	GTTTCCAGGTTAGAGATATCGAGTTTGTGTTTCAAGGTCTTGCAAT-----A--AATC
A12_BOLC.VG2.A	3332	TAATGATGTTTAAACTCATCTCTGCGCTCATAACCTGCTCAGGCACCTGTTTCAGTTTAC
GD33_BOLC.VG2.A	3372	TAATGATGTTTAAACTCATCTCTGCGCTCATAACCTGCTCAGGCACCTGTTTCAGTTTAC
A12_BOLC.VG2.B	3238	TAATGATGTTTAAACTCATCTCTGCGCTCATAACCTGCTCAGGCACCTGTTTCAGTTTAC
GD33_BOLC.VG2.B	3237	TAATGATGTTTAAACTCATCTCTGCGCTCATAACCTGCTCAGGCACCTGTTTCAGTTTAC
A12_BOLC.VG2.A	3392	TGATGTGGACACTGCCCTTAGCGGCAAGGACTGCGCTGATGGTAGAAGTATACCCACGTA
GD33_BOLC.VG2.A	3432	TGATGTGGACACTGCCCTTAGCGGCAAGGACTGCGCTGATGGTAGAAGTATACCCACGTA
A12_BOLC.VG2.B	3298	TGATGTGGACACTGCCCTTAGCGGCAAGGACTGCGCTGATGGTAGAAGTATACCCACGTA
GD33_BOLC.VG2.B	3297	TGATGTGGACACTGCCCTTAGCGGCAAGGACTGCGCTGATGGTAGAAGTATACCCACGTA
A12_BOLC.VG2.A	3452	TGCTCAAATCCTTCATTCATGCTTTTGACCATAAGATAAAGCTCTGTTGATGGTTTCTTC
GD33_BOLC.VG2.A	3492	TGCTCAAATCCTTCATTCATGCTTTTGACCATAAGATAAAGCTCTGTTGATGGTTTCTTC
A12_BOLC.VG2.B	3358	TGCTCAAATCCTTCATTCATGCTTTTGACCATAAGATAAAGCTCTGTTGATGGTTTCTTC
GD33_BOLC.VG2.B	3357	TGCTCAAATCCTTCATTCATGCTTTTGACCATAAGATAAAGCTCTGTTGATGGTTTCTTC
A12_BOLC.VG2.A	3512	CTTTTCTTTTGGIAAATTTCAGATATCTGCTTCCAGAACATGTAGGCTCATGCAATTTGCG
GD33_BOLC.VG2.A	3552	CTTTTCTTTTGGIAAATTTCAGATATCTGCTTCCAGAACATGTAGGCTCATGCAATTTGCG
A12_BOLC.VG2.B	3418	CTTTTCTTTTGGIAAATTTCAGATATCTGCTTCCAGAACATGTAGGCTCATGCAATTTGCG
GD33_BOLC.VG2.B	3417	CTTTTCTTTTGGIAAATTTCAGATATCTGCTTCCAGAACATGTAGGCTCATGCAATTTGCG
A12_BOLC.VG2.A	3572	AATGTCTTACTCAGCTCATACTGATCTAAATATCAAATTTTCAGTCCCACCGCAGCAGGTA
GD33_BOLC.VG2.A	3612	AATGTCTTACTCAGCTCATACTGATCTAAATATCAAATTTTCAGTCCCACCGCAGCAGGTA
A12_BOLC.VG2.B	3478	AATGTCTTACTCAGCTCATACTGATCTAAATATCAAATTTTCAGTCCCACCGCAGCAGGTA
GD33_BOLC.VG2.B	3477	AATGTCTTACTCAGCTCATACTGATCTAAATATCAAATTTTCAGTCCCACCGCAGCAGGTA
A12_BOLC.VG2.A	3632	GAGTTTGGAGTCCCTCGCAAATGTGCGCTCTCCATTGTTTATACTGCTCTTACTGTTGGAAGC
GD33_BOLC.VG2.A	3672	GAGTTTGGAGTCCCTCGCAAATGTGCGCTCTCCATTGTTTATACTGCTCTTACTGTTGGAAGC
A12_BOLC.VG2.B	3538	GAGTTTGGAGTCCCTCGCAAATGTGCGCTCTCCATTGTTTATACTGCTCTTACTGTTGGAAGC
GD33_BOLC.VG2.B	3537	GAGTTTGGAGTCCCTCGCAAATGTGCGCTCTCCATTGTTTATACTGCTCTTACTGTTGGAAGC
A12_BOLC.VG2.A	3692	CTTATGGTTGAATAGTTCTTCATGTGTTTATTCCTCTTACGCACTCAGGGACTACACA
GD33_BOLC.VG2.A	3732	CTTATGGTTGAATAGTTCTTCATGTGTTTATTCCTCTTACGCACTCAGGGACTACACA
A12_BOLC.VG2.B	3598	CTTATGGTTGAATAGTTGATTCATGTGTTTATTCCTCTTACGCACTCAGGGACTACACA
GD33_BOLC.VG2.B	3597	CTTATGGTTGAATAGTTGATTCATGTGTTTATTCCTCTTACGCACTCAGGGACTACACA
A12_BOLC.VG2.A	3751	AATCCATACCTTCCGGTGAATCAAACCGCTATGGACGGTTCTATGCAGGTATATTTTCC
GD33_BOLC.VG2.A	3791	AATCCATACCTTCCGGTGAATCAAACCGCTATGGACGGTTCTATGCAGGTATATTTTCC
A12_BOLC.VG2.B	3658	AATCCATACCTTCCGGTGAATCAAACCGCTATGGACGGTTCTATGCAGGTATATTTTCC
GD33_BOLC.VG2.B	3657	AATCCATACCTTCCGGTGAATCAAACCGCTATGGACGGTTCTATGCAGGTATATTTTCC
A12_BOLC.VG2.A	3811	TTTGATCTATCTTATCTCCAACAGCTTATATTATTCTTACACTGAATGCGTTTGTATTA
GD33_BOLC.VG2.A	3851	TTTGATCTATCTTATCTCCAACAGCTTATATTATTCTTACACTGAATGCGTTTGTATTA
A12_BOLC.VG2.B	3718	TCTT-TCTATCTTATCTCCAACAGCTTATCAAATTCTTACATTGAATGCGTT-----

GD33_BOLC.VG2.B 3717 TCTT-TCTAGTCT-ATCTCCAACAGCTTAGACAATTCCTTACATTGAATGCGTT-----

A12_BOLC.VG2.A 3871 CGTTACTGATCGTCTTTGTCCATCACTGCTTCAGCCTGCTTTGGGTGCTGATGGAAAGA
GD33_BOLC.VG2.A 3911 CGTTACTGATCGTCTTTGTCCATCACTGCTTCAGCCTGCTTTGGGTGCTGATGGAAAGA
A12_BOLC.VG2.B 3769 ----ACTGATTGTCTTTGAC-ATCATTGCTTCAGCCTGCTTTGGGTGCTGATGGAAAGA
GD33_BOLC.VG2.B 3768 ----ACTGATTGTCTTTGAC-ATCATTGCTTCAGCCTGCTTTGGGTGCTGATGGAAAGA

A12_BOLC.VG2.A 3931 GGGTTGAAACTCAGAGCAACGTCCTCTTGCTTTGATTGAGAATATGCAGTACGCTGTCA
GD33_BOLC.VG2.A 3971 GGGTTGAAACTCAGAGCAACGTCCTCTTGCTTTGATTGAGAATATGCAGTACGCTGTCA
A12_BOLC.VG2.B 3824 GGGTAGAAACTCAGAGCAACGTCCTCTTGCTTTGATTGAGAATATGCAATACGCTGTCA
GD33_BOLC.VG2.B 3823 GGGTAGAAACTCAGAGCAACGTCCTCTTGCTTTGATTGAGAATATGCAATACGCTGTCA

A12_BOLC.VG2.A 3991 CCGTGGATGTTCTTCACACGGTGAGGAACAACAT-ATGCCCTGTTTCTTCATCTCATCAT
GD33_BOLC.VG2.A 4031 CCGTGGATGTTCTTCACACGGTGAGGAACAACAT-ATGCCCTGTTTCTTCATCTCATCAT
A12_BOLC.VG2.B 3884 C-----ACGGTGAGGAACAACAT-ATGCACGTTTCTGCATCTCAGCAT
GD33_BOLC.VG2.B 3883 C-----ACGGTGAGGAACAACAT-ATGCACGTTTCTGCATCTCAGCAT

A12_BOLC.VG2.A 4050 GCTTTCATATCTCAAAATTAACATAT-CCCTTGCTTCTCATAGGTGTTTTCCGCTTATGG
GD33_BOLC.VG2.A 4090 GCTTTCATATCTCAAAATTAACATAT-CCCTTGCTTCTCATAGGTGTTTTCCGCTTATGG
A12_BOLC.VG2.B 3927 CTTTTCATTTCTCAAAATTAATATTTCCCTTGCTTCTCATAGGTGTTTTCCGCTTATGG
GD33_BOLC.VG2.B 3926 CTTTTCATTTCTCAAAATTAATATTTCCCTTGCTTCTCATAGGTGTTTTCCGCTTATGG

A12_BOLC.VG2.A 4109 AACTGTGCAGAAGATTGCAATATTTGAGAAAATGCTTCAACGCAAGCCTTAATTCAATA
GD33_BOLC.VG2.A 4149 AACTGTGCAGAAGATTGCAATATTTGAGAAAATGCTTCAACGCAAGCCTTAATTCAATA
A12_BOLC.VG2.B 3987 AACTGTGCAGAAGATTGCAATATTTGAGAAAATGCTTCAACGCAAGCCTTAATTCAATA
GD33_BOLC.VG2.B 3986 AACTGTGCAGAAGATTGCAATATTTGAGAAAATGCTTCAACGCAAGCCTTAATTCAATA

A12_BOLC.VG2.A 4169 CTCTGGTACATGACCTTGATATCTGAATACATATATGATTCTACTATACTTCTTTTGGT
GD33_BOLC.VG2.A 4209 CTCTGGTACATGACCTTGATATCTGAATACATATATGATTCTACTATACTTCTTTTGGT
A12_BOLC.VG2.B 4047 CTCTGGTACATGACCTTGATGATCTGAATACATATATTTATTCTACTATACTTCTTTTGGT
GD33_BOLC.VG2.B 4046 CTCTGGTACATGACCTTGATGATCTGAATACATATATTTATTCTACTATACTTCTTTTGGT

A12_BOLC.VG2.A 4229 TATATGGTGAATCTTCTTATGTGTATGCAAGAGAAGAGCCTAAGAATTTGAATATAGTC
GD33_BOLC.VG2.A 4269 TATATGGTGAATCTTCTTATGTGTATGCAAGAGAAGAGCCTAAGAATTTGAATATAGTC
A12_BOLC.VG2.B 4107 TATATGGCGAATCTTCTTATGTGTATGCT-----TAAGAATTTGAAT--AGTC
GD33_BOLC.VG2.B 4106 TATATGGCGAATCTTCTTATGTGTATGCT-----TAAGAATTTGAAT--AGTC

A12_BOLC.VG2.A 4289 ATCTTTTGCTTCATGTTCCGTTTCTTATGATTCTTCTGTAGACATTTCAACGGCGACATG
GD33_BOLC.VG2.A 4329 ATCTTTTGCTTCATGTTCCGTTTCTTATGATTCTTCTGTAGACATTTCAACGGCGACATG
A12_BOLC.VG2.B 4152 ATCTTTTGCTTCATGTTCCGTTTCTTATGATTCTTCTGTAGACATTTCAACGGCGACATG
GD33_BOLC.VG2.B 4151 ATCTTTTGCTTCATGTTCCGTTTCTTATGATTCTTCTGTAGACATTTCAACGGCGACATG

A12_BOLC.VG2.A 4349 GCGAAAGAAGCACTGGAGGGACACTGCATATATGACGG-AGGCTACTGTAAGCTTCGACT
GD33_BOLC.VG2.A 4389 GCGAAAGAAGCACTGGAGGGACACTGCATATATGACGG-AGGCTACTGTAAGCTTCGACT
A12_BOLC.VG2.B 4212 GCGAAAGAAGCACTGGAGGGACACTGCATATATGACGGAGGCTACTGTAAGCTTCGACT
GD33_BOLC.VG2.B 4211 GCGAAAGAAGCACTGGAGGGACACTGCATATATGACGGAGGCTACTGTAAGCTTCGACT

A12_BOLC.VG2.A 4408 AACTTACTCTCGGCATACTGATCTCAATGTAAAGGTACATAAGATCAGTTGCTTT-----
GD33_BOLC.VG2.A 4448 AACTTACTCTCGGCATACTGATCTCAATGTAAAGGTACATAAGATCAGTTGCTTT-----
A12_BOLC.VG2.B 4272 AACATACTC--GTCATACTGATCTCAATGTAAAGGTACATAAGATCAGTTGCTTTCTTGA
GD33_BOLC.VG2.B 4271 AACATACTC--GTCATACTGATCTCAATGTAAAGGTACATAAGATCAGTTGCTTTCTTGA

A12_BOLC.VG2.A 4463 ---TGTACCCAAACAAAAA-GTTCGTTAAAGAGAATAAACAATTTCTGCAAGAAG
GD33_BOLC.VG2.A 4503 ---TGTACCCAAACAAAAA-GTTCGTTAAAGAGAATAAACAATTTCTGCAAGAAG
A12_BOLC.VG2.B 4330 TTGTGTACCCAAACAAAAAGCTTGGGAAAGAGAATAAACAATTTCTGCTGCAAGAAG

GD33_BOLC.VG2.B 4329 TT TGTGCACCCAAACAAAAAGCTTGGGAAAGAGAATAATAACTTATGCTGCAAGAAC

A12_BOLC.VG2.A 4519 CGCATGCT----TTTTTCTCATCAACAATTG--AGTGAAATCTGAATTCGACATTCTAGT

GD33_BOLC.VG2.A 4559 CGCATGCT----TTTTTCTCATCAACAATTG--AGTGAAATCTGAATTCGACATTCTAGT

A12_BOLC.VG2.B 4390 CTATGCTATTTT TTTTCTCTTCAACAATTG--T TGAATCTGAATTCGACATTCTAGT

GD33_BOLC.VG2.B 4389 CTATGCTATTTT TTTTCTCTTCAACAATTG--T TGAATCTGAATTCGACATTCTAGT

A12_BOLC.VG2.A 4573 TTATTCAACTAATGATTTT--CGTCTTCTTTCTTCAGGCATTTAGCGACAAAAGCAGAGAC

GD33_BOLC.VG2.A 4613 TTATTCAACTAATGATTTT--CGTCTTCTTTCTTCAGGCATTTAGCGACAAAAGCAGAGAC

A12_BOLC.VG2.B 4443 TTATTCAACTAATGATTTT--CGTCTTCTTTCTTCAGGCATTTAGTGACAAAAGCAGAGAC

GD33_BOLC.VG2.B 4447 TTATTCAACTAATGATTTT--CGTCTTCTTTCTTCAGGCATTTAGCGACAAAAGCAGAGAC

A12_BOLC.VG2.A 4632 TACACACTGCCTGATCTAAGCCAACCTGCTGGGCCAAAAGGTTCCAGGAGTGGGTGCAGCT

GD33_BOLC.VG2.A 4672 TACACACTGCCTGATCTAAGCCAACCTGCTGGGCCAAAAGGTTCCAGGAGTGGGTGCAGCT

A12_BOLC.VG2.B 4508 TACACACTGCCTGATCTAAGCCAACCTGCTGGGCCAAAAGGTTCCAGGAGTGGGTGCCTGCT

GD33_BOLC.VG2.B 4507 TACACACTGCCTGATCTAAGCCAACCTGCTGGGCCAAAAGGTTCCAGGAGTGGGTGCCTGCT

A12_BOLC.VG2.A 4692 AGTGGGCCAACAGATGGTTGGCCCAATGGGCAGGTGCAGACTCAATACATGG-----

GD33_BOLC.VG2.A 4732 AGTGGGCCAACAGATGGTTGGCCCAATGGGCAGGTGCAGACTCAATACATGG-----

A12_BOLC.VG2.B 4568 AGTGGGCCAACAGATGGTTGGCCCAATGGGCAGGTGCAGACTCAATACATGG-----

GD33_BOLC.VG2.B 4567 AGTGGGCCAACAGATGGTTGGCCCAATGGGCAGGTGCAGACTCAATACATGG-----

A12_BOLC.VG2.A 4745 --AAGTTCTATATGTACCCACC--ACTGATCCACAGGAGCTTCACCTTCTCTGGTC

GD33_BOLC.VG2.A 4785 --AAGTTCTATATGTACCCACC--ACTGATCCACAGGAGCTTCACCTTCTCTGGTC

A12_BOLC.VG2.B 4621 --AAGTTCTATATGTACCCACTGTAGCAAGGTA--GATTTTGATAAAGCGCGTT

GD33_BOLC.VG2.B 4620 --AAGTTCTATATGTACCCACTGTAGCAAGGTA--GATTTTGATAAAGCGCGTT

A12_BOLC.VG2.A 4801 -ATGCTC---CTTATTATGTTGATCCATTA--TACTGT---TTTCTCTTTATATAA-ATTT

GD33_BOLC.VG2.A 4841 -ATGCTC---CTTATTATGTTGATCCATTA--TACTGT---TTTCTCTTTATATAA-ATTT

A12_BOLC.VG2.B 4679 TATTTTAAATTTTTTTCATTGACAAATTTAGTAAATGCTATTTTCATAT-ATTT

GD33_BOLC.VG2.B 4678 TATTTTAAATTTTTTTCATTGACAAATTTAGTAAATGCTATTTTCATAT-ATTT

A12_BOLC.VG2.A 4853 CGGATTATTTTGAAGACTAGTTTCCATGTTATAGGTTTCAAGTAAAGTGTGTTCTT

GD33_BOLC.VG2.A 4893 CGGATTATTTTGAAGACTAGTTTCCATGTTATAGGTTTCAAGTAAAGTGTGTTCTT

A12_BOLC.VG2.B 4733 -GTGTTTTTTTATAAAAGCTTAAACTTTTTTTCGTTATTAT---CGTATT-TTATT

GD33_BOLC.VG2.B 4737 -GTGTTTTTTTATAAAAGCTTAAACTTTTTTTCGTTATTAT---CGTATT-TTATT

A12_BOLC.VG2.A 4913 TTATTTCGTTAAATAACA---TGGGAAAACCAATAATATTATAGTACTTTAATTTTA

GD33_BOLC.VG2.A 4953 TTATTTCGTTAAATAACA---TGGGAAAACCAATAATATTATTTTATTTATTTAT

A12_BOLC.VG2.B 4793 TTAAATGCTATTATTTT---TTTAAATTTAAACTTTATTTCGTTAATGAATTAATTG

GD33_BOLC.VG2.B 4792 TTAAATGCTATTATTTT---TTTAAATTTAAACTTTATTTCGTTAATGAATTAATTG

A12_BOLC.VG2.A 4970 ACATTGTATTTTTCTATTTCAGTTTGTGTGAGCTTCCAAAGTTTAAACCTT--

GD33_BOLC.VG2.A 4850 ATATAACTCTGATAAATTAATTTTATTATGTGGTTAATATTTTATTAAATAAATTATA

A12_BOLC.VG2.B 4849 ATATAACTCTGATAAATTAATTTTATTATGTGGTTAATATTTTATTAAATAAATTATA

GD33_BOLC.VG2.B 5023 CTTTTTCTCTCTTTTAACTGTTTCATTCCTTATTTTGTGAAGATCTGCAGCTT

GD33_BOLC.VG2.A 4910 TACTTTTATAAGATTTTATTTTTCATTAATTTTATGAATCTT

A12_BOLC.VG2.B 4909 TACTTTTATAAGATTTTATTTTTCATTAATTTTATGAATCTT

GD33_BOLC.VG2.B 5086 ACAGTTTAGTACTTTTGTATCC

GD33_BOLC.VG2.A 4970 AAAATTATATTAGAAAAGAAAGA

A12_BOLC.VG2.B 4969 AAAATTATATTAGAAAAGAAAGA

Alineación de secuencias de proteínas para proteínas ortólogas de Brassica oleracea VG2 a partir de GD33 y A12. Las secuencias de las proteínas se predijeron utilizando el programa bioinformático localizado en la red FGENESH (<http://mendel.cs.rhul.ac.uk/mendel.php?topic=faen-file>).

5

A12_BOLC.VG2.A 1 --MSSQFRYTQTPSKVVHLNLPWECVEEELIDLCKRFGKIVNTKTNVGANRNQAFVE

GD33_BOLC.VG2.A 1 --MSSQFRYTQTPSKVVHLNLPWECVEEELIDLCKRFGKIVNTKTNVGANRNQAFVE

A12_BOLC.VG2.B 1 --MASSQFRYTQTPSKVVHLNLPWECVEEELIDLCKRFGKIVNTKTNVGANRNQAFVE

GD33_BOLC.VG2.B 1 --MASSQFRYTQTPSKVVHLNLPWECVEEELIDLCKRFGKIVNTKTNVGANRNQAFVE

A12_BOLC.VG2.A 53 FGDVNQAI SMVSYASSSEPAQIRGKT VYIQYSNRHEIVNNQSFGEVPGNVLLVTFEGVQ

GD33_BOLC.VG2.A 53 FGDVNQAI SMVSYASSSEPAQIRGKT VYIQYSNRHEIVNNQSFGEVPGNVLLVTFEGVQ

A12_BOLC.VG2.B 59 FGDVNQAI SMVSYASSSEPAQIRGKT VYIQYSNRHEIVNNQSFGEVPGNVLLVTFEGVQ

GD33_BOLC.VG2.B 59 FGDVNQAI SMVSYASSSEPAQIRGKT VYIQYSNRHEIVNNQSFGEVPGNVLLVTFEGVQ

ES 2 989 717 T3

A12_BOLC.VG2.A	118	SHHVCIDVIHLVFSAYGFVHKIATFEKAAGFQALVQFTD VDTALAARTALDGRSIPKYL
GD33_BOLC.VG2.A	118	SHHVCIDVIHLVFSAYGFVHKIATFEKAAGFQALVQFTD VDTALAARTALDGRSIPKYL
A12_BOLC.VG2.B	119	SHHVCIDVIHLVFSAYGFVHKIATFEKAAGFQALVQFTD VDTALAARTALDGRSIPKYL
GD33_BOLC.VG2.B	119	SHHVCIDVIHLVFSAYGFVHKIATFEKAAGFQALVQFTD VDTALAARTALDGRSIPKYL
A12_BOLC.VG2.A	173	PEHVGSCNLRMSYSAHTDLNIFQSHRSRDYTNPYLFPVNQTAMDGSMQPALGADGKVE
GD33_BOLC.VG2.A	173	PEHVGSCNLRMSYSAHTDLNIFQSHRSRDYTNPYLFPVNQTAMDGSMQPALGADGKVE
A12_BOLC.VG2.B	179	PEHVGSCNLRMSYSAHTDLNIFQSHRSRDYTNPYLFPVNQTAMDGSMQ-----
GD33_BOLC.VG2.B	179	PEHVGSCNLRMSYSAHTDLNIFQSHRSRDYTNPYLFPVNQTAMDGSMQ-----
A12_BOLC.VG2.A	233	QSNVLLALIENMQYAVTVDLHTVFSAYGTQKIAIFEKNGSTQALIQYSDISTATMAKE
GD33_BOLC.VG2.A	233	QSNVLLALIENMQYAVTVDLHTVFSAYGTQKIAIFEKNGSTQALIQYSDISTATMAKE
A12_BOLC.VG2.B	227	-----VFSAYGTQKIAIFEKNGSTQALIQYSDIPTATAKE
GD33_BOLC.VG2.B	227	-----VFSAYGTQKIAIFEKNGSTQALIQYSDIPTATAKE
A12_BOLC.VG2.A	293	ALEGHCIYDGGYCKLRLYSRHTDLNVKAFSDKSRDYTLPDLSQLVGQKVPGVAAASGPT
G GD33_BOLC.VG2.A	298	ALEGHCIYDGGYCKLRLYSRHTDLNVKAFSDKSRDYTLPDLSQLVGQKVPGVAAASGPT
A12_BOLC.VG2.B	264	ALEGHCIYDC-----RLLDKSGDYTLPDLSQLVGQKVPGVAAASGPT
GD33_BOLC.VG2.B	264	ALEGHCIYDC-----RLLDKSGDYTLPDLSQLVGQKVPGVAAASGPT
A12_BOLC.VG2.A	353	DGWP--NGQVQTQYMGSSYMYEPADPTGASPSGHPPIYG
GD33_BOLC.VG2.A	353	DGWP--NGQVQTQYMGSSYMYEPADPTGASPSGHPPIYG
A12_BOLC.VG2.B	306	DCWH--NGQVQTQYTRSSYMYEL-----
GD33_BOLC.VG2.B	306	DCWH--NGQVQTQYTRSSYMYEL-----

Alineación de secuencias de ADN para los genes de Brassica oleracea VG1

A12 1 AAAAAGCAGGCTAAGAGTGTAAGCGACCCATTTACATCGGATCTAAAAATCATATAATC
GD33 1 AAAAAGCAGGCTAAGAGTGTAAGCGACCCATTTACATCGGATCTAAAAATCAATAATC

A12 61 TAGAGTCTGACAAGAATCAACGTTGATAGAGAGTCCGAGAGAATAGTACAGAGCCAGTCG
GD33 61 TAGAGTCTGACAAGAATCAACGTTGATAGAGAGTCCGAGAGAATAGTACAGAGCCAGTCG

A12 121 TCGAGGTAGAAATGTTTGTCTGAGCCCAAATGGTATAGTTGATATTGGGATTCCAGAAC
GD33 121 TCGAGGTAGAAATGTTTGTCTGAGCCCAAATGGTATAGTTGATATTGGGATTCCAGAAC

A12 181 TTGTTATCACCCTATCCATCTCTTAGCGGCCACCTCAGTCACCGGAGCTGCCGCAAGA
GD33 181 TTGTTATCACCCTATCCATCTCTTAGCGGCCACCTCAGTCACCGGAGCTGCCGCAAGA

A12 241 AAAGCCAGAACCATTGCGCGCAGCCATCAACGCAAGTCTAGCCATAGATCACTCTAGAAAG
GD33 241 AAAGCCAGAACCATTGCGCGCAGCCATCAACGCAAGTCTAGCCATAGATCACTCTAGAAAG

A12 301 AAAAAGAAGCTCAGTGCTTATGACTCTTAGTCTTATCTCTCAAGAAATAGCAGTTTTAT
GD33 301 AAAAAGAAGCTCAGTGCTTATGACTCTTAGTCTTATCTCTCAAGAAATAGCAGTTTTAT

A12 361 GATGAATGTATATGAGCGAGTGGGGCTGTTTCCAAGTCAAATCAGACATACAAATTTCT
GD33 361 GATGA-----GTGGGGCTGTTTCCAAGTCAAATCAGACATACAAATTTCT

A12 421 CGTTTTACGGGGAAAAGATCTCTCAAGATCCGTAATTTCTTCAACTAACGTTTGAATTT
GD33 407 CGTTTTACGGGGAAAAGATCTCTCAAGATCCGTAATTTCTTCAACTAACGTTTGAATTT

A12 481 TATTACTACTTTAATCAACCTCGATGTTGACGATGACGTACACGTTAATTAAAGGTACGC
GD33 467 TATTACTACTTTAATCAACCTCGATGTTGACGATGACGTACACGTTAATTAAAGGTACGC

A12 541 GTTATATAGTTAAAGGAGTAATTAGATTTTAAATGCTGGCATTATTATTATTTTCTCTG
GD33 527 GTTATATAGTTAAAGGAGTAATTAGATTTTAAATGCTGGCATTATTATTATTTTCTCTG

A12 601 GT----ATATT-----TTTAAAGGAATATTTTCTACCTTTTAGGCGGGAATTTATT
GD33 587 GTTATTACATTATTCTGTTTTTAGTAATATTTTCTACCTTTTAGGCGGGAATTTATT

A12 650 CCATTTTCTCTTTCAAGGGTTCAGAGATTCAATTCCTTTATTAAACATACGAAATGAGTA
GD33 647 CCATTTTCTCTTTCAAGGGTTCAGAGATTCAATTCCTTTATTAAACATACGAAATGAGTA

A12 710 TATAGTATCTTTTAAATTTAAATAATTAAAATAAAAATAAATTACCATACAGTGATA
GD33 707 TATAGTATCTTTTAAATTTAAATAATTAAAATAAAAATAAATTACTATAACAGTGATA

A12 770 CATATCAAATTTCCAAAACCTAATGTGAGTACTTATATTAGTTGTACTTCTTAGTAAAAT
GD33 767 CATATCAAATTTCCAAAACCTAATGTGAGTACTTATATTAGTTGTACTTCTTAGTAAAAT

ES 2 989 717 T3

A12 830 AACGAGATATGTTATGGCTCGTCGCATACA CAAAGCAGAGCATATGTTCCCGCAATAAC
GD33 827 AACGAGATATGTTATGGCTCGTCGCATACA CAAAGCAGAGCATATGTTCCCGCAATAAC

A12 890 ATGCACACGCAACTGATAGCACGTCGTAACCATTTCGCGAACCAACCAATGGATATTGT
GD33 887 ATGCACACGCAACTGATAGCACGTCGTAACCATTTCGCGAACCAACCAATGGATATTGT

A12 950 AATTCGATAAGACGACGTCGTGTTGTTGACCACGAATATTTCTTAAATTTCTTTTATCT
GD33 947 AATTCGATAAGACGACGTCGTGTTGTTGACCACGAATATTTCTTAAATTTCTTTTATCT

A12 1010 TTTGTTTTTGACGTCATC ACATTGGTTTATGTCCACACGATATCAGTCATAAAATCCACG
GD33 1007 TTTGTTTTTGACGTCATC ACATTGGTTTATGTCCACACGATATCAGTCATAAAATCCACG

A12 1070 TAGGGAACGACTATCCATTACACCCATCCACGTTCTCTATAACAGCTTCTAGTTCCATT
GD33 1067 TAGG AACGACTATCCATTACACCCATCCACGTTCTCTATAACAGCTTCTAGTTCCATT

A12 1130 AGCTGAACGAACAATAGAAACAGAGCATCGAACTGAGAAAGAAGAAGAAGGAAGCA
GD33 1126 AGCTGAACGAACAATAGAAACAGAGCATCGAACTGAGAAAGAAGAAGAAGGAAGCA

A12 1190 AGAGAAATGGCGGCAGCGGCCATGGCCGTTTCATCTCCCGAAACATTCATCGTTTCTTCT
GD33 1186 AGAGAAATGGCGGCAGCGGCCATGGCCGTTTCATCTCCCGAAACATTCATCGTTTCTTCT

A12 1250 AATCCCAAGCTTCCATTG ATCAAAACTCTAACTTCTCGGTGTGTCCTTAAAGATCGGT
GD33 1246 AATCCCAAGCTTCCATTG ATCAAAACTCTAACTTCTCGGTGGTTCCTTAAAGATCGGT

A12 1310 CGATCCCATGTCCTT AACCGGAAAATGAAAGGTCCGTTACGGTTTCAGCTGCTTCGACG
GD33 1306 CGATCCCATGTCCTT AACCGGAAAATGAAAGGTCCGTTACGGTTTCAGCTGCTTCGACG

A12 1370 GTCGAAGGCGATCGAAGCAAACAGTTTTACATAAACTTCACTGGATTCCCATTTCCTCTT
GD33 1366 GTCGAAGGCGATCGAAGCAAACAGTTTTACATAAACTTCACTGGATTCCCATTTCCTCTT

A12 1430 GGTCTTTTCTTAAACGGCGCACCATCAGAACCGAGGTAGGCTTTCTCCATCCCTCTTG
GD33 1426 GGTCTTTTCTTAAACGGCGCACCATCAGAACCGAGGTAGGCTTTCTCCATCCCTCTTG

A12 1490 AGTTTTGATTTGAACTCGTTGAGAATCCCATATGGAT---GTTATGCAGGCGGTTAAAGG
GD33 1486 AGTTTTGATTTGAACTCGTTGAGAATCCCATATGGATAATGTTATGCAGGCGGTTAAAGG

A12 1547 AAGCATATGGATGTTTGAACAAGAACAAGCTTTAGGTTTCAGCAGTGTCTCTACCAATAT
GD33 1546 AAGCATATGGATGTTTGAACAAGAACAAGCTTTAGGTTTCAGCAGTGTCTCTACCAATAT

A12 1607 AAGAATGACTGTCATCACTCAAATCCGGTGGCTTATGGGTTTCATGCCCTATTGCTCC
GD33 1606 AAGAATGACTGTCATCACTCAAATCCGGTGGCTTATGGGTTTCATGCCCTATTGCTCC

ES 2 989 717 T3

A12 1667 CACCAAAGAGTGTATTGAGTTCCTTTTATTTCATTATCCTTGATAAAGTATCATCC
GD33 1666 CACCAAAGAGTGTATTGAGTTCCTTTTATTTCATTATC-----AGTATCATCC

A12 1727 TTTCTTATCATTTGATAA--TAACCAATGTCTAAACTCTTTTGGGTGTTTTGCAAAAC
GD33 1717 TTTCTTATCATTTGATAAATTAACCAATGTCTAAACTCTTTTGGGTGTTTTGCTTAC

A12 1785 TTTTCTTTGAGCAGCTTATTGAGGAGTTGGGAGCTCCGGTTGAGTACATTGTCCTCC
GD33 1777 TTTTCTTTGAGCAGCTTATTGAGGAGTTGGGAGCTCCGGTTGAGTACATCGTCCTCC

A12 1845 AACCTTTGCTTACGAGCACAAGATCTTCGTCGGTCCCTTCTCTAGAAAGTCCCCAAGGC
GD33 1837 AACCTTCGCTTACGAGCACAAGATCTTCGTCGGTCCCTTCTCTAGAAAGTCCCCAAGGC

A12 1905 TCAAGTATGGGTGGCGCCAAGACAATGGAGCTGGCCACTGAACTTACCACTCGAGTTTTT
GD33 1897 TCAAGTATGGGTGGCGCCAAGACAATGGAGCTGGCCACTGAACTTACCACTCGAGTTTTT

A12 1965 CGGTATCTTTTCGCGCTAAACCATTAAGACGGGACTTTCTACCCCGTGGGCTGATGA
GD33 1957 CGGTATCTTTTCGCGCTAAACCATTAAGACGGGACTTTCTACCCCGTGGGCTGATGA

A12 2025 GATCGAGCAGAAAGTCTTAAGCTCTCCTGAAGTCGGTACGTTCTTCACTTTCTCAATCT
GD33 2017 GATCGAGCAGAAAGTCTTAAGCTCTCCTGAAGTCGGTACGTTCTTCACTTTCTCAATCT

A12 2085 TGTGGGCCCTCCTTAGCCAAGCCTCAGTAAACATTTCGCTAAACCCGCACTCTCTCTCT
GD33 2077 TGTGGGCCCTCCTTAGCCAAGCCTCA-TAAACATTTCGCTCAACCC---CTCACC---

A12 2145 CTTCCCTCCAAATATTTGAAAAAGTTATATATACATAGCCTCCCAAGTTATTAAAAAA
GD33 2129 ---CCACCCAAATTTTTGAAAAAGTTATATATACATAGCC-CCAAATTTATTAAAAAA

A12 2205 AACGTTTTGTCTCCCAAACCTATTGAAATCTTTATCAAATTCTCTATCCCTCCAAATCC
GD33 2185 ATC-----TATTGAAATCTTTATTAAATTCTCTATCCCTCAATCT

A12 2265 GAGGGTCGGTCCTGTACTTATCAATAA-TAATATACAGGAATAGGACCGTATGTGGAAG
GD33 2229 GAGGGTCGGTCCTGTACTTATCAATAAATAATATACAGGAATTGGACCGTATGTGGAAG

A12 2324 TAGCGTTCTACCATAAGCGTTCAAGAACTCTATTAGTCACTGATGCTGTGATCTTCGTCC
GD33 2289 TAGCGTTCTACCATAAGCGTTCAAGAACTCTATTAGTCACTGACGCTGTGTGATCTTCGTCC

A12 2384 CACGAAGCCGCCATCGAGTATCAGCAGCGAGTCTTGCTGGCTCTGCTAAGAACGGAC
GD33 2349 CAAGAAGCCGCCATCGAGTATCAGCAGCGAGTCTTGCTGGCTCTGCTAAGAACGGAC

A12 2444 TGGCTGTGAAGATACTTAGCAAAGGCAAACAAGTACCTATGACCTGTCGTTGATACCC
GD33 2409 TGGCTGTGAAGATACTTAGCAAAGGCAAACAAGTACCTATGACCTGTCGTTGATACCC

A12 2504 CAAACACCCGCCAAAAAGGTCAGCCATCAGCTTCTTGAAATTCAAATTCATCAAACAAA
GD33 2469 CAAACACCCGCCAAAAAGGTCAGCCATCAGCTTCTTGAAATTCAAATTCATCAAACAAA

A12 2564 GATCTCGATGGGTGTCTTTTAAAGTGCAGGATGGGAAAGAATGGTGCTGCAAAATCCTGTT
GD33 2528 GATCTCGATGGGTGTCTTTTAAAGTGCAGGATGGGAAAGAATGGTGCTGCAAAATCCTGTT

A12 2624 TCTTGGTCCCTCTAATCTCTTGGAGCCAAACGCGAGCTTCCGCAAAATGTCACAGAAGCT
GD33 2588 TCTCGGCCCCTCTAATCTCTTGGAGCCAAACGCGAGCTTCCGCAAAATGTCACAGAAGCT

A12 2684 GATCGTTTCTCCCATTTGTCAAGACTCTGGTCTTTAGCAAAGTCCCTGAGAAGGTGAGGGA
GD33 2648 GATCGTTTCTCCCATTTGTCAAGACTCTGGTCTTTAGCAAAGTCCCTGAGAAGGTGAGGGA

A12 2744 CTGGATCGATGAGATAGCGAGTGACTGGAGTTCAAGAGGATAATCCAGCTCATTTTCCA
GD33 2708 CTGGATCGATGAGATAGCGAGTGACTGGAGTTCAAGAGGATAATCCAGCTCATTTTCCA

A12 2804 GGCTCCGCTAAACGCGGGAGGTGAGATTTCTAGCTGCGTTTGGGTTTCTTGATGATCT
GD33 2768 GGCTCCGCTAAACGCGGGAGGTGAGATTTCTAGCTGCGTTTGGGTTTCTTGATGATCT

A12 2864 TCTAGGGGAAAGATATGTGAACCGTCTCTCTGCTCTCTGTTCTCTTCACTTCGCTGAT
GD33 2828 TCTAGGGGAAAGATATGTGAACCGTCTCTCTGCTCTCTGTTCTCTTCACTTCGCTGAT

A12 2924 GGGTAAAGCAGCGAGCTATTTTCCCTCCGGATGATATGAGACTCTCTCTTCTCTTGATCA
GD33 2888 GGGTAAAGCAGCGAGCTATTTTCCCTCCGGATGATATGAGACTCTCTCTTCTCTTGATCA

A12 2984 GTTCTTAGTCTCTGTTGGTGCTGTTAAGAAGACCGTCTCTGCTAGAAAACGAAGATGACG
GD33 2948 GTTCTTAGTCTCTGTTGGTGCTGTTAAGAAGACCGTCTCTGCTAGAAAACGAAGATGACG

A12 3044 GAACAAACCCAGCTTTCT
GD33 3008 GAACAAACCCAGCTTTCT

Alineación de secuencias de proteínas para proteínas ortólogas de Brassica oleracea VG1 a partir de A12 y GD33. Las secuencias de las proteínas se predijeron utilizando el programa bioinformático localizado en la red FGENESH.

5

A12 1 MAAAMAVHLPKHSSFLNPKLPINQNSNFLGVSLKGRPMSVNRKMRGPEVTVSAAST--
GD33 1 MAAAMAVHLPKHSSFLNPKLPESDQNSNFLGGSILKGRMSINRRTKGPVMVSAAST--

A12 59 -----VEGDRSKQFYINFTGFFPLGPFNLNRRRTIRTEAVKGSIWMFEQEALGFSSVST
GD33 59 -----VEGDRSKQFYINFTGFFPLGPFNLNRRRTIRTEAVKGSIWMFEQEALGFSSVST

A12 113 NIRMVILKSGGLWVHAPIAPTKECIQLIEELGAPVEYIVLPTFAYEHKIFVGPFSRKF
GD33 113 NIRMVILKSGGLWVHAPIAPTKECIQLIEELGAPVEYIVLPTFAYEHKIFVGPFSRKF

A12 173 PKAQVWVAPRQWSWPLNLPLEFFGIFRAKTIKDGLSTPWADIEQKVLSSPEVGIGPYV
GD33 173 PKAQVWVAPRQWSWPLNLPLEFFGIFRAKTIKDGLSTPWADIEQKVLSSPEVGIGPYV

A12 233 EVAFYHKRSRTLLVTDAVIFVPRKPSSISSESLASAKNGLAVKILSKGKQVPDDEVVD
GD33 233 EVAFYHKRSRTLLVTDAVIFVPRKPSSISSESLASAKNGLAVKILSKGKQVPNDPVVD

A12 293 TPNTRQKGWERMVLQILFLGPSNLLEPNASFAQMSQKLIVSPIVKTIVFSKVPEKVRDWI
GD33 293 TPNTRQKGWERMVLQILFLGPSNLLEPNASFAQMSQKLIVSPIVKTIVFSKVPEKVRDWI

A12 353 DEIASDWRFKRIIPAHFEAPINAGRSEFLAAFGLDDLLGERYVNRPPSLSVLFTSLMGK
GD33 353 DEIASDWRFKRIIPAHFEAPINAGRSEFLAAFGLDDLLGERYVNRPPSLSVLFTSLMGK

A12 413 AASYFFPDDMRTLSSLDQFLVSVGAVKKTIVSGRKRR
GD33 413 AASYFFPDDMRTLSSLDQFLVSVGAVKKTIVSGRKRR

REIVINDICACIONES

- 5 1. Una planta de Brassica híbrida cultivada o parte de planta de la misma que contiene dentro de su genoma una introgresión que comprende un polinucleótido, en particular un polinucleótido aislado, que comprende una molécula de ácido nucleico que comprende una secuencia de nucleótidos como se representa en la SEQ ID NO: 3 (GD33 versión de BolC.VG1.a); en la que dicha molécula de ácido nucleico, tras la expresión en la planta de Brassica cultivada o parte de planta de la misma, da lugar a un vigor aumentado de las semillas en comparación con una planta de Brassica cultivada o parte de planta de la misma que no comprende dicho polinucleótido.
- 10 2. Una planta de Brassica híbrida cultivada o parte de planta de la misma que contiene dentro de su genoma una introgresión que comprende un polinucleótido, en particular un polinucleótido aislado, que comprende una molécula de ácido nucleico seleccionada del grupo que consiste en:
 - a. molécula de ácido nucleico que comprende una secuencia de nucleótidos que tiene al menos un 95 % de identidad de secuencia con la secuencia representada en la SEQ ID NO:3;
 - 15 b. molécula de ácido nucleico que comprende una secuencia de nucleótidos que tiene al menos un 98 % de identidad de secuencia con la secuencia representada en la SEQ ID NO:3;en la que dicha molécula de ácido nucleico como se define en cualquiera de a) – b), tras la expresión en una planta de Brassica cultivada o parte de planta de la misma, da lugar a un vigor aumentado de las semillas en comparación con una planta de Brassica cultivada o parte de planta de la misma que no comprende dicho polinucleótido.
- 20 3. La planta de Brassica cultivada o parte de planta de la misma de acuerdo con cualquiera de las reivindicaciones precedentes, en la que el fenotipo de vigor aumentado de las semillas se caracteriza por un fenotipo adicional seleccionado del grupo que comprende: velocidad aumentada de germinación, velocidad aumentada de brote de plántulas, uniformidad aumentada de germinación de las semillas, uniformidad aumentada de brote de plántulas, porcentaje aumentado de germinación de las semillas, tolerancia aumentada de la semilla frente al entorno externo y/o condiciones maternas, sensibilidad reducida a ABA o contenido reducido de ABA.
- 25 4. Una planta de Brassica cultivada o parte de planta de la misma de acuerdo con cualquiera de las reivindicaciones precedentes que se selecciona del grupo que comprende *Brassica oleracea*, *Brassica napus*, *Brassica rapa*, *Brassica campestris*, *Brassica juncea*, *Brassica nigra*, *Brassica pekinensis*, *Brassica chinensis*, *Brassica rosularis*, *Eruca vesicaria*, *Eruca sativa*, *Raphanus sativus*, *Lepidium sativum*, *Nasturtium officinale*, *Wasabia japónica*.
- 30 5. Una semilla de Brassica cultivada producida por la planta de Brassica cultivada o parte de planta de la misma de acuerdo con cualquiera de las reivindicaciones precedentes, en la que la semilla de Brassica cultivada comprende el polinucleótido, en la que dicha semilla está recubierta con cualquier tipo de recubrimiento.
- 35

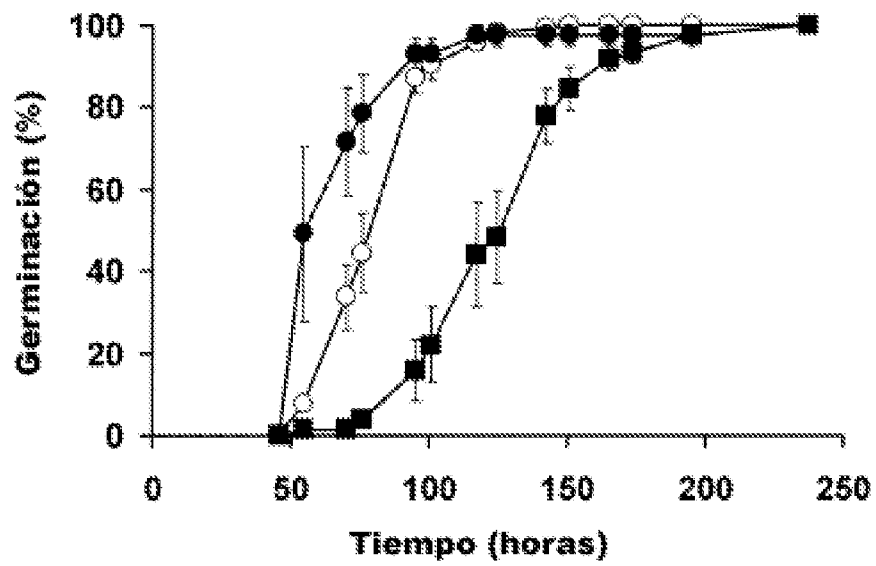


Fig. 1

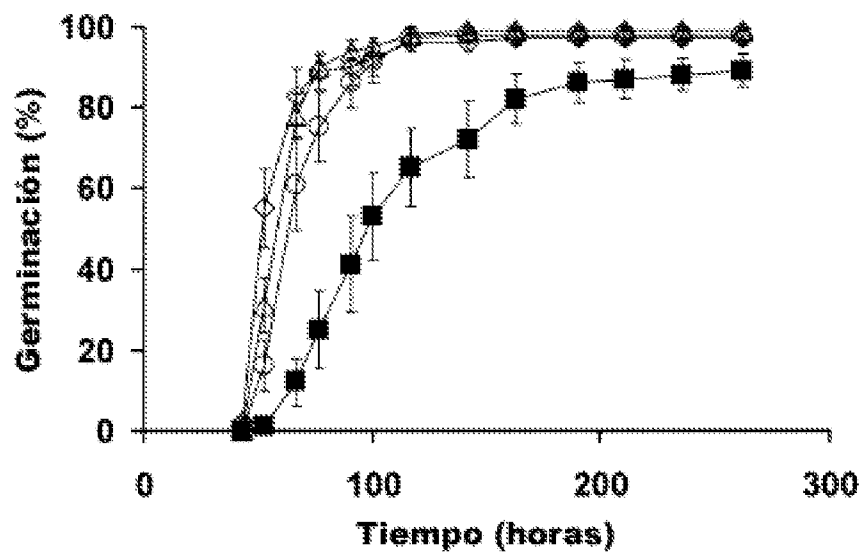


Fig. 2

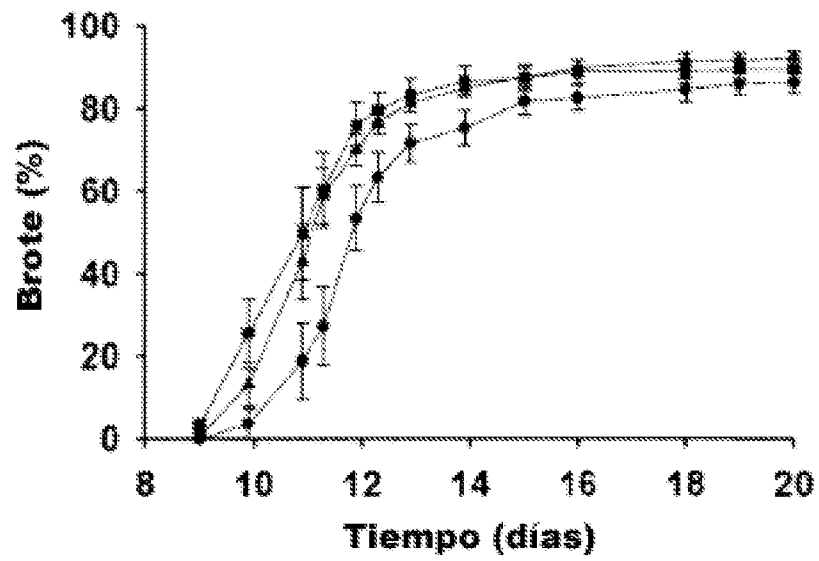


Fig. 3

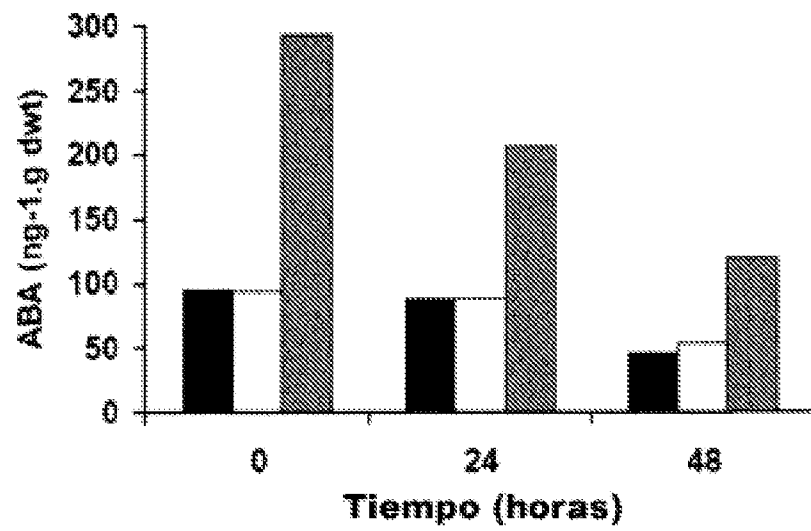


Fig. 4

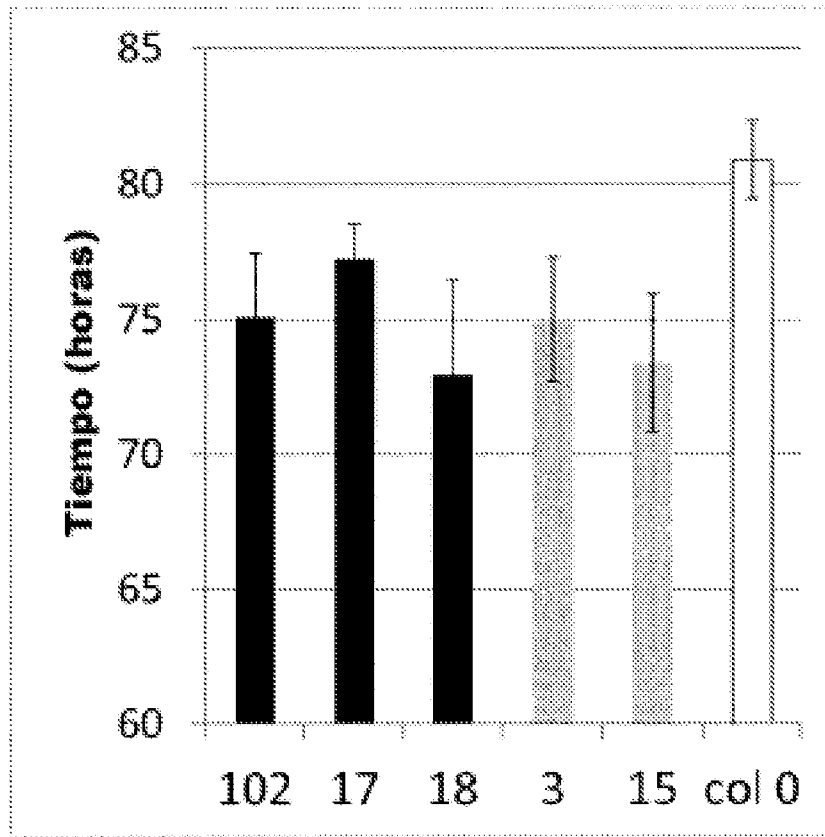


Fig. 5

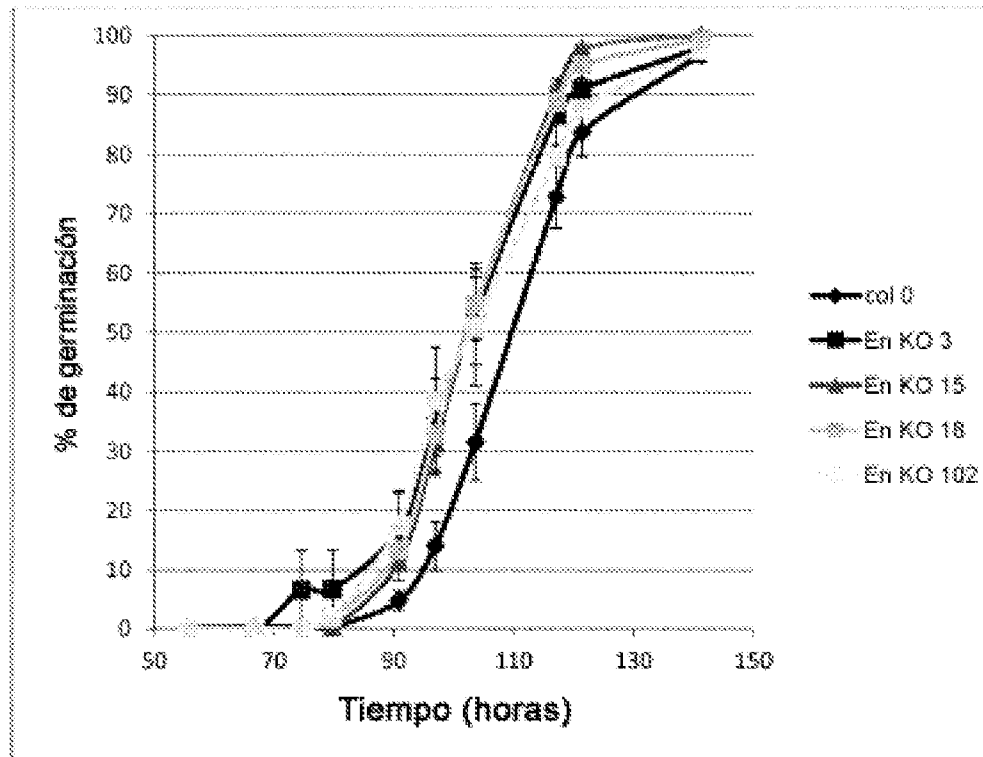


Fig. 6