

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 905 103**

51 Int. Cl.:

C07K 14/435 (2006.01)

C12N 15/81 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **12.11.2018 PCT/EP2018/080876**

87 Fecha y número de publicación internacional: **23.05.2019 WO19096718**

96 Fecha de presentación y número de la solicitud europea: **12.11.2018 E 18803387 (2)**

97 Fecha y número de publicación de la concesión europea: **08.12.2021 EP 3710469**

54 Título: **Polipéptidos con especificidad de transporte de arabinosa mejorada**

30 Prioridad:

14.11.2017 EP 17201608

28.11.2017 EP 17203961

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

07.04.2022

73 Titular/es:

DSM IP ASSETS B.V. (100.0%)

Het Overloon 1

6411 TE Heerlen, NL

72 Inventor/es:

**VAN MARIS, ANTONIUS JEROEN ADRIAAN;
PRONK, JACOBUS THOMAS y
VERHOEVEN, MAARTEN DIRK**

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 905 103 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

Polipéptidos con especificidad de transporte de arabinosa mejorada

Campo

5 La invención se refiere a polipéptidos variantes con especificidad de transporte de arabinosa mejorada y actividad de transporte de glucosa reducida, a una levadura recombinante que comprende una secuencia de ácido nucleico que codifica dichos polipéptidos, y a su uso para producir un producto de fermentación tal como etanol.

Antecedentes

10 Al reemplazar funcionalmente los compuestos derivados de combustibles fósiles, la producción microbiana de productos químicos y combustibles para el transporte puede contribuir a una transición hacia una economía mundial sostenible y con bajas emisiones de carbono. La producción industrial total de etanol combustible, que alcanzó aprox. 100 mil millones de litros en 2015, se prevé que aumente aún más. La levadura *Saccharomyces cerevisiae* es la fábrica de células microbianas establecida para la conversión de unidades de hexosa derivadas de almidón y sacarosa en etanol, ya que combina un alto rendimiento y productividad de etanol con robustez en las condiciones del procedimiento. Los esfuerzos en la mejora de la cepa de levadura y la optimización del procedimiento de producción de bioetanol a base de almidón de maíz y azúcar de caña han mejorado aún más el rendimiento y la productividad del producto. Además, estudios intensivos de ingeniería metabólica y evolutiva han producido cepas de levadura capaces de fermentar eficientemente los azúcares de pentosa xilosa y arabinosa, allanando así el camino para la producción de bioetanol de "segunda generación" a base de levadura a partir de hidrolizados lignocelulósicos.

20 El documento WO 2014/195520 se refiere a nuevos polipéptidos de permeasa con especificidad de azúcar alterada y/o actividad de transporte de azúcar, y describe un polipéptido que tiene una o más sustituciones correspondientes a F85T/A/C/H/M/N, T89H/D/C/A/N, V187T/C/D/E/F/G/H/K/M/Q/R/S/W/Y, A191N/G/P/C/D/F/H/I/K/Q/V/W/Y, Y214V/Q/G/L/M/N/P/R/S, Q215L/M/I/A/C/D/F/S, 1218K/N/A/C/ D/E/H/L, T219G/F/D/N, I222D/C/H/E/G/P/S/V/Y, Y226D/M/R/C/G/H/I/L/P/Q/T/W, Q338C/E/F/D/G/ I/P/R/S/T/V/W/Y, M339N/V/D/E/H/K/P/Q/R/Y, Q341T/M/N/S/Y, Q342Y/S/C/A/D/F/G/H/I/K/L/N/P /R/T/V/W, L343K/R/P/C/D/E/G/H/I/M/N/Q/S/T/Y, N346W/Y/F/G/H/K/L/M/R/T, N347A/I/E/H/L, Y350R/L/T/D/K/M/N/V, G373L/E/N/F/K/M/R/V/W/Y, N376I/T/M/A/C/E/F/H/K/L/P/Q/R/V/W/Y, T380M/F/H/N/P/Q/R/S/V, S383Q/T/C/D/E/F/I/K/L/M/R/V/W/Y, F444Y/D/E/H/K/N/P/Q/R/W, F446Q/L/H/D/E/I/M/N, T448M/H/D/E/F/K/R/W/Y, T449F/M/N/S/D/E/G/H/I/K/L/P/Q/R/V/W/Y, T451I/E/C/D/F/H/K/L/M/N/Q/R/S/T/V/W/Y, W455T/N/A/C/E/L/P/Q/T, N478P/F/V/G/I/K/R/W, o W479T/S/Q/D/E/F/G/H/I/M/N/P/R/V/Y de SEQ ID NO:59, en el que el polipéptido es un miembro de la Superfamilia de Facilitadores Principales (MFS).

30 Wang et al, "Identification of Important Amino Acids in Gal2p for improving the L-arabinose Transport and Metabolism in *Saccharomyces cerevisiae*", *Frontiers in Microbiology*, Vol. 8, (julio de 2017), páginas 1-11, estudian restos de aminoácidos cruciales para el transporte de L-arabinosa. Los restos de aminoácidos polares o aromáticos en diferentes secuencias transmembránicas (TMS) de Gal2p se escogieron para un estudio más en profundidad, incluyendo F85, T89, F223, N346, N347, F350, Y351, N376, y Y446. Wang et al concluyen que el sitio F85 es un resto clave para mejorar la actividad de transporte de L-arabinosa de Gal2p.

35 Se sabe que el transportador de hexosa Gal2 media el transporte de arabinosa (Becker J, Boles E. 2003, A modified *Saccharomyces cerevisiae* strain that consumes L-Arabinose and produces ethanol, *Appl Environ Microbiol.* 69: 4144-4150). Sin embargo, la afinidad por los respectivos azúcares de pentosa es aproximadamente 10 a 100 veces menor que por los respectivos azúcares de hexosa. La falta de un transportador dedicado de xilosa o arabinosa en las células de levadura recombinantes limita así la capacidad de cointilización de hexosas y pentosas en mezclas de azúcares, y prohíbe un alto flujo catabólico de pentosas. Como consecuencia, la conversión de azúcares de biomasa puede considerarse bifásica: en la primera fase, tiene lugar una conversión relativamente rápida de hexosas (glucosa), mientras que en la segunda fase, que comienza cuando las hexosas se han agotado del medio, comienza la fermentación de pentosas, pero a una velocidad mucho menor en comparación con la velocidad de conversión de hexosas.

Por lo tanto, es un deseo sentido desde hace mucho tiempo expresar transportadores de azúcares específicos de arabinosa, es decir, sin interferencia de glucosa (especificidad de arabinosa) y alta afinidad por la arabinosa, para mantener la capacidad de convertir hexosas aproximadamente al mismo nivel.

Breve descripción de los dibujos

50 Figura 1. Experimentos de absorción usando [¹⁴C]-L-arabinosa 50 mM en presencia de un intervalo de concentraciones de D-glucosa usando la cepa DS68625 que expresa variantes de Gal2: N376T+T89I (círculos en blanco), N376T (círculos en negro), T89I (cuadrados en blanco) y Gal2 de tipo salvaje (cuadrados en negro).

55 Figura 2. Experimentos de absorción usando [¹⁴C]-L-arabinosa 50 mM en presencia de un intervalo de concentraciones de D-glucosa usando la cepa DS68625 que expresa variantes de Gal2: N376T+T89I (círculos en blanco), N376T (círculos en negro), T89I (cuadrados en blanco) y Gal2 de tipo salvaje (cuadrados en negro).

Figura 3. Experimentos de absorción usando diversas concentraciones de [¹⁴C]-L-arabinosa (arriba) o [¹⁴C]-D-glucosa (abajo) para determinar K_m y V_{max} de la cepa DS68625 que expresa variantes de Gal2: T89I (círculos en blanco), N376T+T89I (círculos en negro), N376T (cuadrados en blanco), Gal2 de tipo salvaje (cuadrados en negro), N376S (diamantes en negro), y N376I (diamantes en blanco).

5 Descripción detallada

Un objeto de la invención es proporcionar nuevos polipéptidos de transporte de hexosa variantes que tienen una afinidad reducida por la glucosa. Otro objeto de la invención es proporcionar nuevos polipéptidos de transporte de hexosa variantes que tienen alta afinidad por la arabinosa. Uno o más de estos objetos se logran según la invención.

10 A lo largo de la presente memoria descriptiva y las reivindicaciones adjuntas, las palabras “comprender” e “incluir”, y variaciones tales como “comprende”, “que comprende”, “incluye” y “que incluye” deben interpretarse de manera inclusiva. Es decir, estas palabras están destinadas a transmitir la posible inclusión de otros elementos o números enteros no citados específicamente, cuando el contexto lo permita. Los artículos “un” y “una” se usan aquí para hacer referencia a uno o más de uno (es decir, uno o al menos uno) del objeto gramatical del artículo. A modo de ejemplo, “un aminoácido” puede significar un aminoácido o más de un aminoácido.

15 La secuencia de aminoácidos alineada con la secuencia de aminoácidos expuesta en la SEC ID NO: X (cuando se refiere a un polipéptido variante) significa que la secuencia de aminoácidos variante y la secuencia de aminoácidos expuesta en la SEC ID NO: X están alineadas por un método adecuado que permite la comparación de las secuencias entre sí e identificaciones de las posiciones en la secuencia de aminoácidos de la variante en la que está presente el mismo aminoácido (posición idéntica), o está presente otro aminoácido (sustitución), o están presentes uno o más aminoácidos adicionales (inserción o extensión), o no hay aminoácidos presentes (supresión o truncamiento), si se compara con la secuencia de aminoácidos expuesta en SEQ ID NO: X.

20 Un método adecuado que permite la comparación de dos secuencias de aminoácidos puede ser cualquier método de alineación de secuencias por pares adecuado conocido por los expertos en la técnica, preferiblemente un método de alineación de secuencias por pares global. Un método preferido de alineación de secuencias por pares global es el método EMBOSS Needle basado en el algoritmo de alineación de Needleman-Wunsch (con el objetivo de encontrar la alineación óptima (incluyendo los espacios) de las dos secuencias a lo largo de toda su longitud) (Needleman, S. B. y Wunsch, C. D. (1970) J. Mol. Biol. 48, 443-453) como se describe aquí. En una realización, la secuencia de aminoácidos se alinea con la secuencia de aminoácidos expuesta en SEQ ID NO: x usando el programa NEEDLE del paquete EMBOSS, usando EBLOSUM62 como matriz de sustitución, con una penalización por apertura de espacio de 10 y una penalización de extensión de espacio de 0,5.

25 Una “levadura” o “célula de levadura”, como se define aquí, es una levadura adecuada para manipulación genética, y que puede cultivarse a densidades celulares útiles para la producción industrial de un producto diana. Una levadura o una célula de levadura se puede encontrar en la naturaleza o una célula derivada de una célula parental después de la manipulación genética o mutagénesis clásica.

35 La expresión “secuencia de control”, como se usa aquí, se refiere a componentes implicados en la regulación de la expresión de una secuencia codificante en un organismo específico o in vitro. Ejemplos de secuencias de control son secuencias de inicio de la transcripción, secuencias de terminación, promotores, líderes, péptidos señal, propéptidos, prepropéptidos, o secuencias potenciadoras; secuencias de Shine-Delgarno, secuencias represoras o activadoras; señales de procesamiento de ARN eficientes, tales como señales de ajuste y poliadenilación; secuencias que estabilizan el ARNm citoplásmico; secuencias que mejoran la eficacia de la traducción (por ejemplo, sitios de unión a ribosomas); secuencias que mejoran la estabilidad de las proteínas; y cuando se desee, secuencias que potencian la secreción de proteínas.

40 El término “cultivo” se refiere a un método de multiplicación de microorganismos en un medio nutritivo y en condiciones adecuadas para el crecimiento y/o propagación de dicho microorganismo y/o la producción de un compuesto de interés por el microorganismo. Estos métodos son conocidos en la técnica. Cuando el microorganismo es capaz de expresar/producir un compuesto de interés, por ejemplo, los microorganismos pueden cultivarse mediante cultivo en matraz agitado, fermentación a pequeña o gran escala (incluyendo fermentaciones continuas, por lotes, por lotes alimentados, o en estado sólido) en fermentadores de laboratorio o industriales realizada en un medio adecuado y en condiciones que permitan expresar y/o aislar el compuesto de interés. Normalmente, el cultivo comprenderá una fase de crecimiento dirigida principalmente a la formación de biomasa, y una fase de producción dirigida principalmente a la producción del compuesto de interés. La fase de crecimiento y la fase de producción pueden superponerse hasta cierto punto. Un medio nutriente adecuado comprende fuentes de carbono, fuentes de nitrógeno, y compuestos adicionales (tales como sales inorgánicas (por ejemplo, fosfato), oligoelementos, y/o vitaminas) (véase, por ejemplo, Bennett, J. W. y LaSure, L., eds., More Gene Manipulations in Fungi, Academic Press, CA, 1991), y se puede realizar en condiciones aeróbicas o anaeróbicas.

55 La expresión “derivado de” también incluye las expresiones “procede de”, “obtenido de”, “obtenible de”, “aislado de”, y “creado a partir de”, y normalmente indica que un material especificado encuentra su origen en otro material especificado, o tiene características que pueden describirse con referencia al otro material especificado. Como se usa

aquí, una sustancia (por ejemplo, una molécula de ácido nucleico o polipéptido) “derivada de” un microorganismo significa preferiblemente que la sustancia es nativa de ese microorganismo.

5 El término “expresión” incluye cualquier etapa implicada en la producción de (a) polipéptido o polipéptidos, que incluye, pero no se limita a, transcripción, modificación postranscripcional, traducción, modificación postraducciona, y secreción. Una reducción o supresión de la producción de B significa una limitación a x% o menos de B producido mediante la conversión enzimática de A. Esto se puede lograr con una enzima/proteína/célula/gen como se describe aquí. Un aumento en la producción de B significa un aumento de al menos un x% de B producido a través de la conversión enzimática de A en comparación con la cantidad de B obtenida en un procedimiento que usa una célula no modificada/proteína de tipo salvaje/enzima/gen. La reducción o el aumento de la expresión génica se puede medir mediante diversos métodos, tales como, por ejemplo, tecnología de transferencia Northern, Southern o Western, como se conoce en la técnica. Las expresiones “aumento de actividad” o “sobrexpresión” se usan indistintamente aquí.

10 Un casete de expresión comprende un polinucleótido que codifica un polipéptido, unido operativamente a las secuencias de control apropiadas que permiten la expresión del polinucleótido en una célula o in vitro.

15 El casete de expresión puede ser un vector de replicación autónoma (por ejemplo, un plásmido), es decir, por ejemplo, un vector cuya replicación es independiente de la replicación del genoma. Alternativamente, el casete puede ser aquel que, cuando se introduce en la célula, se integra total o parcialmente en el genoma de la célula. En los últimos casos, puede comprender una o más secuencias de direccionamiento para dirigir la integración en el genoma.

El casete de expresión puede contener o no uno o más marcadores seleccionables, que permiten una fácil selección de células transformadas.

20 La expresión “fragmento polipeptídico” se define aquí como un polipéptido que tiene uno o más aminoácidos suprimidos del término amino y/o carboxilo del polipéptido parental.

25 La expresión “polipéptido maduro” se define aquí como un polipéptido en su forma o formas finales, y se obtiene después de la traducción de un ARNm en polipéptido, modificaciones postraduccionales de dicho polipéptido dentro o fuera de la célula. La modificación postraducciona incluye procesamiento N-terminal, truncamiento C-terminal, glicosilación, fosforilación, y eliminación de secuencias líder tales como péptidos señal, propéptidos y/o prepropéptidos, como se define aquí, por escisión.

30 La expresión “de origen natural”, como se usa aquí, se refiere a procedimientos, eventos o productos que ocurren en su forma relevante en la naturaleza. Por el contrario, “de origen no natural” se refiere a procedimientos, eventos o productos cuya existencia o forma implica la mano del hombre. La expresión “de origen no natural” es aquí sinónimo de “hecho por el hombre”. Generalmente, la expresión “de origen natural”, con respecto a polipéptidos o ácidos nucleicos, se puede usar de manera intercambiable con la expresión “de tipo salvaje” o “nativo”. Se refiere a polipéptidos o ácidos nucleicos que codifican un polipéptido, que tienen una secuencia de aminoácidos o una secuencia polinucleotídica, respectivamente, idéntica a la que se encuentra en la naturaleza. Los polipéptidos de origen natural incluyen polipéptidos nativos, tales como los polipéptidos expresados de forma natural o encontrados en una célula particular. Los polinucleótidos de origen natural incluyen polinucleótidos nativos tales como los polinucleótidos que se encuentran naturalmente en el genoma de una célula particular. Además, una secuencia que es de tipo salvaje o de origen natural puede referirse a una secuencia de la que deriva una secuencia variante o sintética.

40 La expresión “construcción de ácido nucleico” se denomina aquí una molécula de ácido nucleico, ya sea monocatenaria o bicatenaria, que deriva de un gen de origen natural o que se ha modificado para contener segmentos de ácido nucleico que se combinan yuxtaponen de una manera que de otro modo no existiría en la naturaleza. Las construcciones de ácido nucleico pueden aislarse, obtenerse sintéticamente o mutagenizarse. La expresión construcción de ácido nucleico es sinónimo de la expresión “casete de expresión” cuando la construcción de ácido nucleico contiene todas las secuencias de control requeridas para la expresión de una secuencia codificante, en el que dichas secuencias de control están unidas operativamente a dicha secuencia codificante.

45 La expresión “unido operativamente”, como se usa aquí, se refiere a dos o más componentes tales como secuencias de ácido nucleico o secuencias polipeptídicas que están unidas físicamente y están en una relación funcional entre sí que les permite funcionar de la manera deseada. Por ejemplo, un promotor está operativamente unido a una secuencia codificante si el promotor puede regular la transcripción o expresión de una secuencia codificante, en cuyo caso la secuencia codificante debe entenderse como que está “bajo el control” del promotor.

La expresión “polipéptido parental” se refiere al polipéptido con respecto al cual otro polipéptido difiere al sustituir, añadir o suprimir uno o más aminoácidos.

55 Posición definida con referencia a SEQ ID NO: x significa que la posición en la secuencia de aminoácidos según la descripción en la que ha tenido lugar una modificación se da con respecto a la posición del aminoácido correspondiente en la secuencia de aminoácidos según SEQ ID NO: x cuando las dos secuencias se alinean usando un método de alineación como se describe aquí.

El término “recombinante”, cuando se usa en referencia a un ácido nucleico o proteína, indica que el ácido nucleico o proteína se ha modificado en su secuencia, si se compara con su forma nativa, por intervención humana. El término “recombinante”, cuando se refiere a una célula, indica que el genoma de la célula se ha modificado en su secuencia, si se compara con su forma nativa, por intervención humana. El término “recombinante” es sinónimo de “modificado genéticamente”.

Un marcador seleccionable es un gen que permite la selección de células transformadas con dicho gen y que proporciona resistencia a biocidas o virus, resistencia a metales pesados, prototrofia a auxótrofos, y similares.

Los marcadores seleccionables preferidos incluyen, pero no se limitan a, aquellos que confieren resistencia a fármacos o que complementan un defecto en la célula. Incluyen, por ejemplo, genes marcadores versátiles que se pueden usar para la transformación de la mayoría de hongos filamentosos y levaduras, tales como genes de acetamidasa o ADNc, o genes que proporcionan resistencia a antibióticos.

Alternativamente, se pueden usar marcadores de selección específicos, tales como marcadores auxotróficos, que requieren cepas mutantes correspondientes. Preferiblemente, el marcador de selección se suprime de la célula transformada después de la introducción de la construcción de expresión, para obtener células transformadas que están libres de genes marcadores de selección.

La expresión marcador seleccionable se extiende a un gen marcador usado para el cribado, es decir, un gen marcador que, una vez introducido en una célula, confiere a la célula un fenotipo visible y hace que la célula se vea diferente. Un ejemplo de marcador para cribado es un gen que codifica una proteína fluorescente que hace que las células emitan fluorescencia bajo una fuente de luz apropiada.

Para los fines de esta descripción, se define aquí que para determinar el porcentaje de homología de secuencia o identidad de secuencia de dos secuencias de aminoácidos o de dos secuencias de ácido nucleico, las secuencias se alinean para fines de comparación óptimos. Con el fin de optimizar el alineamiento entre las dos secuencias, se pueden introducir espacios en cualquiera de las dos secuencias que se comparan. Tal alineación se puede llevar a cabo en toda la longitud de las secuencias que se comparan. Alternativamente, la alineación se puede llevar a cabo en una longitud más corta, por ejemplo en alrededor de 20, alrededor de 50, alrededor de 100 o más ácidos nucleicos/basados en aminoácidos. La identidad de secuencia es el porcentaje de coincidencias idénticas entre las dos secuencias sobre la región alineada dada a conocer.

Una comparación de secuencias y la determinación del porcentaje de identidad de secuencia entre dos secuencias se pueden realizar usando un algoritmo matemático. La persona experta será consciente del hecho de que están disponibles varios programas informáticos diferentes para alinear dos secuencias y determinar la identidad entre dos secuencias (Kruskal, J. B. (1983) An overview of sequence comparison en D. Sankoff and J. B. Kruskal, (ed.), Time warps, string edits and macromolecules: the theory and practice of sequence comparison, p. 1-44 Addison Wesley). El porcentaje de identidad de secuencia entre dos secuencias de aminoácidos o entre dos secuencias nucleotídicas puede determinarse usando el algoritmo de Needleman y Wunsch para el alineamiento de dos secuencias. (Needleman, S. B. y Wunsch, C. D. (1970) J. Mol. Biol. 48, 443-453). El algoritmo puede alinear tanto las secuencias de aminoácidos como las secuencias nucleotídicas. El algoritmo de Needleman-Wunsch se ha implementado en el programa informático NEEDLE. Para los fines de esta descripción, se utilizó el programa NEEDLE del paquete EMBOSS (versión 2.8.0 o superior, EMBOSS: The European Molecular Biology Open Software Suite (2000) Rice, P. Longden, I. y Bleasby, A. Trends in Genetics 16, (6) p. 276—277, <http://emboss.bioinformatics.nl/>). Para las secuencias proteicas, se usa EBLOSUM62 para la matriz de sustitución. Para la secuencia nucleotídica, se usa EDNAFULL. Los parámetros opcionales usados son una penalización por apertura de espacio de 10 y una penalización por extensión de espacio de 0,5. La persona experta apreciará que todos estos parámetros diferentes producirán resultados ligeramente diferentes, pero que el porcentaje de identidad global de dos secuencias no se altera significativamente cuando se usan algoritmos diferentes.

Después de la alineación por el programa NEEDLE como se describió anteriormente, el porcentaje de identidad de secuencia entre una secuencia de consulta y una secuencia de la descripción se calcula de la siguiente manera: Número de posiciones correspondientes en la alineación que muestran un aminoácido idéntico o un nucleótido idéntico en ambas secuencias dividido entre la longitud total de la alineación después de restar el número total de espacios en la alineación. La identidad, definida según el presente documento, se puede obtener de NEEDLE usando la opción NOBRIEF, y se etiqueta en el resultado del programa como “identidad más larga”.

Las secuencias de ácidos nucleicos y proteínas de la presente descripción pueden usarse además como una “secuencia de consulta” para realizar una búsqueda en bases de datos públicas para, por ejemplo, identificar otros miembros de la familia o secuencias relacionadas. Estas búsquedas se pueden realizar usando los programas BLASTN y BLASTX (versión 2.0) de Altschul, et al. (1990) J. Mol. Biol. 215:403—10. Las búsquedas de nucleótidos BLAST se pueden realizar con el programa BLASTN, puntuación = 100, longitud de palabra = 12, para obtener secuencias nucleotídicas homólogas a las moléculas de ácido nucleico de la descripción. Las búsquedas de proteínas BLAST se pueden realizar con el programa BLASTX, puntuación = 50, longitud de palabra = 3, para obtener secuencias de aminoácidos homólogas a las moléculas proteicas de la descripción. Para obtener alineaciones con espacios con fines de comparación, se puede usar Gapped BLAST, como se describe en Altschul et al., (1997) Nucleic

Acids Res. 25(17): 3389-3402. Cuando se usan los programas BLAST y Gapped BLAST, se pueden usar los parámetros predeterminados de los programas respectivos (por ejemplo, BLASTX y BLASTN). Véase la página de inicio del National Center for Biotechnology Information en <http://www.ncbi.nlm.nih.gov/>.

5 Como se usa aquí, los términos “variante”, “derivado”, “mutante”, u “homólogo” se pueden usar indistintamente. Pueden referirse a polipéptidos o ácidos nucleicos. Las variantes incluyen sustituciones, inserciones, supresiones, truncamientos, transversiones, y/o inversiones, en una o más localizaciones con respecto a una secuencia de referencia. Las variantes se pueden obtener, por ejemplo, mediante mutagénesis de saturación de sitio, mutagénesis de barrido, mutagénesis de inserción, mutagénesis aleatoria, mutagénesis dirigida de sitio, y evolución dirigida, así como otros enfoques diversos de recombinación. Los polipéptidos variantes pueden diferir de un polipéptido de referencia por un pequeño número de restos de aminoácidos, y pueden definirse por su nivel de homología/identidad de secuencia de aminoácidos primaria con un polipéptido de referencia. Preferiblemente, los polipéptidos variantes tienen al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o incluso al menos 99% de identidad de secuencia de aminoácidos con un polipéptido de referencia. Los métodos para determinar el porcentaje de identidad se conocen en la técnica y se describen aquí. Generalmente, las variantes conservan la naturaleza característica del polipéptido de referencia, pero tienen propiedades alteradas en algunos aspectos específicos. Por ejemplo, una variante puede tener un óptimo de pH modificado, una capacidad de unión de sustrato modificada, una resistencia modificada a la degradación enzimática u otra degradación, una mayor o menor actividad, una estabilidad modificada a temperatura u oxidativa, pero conserva su funcionalidad característica. Las variantes incluyen además polipéptidos con modificaciones químicas que cambian las características de un polipéptido de referencia.

Con respecto a los ácidos nucleicos, los términos se refieren a un ácido nucleico que codifica un polipéptido variante, que tiene un grado especificado de homología/identidad con un ácido nucleico de referencia, o que se hibrida en condiciones rigurosas con un ácido nucleico de referencia o el complemento del mismo. Preferiblemente, un ácido nucleico variante tiene al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o incluso al menos 99% de identidad de secuencia de ácido nucleico con un ácido nucleico de referencia. Los métodos para determinar el porcentaje de identidad se conocen en la técnica y se describen aquí.

Los términos permeasa, facilitador, transportador, o proteína de transporte, o términos relacionados, describen proteínas con múltiples dominios que atraviesan la membrana que exhiben una función en el transporte de moléculas a través de una membrana. Este transporte puede ser provocado por diferentes mecanismos: unipuerto (transporte de una molécula), simpuerto (cotransporte simultáneo de dos moléculas diferentes en la misma dirección), antiuerto (transporte simultáneo de dos moléculas en direcciones opuestas), difusión facilitada, translocación de grupo (por ejemplo, sistemas de fosfotransferasa), y transporte primario (por ejemplo, transporte impulsado por hidrólisis de trifosfato de adenosina o por luz).

35 La familia de transportadores de azúcar en la levadura se puede dividir en cinco grupos: permeasas de hexosa (genes *HXT*, *GAL2*), permeasas de disacáridos, permeasas de mioinositol, receptores de azúcar, y un grupo adicional de transportadores cuyo sustrato se ha determinado (Nelissen et al., 1997, FEMS Yeast Rev, 21, p. 113-134).

En un primer aspecto, la invención proporciona una variante de un polipéptido parental que es preferiblemente un transportador de hexosa, en la que el polipéptido parental comprende la secuencia de aminoácidos de SEQ ID NO: 1, y en la que la variante comprende una secuencia de aminoácidos que, cuando se alinea con la secuencia de aminoácidos expuesta en SEQ ID NO: 1, comprende una sustitución de aminoácidos N376 y T89, definiéndose las posiciones de dichos aminoácidos con referencia a la secuencia de aminoácidos expuesta en SEQ ID NO: 1, en la que la sustitución de aminoácido T89 comprende una cualquiera de T89I, T89V, T89L, T89S, T89M, o T89F.

45 El documento WO2014/195376 describe variantes de polipéptidos transportadores de hexosa que pertenecen a la Superfamilia de Facilitadores Principales (MFS), incluyendo Gal2 que tiene sustituciones en la posición N376 con mayor preferencia por xilosa. Sin embargo, el documento WO2014/195376 no dice nada sobre las propiedades de transporte de arabinosa de las variantes N376 de Gal2, y no describe o sugiere que la sustitución de la treonina en la posición 89 por isoleucina da como resultado un transportador de hexosa que tiene mayor afinidad por L-arabinosa y que ha perdido toda afinidad por glucosa (véase Tabla 3).

50 El polipéptido parental es un miembro de la Superfamilia de Facilitadores Principales (MFS). Dos familias de transportadores principales cuyas proteínas se encuentran en todos los organismos vivos son la superfamilia del casete de unión de ATP (ABC) y la superfamilia de facilitadores principales (MFS), también conocida como familia uniportador-simportador-antiportador. Mientras que las permeasas de la familia ABC consisten en múltiples componentes y son transportadores activos primarios, capaces de transportar tanto moléculas pequeñas como macromoléculas solo después de generar energía a través de la hidrólisis de ATP, los transportadores MFS consisten en un único polipéptido de un portador secundario que facilita el transporte de pequeños solutos en respuesta a un gradiente iónico quimiosmótico. La superfamilia ABC y las proteínas MFS representan casi la mitad de los transportadores de solutos codificados dentro de los genomas microbianos (revisado por Pao et al, 1998, Microbiol Mol Biol Rev.; 62 pp.1-34, y Saier et al, 1999, J Mol Microbiol Biotechnol, 1 p. 257-279).

El polipéptido según SEQ ID NO: 1 es Gal2, que es un transportador de difusión facilitado. Pertenece a la superfamilia de facilitadores principales, la familia de los transportadores de azúcar (TC 2.A.1.1). El "polipéptido de permeasa" también se designa aquí como "permeasa polipeptídica" o "polipéptido". "Polinucleótido de polipéptido de permeasa" es aquí un polinucleótido que codifica el polipéptido de permeasa.

5 En una realización, el polipéptido variante tiene actividad de transporte de xilosa y/o actividad de transporte de arabinosa, y/o actividad de transporte de galactosa.

El polipéptido variante de la invención puede tener una o más actividades alternativas y/o adicionales distintas de la actividad de permeasa de azúcar.

10 Como se expuso anteriormente, un polipéptido variante de la invención típicamente tendrá actividad de permeasa de azúcar. Sin embargo, un polipéptido variante de la invención puede tener una o más de las actividades expuestas anteriormente, además de o como alternativa a esa actividad.

15 El polipéptido variante no es necesariamente un polipéptido según SEQ ID NO: 1. Más bien, el polipéptido variante comprende aminoácidos sustituidos, que, en SEQ ID NO: 1, están en las posiciones 89 y 376. Sin embargo, en un polipéptido de permeasa de hexosa que tiene otra secuencia de aminoácidos que SEQ ID NO: 1, tales posiciones pueden tener una numeración diferente, pero aún corresponden a las posiciones 89 o 376 de SEQ ID NO: 1.

20 Las posiciones 89 y 376, para (otros) polipéptidos de permeasa de hexosa, pueden identificarse comparando estructuras cristalinas en 3D, o alineando las secuencias de aminoácidos, como se explica, entre otros, en el documento WO2014/195376, en particular en la página 17, línea 22 y siguientes. Por ejemplo, en el polipéptido transportador de hexosa Hxt11 (SEQ ID NO: 10), la posición N376 del polipéptido de SEQ ID NO: 1 corresponde a N366, y la posición T89 corresponde a T79.

En una realización, el polipéptido variante comprende la secuencia de aminoácidos expuesta en SEQ ID NO: 1.

25 En otra realización, la sustitución del aminoácido N376 comprende una cualquiera de N376T, N376C, N376V, N376M, N376L, N376I, o N376F. Es decir, el polipéptido variante tiene un aminoácido en la posición 376 que es una treonina, una cisteína, una valina, una metionina, una leucina, una isoleucina, o una fenilalanina. Preferiblemente, la posición 376 es T, S, o I.

La sustitución de aminoácido T89 comprende una cualquiera de T89I, T89V, T89L, T89S, T89M, o T89F.

Es decir, el polipéptido variante tiene un aminoácido en la posición 89 que es una isoleucina, una valina, una leucina, una serina, una metionina, o una fenilalanina. Preferiblemente, la posición 89 es una isoleucina.

30 En una realización preferida, la sustitución del aminoácido N376 comprende una cualquiera de N376T, N376S, o N376I, y la sustitución de aminoácido T89 comprende T89I.

35 En una realización, el polipéptido variante tiene al menos 70% de identidad de secuencia con la secuencia de aminoácidos del polipéptido parental, o tiene al menos 70% de identidad de secuencia de un polipéptido con una secuencia de aminoácidos correspondiente a SEQ ID NO: 1. La variante puede tener al menos 80% u 85% de identidad de secuencia con la secuencia de aminoácidos del polipéptido parental, o al menos 80% u 85% de identidad de secuencia de un polipéptido con una secuencia de aminoácidos correspondiente a SEQ ID NO: 1. La variante puede tener al menos 90% de identidad de secuencia con la secuencia de aminoácidos del polipéptido parental, o al menos 90% de identidad de secuencia de un polipéptido con una secuencia de aminoácidos correspondiente a SEQ ID NO: 1. La variante puede tener al menos 95% de identidad de secuencia con la secuencia de aminoácidos del polipéptido parental, o al menos 95% de identidad de secuencia de un polipéptido con una secuencia de aminoácidos correspondiente a SEQ ID NO: 1. La variante puede tener al menos 98% de identidad de secuencia con la secuencia de aminoácidos del polipéptido parental, o al menos 98% de identidad de secuencia de un polipéptido con una secuencia de aminoácidos correspondiente a SEQ ID NO: 1. La variante puede tener al menos 99% de identidad de secuencia con la secuencia de aminoácidos del polipéptido parental, o al menos 99% de identidad de secuencia de un polipéptido con una secuencia de aminoácidos correspondiente a SEQ ID NO: 1. El polipéptido variante es preferiblemente un polipéptido hecho por el hombre.

La invención proporciona además una secuencia de ácido nucleico que codifica el polipéptido variante de la invención, preferiblemente en la que dicha secuencia de ácido nucleico está hecha por el hombre.

50 La invención también proporciona una construcción de ácido nucleico que comprende la secuencia de ácido nucleico de la invención unida operativamente a una o más secuencias de control capaces de dirigir la expresión de la variante polipeptídica en una célula adecuada.

La invención también proporciona un casete de expresión recombinante que comprende la construcción de ácido nucleico de la invención.

La invención también proporciona una levadura recombinante que comprende el ácido nucleico de la invención, la construcción de ácido nucleico de la invención, y/o el casete de expresión de la invención. La levadura se transforma

con una secuencia de ácido nucleico, una construcción de ácido nucleico que comprende dicha secuencia de ácido nucleico, o un casete de expresión recombinante que comprende dicha construcción de ácido nucleico. En una realización, tal secuencia de ácido nucleico codifica un polipéptido variante que es un mutante de un polipéptido que es nativo en la célula hospedante no transformada. En una realización, el polipéptido que es nativo en la célula hospedante no transformada es un polipéptido transportador escogido de la lista que consiste en Gal2, Hxt1, Hxt2, Hxt3, Hxt4, Hxt5, Hxt6, Hxt7, Hxt8, Hxt9, Hxt10, Hxt11, Hxt12, Hxt13, Hxt14, Hxt15, Hxt16 y Hxt17.

La levadura recombinante puede seleccionarse de Saccharomycetaceae, en particular del grupo de *Saccharomyces cerevisiae*; *Kluyveromyces*, tal como *Kluyveromyces marxianus*; *Pichia*, tal como *Pichia/Scheffersomyces stipitis* o *Pichia angusta*; *Zygosaccharomyces*, tal como *Zygosaccharomyces bailii*; y *Brettanomyces*, tal como *Brettanomyces intermedius*, *Issatchenkia*, tal como *Issatchenkia orientalis* y *Hansenula/Ogateae*.

La levadura recombinante puede someterse a ingeniería evolutiva para mejorar sus propiedades. Los procedimientos de ingeniería evolutiva son procedimientos conocidos. La ingeniería evolutiva es un procedimiento en el que los fenotipos industrialmente relevantes de un microorganismo, en este caso la levadura recombinante, se pueden acoplar a la tasa de crecimiento específica y/o la afinidad por un nutriente, mediante un procedimiento de selección natural establecida racionalmente. La ingeniería evolutiva se describe con detalle, por ejemplo, en Kuijper, M, et al, FEMS, Eukaryotic cell Research 5(2005) 925-934, documentos WO2008/041840 y WO2009/112472. Después de la ingeniería evolutiva, se aísla la célula recombinante resultante. El aislamiento se puede ejecutar de cualquier manera conocida, por ejemplo por separación de células de un caldo de células recombinantes usado en la ingeniería evolutiva, por ejemplo tomando una muestra de células, o por filtración o centrifugación.

En una realización, la levadura recombinante comprende uno o más genes que codifican una L-arabinosa isomerasa (E.C. 5.3.1.4; araA); uno o más genes que codifican una L-ribulocinasa (E.C. 2.7.1.16; araB); y uno o más genes que codifican una L-ribulosa-5-P4-epimerasa (E.C. 5.1.3.4; araD); genes los cuales permiten que dicha levadura fermente arabinosa. Los genes para araA, araB y araD pueden ser de *Lactobacillus plantarum*, como se describe, por ejemplo, en el documento WO2008/041840. Alternativamente, se puede usar un gen araA de *Bacillus subtilis* y genes araB y araD de *Escherichia coli* como se describe, por ejemplo, en el documento EP1499708. Alternativamente, los genes araA, araB y araD pueden derivar de al menos uno del género *Clavibacter*, *Arthrobacter* y/o *Gramella*, en particular uno de *Clavibacter michiganensis*, *Arthrobacter aureescens*, y/o *Gramella forsetii*, como se describe, por ejemplo, en el documento WO2009/011591.

En una realización, la levadura recombinante comprende un gen que codifica una L-arabinosa isomerasa que tiene una secuencia de aminoácidos según SEQ ID NO: 5, o un homólogo funcional de la misma que tiene al menos 60%, preferiblemente al menos 65%, al menos 70%, al menos al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, al menos 98%, al menos 99% de identidad con SEQ ID NO: 5.

En una realización, la levadura recombinante comprende un gen que codifica una L-ribulocinasa que tiene una secuencia de aminoácidos según SEQ ID NO: 6, o un homólogo funcional de la misma que tiene al menos 60%, preferiblemente al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, al menos 98%, al menos 99% de identidad con SEQ ID NO: 6.

En una realización, la levadura recombinante comprende un gen que codifica una L-ribulosa-5-P4-epimerasa que tiene una secuencia de aminoácidos según SEQ ID NO: 7, o un homólogo funcional de la misma que tiene al menos 60%, preferiblemente al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, al menos 98%, al menos 99% de identidad con SEQ ID NO: 7.

En una realización, la levadura recombinante comprende uno o más genes que codifican una xilosa isomerasa (EC 5.3.1.5; tal como xylA) y uno o más genes que codifican una xilulosa cinasa (E.C. 2.7.1.17; tal como XKS1); o uno o más genes que codifican una xilosa reductasa (E.C. 1.1.1.307; tal como XYL1 de *Pichia/Scheffersomyces stipitis*) y uno o más genes que codifican una xilitol deshidrogenasa (E.C. 1.1.1.B19; tal como XYL2 de *Scheffersomyces stipitis*) para permitir que la levadura recombinante fermente xilosa.

En una realización, la levadura recombinante comprende un gen que codifica una xilosa isomerasa que tiene una secuencia de aminoácidos según SEQ ID NO: 8, o un homólogo funcional de la misma que tiene al menos 60%, preferiblemente al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, al menos 98%, al menos 99% de identidad con SEQ ID NO: 8.

En una realización, la levadura recombinante comprende un gen que codifica una xilulosa cinasa que tiene una secuencia de aminoácidos según SEQ ID NO: 9, o un homólogo funcional de la misma que tiene al menos 60%, preferiblemente al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, al menos 98%, al menos 99% de identidad con SEQ ID NO: 9.

Una "xilosa isomerasa" (EC 5.3.1.5) se define aquí como una enzima que cataliza la isomerización directa de D-xilosa en D-xilulosa y/o viceversa. La enzima también se conoce como una D-xilosa cetoisomerasa. Una xilosa isomerasa aquí también puede ser capaz de catalizar la conversión entre D-glucosa y D-fructosa (y en consecuencia, por lo tanto puede denominarse como una glucosa isomerasa). Una xilosa isomerasa aquí puede requerir un catión bivalente, tal

como magnesio, manganeso o cobalto, como cofactor. Los ensayos para medir la actividad de la xilosa isomerasa y la xilulosa cinasa se describen, entre otros, en el documento WO2003/062430.

En una realización, la levadura recombinante comprende además la sobreexpresión de uno o más genes de la rama no oxidativa de la ruta de las pentosas fosfatos.

- 5 En una realización, el uno o más genes de la ruta de las pentosas fosfatos que se sobreexpresan codifican una enzima seleccionada de la lista de una transaldolasa (EC 2.2.1.2), una transcetolasa (EC 2.2.1.1), una ribosa-5-fosfato isomerasa (EC 5.3.1.6), y una D-ribulosa-5-fosfato 3-epimerasa (EC 5.1.3.1).

En otra realización, el uno o más genes de la ruta de las pentosas fosfatos que se sobreexpresan se seleccionan de la lista de *TAL1*, *TAL2*, *NQM1*, *TKL1*, *TKL2*, *RPE1* y *RK11*.

- 10 También se describe un método para diseñar una secuencia variante de una secuencia polipeptídica parental, método el cual comprende:

a) seleccionar una secuencia polipeptídica parental, preferiblemente un polipéptido parental con una secuencia de aminoácidos según SEQ ID NO: 1;

- 15 b) sustituir en dicha secuencia polipeptídica los aminoácidos N376 y T89, estando definida la posición de dichos aminoácidos con referencia a la SEQ ID NO: 1.

En una realización, en la etapa b), las sustituciones de los aminoácidos N376 dan como resultado una cualquiera de N376T, N376C, N376V, N376M, N376L, N376I, o N376F, y la sustitución del aminoácido T89 da como resultado una cualquiera de T89I, T89V, T89L, T89S, T89M, o T89F.

- 20 Preferiblemente, la sustitución del aminoácido N376 da como resultado una cualquiera de N376T, N376S o N376I, y la sustitución del aminoácido T89 da como resultado T89I.

Las técnicas para diseñar construcciones de ácidos nucleicos que codifican polipéptidos variantes, casetes de expresión de recombinación, para transformar células de levadura con ellos, se describen, entre otros, en el documento WO2014/195376.

- 25 La invención también proporciona un procedimiento para la producción de un producto de fermentación, preferiblemente etanol, que comprende:

- fermentar una composición que comprende una biomasa lignocelulósica, en particular que comprende glucosa y arabinosa, preferiblemente que comprende también galactosa, en condiciones anaeróbicas en presencia de una levadura recombinante según la invención; y
- recuperar el producto de fermentación.

- 30 El producto de fermentación puede ser cualquiera de etanol, butanol, ácido láctico, di-terpeno, di-terpeno glicosilado, ácido 3-hidroxipropiónico, ácido acrílico, ácido acético, ácido succínico, ácido cítrico, ácido fumárico, ácido málico, ácido itacónico, un aminoácido, 1,3-propanodiol, etileno, glicerol, un antibiótico β -lactámico, o una cefalosporina. El etanol es un producto de fermentación preferido.

- 35 Las condiciones anaeróbicas se definen aquí como condiciones sin oxígeno, o en las que la levadura recombinante esencialmente no consume oxígeno, y normalmente corresponden a un consumo de oxígeno de menos de 5 mmol/l.h, en particular a un consumo de oxígeno de menos de 2,5 mmol/l.h, o menos de 1 mmol/l.h. Más preferiblemente, se consume 0 mmol/l/h (es decir, el consumo de oxígeno no es detectable. Esto generalmente corresponde a una concentración de oxígeno disuelto en el caldo de cultivo de menos del 5% de saturación de aire, en particular a una concentración de oxígeno disuelto de menos de 1% de saturación de aire, o menos del 0,2% de saturación de aire. El procedimiento también se puede realizar en condiciones microaeróbicas.

- 40 En una realización, la biomasa lignocelulósica comprende lignocelulosa y/o hemicelulosa. La composición puede ser un hidrolizado de biomasa. Dicho hidrolizado de biomasa puede ser un hidrolizado de biomasa lignocelulósica. Aquí, lignocelulosa incluye hemicelulosa y partes de hemicelulosa de biomasa. La lignocelulosa también puede incluir fracciones lignocelulósicas de biomasa. Los materiales lignocelulósicos adecuados se pueden encontrar en la siguiente lista: podas de huertos, chaparral, desechos de molinos, desechos de madera urbana, desechos municipales, desechos de tala, aclareos forestales, cultivos leñosos de rotación corta, desechos industriales, paja de trigo, paja de avena, paja de arroz, paja de cebada, paja de centeno, paja de lino, cáscaras de soja, cáscaras de arroz, paja de arroz, pienso de gluten de maíz, cáscaras de avena, caña de azúcar, rastrojo de maíz, tallos de maíz, mazorcas de maíz, cáscaras de maíz, pasto varilla, miscanthus, sorgo dulce, tallos de cánola, tallos de soja, hierba de la pradera, 45 hierba gama, cola de zorro; pulpa de remolacha azucarera, pulpa de cítricos, cáscaras de semillas, desechos animales celulósicos, recortes de césped, algodón, algas marinas, árboles, madera blanda, madera dura, álamo, pino, arbustos, hierbas, trigo, paja de trigo, bagazo de caña de azúcar, maíz, cáscaras de maíz, hojuelas de maíz, grano de maíz, fibra de granos, productos y subproductos de la molienda húmeda o seca de granos, residuos sólidos urbanos, papel de desecho, residuos de jardín, material herbáceo, residuos agrícolas, residuos forestales, residuos sólidos urbanos,

5 papel de desecho, pulpa, residuos de molino de papel, ramas, arbustos, cañas, maíz, hojas de maíz, un cultivo energético, bosque, una fruta, una flor, un grano, una hierba, un cultivo herbáceo, una hoja, corteza, una aguja, un tronco, una raíz, un árbol joven, un arbusto, pasto varilla, un árbol, una verdura, cáscara de fruta, una vid, pulpa de remolacha azucarera, harinillas de trigo, cáscaras de avena, madera dura o blanda, material de desecho orgánico
 10 generado a partir de un procedimiento agrícola, desechos de madera forestal, o una combinación de dos o más de los mismos. La lignocelulosa, que puede considerarse como una materia prima renovable potencial, generalmente comprende los polisacáridos celulosa (glucanos) y hemicelulosas (xilanos, heteroxilanos y xiloglucanos). Además, algo de hemicelulosa puede estar presente como glucomananos, por ejemplo en materias primas derivadas de la madera. La hidrólisis enzimática de estos polisacáridos a azúcares solubles, incluyendo tanto monómeros como
 15 multímeros, por ejemplo glucosa, celobiosa, xilosa, arabinosa, galactosa, fructosa, manosa, ramnosa, ribosa, ácido galacturónico, ácido glucurónico, y otras hexosas y pentosas, se produce bajo la acción de diferentes enzimas que actúan en concierto.

15 La biomasa lignocelulósica puede comprender pectina y/u otras sustancias pécticas tales como arabinanos, que pueden constituir una proporción considerable de la masa seca de las paredes celulares típicas de los tejidos vegetales no leñosos (alrededor de un cuarto a la mitad de la masa seca pueden ser pectinas). El material lignocelulósico puede pretratarse. El pretratamiento puede comprender exponer el material lignocelulósico a un ácido, una base, un disolvente, calor, un peróxido, ozono, corte mecánico en tiras, trituración, molienda, o despresurización rápida, o una combinación de dos o más de los mismos. Este pretratamiento químico a menudo se combina con un pretratamiento térmico, por ejemplo entre 150-220°C durante 1 a 30 minutos.

20 En otra realización, dicha composición es un hidrolizado de rastrojo de maíz pretratado. Otra composición preferida es un hidrolizado de fibra de maíz, que opcionalmente se trata previamente.

La composición comprende preferiblemente glucosa y arabinosa, preferiblemente también galactosa. En una realización, se consume al menos parte de la glucosa y al menos parte de la arabinosa, preferiblemente se consume al menos parte de la glucosa y al menos parte de la arabinosa y al menos parte de la galactosa.

25 La invención también proporciona el uso de la célula de levadura recombinante de la invención para la producción de un producto de fermentación, preferiblemente etanol, a partir de una composición que comprende glucosa, arabinosa, y preferiblemente también que comprende galactosa.

Ejemplo

Materiales y métodos

30 Cepas y mantenimiento

La construcción de la cepa DS68625 de *S. cerevisiae* de supresión del transportador de hexosa usada para experimentos de absorción de azúcar en el ejemplo siguiente se describió previamente en Shin et al., 2015 (An engineered cryptic Hxt11 sugar transporter facilitates glucose-xylose co-consumption in *Saccharomyces cerevisiae*, *Biotechnology for Biofuels*, volumen 8, p. 176). El genotipo de DS68625 se describe como:

35 *Mat a, ura3-52, leu2-112, gre3::loxP, loxP-Ptpi:TAL1, loxP-Ptpi::RKI1, loxP-Ptpi-TKL1, loxP-Ptpi-RPE1, delta::Padh1XKS1Tcyc1-LEU2, delta::URA3-Ptpi-xylA-Tcyc1, his3::loxP, hxt2::loxP-kanMX-loxP, hxt367::loxP-hphMX-loxP, hxt145::loxP-natMX-loxP, gal2::loxP-zeoMX-loxP*

40 Los cultivos madre se cultivaron en matraces de agitación en medio de extracto de levadura/peptona (YP) (10 g/l de extracto de levadura Bacto, Becton Dickinson, Franklin Lakes, NJ, y 20 g/l de Bacto Peptona, Becton Dickinson). Este medio se suplementó con 20 g/l de maltosa. Los cultivos madre congelados se prepararon añadiendo glicerol (30% vol/vol) a cultivos madre, después de lo cual se almacenaron alícuotas de 1 ml a -80°C.

Cultivo y medios

45 Los estudios de caracterización de cepas se realizaron en medio sintético (SM) preparado como se describió anteriormente (Verduyn C, Postma E, Scheffers WA, Van Dijken JP: Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* 1992, 8:501-517). La fuente de carbono y las disoluciones de vitamina se añadieron después de esterilizar en autoclave el medio durante 20 min a 121°C. La disolución de D-xilosa al 50% p/v se esterilizó en autoclave por separado a 110°C durante 20 min. Antes de la inoculación, se añadieron 20 g/l de D-xilosa (SMX) o 20 g/l de maltosa (SMM) a SM como fuente de carbono. El medio SMX se usó para propagar transformantes derivados de DS68625 para ensayos de absorción de
 50 azúcar. El medio SMM se usó en placas para la selección después de la transformación.

Cultivos aeróbicos en matraces de agitación se hicieron crecer en un agitador orbital a 200 rpm ajustado a 30°C, usando matraces de 500 ml que contenían 100 ml de medio. Para las placas, se añadió agar (BD) 20 g/l antes de esterilizar en autoclave a 121°C durante 20 min.

ES 2 905 103 T3

Construcción de plásmidos y cepas

La secuencia de proteínas para Gal2 que se origina a partir de CEN.PK113-7D se proporciona como SEQ ID NO: 1. La secuencia de ADN codificante para *GAL2* (YLR081W) que se encuentra en CEN.PK113-7D se proporciona como SEQ ID NO: 2. SEQ ID NO: 2 se puede sintetizar artificialmente en una empresa de síntesis de ADN, y se puede usar como molde para mutagénesis dirigida al sitio para generar variantes de Gal2 específicas. Las variantes de Gal2 ensayadas en el ejemplo siguiente se dan en la Tabla 1.

Tabla 1: Cambios de aminoácidos en variantes de Gal2

Variante	Aminoácido en la posición 89	Aminoácido en la posición 376
Gal2 (tipo salvaje)	T	N
Gal2 ^{N376I}	T	I
Gal2 ^{N376S}	T	S
Gal2 ^{N376T}	T	T
Gal2 ^{N376T, T89I}	I	T
Gal2 ^{T89I}	I	N

Los plásmidos construidos y usados en este estudio se muestran en la Tabla 2. El ADN plasmídico se aisló de cultivos de *E. coli* usando un kit GenElute Plasmid (Sigma-Aldrich, St. Louis, MO). La amplificación por PCR de casetes de expresión y fragmentos plasmídicos se realizó usando ADN polimerasa Phusion High Fidelity (Thermo Scientific, Waltham, MA, USA).

Tabla 2. Plásmidos de expresión de *GAL2*

Plásmido	Característica	Fuente
pRS313-mcs	CEN6/ARS4 HIS3 ampR pHXT7-mcs-tHXT7	Nijland et al., 2014*
pRS313-Gal2	CEN6/ARS4 HIS3 ampR pHXT7-GAL2-tHXT7	Este ejemplo
pRS313-Gal2-TT	CEN6/ARS4 HIS3 ampR pHXT7-GAL2 ^{N376T} -tHXT7	Este ejemplo
pRS313-Gal2-ST	CEN6/ARS4 HIS3 ampR pHXT7-GAL2 ^{N376S} -tHXT7	Este ejemplo
pRS313-Gal2-IT	CEN6/ARS4 HIS3 ampR pHXT7-GAL2 ^{N376I} -tHXT7	Este ejemplo
pRS313-Gal2-TI	CEN6/ARS4 HIS3 ampR pHXT7-GAL2 ^{N376IT89I} -tHXT7	Este ejemplo
pRS313-Gal2-NI	CEN6/ARS4 HIS3 ampR pHXT7-GAL2 ^{T89I} -tHXT7	Este ejemplo

* Nijland JG, Shin HY, de Jong RM, De Waal PP, Klaassen P, Driessen AJ: Engineering of an endogenous hexose transporter into a specific D-xylose transporter facilitates glucose-xylose co-consumption in *Saccharomyces cerevisiae*. *Biotechnol Biofuels* 2014, 7:168.

Para construir plásmidos de expresión, *GAL2* de tipo salvaje y las variantes de *GAL2* mencionadas anteriormente se usaron como molde y se amplificaron usando los cebadores F_Gal2_XbaI (SEQ ID NO: 3) y R_Gal2_Cfr9I (SEQ ID NO: 4). Los fragmentos de PCR resultantes se clonaron posteriormente en el plásmido pRS313-P7T7 de manera similar a como se describe en Nijland et al., 2014 (*Biotechnology for Biofuels*, vol. 7, p. 168). Todos los genes se clonaron en pRS313-P7T7, y posteriormente se confirmaron mediante secuenciación de Sanger (Baseclear, Leiden, Países Bajos).

Las cepas de *S. cerevisiae* se transformaron como se describió por Gietz y Woods (Gietz RD, Schiestl RH, Willems AR, Woods RA: Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast* 1995, 11). Las variantes de Gal2 y el plásmido pRS313-P7T7-mcs (como plásmido vacío/control) se transformaron en la cepa DS68625 de supresión del transportador de hexosa. Los transformantes se seleccionaron en placas de agar SMM en base a la complementación mediante los plásmidos de expresión de *GAL2* (que portan el marcador *HIS3*) de la auxotrofia de histidina de DS68625. Las colonias positivas resultantes se seleccionaron para cada construcción, y se denominaron DS68625-Gal2, DS68625-Gal2 N376I, DS68625-Gal2 N376S, DS68625-Gal2 N376T, DS68625-Gal2 N376T / T89I, DS68625-Gal2 T89I y DS68625-mcs.

Ensayos de transporte de azúcar

Las cepas se cultivaron previamente en matraces de agitación aeróbica en SMX, después de lo cual las células se recogieron por centrifugación (3000 g, 3 min), se lavaron, y se resuspendieron en SM sin azúcar. Los experimentos de absorción se iniciaron añadiendo [¹⁴C] L-arabinosa o [¹⁴C] D-glucosa (ARC St. Louis, MO, USA) a la suspensión celular con concentraciones variables entre 0,2 y 500 mM. [¹⁴C] L-arabinosa y [¹⁴C] D-glucosa se añadieron 50-60 mCi/mmol a una concentración de 0,1 mCi/ml, y se marcaron en el primer átomo de carbono. En los puntos de tiempo establecidos, la absorción se detuvo añadiendo 5 ml de LiCl 0,1 M enfriado con hielo, filtrando sobre filtros de membrana HV de 0,45 µm (Millipore, Francia), y lavando con 5 ml de LiCl 0,1 M enfriado con hielo. A continuación, se contó la radiactividad en los filtros usando un contador de centelleo de líquidos (PerkinElmer, Waltham, MA, USA). En el cóctel de centelleo Ultima Gold MV (PerkinElmer, Waltham, MA, USA). Los experimentos de inhibición de D-glucosa se midieron usando 50 mM de [¹⁴C] L-arabinosa, con [¹⁴C] D-glucosa añadida en concentraciones entre 50 y 500 mM.

Resultados

El efecto inhibitorio de D-glucosa en el transporte de L-arabinosa de las posiciones N376 y T89 se evaluó en DS68625 de *S. cerevisiae*. Para examinar en qué medida D-glucosa inhibió la absorción de L-arabinosa, se realizaron experimentos de absorción de [¹⁴C]-L-arabinosa a diferentes concentraciones de D-glucosa (Figura 1). La absorción de L-arabinosa por Gal2 de tipo salvaje fue severamente inhibida por D-glucosa. Las sustituciones Gal2^{N376T} y Gal2^{N376T/T89I} disminuyeron significativamente la inhibición de la absorción de L-arabinosa por D-glucosa, mientras la inhibición de D-glucosa no se vio afectada significativamente por solo la sustitución Gal2^{T89I} (Figura 1).

En ensayos de transporte con [¹⁴C]-L-arabinosa o [¹⁴C]-D-glucosa, las sustituciones Gal2^{N37I}, Gal2^{N376S} y Gal2^{N376T} produjeron valores de K_m más bajos para L-arabinosa que Gal2 de tipo salvaje, mientras que sus valores de K_m para D-glucosa fueron uno a dos órdenes de magnitud mayores que los de Gal2 de tipo salvaje. Como resultado, la relación de los valores de K_m para L-arabinosa frente a D-glucosa fue 2 órdenes de magnitud menor para las variantes de Gal2 de expresión de cepas con sustituciones en la posición 376. Para las variantes de Gal2 que solo portaban una sustitución de aminoácido en esta posición, las capacidades de transporte (V_{max}) para L-arabinosa y D-glucosa diferían en menos de dos veces de las de Gal2 de tipo salvaje (Tabla 3). Estos cambios en la cinética de transporte son consistentes con una inhibición competitiva fuertemente reducida del transporte de L-arabinosa por D-glucosa. Una sola sustitución Gal2^{T89I} solo aumentó ligeramente la K_m para el transporte de D-glucosa, mientras que disminuyó la K_m para L-arabinosa 335 a 99 mM. Para ambos azúcares, la sustitución Gal2^{T89I} provocó una reducción de 2-3 veces de V_{max} (Tabla 3). Estas propiedades cinéticas sugieren que Gal2^{T89I} puede conferir, por sí mismo, una ventaja selectiva a concentraciones de L-arabinosa extracelulares más bajas. Sorprendentemente, una variante de Gal2 que albergaba tanto las sustituciones Gal2^{N376T} como Gal2^{T89I} ya no transportó D-glucosa, mientras que K_m y V_{max} para L-arabinosa fueron similares a los de Gal2^{T89I} (Tabla 3).

Tabla 3. Valores de K_m y V_{max} para L-arabinosa y D-glucosa para variantes de Gal2 con sustituciones de aminoácidos en la posición N376 y T89. Las cinéticas de transporte se midieron mediante estudios de absorción con azúcares radiactivos tras la expresión de alelos de GAL2 en *S. cerevisiae* DS68625. Los valores son la media y la desviación media de dos conjuntos independientes de experimentos de absorción. Los datos usados para calcular K_m y V_{max} se muestran en la Figura 3.

Variante de Gal2	Km (mM)		Relación de Km Ara/Glc	Vmax (nmol (mg de biomasa)/min)	
	L-arabinosa	D-glucosa		L-arabinosa	D-glucosa
Gal2 ¹	335 ± 21	1,9 ± 0,2	176	75 ± 5	26 ± 1
Gal2 ^{N376I}	117 ± 16	101 ± 47	1	39 ± 3	32 ± 18
Gal2 ^{N376S}	186 ± 33	38 ± 1	5	64 ± 2	28 ± 1
Gal2 ^{N376T}	171 ± 17	57 ± 1	3	65 ± 2	17 ± 4
Gal2 ^{T89I + N376T}	103 ± 40	-	-	30 ± 2	*)
Gal2 ^{T89I}	99 ± 18	7 ± 0,2	15	22 ± 3	13 ± 0,1

*) El límite de detección para determinar la velocidad de absorción de d-glucosa Vmax es 1,8 nmol⁻¹ (mg de biomasa)⁻¹ min⁻¹.

LISTADO DE SECUENCIAS

40 <110> DSM IP Assets B.V.

<120> POLIPÉPTIDOS CON ESPECIFICIDAD DE TRANSPORTE DE ARABINOSA MEJORADA

<130> 32618-WO-PCT

<160> 10

<170> BiSSAP 1.3.6

5

<210> 1

<211> 574

<212> PRT

<213> *Saccharomyces cerevisiae*

10

<220>

<223> Secuencia de aminoácidos de *S. cerevisiae* WT Gal2p CEN.PK113-7D

<400> 1

```

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

```

ES 2 905 103 T3

				325						330					335
Val	Gln	Met	Phe	Gln	Gln	Leu	Thr	Gly	Asn	Asn	Tyr	Phe	Phe	Tyr	Tyr
			340					345					350		
Gly	Thr	Val	Ile	Phe	Lys	Ser	Val	Gly	Leu	Asp	Asp	Ser	Phe	Glu	Thr
		355					360					365			
Ser	Ile	Val	Ile	Gly	Val	Val	Asn	Phe	Ala	Ser	Thr	Phe	Phe	Ser	Leu
	370					375					380				
Trp	Thr	Val	Glu	Asn	Leu	Gly	Arg	Arg	Lys	Cys	Leu	Leu	Leu	Gly	Ala
385					390					395					400
Ala	Thr	Met	Met	Ala	Cys	Met	Val	Ile	Tyr	Ala	Ser	Val	Gly	Val	Thr
			405						410					415	
Arg	Leu	Tyr	Pro	His	Gly	Lys	Ser	Gln	Pro	Ser	Ser	Lys	Gly	Ala	Gly
			420					425					430		
Asn	Cys	Met	Ile	Val	Phe	Thr	Cys	Phe	Tyr	Ile	Phe	Cys	Tyr	Ala	Thr
		435				440						445			
Thr	Trp	Ala	Pro	Val	Ala	Trp	Val	Ile	Thr	Ala	Glu	Ser	Phe	Pro	Leu
	450					455					460				
Arg	Val	Lys	Ser	Lys	Cys	Met	Ala	Leu	Ala	Ser	Ala	Ser	Asn	Trp	Val
465					470					475					480
Trp	Gly	Phe	Leu	Ile	Ala	Phe	Phe	Thr	Pro	Phe	Ile	Thr	Ser	Ala	Ile
				485					490					495	
Asn	Phe	Tyr	Tyr	Gly	Tyr	Val	Phe	Met	Gly	Cys	Leu	Val	Ala	Met	Phe
			500					505					510		
Phe	Tyr	Val	Phe	Phe	Phe	Val	Pro	Glu	Thr	Lys	Gly	Leu	Ser	Leu	Glu
		515				520						525			
Glu	Ile	Gln	Glu	Leu	Trp	Glu	Glu	Gly	Val	Leu	Pro	Trp	Lys	Ser	Glu
	530				535					540					
Gly	Trp	Ile	Pro	Ser	Ser	Arg	Arg	Gly	Asn	Asn	Tyr	Asp	Leu	Glu	Asp
545					550					555					560
Leu	Gln	His	Asp	Asp	Lys	Pro	Trp	Tyr	Lys	Ala	Met	Leu	Glu		
				565					570						

<210> 2

<211> 1725

<212> ADN

<213> *Saccharomyces cerevisiae*

<220>

<223> Secuencia de ácido nucleico del gen GAL2 de *S. cerevisiae* CEN.PK113-7D

<400> 2

atggcagttg aggagaacaa tatgcctggt gtttcacagc aacccaagc tggatgaagc	60
gtgatctctt cactcagtaa agattcccat ttaagcgac aatctcaaaa gtattccaat	120
gatgaattga aagccggtga gtcagggcct gaaggctccc aaagtgttcc tatagagata	180
ccaagaagc ccatgtctga atatgttacc gtttccttgc tttgtttgtg tgttgccttc	240
ggcggcttca tgtttggctg ggataccggt actatttctg ggtttgttgt ccaaacagac	300
tttttgagaa ggtttggtat gaaacataag gatggtacc actatttgtc aaacgtcaga	360
acaggtttaa tcgtcgccat tttcaatatt ggctgtgcct ttggtggtat tatactttcc	420
aaaggtggag atatgtatgg ccgtaaaaag ggtccttcga ttgtcgtctc ggtttatata	480
gttggatatta tcattcaaat tgctctatc aacaagtgtt accaatattt cattggtaga	540
atcatatctg gtttgggtgt cggcggcatc gccgtcttat gtcctatggt gatctctgaa	600
attgctccaa agcacttgag aggcacacta gtttcttgtt atcagctgat gattactgca	660

5

ES 2 905 103 T3

ggtatctttt tgggctactg tactaattac ggtacaaaga gctattcgaa ctcagttcaa 720
 tggagagttc cattagggct atgtttcgtt tggtcattat ttatgattgg cgctttgacg 780
 ttagttcctg aatccccacg ttattttatgt gaggtgaata aggtagaaga cgccaagctt 840
 tccattgcta agtctaacaa ggtgtcacca gaggatcctg ccgtccaggc agagttagat 900
 ctgatcatgg ccggtataga agctgaaaaa ctggctggca atgcgtcctg gggggaatta 960
 ttttccacca agaccaaagt atttcaacgt ttgttgatgg gtgtatttgt tcaaagttc 1020
 caacaattaa ccggtaacaa ttattttttc tactacggta ccgttatttt caagtcagtt 1080
 ggcttgatg attcctttga aacatccatt gtcattgggtg tagtcaactt tgcctccact 1140
 ttcttttagtt tgtggactgt cgaaaacttg gggcgtcgta aatgtttact tttgggcgct 1200
 gccactatga tggcttgat ggtcatctac gcctctgttg gtgttaccag attatattct 1260
 cacggtaaaa gccagccatc ttctaaaggt gccggtaact gtatgattgt ctttacctgt 1320
 ttttatattt tctgttatgc cacaacctgg gcgccagttg cctgggtcat cacagcagaa 1380
 tcattcccac tgagagtcaa gtcgaaatgt atggcgttgg cctctgcttc caattgggta 1440
 tgggggttct tgattgcatt tttcacccca ttcattcacat ctgccattaa cttctactac 1500
 ggttatgtct tcatgggctg tttggttgcc atgttttttt atgtcttttt ctttgttcca 1560
 gaaactaaag gcctatcgtt agaagaaatt caagaattat gggagaagg tgttttacct 1620
 tggaaatctg aaggctggat tccttcatcc agaagaggta ataattacga ttttagaggat 1680
 ttacaacatg acgacaaacc gtggtacaag gccatgctag aataa 1725

- 5
 <210> 3
 <211> 35
 <212> ADN
 <213> Secuencia artificial
- <220>
 <223> F Gal2 XbaI
- 10
 <400> 3
 actcgtctag aatggcagtt gaggagaaca atatg 35
- <210> 4
 <211> 34
 <212> ADN
 <213> Secuencia artificial
- 15
 <220>
 <223> R Gal2 Cfr9I
- <400> 4
 gcagcccggg ttatcagact tcttcgtgag tggc 34
- 20
 <210> 5
 <211> 474
 <212> PRT
 <213> Lactobacillus plantarum
- <220>
 <223> L-arabinosa isomerasa (araA) de L. plantarum

ES 2 905 103 T3

<400> 5

```

Met Leu Ser Val Pro Asp Tyr Glu Phe Trp Phe Val Thr Gly Ser Gln
1      5      10      15
His Leu Tyr Gly Glu Glu Gln Leu Lys Ser Val Ala Lys Asp Ala Gln
20      25      30
Asp Ile Ala Asp Lys Leu Asn Ala Ser Gly Lys Leu Pro Tyr Lys Val
35      40      45
Val Phe Lys Asp Val Met Thr Thr Ala Glu Ser Ile Thr Asn Phe Met
50      55      60
Lys Glu Val Asn Tyr Asn Asp Lys Val Ala Gly Val Ile Thr Trp Met
65      70      75      80
His Thr Phe Ser Pro Ala Lys Asn Trp Ile Arg Gly Thr Glu Leu Leu
85      90      95
Gln Lys Pro Leu Leu His Leu Ala Thr Gln Tyr Leu Asn Asn Ile Pro
100     105     110
Tyr Ala Asp Ile Asp Phe Asp Tyr Met Asn Leu Asn Gln Ser Ala His
115     120     125
Gly Asp Arg Glu Tyr Ala Tyr Ile Asn Ala Arg Leu Gln Lys His Asn
130     135     140
Lys Ile Val Tyr Gly Tyr Trp Gly Asp Glu Asp Val Gln Glu Gln Ile
145     150     155     160
Ala Arg Trp Glu Asp Val Ala Val Ala Tyr Asn Glu Ser Phe Lys Val
165     170     175
Lys Val Ala Arg Phe Gly Asp Thr Met Arg Asn Val Ala Val Thr Glu
180     185     190
Gly Asp Lys Val Glu Ala Gln Ile Lys Met Gly Trp Thr Val Asp Tyr
195     200     205
Tyr Gly Ile Gly Asp Leu Val Glu Glu Ile Asn Lys Val Ser Asp Ala
210     215     220
Asp Val Asp Lys Glu Tyr Ala Asp Leu Glu Ser Arg Tyr Glu Met Val
225     230     235     240
Gln Gly Asp Asn Asp Ala Asp Thr Tyr Lys His Ser Val Arg Val Gln
245     250     255
Leu Ala Gln Tyr Leu Gly Ile Lys Arg Phe Leu Glu Arg Gly Gly Tyr
260     265     270
Thr Ala Phe Thr Thr Asn Phe Glu Asp Leu Trp Gly Met Glu Gln Leu
275     280     285
Pro Gly Leu Ala Ser Gln Leu Leu Ile Arg Asp Gly Tyr Gly Phe Gly
290     295     300
Ala Glu Gly Asp Trp Lys Thr Ala Ala Leu Gly Arg Val Met Lys Ile
305     310     315     320
Met Ser His Asn Lys Gln Thr Ala Phe Met Glu Asp Tyr Thr Leu Asp
325     330     335
Leu Arg His Gly His Glu Ala Ile Leu Gly Ser His Met Leu Glu Val
340     345     350
Asp Pro Ser Ile Ala Ser Asp Lys Pro Arg Val Glu Val His Pro Leu
355     360     365
Asp Ile Gly Gly Lys Asp Asp Pro Ala Arg Leu Val Phe Thr Gly Ser
370     375     380
Glu Gly Glu Ala Ile Asp Val Thr Val Ala Asp Phe Arg Asp Gly Phe
385     390     395     400
Lys Met Ile Ser Tyr Ala Val Asp Ala Asn Lys Pro Glu Ala Glu Thr
405     410     415

Pro Asn Leu Pro Val Ala Lys Gln Leu Trp Thr Pro Lys Met Gly Leu
420     425     430
Lys Lys Gly Ala Leu Glu Trp Met Gln Ala Gly Gly Gly His His Thr
435     440     445
Met Leu Ser Phe Ser Leu Thr Glu Glu Gln Met Glu Asp Tyr Ala Thr
450     455     460
Met Val Gly Met Thr Lys Ala Phe Leu Lys
465     470

```

5

<210> 6
 <211> 533
 <212> PRT
 <213> Lactobacillus plantarum

<220>

10

<223> L-ribulocinasa (araB) de L. plantarum

ES 2 905 103 T3

<400> 6

Met Asn Leu Val Glu Thr Ala Gln Ala Ile Lys Thr Gly Lys Val Ser
1 5 10 15
Leu Gly Ile Glu Leu Gly Ser Thr Arg Ile Lys Ala Val Leu Ile Thr
20 25 30
Asp Asp Phe Asn Thr Ile Ala Ser Gly Ser Tyr Val Trp Glu Asn Gln
35 40 45
Phe Val Asp Gly Thr Trp Thr Tyr Ala Leu Glu Asp Val Trp Thr Gly
50 55 60
Ile Gln Gln Ser Tyr Thr Gln Leu Ala Ala Asp Val Arg Ser Lys Tyr
65 70 75 80
His Met Ser Leu Lys His Ile Asn Ala Ile Gly Ile Ser Ala Met Met
85 90 95
His Gly Tyr Leu Ala Phe Asp Gln Gln Ala Lys Leu Leu Val Pro Phe
100 105 110
Arg Thr Trp Arg Asn Asn Ile Thr Gly Gln Ala Ala Asp Glu Leu Thr
115 120 125
Glu Leu Phe Asp Phe Asn Ile Pro Gln Arg Trp Ser Ile Ala His Leu
130 135 140
Tyr Gln Ala Ile Leu Asn Asn Glu Ala His Val Lys Gln Val Asp Phe
145 150 155 160
Ile Thr Thr Leu Ala Gly Tyr Val Thr Trp Lys Leu Ser Gly Glu Lys
165 170 175
Val Leu Gly Ile Gly Asp Ala Ser Gly Val Phe Pro Ile Asp Glu Thr
180 185 190
Thr Asp Thr Tyr Asn Gln Thr Met Leu Thr Lys Phe Ser Gln Leu Asp
195 200 205
Lys Val Lys Pro Tyr Ser Trp Asp Ile Arg His Ile Leu Pro Arg Val
210 215 220
Leu Pro Ala Gly Ala Ile Ala Gly Lys Leu Thr Ala Ala Gly Ala Ser
225 230 235 240
Leu Leu Asp Gln Ser Gly Thr Leu Asp Ala Gly Ser Val Ile Ala Pro
245 250 255
Pro Glu Gly Asp Ala Gly Thr Gly Met Val Gly Thr Asn Ser Val Arg
260 265 270
Lys Arg Thr Gly Asn Ile Ser Val Gly Thr Ser Ala Phe Ser Met Asn
275 280 285
Val Leu Asp Lys Pro Leu Ser Lys Val Tyr Arg Asp Ile Asp Ile Val
290 295 300
Met Thr Pro Asp Gly Ser Pro Val Ala Met Val His Val Asn Asn Cys
305 310 315 320
Ser Ser Asp Ile Asn Ala Trp Ala Thr Ile Phe His Glu Phe Ala Ala
325 330 335
Arg Leu Gly Met Glu Leu Lys Pro Asp Arg Leu Tyr Glu Thr Leu Phe
340 345 350
Leu Glu Ser Thr Arg Ala Asp Ala Asp Ala Gly Gly Leu Ala Asn Tyr
355 360 365
Ser Tyr Gln Ser Gly Glu Asn Ile Thr Lys Ile Gln Ala Gly Arg Pro
370 375 380
Leu Phe Val Arg Thr Pro Asn Ser Lys Phe Ser Leu Pro Asn Phe Met
385 390 395 400
Leu Thr Gln Leu Tyr Ala Ala Phe Ala Pro Leu Gln Leu Gly Met Asp
405 410 415
Ile Leu Val Asn Glu Glu His Val Gln Thr Asp Val Met Ile Ala Gln
420 425 430
Gly Gly Leu Phe Arg Thr Pro Val Ile Gly Gln Gln Val Leu Ala Asn
435 440 445
Ala Leu Asn Ile Pro Ile Thr Val Met Ser Thr Ala Gly Glu Gly Gly
450 455 460
Pro Trp Gly Met Ala Val Leu Ala Asn Phe Ala Cys Arg Gln Thr Ala
465 470 475 480
Met Asn Leu Glu Asp Phe Leu Asp Gln Glu Val Phe Lys Glu Pro Glu
485 490 495
Ser Met Thr Leu Ser Pro Glu Pro Glu Arg Val Ala Gly Tyr Arg Glu
500 505 510
Phe Ile Gln Arg Tyr Gln Ala Gly Leu Pro Val Glu Ala Ala Ala Gly
515 520 525
Gln Ala Ile Lys Tyr
530

ES 2 905 103 T3

<210> 7
 <211> 242
 <212> PRT
 <213> Lactobacillus plantarum

5 <220>
 <223> L-ribulosa-5-P4-epimerasa (araD) de L. plantarum

<400> 7
 Met Leu Glu Ala Leu Lys Gln Glu Val Tyr Glu Ala Asn Met Gln Leu
 1 5 10 15
 Pro Lys Leu Gly Leu Val Thr Phe Thr Trp Gly Asn Val Ser Gly Ile
 20 25 30
 Asp Arg Glu Lys Gly Leu Phe Val Ile Lys Pro Ser Gly Val Asp Tyr
 35 40 45
 Gly Glu Leu Lys Pro Ser Asp Leu Val Val Val Asn Leu Gln Gly Glu
 50 55 60
 Val Val Glu Gly Lys Leu Asn Pro Ser Ser Asp Thr Pro Thr His Thr
 65 70 75 80
 Val Leu Tyr Asn Ala Phe Pro Asn Ile Gly Gly Ile Val His Thr His
 85 90 95
 Ser Pro Trp Ala Val Ala Tyr Ala Ala Ala Gln Met Asp Val Pro Ala
 100 105 110
 Met Asn Thr Thr His Ala Asp Thr Phe Tyr Gly Asp Val Pro Ala Ala
 115 120 125
 Asp Ala Leu Thr Lys Glu Glu Ile Glu Ala Asp Tyr Glu Gly Asn Thr
 130 135 140
 Gly Lys Thr Ile Val Lys Thr Phe Gln Glu Arg Gly Leu Asp Tyr Glu
 145 150 155 160
 Ala Val Pro Ala Ser Leu Val Ser Gln His Gly Pro Phe Ala Trp Gly
 165 170 175
 Pro Thr Pro Ala Lys Ala Val Tyr Asn Ala Lys Val Leu Glu Val Val
 180 185 190
 Ala Glu Glu Asp Tyr His Thr Ala Gln Leu Thr Arg Ala Ser Ser Glu
 195 200 205
 Leu Pro Gln Tyr Leu Leu Asp Lys His Tyr Leu Arg Lys His Gly Ala
 210 215 220

 Ser Ala Tyr Tyr Gly Gln Asn Asn Ala His Ser Lys Asp His Ala Val
 225 230 235 240
 Arg Lys

10 <210> 8
 <211> 437
 <212> PRT
 <213> Piromyces sp. E2

15 <220>
 <223> Xilosa isomerasa (XylA) de Piromyces sp. E2

ES 2 905 103 T3

<400> 8

Met Ala Lys Glu Tyr Phe Pro Gln Ile Gln Lys Ile Lys Phe Glu Gly
 1 5 10 15
 Lys Asp Ser Lys Asn Pro Leu Ala Phe His Tyr Tyr Asp Ala Glu Lys
 20 25 30
 Glu Val Met Gly Lys Lys Met Lys Asp Trp Leu Arg Phe Ala Met Ala
 35 40 45
 Trp Trp His Thr Leu Cys Ala Glu Gly Ala Asp Gln Phe Gly Gly Gly
 50 55 60
 Thr Lys Ser Phe Pro Trp Asn Glu Gly Thr Asp Ala Ile Glu Ile Ala
 65 70 75 80
 Lys Gln Lys Val Asp Ala Gly Phe Glu Ile Met Gln Lys Leu Gly Ile
 85 90 95
 Pro Tyr Tyr Cys Phe His Asp Val Asp Leu Val Ser Glu Gly Asn Ser
 100 105 110
 Ile Glu Glu Tyr Glu Ser Asn Leu Lys Ala Val Val Ala Tyr Leu Lys
 115 120 125
 Glu Lys Gln Lys Glu Thr Gly Ile Lys Leu Leu Trp Ser Thr Ala Asn
 130 135 140
 Val Phe Gly His Lys Arg Tyr Met Asn Gly Ala Ser Thr Asn Pro Asp
 145 150 155 160
 Phe Asp Val Val Ala Arg Ala Ile Val Gln Ile Lys Asn Ala Ile Asp
 165 170 175
 Ala Gly Ile Glu Leu Gly Ala Glu Asn Tyr Val Phe Trp Gly Gly Arg
 180 185 190
 Glu Gly Tyr Met Ser Leu Leu Asn Thr Asp Gln Lys Arg Glu Lys Glu
 195 200 205
 His Met Ala Thr Met Leu Thr Met Ala Arg Asp Tyr Ala Arg Ser Lys
 210 215 220
 Gly Phe Lys Gly Thr Phe Leu Ile Glu Pro Lys Pro Met Glu Pro Thr
 225 230 235 240
 Lys His Gln Tyr Asp Val Asp Thr Glu Thr Ala Ile Gly Phe Leu Lys
 245 250 255
 Ala His Asn Leu Asp Lys Asp Phe Lys Val Asn Ile Glu Val Asn His
 260 265 270
 Ala Thr Leu Ala Gly His Thr Phe Glu His Glu Leu Ala Cys Ala Val
 275 280 285
 Asp Ala Gly Met Leu Gly Ser Ile Asp Ala Asn Arg Gly Asp Tyr Gln
 290 295 300
 Asn Gly Trp Asp Thr Asp Gln Phe Pro Ile Asp Gln Tyr Glu Leu Val
 305 310 315 320
 Gln Ala Trp Met Glu Ile Ile Arg Gly Gly Phe Val Thr Gly Gly
 325 330 335
 Thr Asn Phe Asp Ala Lys Thr Arg Arg Asn Ser Thr Asp Leu Glu Asp
 340 345 350
 Ile Ile Ile Ala His Val Ser Gly Met Asp Ala Met Ala Arg Ala Leu
 355 360 365
 Glu Asn Ala Ala Lys Leu Leu Gln Glu Ser Pro Tyr Thr Lys Met Lys
 370 375 380
 Lys Glu Arg Tyr Ala Ser Phe Asp Ser Gly Ile Gly Lys Asp Phe Glu
 385 390 395 400
 Asp Gly Lys Leu Thr Leu Glu Gln Val Tyr Glu Tyr Gly Lys Lys Asn
 405 410 415
 Gly Glu Pro Lys Gln Thr Ser Gly Lys Gln Glu Leu Tyr Glu Ala Ile
 420 425 430
 Val Ala Met Tyr Gln
 435

5

<210> 9
 <211> 600
 <212> PRT
 <213> Saccharomyces cerevisiae

10

<220>
 <223> Xilulocinasa de S. cerevisiae (Xkslp)

ES 2 905 103 T3

<400> 9

```

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

```

Ile Cys Tyr Cys Asn Gly Ser Leu Ala Arg Glu Arg Ile Arg Asp Glu
 355 360 365
 Leu Asn Lys Glu Arg Glu Asn Asn Tyr Glu Lys Thr Asn Asp Trp Thr
 370 375 380
 Leu Phe Asn Gln Ala Val Leu Asp Asp Ser Glu Ser Ser Glu Asn Glu
 385 390 395 400
 Leu Gly Val Tyr Phe Pro Leu Gly Glu Ile Val Pro Ser Val Lys Ala
 405 410 415
 Ile Asn Lys Arg Val Ile Phe Asn Pro Lys Thr Gly Met Ile Glu Arg
 420 425 430
 Glu Val Ala Lys Phe Lys Asp Lys Arg His Asp Ala Lys Asn Ile Val
 435 440 445
 Glu Ser Gln Ala Leu Ser Cys Arg Val Arg Ile Ser Pro Leu Leu Ser
 450 455 460
 Asp Ser Asn Ala Ser Ser Gln Gln Arg Leu Asn Glu Asp Thr Ile Val
 465 470 475 480
 Lys Phe Asp Tyr Asp Glu Ser Pro Leu Arg Asp Tyr Leu Asn Lys Arg
 485 490 495
 Pro Glu Arg Thr Phe Phe Val Gly Gly Ala Ser Lys Asn Asp Ala Ile
 500 505 510
 Val Lys Lys Phe Ala Gln Val Ile Gly Ala Thr Lys Gly Asn Phe Arg
 515 520 525
 Leu Glu Thr Pro Asn Ser Cys Ala Leu Gly Gly Cys Tyr Lys Ala Met
 530 535 540
 Trp Ser Leu Leu Tyr Asp Ser Asn Lys Ile Ala Val Pro Phe Asp Lys
 545 550 555 560
 Phe Leu Asn Asp Asn Phe Pro Trp His Val Met Glu Ser Ile Ser Asp
 565 570 575
 Val Asp Asn Glu Asn Trp Asp Arg Tyr Asn Ser Lys Ile Val Pro Leu
 580 585 590
 Ser Glu Leu Glu Lys Thr Leu Ile
 595 600

<210> 10

<211> 567

<212> PRT

<213> *Saccharomyces cerevisiae*

<220>

<223> Secuencia de aminoácidos de *S. cerevisiae* Hxt11p

<400> 10

Met Ser Gly Val Asn Asn Thr Ser Ala Asn Glu Leu Ser Thr Thr Met
 1 5 10 15
 Ser Asn Ser Asn Ser Ala Val Gly Ala Pro Ser Val Lys Thr Glu His
 20 25 30
 Gly Asp Ser Lys Asn Ser Leu Asn Leu Asp Ala Asn Glu Pro Pro Ile
 35 40 45
 Asp Leu Pro Gln Lys Pro Leu Ser Ala Tyr Thr Thr Val Ala Ile Leu
 50 55 60
 Cys Leu Met Ile Ala Phe Gly Gly Phe Ile Phe Gly Trp Asp Thr Gly
 65 70 75 80
 Thr Ile Ser Gly Phe Val Asn Leu Ser Asp Phe Ile Arg Arg Phe Gly
 85 90 95
 Gln Lys Asn Asp Lys Gly Thr Tyr Tyr Leu Ser Lys Val Arg Met Gly
 100 105 110
 Leu Ile Val Ser Ile Phe Asn Ile Gly Cys Ala Ile Gly Gly Ile Val
 115 120 125
 Leu Ser Lys Val Gly Asp Ile Tyr Gly Arg Arg Ile Gly Leu Ile Thr
 130 135 140
 Val Thr Ala Ile Tyr Val Val Gly Ile Leu Ile Gln Ile Thr Ser Ile
 145 150 155 160

5

ES 2 905 103 T3

Asn Lys Trp Tyr Gln Tyr Phe Ile Gly Arg Ile Ile Ser Gly Leu Gly
 165 170 175
 Val Gly Gly Ile Ala Val Leu Ser Pro Met Leu Ile Ser Glu Val Ala
 180 185 190
 Pro Lys His Ile Arg Gly Thr Leu Val Gln Leu Tyr Gln Leu Met Gly
 195 200 205
 Thr Met Gly Ile Phe Leu Gly Tyr Cys Thr Asn Tyr Gly Thr Lys Asn
 210 215 220
 Tyr His Asn Ala Thr Gln Trp Arg Val Gly Leu Gly Leu Cys Phe Ala
 225 230 235 240
 Trp Ala Thr Phe Met Val Ser Gly Met Met Phe Val Pro Glu Ser Pro
 245 250 255
 Arg Tyr Leu Ile Glu Val Gly Lys Asp Glu Glu Ala Lys Arg Ser Leu
 260 265 270
 Ser Lys Ser Asn Lys Val Ser Val Asp Asp Pro Ala Leu Leu Val Glu
 275 280 285
 Tyr Asp Thr Ile Lys Ala Gly Ile Glu Leu Glu Lys Leu Ala Gly Asn
 290 295 300
 Ala Ser Trp Ser Glu Leu Leu Ser Thr Lys Thr Lys Val Phe Gln Arg
 305 310 315 320
 Val Leu Met Gly Val Met Ile Gln Ser Leu Gln Gln Leu Thr Gly Asp
 325 330 335
 Asn Tyr Phe Phe Tyr Tyr Gly Thr Thr Ile Phe Lys Ser Val Gly Leu
 340 345 350
 Lys Asp Ser Phe Gln Thr Ser Ile Ile Ile Gly Val Val Asn Phe Phe
 355 360 365
 Ser Ser Phe Ile Ala Val Tyr Thr Ile Glu Arg Phe Gly Arg Arg Thr
 370 375 380
 Cys Leu Leu Trp Gly Ala Ala Ser Met Leu Cys Cys Phe Ala Val Phe
 385 390 395 400
 Ala Ser Val Gly Val Thr Lys Leu Trp Pro Gln Gly Ser Ser His Gln
 405 410 415
 Asp Ile Thr Ser Gln Gly Ala Gly Asn Cys Met Ile Val Phe Thr Met
 420 425 430
 Phe Phe Ile Phe Ser Phe Ala Thr Thr Trp Ala Gly Gly Cys Tyr Val
 435 440 445
 Ile Val Ser Glu Thr Phe Pro Leu Arg Val Lys Ser Arg Gly Met Ala
 450 455 460
 Ile Ala Thr Ala Ala Asn Trp Met Trp Gly Phe Leu Ile Ser Phe Phe
 465 470 475 480
 Thr Pro Phe Ile Thr Gly Ala Ile Asn Phe Tyr Tyr Gly Tyr Val Phe
 485 490 495
 Leu Gly Cys Leu Val Phe Ala Tyr Phe Tyr Val Phe Phe Phe Val Pro
 500 505 510
 Glu Thr Lys Gly Leu Thr Leu Glu Glu Val Asn Thr Met Trp Leu Glu
 515 520 525
 Gly Val Pro Ala Trp Lys Ser Ala Ser Trp Val Pro Pro Glu Arg Arg
 530 535 540
 Thr Ala Asp Tyr Asp Ala Asp Ala Ile Asp His Asp Asn Arg Pro Ile
 545 550 555 560
 Tyr Lys Arg Phe Phe Ser Ser
 565

REIVINDICACIONES

1. Una variante de un polipéptido parental que es preferiblemente un transportador de hexosa, en la que el polipéptido parental comprende la secuencia de aminoácidos de SEQ ID NO: 1, y en la que la variante comprende una secuencia de aminoácidos que, cuando se alinea con la secuencia de aminoácidos de SEQ ID NO: 1, comprende una sustitución de los aminoácidos N376 y T89, definiéndose las posiciones de dichos aminoácidos con referencia a la secuencia de aminoácidos de SEQ ID NO: 1, en la que la sustitución del aminoácido T89 comprende una cualquiera de T89I, T89V, T89L, T89S, T89M o T89F.
2. Una variante según la reivindicación 1, en la que la sustitución del aminoácido N376 comprende una cualquiera de N376T, N376C, N376V, N376M, N376L, N376I o N376F.
3. Una variante según una cualquiera de las reivindicaciones anteriores, en la que la sustitución del aminoácido N376 comprende una cualquiera de N376T, N376S o N376I, y en la que la sustitución del aminoácido T89 comprende T89I.
4. Una variante según una cualquiera de las reivindicaciones anteriores, que tiene al menos 80% de identidad de secuencia con la secuencia de aminoácidos del polipéptido parental.
5. Una variante según una cualquiera de las reivindicaciones anteriores, que es un polipéptido hecho por el hombre.
6. Una secuencia de ácido nucleico que codifica la variante según una cualquiera de las reivindicaciones 1 a 5, preferiblemente en la que dicha secuencia de ácido nucleico está hecha por el hombre.
7. Una construcción de ácido nucleico que comprende la secuencia de ácido nucleico de la reivindicación 6 unida operativamente a una o más secuencias de control capaces de dirigir la expresión de la variante polipeptídica en una célula adecuada.
8. Un casete de expresión recombinante que comprende la construcción de ácido nucleico de la reivindicación 7.
9. Una levadura recombinante que comprende la secuencia de ácido nucleico de la reivindicación 6, la construcción de ácido nucleico de la reivindicación 7, y/o el casete de expresión de la reivindicación 8.
10. Una levadura recombinante según la reivindicación 9, que comprende además uno o más genes que codifican una L-arabinosa isomerasa (E.C. 5.3.1.4; araA); uno o más genes que codifican una L-ribulocinasa (E.C. 2.7.1.16; araB); y uno o más genes que codifican una L-ribulosa-5-P-4-epimerasa (E.C. 5.1.3.4; araD).
11. Una levadura recombinante según la reivindicación 9 o 10, que comprende además uno o más genes que codifican una xilosa isomerasa (EC 5.3.1.5), tal como xylA, y uno o más genes que codifican una xilulosa cinasa (E.C. 2.7.1.17), tal como XKS1.
12. Una levadura recombinante según una cualquiera de las reivindicaciones 9-11, que comprende además la sobreexpresión de uno o más genes de la rama no oxidativa de la ruta de las pentosas fosfatos.
13. Una levadura recombinante según la reivindicación 12, en la que el uno o más genes de la ruta de las pentosas fosfatos que se sobreexpresan codifican una enzima seleccionada de la lista de una transaldolasa (EC 2.2.1.2), una transcetolasa (EC 2.2.1.1), una ribosa-5-fosfato isomerasa (EC 5.3.1.6), y una D-ribulosa-5-fosfato 3-epimerasa (EC 5.1.3.1).
14. Una levadura recombinante según una cualquiera de las reivindicaciones 12 a 13, en la que el uno o más genes de la ruta de las pentosas fosfatos que se sobreexpresan se seleccionan de la lista de TAL1, NQM1, TKL1, TKL2, RPE1, y RKI1.
15. Procedimiento para la producción de etanol, que comprende:
 - fermentar una composición que comprende una biomasa lignocelulósica, en particular que comprende glucosa y arabinosa, preferiblemente que también comprende galactosa, en condiciones anaeróbicas en presencia de una levadura recombinante según una cualquiera de las reivindicaciones 9-14; y
 - recuperar el etanol.
16. Procedimiento según la reivindicación 15, en el que la biomasa lignocelulósica comprende lignocelulosa y/o hemicelulosa y/o pectina.

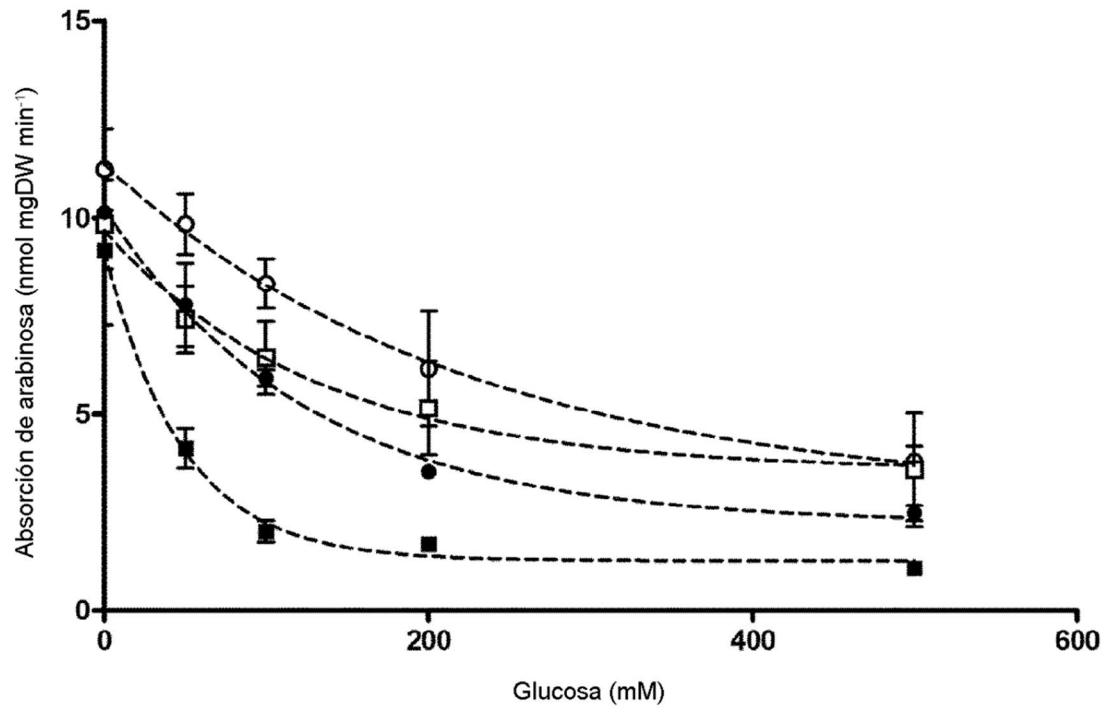


Fig. 1

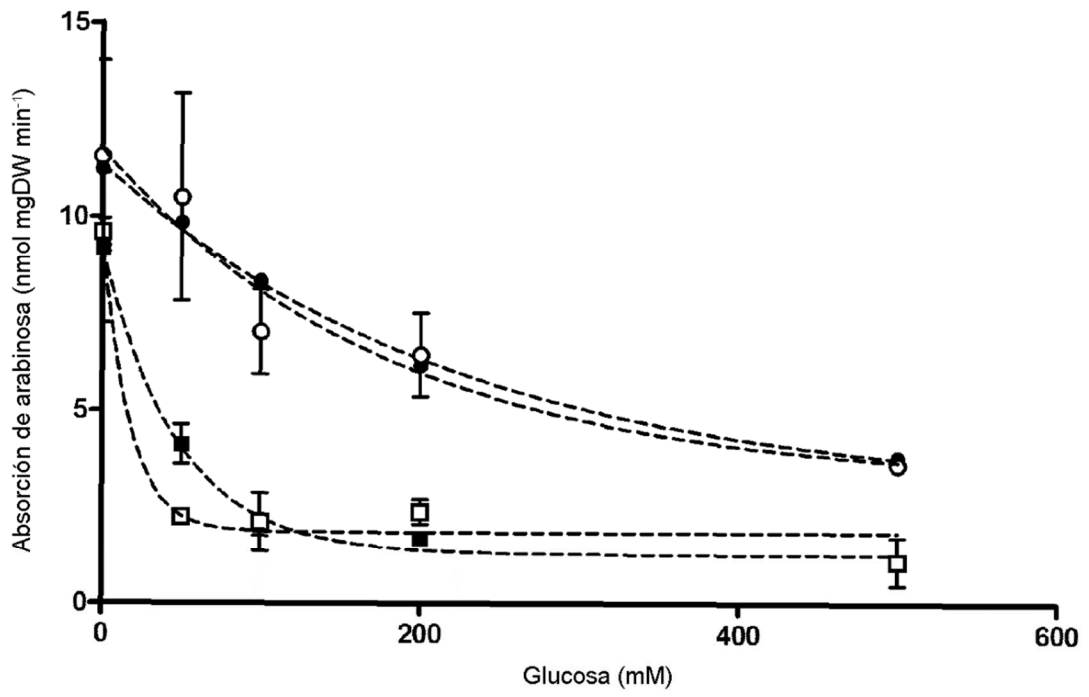


Fig. 2

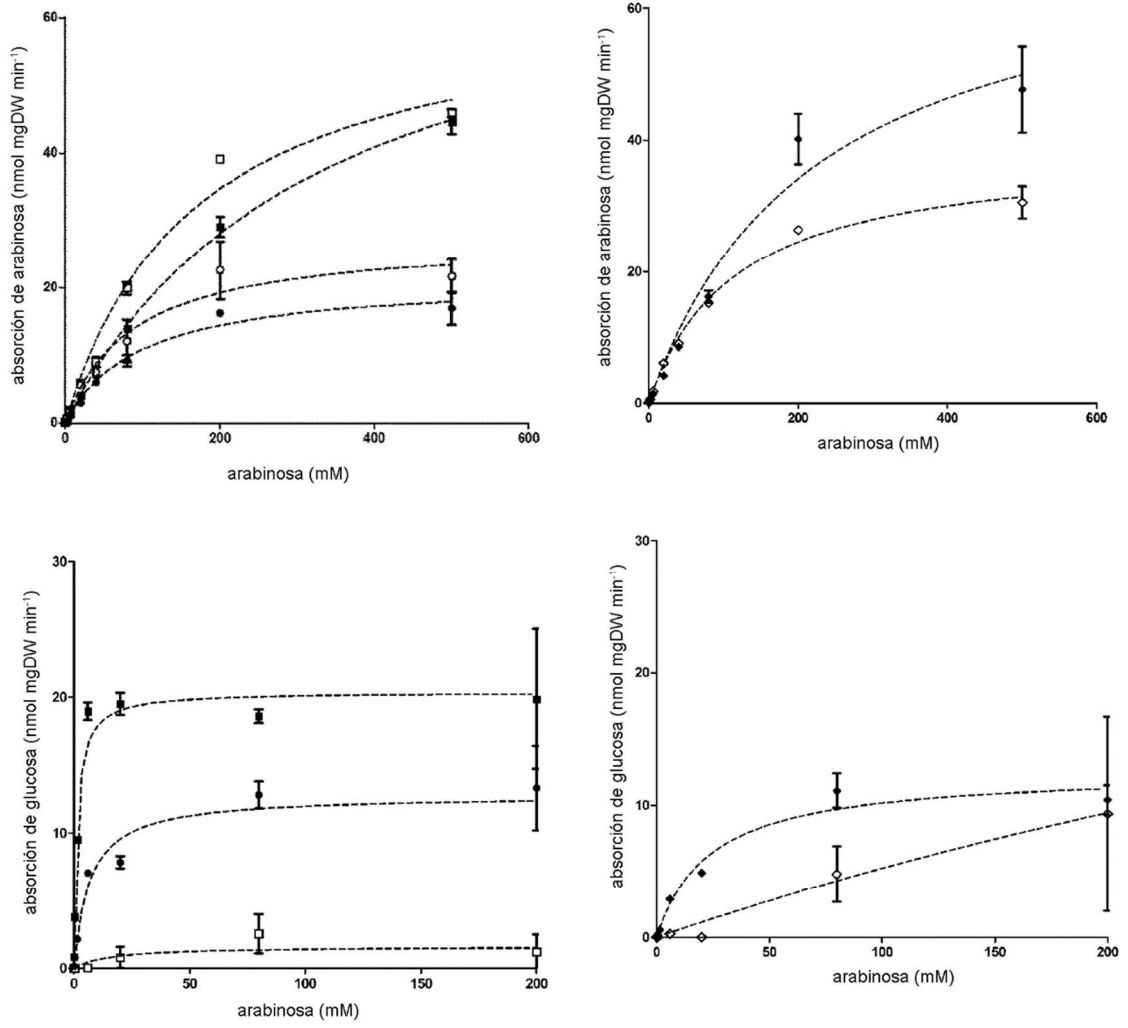


Fig. 3