

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 601 052**

51 Int. Cl.:

C12N 1/18 (2006.01)

C12P 7/08 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **01.10.2007 PCT/NL2007/000246**

87 Fecha y número de publicación internacional: **10.04.2008 WO08041840**

96 Fecha de presentación y número de la solicitud europea: **01.10.2007 E 07834562 (6)**

97 Fecha y número de publicación de la concesión europea: **10.08.2016 EP 2069476**

54 Título: **Ingeniería metabólica de células de levadura que fermentan arabinosa**

30 Prioridad:

02.10.2006 EP 06121633

02.10.2006 US 848357 P

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

14.02.2017

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;

WISSELINK, HENDRIK, WOUTER;

VAN DIJKEN, JOHANNES, PIETER;

WINKLER, AARON, ADRIAAN y

DE WINDE, JOHANNES, HENDRIK

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 601 052 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

Ingeniería metabólica de células de levadura que fermentan arabinosa

Campo de la invención

5 La invención se refiere a una célula de levadura que tiene la capacidad de usar L-arabinosa y/o de convertir L-arabinosa en L-ribulosa y/o xilulosa-5-fosfato y/o en un producto de fermentación deseado y a un proceso para producir un producto de fermentación en el que se usa esta célula.

Antecedentes de la invención

10 El etanol para combustible es conocido como una valiosa alternativa a los combustibles fósiles. La producción de etanol económicamente viable a partir de la fracción de hemicelulosa de la biomasa vegetal requiere la conversión fermentativa simultánea de tanto pentosas como hexosas a velocidades comparables y con altos rendimientos. Las levaduras, en particular *Saccharomyces spp.*, son los candidatos más apropiados para este proceso, ya que pueden crecer y fermentar rápido sobre las hexosas, tanto aerobia como anaerobiamente. Además, son mucho más resistentes al entorno tóxico de los hidrolizados de lignocelulosa que las bacterias (genéticamente modificadas).

15 El documento EP 1 499 708 describe un proceso de producción de cepas de *S. cerevisiae* capaces de producir etanol a partir de L-arabinosa. Estas cepas se modificaron introduciendo el gen *araA* (L-arabinosa isomerasa) de *Bacillus subtilis*, los genes *araB* (L-ribulocinasa) y *araD* (L-ribulosa-5-P4-epimerasa) de *Escherichia coli*. Además, estas cepas estaban tanto llevando mutaciones adicionales en su genoma como expresando en exceso un gen de TAL1 (transaldolasa). Sin embargo, estas cepas tienen varios inconvenientes. Fermentan arabinosa en condiciones de oxígeno limitado. Además, tienen una baja velocidad de producción de etanol de $0,05 \text{ g} \cdot \text{g}^{-1} \cdot \text{h}^{-1}$ (Becker y Boles, 2003). Además, estas cepas no son capaces de usar L-arabinosa en condiciones anaerobias. Finalmente, estas cepas de *S. cerevisiae* tiene un antecedente no mutante, por tanto no pueden usarse para co-fermentar varios azúcares C5.

Los documentos WO 03/062430 y WO 06/009434 desvelan cepas de levadura capaces de convertir xilosa en etanol. Estas cepas de levadura son capaces de isomerizar directamente xilosa en xilulosa.

25 Sin embargo, existe la necesidad de cepas alternativas que produzcan etanol, que den mejor rendimiento y sean más robustas y resistentes a condiciones de producción relativamente agresivas.

Descripción de las figuras

Figura 1. Mapas de plásmidos de pRW231 y pRW243.

30 Figura 2. Patrón de crecimiento de cultivos en matraz oscilante de la cepa RWB219 (○) y IMS0001 (●) en medio sintético que contiene 0,5 % de galactosa (A) y 0,1 % de galactosa + 2 % de L-arabinosa (B). Los cultivos se cultivaron durante 72 horas en medio sintético con galactosa (A) y luego se transfirieron a medio sintético con galactosa y arabinosa (B). El crecimiento se determinó midiendo la DO_{660} .

35 Figura 3. Velocidad de crecimiento durante transferencias en serie de IMS0001 de *S. cerevisiae* en cultivos de matraz oscilante que contienen medio sintético con 2 % (peso/volumen) de L-arabinosa. Cada punto de datos representa la velocidad de crecimiento estimada a partir de la DO_{660} medida durante el crecimiento (exponencial). Los círculos negros y blancos representan experimentos de transferencia en serie por duplicado.

Figura 4. Velocidad de crecimiento durante una fermentación en SBR anaerobia de IMS0001 de *S. cerevisiae* en medio sintético con 2 % (peso/volumen) de L-arabinosa. Cada punto de datos representa la velocidad de crecimiento estimada a partir del perfil de CO_2 (línea continua) durante el crecimiento exponencial.

40 Figura 5. Consumo de azúcar y formación de productos durante las fermentaciones discontinuas anaerobias de la cepa IMS0002. Las fermentaciones se realizaron en 1 medio sintético complementado con: 20 g l^{-1} de arabinosa (A); 20 g l^{-1} de glucosa y 20 g l^{-1} de arabinosa (B); 30 g l^{-1} de glucosa, 15 g l^{-1} de xilosa y 15 g l^{-1} de arabinosa (C); Consumo de azúcar y formación de productos durante las fermentaciones discontinuas anaerobias con una mezcla de cepas IMS0002 y RWB218. Las fermentaciones se realizaron en 1 litro de medio sintético complementado con 30 g l^{-1} de glucosa, 15 g l^{-1} de xilosa y 15 g l^{-1} de arabinosa (D). Símbolos: glucosa (●); xilosa (○); arabinosa (■); etanol calculado a partir de la producción acumulada de CO_2 (□); etanol medido por HPLC (▲); producción acumulada de CO_2 (Δ); xilitol (▼)

50 Figura 6. Consumo de azúcar y formación de productos durante una fermentación discontinua anaerobia de células de la cepa IMS0002 seleccionadas para el crecimiento anaerobio sobre xilosa. La fermentación se realizó en 1 litro de medio sintético complementado con 20 g l^{-1} de xilosa y 20 g l^{-1} de arabinosa. Símbolos: xilosa (○); arabinosa (■); etanol medido por HPLC (▲); producción acumulada de CO_2 (Δ); xilitol (▼).

Figura 7. Consumo de azúcar y formación de productos durante una fermentación discontinua anaerobia de la cepa IMS0003. La fermentación se realizó en 1 litro de medio sintético complementado con: 30 g l^{-1} de glucosa,

15 g l⁻¹ de xilosa y 15 g l⁻¹ de arabinosa. Símbolos: glucosa (●); xilosa (○); arabinosa (■); etanol calculado a partir de la producción acumulada de CO₂ (□); etanol medido por HPLC (▲); producción acumulada de CO₂ (Δ);

Descripción de la invención

5 Célula de levadura

En un primer aspecto, la invención se refiere a una célula de levadura capaz de expresar las siguientes secuencias de nucleótidos, por lo que la expresión de estas secuencias de nucleótidos confiere a la célula de levadura la capacidad de usar L-arabinosa y/o de convertir L-arabinosa en L-ribulosa y/o xilulosa-5-fosfato y/o en etanol:

10 (a) una secuencia de nucleótidos que codifica una arabinosa isomerasa (*araA*), en la que dicha secuencia de nucleótidos está seleccionada del grupo que consiste en:

(i) secuencias de nucleótidos que codifican una *araA*, comprendiendo dicha *araA* una secuencia de aminoácidos que tiene al menos el 80 % de identidad de secuencias con la secuencia de aminoácidos de SEQ ID NO: 1.

15 (ii) secuencias de nucleótidos que comprenden una secuencia de nucleótidos que tiene al menos el 80 % de identidad de secuencias con la secuencia de nucleótidos de SEQ ID NO: 2.

(iii) secuencias de nucleótidos cuya hebra complementaria se hibrida con una molécula de ácido nucleico de secuencia de (i) o (ii);

(b) una secuencia de nucleótidos que codifica una L-ribulocinasa (*araB*), en la que dicha secuencia de nucleótidos está seleccionada del grupo que consiste en:

20 (i) secuencias de nucleótidos que codifican una *araB*, comprendiendo dicha *araB* una secuencia de aminoácidos que tiene al menos el 80 % de identidad de secuencias con la secuencia de aminoácidos de SEQ ID NO: 3.

(ii) secuencias de nucleótidos que comprenden una secuencia de nucleótidos que tiene al menos el 80 % de identidad de secuencias con la secuencia de nucleótidos de SEQ ID NO: 4.

25 (iii) secuencias de nucleótidos cuya hebra complementaria se hibrida con una molécula de ácido nucleico de secuencia de (i) o (ii);

(c) una secuencia de nucleótidos que codifica una L-ribulosa-5-P-4-epimerasa (*araD*), en la que dicha secuencia de nucleótidos está seleccionada del grupo que consiste en:

30 (i) secuencias de nucleótidos que codifican una *araD*, comprendiendo dicha *araD* una secuencia de aminoácidos que tiene al menos el 80 % de identidad de secuencias con la secuencia de aminoácidos de SEQ ID NO: 5.

(ii) secuencias de nucleótidos que comprenden una secuencia de nucleótidos que tiene al menos el 70 % de identidad de secuencias con la secuencia de nucleótidos de SEQ ID NO: 6.

35 (iii) secuencias de nucleótidos cuya hebra complementaria se hibrida con una molécula de ácido nucleico de secuencia de (i) o (ii);

(iv)

40 en la que en los puntos (iii) la hibridación se determina bajo condiciones de hibridación que permiten que una secuencia de ácidos nucleicos de 200 nucleótidos se hibride a una temperatura de 65 °C en una disolución que comprende sal aproximadamente 1 M, y lavar a 65 °C en una disolución de aproximadamente 0,1 M, en la que la hibridación se realiza durante 10 horas y el lavado se realiza una hora con dos cambios de la disolución de lavado.

BLANCO

Identidad y similitud de secuencias

45 La identidad de secuencias se define en el presente documento como una relación entre dos o más secuencias de aminoácidos (polipéptidos o proteínas) o dos o más secuencias de ácidos nucleicos (polinucleótidos), como se determina comparando las secuencias. Normalmente, las identidades o similitudes de secuencia se comparan a lo largo de la longitud completa de las secuencias comparadas. En la materia, "identidad" también significa el grado de relación de secuencias entre secuencias de aminoácidos o de ácidos nucleicos, según sea el caso, como se ha determinado por la coincidencia entre lazos de tales secuencias. La "similitud" entre dos secuencias de aminoácidos se determina comparando la secuencia de aminoácidos y sus sustitutos de aminoácido conservados de un

polipéptido con la secuencia de un segundo polipéptido. La "identidad" y "similitud" pueden ser fácilmente calculadas por diversos métodos, conocidos para aquellos expertos en la materia.

Métodos preferidos para determinar la identidad se diseñan para dar la mayor coincidencia entre las secuencias probadas. Los métodos de determinación de la identidad y similitud están codificados en programas informáticos públicamente disponibles. Métodos de programas informáticos preferidos para determinar la identidad y similitud entre dos secuencias incluyen, por ejemplo, BestFit, BLASTP, BLASTN y FASTA (Altschul, S. F. et al., J. Mol. Biol. 215:403-410 (1990), públicamente disponibles de NCBI y otras fuentes (BLAST Manual, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894). Un algoritmo más preferido usado es EMBOSS (<http://www.ebi.ac.uk/emboss/align>). Parámetros preferidos para la comparación de secuencias de aminoácidos usando EMBOSS son abertura por hueco 10,0, extensión por hueco 0,5, matriz Blosum 62. Parámetros preferidos para la comparación de secuencias de ácidos nucleicos usando EMBOSS son abertura por hueco 10,0, extensión por hueco 0,5, matriz completa de ADN (matriz de identidad de ADN).

Opcionalmente, en la determinación del grado de similitud de aminoácidos, el experto también puede tener en cuenta las llamadas sustituciones de aminoácidos "conservativas", como será evidente para el experto. Sustituciones de aminoácidos conservativas se refieren a la intercambiabilidad de restos que tienen cadenas laterales similares. Por ejemplo, un grupo de aminoácidos que tienen cadenas laterales alifáticas es glicina, alanina, valina, leucina e isoleucina; un grupo de aminoácidos que tienen cadenas laterales de hidroxilo alifáticas es serina y treonina; un grupo de aminoácidos que tienen cadenas laterales que contienen amida es asparagina y glutamina; un grupo de aminoácidos que tienen cadenas laterales aromáticas es fenilalanina, tirosina y triptófano; un grupo de aminoácidos que tienen cadenas laterales básicas es lisina, arginina e histidina; y un grupo de aminoácidos que tienen cadenas laterales que contienen azufre es cisteína y metionina. Grupos de sustitución de aminoácidos conservativa preferidos son: valina-leucina-isoleucina, fenilalanina-tirosina, lisina-arginina, alanina-valina y asparagina-glutamina. Variantes de sustitución de la secuencia de aminoácidos desveladas en el presente documento son aquellas en las que se ha eliminado al menos un resto en las secuencias desveladas y en su lugar se ha insertado un resto diferente. Preferentemente, el cambio de aminoácido es conservativo. Sustituciones conservativas preferidas para cada uno de los aminoácidos que existen de forma natural son las siguientes: Ala por ser; Arg por lys; Asn por gln o his; Asp por glu; Cys por ser o ala; Gln por asn; Glu por asp; Gly por pro; His por asn o gln; Ile por leu o val; Leu por ile o val; Lys por arg; gln o glu; Met por leu o ile; Phe por met, leu o tyr; Ser por thr; Thr por ser; Trp por tyr; Tyr por trp o phe; y, Val por ile o leu.

30 Hibridación de secuencias de ácidos nucleicos

También pueden definirse secuencias de nucleótidos que codifican las enzimas expresadas en la célula de la invención por su capacidad para hibridarse con las secuencias de nucleótidos de SEQ ID NO. 2, 4, 6, 8, 16, 18, 20, 22, 24, 26, 28, 30 respectivamente, bajo condiciones de hibridación rigurosas. Condiciones de hibridación rigurosas se definen en el presente documento como las condiciones que permiten que una secuencia de ácidos nucleicos de 200 nucleótidos se hibride a una temperatura de 65 °C en una disolución que comprende sal 1 M y lavar a 65 °C en una disolución que comprende sal aproximadamente 0,1 M. La hibridación se realiza durante 10 horas y el lavado se realiza durante una hora con dos cambios de la disolución de lavado.

AraA

Una secuencia de nucleótidos preferida que codifica una arabinosa isomerasa (araA) expresada en la célula de la invención está seleccionada del grupo que consiste en:

(a) secuencias de nucleótidos que codifican un polipéptido araA, comprendiendo dicha araA una secuencia de aminoácidos que tiene al menos el 80, 85, 90, 95, 97, 98 o el 99 % de identidad de secuencias con la secuencia de aminoácidos de SEQ ID NO. 1;

(b) secuencias de nucleótidos que comprenden una secuencia de nucleótidos que tiene al menos el 80, 90, 95, 97, 98 o el 99 % de identidad de secuencias con la secuencia de nucleótidos de SEQ ID NO. 2;

(c) secuencias de nucleótidos cuya hebra complementaria se hibrida con una secuencia de la molécula de ácido nucleico de (a) o (b);

La secuencia de nucleótidos que codifica araA puede codificar tanto una araA procariota como una eucariota, es decir, una araA con una secuencia de aminoácidos que es idéntica a la de una araA que se produce naturalmente en el organismo procariota o eucariota. Los presentes inventores han encontrado que la capacidad de una araA particular para conferir a una célula huésped eucariota la capacidad de usar arabinosa y/o de convertir arabinosa en L-ribulosa y/o xilulosa-5-fosfato y/o en un producto de fermentación deseado tal como etanol cuando se co-expresa con araB y araD no depende tanto de si la araA es de origen procariota o eucariota. Más bien, esto depende de la relación de la secuencia de aminoácidos de araA con la de la secuencia SEQ ID NO. 1.

AraB

Una secuencia de nucleótidos preferida que codifica una L-ribulocinasa (AraB) expresada en la célula de la invención está seleccionada del grupo que consiste en:

- 5 (a) secuencias de nucleótidos que codifican un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 80, 85, 90, 95, 97, 98 o el 99 % de identidad de secuencias con la secuencia de aminoácidos de SEQ ID NO. 3;
- (b) secuencias de nucleótidos que comprenden una secuencia de nucleótidos que tiene al menos el 80, 90, 95, 97, 98 o el 99 % de identidad de secuencias con la secuencia de nucleótidos de SEQ ID NO. 4;
- 10 (c) secuencias de nucleótidos cuya hebra complementaria se hibrida con una secuencia de la molécula de ácido nucleico de (a) o (b);

15 La secuencia de nucleótidos que codifica una *araB* puede codificar tanto una *araB* procariota como una eucariota, es decir, una *araB* con una secuencia de aminoácidos que es idéntica a la de una *araB* que se produce naturalmente en el organismo procariota o eucariota. Los presentes inventores han encontrado que la capacidad de una *araB* particular para conferir a una célula huésped eucariota la capacidad de usar arabinosa y/o de convertir arabinosa en L-ribulosa y/o xilulosa-5-fosfato y/o en un producto de fermentación deseado cuando se co-expresa con *araA* y *araD* no depende tanto de si la *araB* es de origen procariota o eucariota. Más bien, esto depende de la relación de la secuencia de aminoácidos de *araB* con la de la secuencia SEQ ID NO. 3.

AraD

20 Una secuencia de nucleótidos preferida que codifica una L-ribulosa-5-P-4-epimerasa (*araD*) expresada en la célula de la invención está seleccionada del grupo que consiste en:

- (d) secuencias de nucleótidos que codifican un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 80, 85, 90, 95, 97, 98 o el 99 % de identidad de secuencias con la secuencia de aminoácidos de SEQ ID NO. 5;
- 25 (e) secuencias de nucleótidos que comprenden una secuencia de nucleótidos que tiene al menos el 80, 85, 90, 95, 97, 98 o el 99 % de identidad de secuencias con la secuencia de nucleótidos de SEQ ID NO. 6;
- (f) secuencias de nucleótidos cuya hebra complementaria se hibrida con una secuencia de la molécula de ácido nucleico de (a) o (b);
- (g) secuencias de nucleótidos cuya secuencia se diferencia de la secuencia de una molécula de ácido nucleico de (c) debido a la degeneración del código genético.

30 La secuencia de nucleótidos que codifica una *araD* puede codificar tanto una *araD* procariota como una eucariota, es decir, una *araD* con una secuencia de aminoácidos que es idéntica a la de una *araD* que se produce naturalmente en el organismo procariota o eucariota. Los presentes inventores han encontrado que la capacidad de una *araD* particular para conferir a una célula huésped eucariota la capacidad de usar arabinosa y/o de convertir arabinosa en L-ribulosa y/o xilulosa-5-fosfato y/o en un producto de fermentación deseado cuando se co-expresa con *araA* y *araB*

35 no depende tanto de si la *araD* es de origen procariota o eucariota. Más bien esto depende de la relación de la secuencia de aminoácidos de *araD* con la de la secuencia SEQ ID NO. 5.

Sorprendentemente, el índice codónico preferido indicó que la expresión de los genes *araA*, *araB* y *araD* de *Lactobacillus plantarum* fue más favorable para la expresión en levadura que los genes *araA*, *araB* y *araD* procariotas descritos en el documento EP 1 499 708.

40 Debe observarse que *L. plantarum* es un organismo Generalmente Considerado como Seguro (GRAS), que es reconocido como seguro por las autoridades de registro alimentario. Por tanto, una secuencia de nucleótidos preferida codifica una *araA*, *araB* o *araD*, respectivamente, que tiene una secuencia de aminoácidos que está relacionada con las secuencias SEQ ID NO: 1, 3 o 5, respectivamente, como se ha definido anteriormente. Una secuencia de nucleótidos preferida codifica una *araA*, *araB* o *araD* fúngica, respectivamente (por ejemplo, de un

45 *Basidiomycete*), más preferentemente una *araA*, *araB* o *araD*, respectivamente, de un hongo anaerobio, por ejemplo, un hongo anaerobio que pertenece a las familias *Neocallimastix*, *Caecomyces*, *Piromyces*, *Orpinomyces* o *Ruminomyces*. Alternativamente, una secuencia de nucleótidos preferida codifica una *araA*, *araB* o *araD* bacteriana, respectivamente, preferentemente de una bacteria Gram-positiva, más preferentemente del género *Lactobacillus*, lo más preferentemente de especies de *Lactobacillus plantarum*. Preferentemente, una, dos o tres o las secuencias de

50 nucleótidos de *araA*, *araB* y *araD* se originan a partir de un género de *Lactobacillus*, más preferentemente una especie de *Lactobacillus plantarum*. La *araA* bacteriana expresada en la célula de la invención no es la *araA* de *Bacillus subtilis* desvelada en el documento EP 1 499 708 y dada como SEQ ID NO: 9. SEQ ID NO: 10 representa la secuencia de ácido nucleótico que codifica SEQ ID NO: 9. Las *araB* y *araD* bacterianas expresadas en la célula de la invención no son las de *Escherichia coli* (*E. coli*) como se ha desvelado en el documento EP 1 499 708 y dada

como SEQ ID NO: 11 y SEQ ID NO: 13. SEQ ID NO: 12 representa la secuencia de ácido nucleótido que codifica SEQ ID NO: 11. SEQ ID NO: 14 representa la secuencia de ácido nucleótido que codifica SEQ ID NO: 13.

5 Para aumentar la probabilidad de que las enzimas araA, araB y araD (bacterianas), respectivamente, se expresen en forma activa en una célula huésped eucariota de la invención tal como levadura, la secuencia de nucleótidos codificante correspondiente puede adaptarse para optimizar su uso de codones al de la célula huésped eucariota elegida. La adaptabilidad de una secuencia de nucleótidos que codifica las enzimas araA, araB y araD (u otras enzimas de la invención, véase más adelante) al uso de codones de la célula huésped elegida puede expresarse como el índice de adaptación de codones (CAI). El índice de adaptación de codones se define en el presente documento como una medición de la adaptabilidad relativa del uso de codones de un gen hacia el uso de codones de genes altamente expresados. La adaptabilidad relativa (w) de cada codón es la relación del uso de cada codón, con la del codón más abundante para el mismo aminoácido. El índice CAI se define como la media geométrica de estos valores de la adaptabilidad relativa. Se excluyen codones no sinónimos y codones de terminación (dependientes del código genético). Los valores de CAI oscilan de 0 a 1, indicando valores más altos una mayor proporción de los codones más abundantes (véase Sharp y Li, 1987, *Nucleic Acids Research* 15: 1281-1295; véase también: Jansen et al., 2003, *Nucleic Acids Res.* 31(8):2242-51). Una secuencia de nucleótidos adaptada tiene preferentemente un CAI de al menos 0,2, 0,3, 0,4, 0,5, 0,6 o 0,7.

20 En una realización preferida, la expresión de las secuencias de nucleótidos que codifican una ara A, una ara B y una ara D como se definen anteriormente en el presente documento confiere a la célula la capacidad de usar L-arabinosa y/o de convertirla en L-ribulosa y/o xilulosa-5-fosfato. Sin desear quedar ligado a teoría alguna, se espera que la L-arabinosa se convierta primero en L-ribulosa, que posteriormente se convierte en xilulosa-5-fosfato que es la principal molécula que entra en la vía de pentosa fosfato. En el contexto de la invención, "usar L-arabinosa" significa preferentemente que la densidad óptica medida a 660 nm (DO_{660}) de células transformadas cultivadas bajo condiciones aerobias o anaerobias en presencia de al menos 0,5 % de L-arabinosa durante al menos 20 días aumenta de aproximadamente 0,5 hasta 1,0 o más. Más preferentemente, la DO_{660} aumenta de 0,5 hasta 1,5 o más. Más preferentemente, las células se cultivan en presencia de al menos 1 %, al menos 1,5 %, al menos 2 % de L-arabinosa. Lo más preferentemente, las células se cultivan en presencia de aproximadamente 2 % de L-arabinosa.

30 En el contexto de la invención, una célula es capaz "de convertir L-arabinosa en L-ribulosa" cuando cantidades detectables de L-ribulosa se detectan en células cultivadas bajo condiciones aerobias o anaerobias en presencia de L-arabinosa (mismas concentraciones preferidas que en el párrafo previo) durante al menos 20 días usando un ensayo adecuado. Preferentemente, el ensayo es HPLC para L-ribulosa.

35 En el contexto de la invención, una célula es capaz "de convertir L-arabinosa en xilulosa-5-fosfato" cuando se detecta un aumento de al menos el 2 % de xilulosa-5-fosfato en células cultivadas bajo condiciones aerobias o anaerobias en presencia de L-arabinosa (mismas concentraciones preferidas que en el párrafo previo) durante al menos 20 días usando un ensayo adecuado. Preferentemente, un ensayo basado en HPLC para xilulosa-5-fosfato se ha descrito en Zaldivar J., et al ((2002), *Appl. Microbiol. Biotechnol.*, 59:436-442). Este ensayo se describe brevemente en la parte experimental. Más preferentemente, el aumento es de al menos el 5 %, 10 %, 15 %, 20 %, 25 % o más.

40 En otra realización preferida, la expresión de las secuencias de nucleótidos que codifican una ara A, ara B y ara D como se define anteriormente en el presente documento confiere a la célula la capacidad de convertir L-arabinosa en un producto de fermentación deseado cuando se cultiva bajo condiciones aerobias o anaerobias en presencia de L-arabinosa (mismas concentraciones preferidas que en el párrafo previo) durante al menos un mes hasta un año. Más preferentemente, una célula es capaz de convertir L-arabinosa en un producto de fermentación deseado cuando cantidades detectables de un producto de fermentación deseado se detectan usando un ensayo adecuado y cuando las células se cultivan en las condiciones dadas en la frase previa. Incluso más preferentemente, el ensayo es HPLC. Incluso más preferentemente, el producto de fermentación es etanol.

50 Una célula para la transformación con las secuencias de nucleótidos que codifican las enzimas araA, araB y araD, respectivamente, como se ha descrito anteriormente, es preferentemente una célula huésped capaz de transporte activo o pasivo de xilosa dentro de e isomerización de xilosa dentro de la célula. La célula es preferentemente capaz de glicólisis activa. La célula puede contener adicionalmente una vía de pentosa fosfato endógena y puede contener actividad de xilulosa cinasa endógena de manera que la xilulosa isomerizada a partir de xilosa pueda metabolizarse a piruvato. La célula contiene preferentemente adicionalmente enzimas para la conversión de piruvato en un producto de fermentación deseado tal como etanol, ácido láctico, ácido 3-hidroxi-propiónico, ácido acrílico, ácido acético, ácido succínico, ácido cítrico, ácido málico, ácido fumárico, un aminoácido, 1,3-propano-diol, etileno, glicerol, butanol, un antibiótico β -lactámico o una cefalosporina. Puede hacerse que la célula sea capaz de producir butanol por introducción de uno o más genes de la vía de butanol como se desvela en el documento WO2007/041269.

60 Una célula preferida es naturalmente capaz de fermentación alcohólica, preferentemente, fermentación alcohólica anaerobia. La célula huésped tiene preferentemente adicionalmente una alta tolerancia al etanol, una alta tolerancia a pH bajo (es decir, capaz de crecer a un pH inferior a 5, 4, 3 o 2,5) y hacia ácidos orgánicos como ácido láctico, ácido acético o ácido fórmico, y productos de degradación de azúcares tales como furfural e hidroximetilfurfural, y

una alta tolerancia a temperaturas elevadas. Cualquiera de estas características o actividades de la célula huésped puede estar naturalmente presente en la célula huésped o puede introducirse o modificarse mediante selección genética o por modificación genética. Una célula huésped adecuada es un microorganismo eucariota como, por ejemplo, un hongo, sin embargo, la más adecuada como célula huésped son levaduras u hongos filamentosos.

5 Levaduras se definen en el presente documento como microorganismos eucariotas e incluyen todas las especies de la subdivisión Eumycotina (Alexopoulos, C. J., 1962, en: *Introductory Mycology*, John Wiley & Sons, Inc., New York) que crecen predominantemente en forma unicelular. Las levaduras pueden tanto crecer por gemación de un talo unicelular como pueden crecer por fisión del organismo. Levaduras preferidas como células huésped pertenecen a uno de los géneros *Saccharomyces*, *Kluyveromyces*, *Candida*, *Pichia*, *Schizosaccharomyces*, *Hansenula*, *Kloeckera*,
10 *Schwanniomyces* o *Yarrowia*. Preferentemente, la levadura es capaz de fermentación anaerobia, más preferentemente fermentación alcohólica anaerobia.

Los hongos filamentosos se definen en el presente documento como microorganismos eucariotas que incluyen todas las formas filamentosas de la subdivisión Eumycotina. Estos hongos se caracterizan por un micelio vegetativo compuesto de quitina, celulosa y otros polisacáridos complejos. Los hongos filamentosos de la presente invención son morfológica, fisiológica y genéticamente distintos de las levaduras. El crecimiento vegetativo por hongos filamentosos es por alargamiento de hifas y el catabolismo de carbono de la mayoría de los hongos filamentosos es aerobio estricto. Hongos filamentosos preferidos como células huésped pertenecen a uno de los géneros *Aspergillus*, *Trichoderma*, *Humicola*, *Acremonium*, *Fusarium* o *Penicillium*.

Con los años se han hecho sugerencias para la introducción de diversos organismos para la producción de bioetanol a partir de azúcares de cultivo. En la práctica, sin embargo, todos los principales procesos de producción de bioetanol han continuado usando las levaduras del género *Saccharomyces* como productor de etanol. Esto es debido a las muchas características atractivas de las especies de *Saccharomyces* para procesos industriales, es decir, una alta tolerancia a ácidos, a etanol y osmo-tolerancia, capacidad de crecimiento anaerobio y, por supuesto, su alta capacidad fermentativa alcohólica. Especies de levadura preferidas como células huésped incluyen *S. cerevisiae*, *S. bulderi*, *S. barnetti*, *S. exiguus*, *S. uvarum*, *S. diastaticus*, *K. lactis*, *K. marxianus*, *K. fragilis*.

En una realización preferida, la célula huésped de la invención es una célula huésped que ha sido transformada con una construcción de ácidos nucleicos que comprende la secuencia de nucleótidos que codifica las enzimas *araA*, *araB* y *araD*, como se han definido anteriormente. En una realización más preferida, la célula huésped se co-transforma con tres construcciones de ácidos nucleicos, comprendiendo cada construcción de ácidos nucleicos la secuencia de nucleótidos que codifica *araA*, *araB* o *araD*. La construcción de ácidos nucleicos que comprende la secuencia codificante de *araA*, *araB* y/o *araD* es capaz de expresión de las enzimas *araA*, *araB* y/o *araD* en la célula huésped. Para este fin, la construcción de ácidos nucleicos puede construirse como se describe en, por ejemplo, el documento WO 03/0624430. La célula huésped puede comprender una única copia, pero preferentemente comprende múltiples copias, de cada construcción de ácidos nucleicos. La construcción de ácidos nucleicos puede mantenerse episómicamente y así comprender una secuencia para la replicación autónoma, tal como una secuencia ARS. Construcciones de ácidos nucleicos episómicas adecuadas pueden, por ejemplo, basarse en la plásmidos 2 μ o pKD1 de levadura (Fleer et al., 1991, *Biotechnology* 9:968-975). Preferentemente, sin embargo, cada construcción de ácidos nucleicos se integra en una o más copias dentro del genoma de la célula huésped. La integración dentro del genoma de la célula huésped puede producirse al azar por recombinación ilegítima, pero preferentemente la construcción de ácidos nucleicos se integra dentro del genoma de la célula huésped por recombinación homóloga, como es muy conocido en la técnica de la genética molecular fúngica (véanse, por ejemplo, los documentos WO 90/14423, EP-A-0 481 008, EP-A-0 635 574 y US 6.265.186). Por consiguiente, en una realización más preferida, la célula de la invención comprende una construcción de ácidos nucleicos que comprende la secuencia codificante de *araA*, *araB* y/o *araD* y es capaz de expresión de las enzimas *araA*, *araB* y/o *araD*. En una realización incluso más preferida, las secuencias codificantes de *araA*, *araB* y/o *araD* están cada una operativamente enlazadas a un promotor que produce la expresión suficiente de las secuencias de nucleótidos correspondientes en una célula para conferir a la célula la capacidad de usar L-arabinosa, y/o para convertir L-arabinosa en L-ribulosa y/o xilulosa-5-fosfato. Preferentemente, la célula es una célula de levadura. Por consiguiente, en otro aspecto, la invención también engloba una construcción de ácidos nucleicos como se ha expuesto brevemente anteriormente en el presente documento. Preferentemente, una construcción de ácidos nucleicos comprende una secuencia de ácidos nucleicos que codifica *araA*, *araB* y/o *araD*. Las secuencias de ácidos nucleicos que codifican *araA*, *araB* o *araD* se han definido todas anteriormente en el presente documento.

Incluso más preferentemente, la expresión de las secuencias de nucleótidos correspondientes en una célula confiere a la célula la capacidad de convertir L-arabinosa en un producto de fermentación deseado como se define después en el presente documento. En una realización incluso más preferida, el producto de fermentación es etanol. Incluso más preferentemente, la célula es una célula de levadura.

Como se usa en el presente documento, el término "operativamente enlazado" se refiere a un enlace de elementos de polinucleótidos (o secuencias codificantes o secuencia de ácidos nucleicos) en una relación funcional. Una secuencia de ácidos nucleicos está "operativamente enlazada" cuando se pone en una relación funcional con otra secuencia de ácidos nucleicos. Por ejemplo, un promotor o potenciador está operativamente enlazado a una secuencia codificante si afecta la transcripción de la secuencia codificante. Operativamente enlazado significa que

las secuencias de ácidos nucleicos que están enlazadas normalmente son contiguas y, si fuera necesario unir dos regiones codificantes de proteína, contiguas y en marco de lectura.

Como se usa en el presente documento, el término "promotor" se refiere a un fragmento de ácido nucleico que funciona controlando la transcripción de uno o más genes, localizados en la dirección 5' con respecto a la dirección de transcripción del sitio de iniciación de la transcripción del gen, y se identifica estructuralmente por la presencia de un sitio de unión para ARN polimerasa dependiente de ADN, sitios de iniciación de la transcripción y cualquier otra secuencia de ADN, que incluye, pero no se limita a, sitios de unión del factor de transcripción, sitios de unión de la proteína represora y activadora, y cualquier otra secuencias de nucleótidos conocida para un experto en la materia para actuar directa o indirectamente para regular la cantidad de transcripción de un promotor. Un promotor "constitutivo" es un promotor que es activo bajo la mayoría de las condiciones ambientales y de desarrollo. Un promotor "inducible" es un promotor que es activo bajo regulación ambiental y de desarrollo.

El promotor que podría usarse para lograr la expresión de las secuencias de nucleótidos que codifican *araA*, *araB* y/o *araD* puede ser no nativo para la secuencia de nucleótidos que codifica la enzima que va a expresarse, es decir, un promotor que es heterólogo a la secuencia de nucleótidos (secuencia codificante) con la que está operativamente enlazado. Aunque un promotor preferentemente es heterólogo para la secuencia codificante con la que está operativamente enlazado, también se prefiere que un promotor sea homólogo, es decir, endógeno para la célula huésped. Preferentemente, un promotor heterólogo (para la secuencia de nucleótidos) es capaz de producir un mayor nivel en estado estacionario del transcrito que comprende la secuencia codificante (o es capaz de producir más moléculas de transcrito, es decir, moléculas de ARNm, por unidad de tiempo) que es un promotor que es nativo para la secuencia codificante, preferentemente en condiciones en las que la arabinosa, o arabinosa y glucosa, o xilosa y arabinosa o xilosa y arabinosa y glucosa están disponibles como fuentes de carbono, más preferentemente como fuentes de carbono principales (es decir, más del 50 % de la fuente de carbono disponible consiste en arabinosa, o arabinosa y glucosa, o xilosa y arabinosa o xilosa y arabinosa y glucosa), lo más preferentemente como fuentes de carbono únicas. Promotores adecuados en este contexto incluyen tanto promotores naturales constitutivos como inducibles, además de promotores manipulados por ingeniería. Un promotor preferido para su uso en la presente invención será además insensible a la represión de catabolitos (glucosa) y/o preferentemente no requerirá arabinosa y/o xilosa para la inducción.

Los promotores que tienen estas características están ampliamente disponibles y son conocidos para el experto. Ejemplos adecuados de tales promotores incluyen, por ejemplo, promotores de genes glicolíticos, tales como los promotores de fosfofructocinasa (PPK), triosa fosfato isomerasa (TPI), gliceraldehído-3-fosfato deshidrogenasa (GPD, TDH3 o GAPDH), piruvato cinasa (PYK), fosfoglicerato cinasa (PGK) de levaduras u hongos filamentosos; más detalles sobre tales promotores de levadura pueden encontrarse en (documento WO 93/03159). Otros promotores útiles son proteína ribosómica que codifica promotores de gen, el promotor del gen de lactasa (LAC4), promotores de alcohol deshidrogenasa (ADH1, ADH4, y similares), el promotor de enolasa (ENO), el promotor de glucosa-6-fosfato isomerasa (PGI1, Hauf et al, 2000) o el promotor del transportador de hexosa (glucosa) (HXT7) o la gliceraldehído-3-fosfato deshidrogenasa (TDH3). La secuencia del promotor de PGI1 se da en SEQ ID NO: 51. La secuencia del promotor de HXT7 se da en SEQ ID NO: 52. La secuencia del promotor de TDH3 se da en SEQ ID NO: 49. Otros promotores, tanto constitutivos como inducibles, y potenciadores o secuencias activantes en la dirección 5' serán conocidos para aquellos expertos en la materia. Los promotores usados en las células huésped de la invención pueden modificarse, si se desea, para afectar sus características de control. Una célula preferida de la invención es una célula eucariota transformada con los genes *araA*, *araB* y *araD* de *L. plantarum*. Más preferentemente, la célula eucariota es una célula de levadura, incluso más preferentemente una cepa de *S. cerevisiae* transformada con los genes *araA*, *araB* y *araD* de *L. plantarum*. Lo más preferentemente, la célula es tanto CBS 120327 como CBS 120328, ambas depositadas en el CBS Institute (Los Países Bajos) el 27 de septiembre de 2006.

El término "homólogo", cuando se usa para indicar la relación entre un ácido nucleico (recombinante) dado o molécula de polipéptido y un organismo huésped dado o célula huésped, se entiende que significa que en la naturaleza el ácido nucleico o molécula de polipéptido se produce por una célula huésped u organismos de la misma especie, preferentemente de la misma variedad o cepa. Si es homóloga a una célula huésped, una secuencia de ácidos nucleicos que codifica un polipéptido normalmente estará operativamente enlazada a otra secuencia promotora o, si es aplicable, otra secuencia señal secretora y/o secuencia terminadora que esté en su entorno natural. Cuando se usa para indicar la relación de dos secuencias de ácidos nucleicos, el término "homólogo" significa que una secuencia de ácidos nucleicos monocatenaria puede hibridarse con una secuencia de ácidos nucleicos monocatenaria complementaria. El grado de hibridación puede depender de varios factores que incluyen la cantidad de identidad entre las secuencias y las condiciones de hibridación tales como temperatura y concentración de sales como se ha presentado anteriormente. Preferentemente, la región de identidad es mayor de aproximadamente 5 pb, más preferentemente la región de identidad es mayor de 10 pb.

El término "heterólogo", cuando se usa con respecto a un ácido nucleico (ADN o ARN) o proteína, se refiere a un ácido nucleico o proteína que no se produce naturalmente como parte del organismo, célula, genoma o secuencia de ADN o ARN en la que está presente, o que se encuentra en una célula o localización o localizaciones en el genoma o secuencia de ADN o ARN que se diferencian de aquellas que se encuentran en la naturaleza. Los ácidos nucleicos heterólogos o proteínas no son endógenos a la célula dentro de la que se introduce, pero han sido obtenidos de otra

célula o producido sintética o recombinantemente. Generalmente, aunque no necesariamente, tales ácidos nucleicos codifican proteínas que normalmente no son producidas por la célula en la que el ADN se transcribe o expresa. Similarmente, el ARN exógeno codifica proteínas normalmente no expresadas en la célula en la que el ARN exógeno está presente. Los ácidos nucleicos heterólogos y proteínas también pueden denominarse ácidos nucleicos extraños o proteínas. Cualquier ácido nucleico o proteína que un experto en la materia reconocería como heterólogo o extraño a la célula en la que se expresa está englobado en el presente documento por el término ácido nucleico heterólogo o proteína. El término heterólogo también se aplica a combinaciones no naturales de secuencias de ácidos nucleicos o de aminoácidos, es decir, combinaciones en las que al menos dos de las secuencias combinadas son extrañas la una con respecto a la otra.

10 Célula eucariota preferida capaz de usar y/o convertir L-arabinosa y xilosa

En una realización más preferida, la célula de la invención que expresa araA, araB y araD es capaz de usar L-arabinosa y/o convertirla en L-ribulosa y/o xilulosa-5-fosfato y/o un producto de fermentación deseado como se ha definido anteriormente en el presente documento y presenta adicionalmente la capacidad de usar xilosa y/o convertir xilosa en xilulosa. La conversión de xilosa en xilulosa es preferentemente una etapa de isomerización de una etapa (isomerización directa de xilosa en xilulosa). Este tipo de célula es, por tanto, capaz de usar tanto L-arabinosa como xilosa. "Usar" xilosa tiene preferentemente el mismo significado que "usar" L-arabinosa como se ha definido anteriormente en el presente documento.

Las definiciones de enzima son como se usan en el documento WO 06/009434, para xilosa isomerasa (EC 5.3.1.5), xilulosa cinasa (EC 2.7.1.17), ribulosa 5-fosfato epimerasa (5.1.3.1), ribulosa 5-fosfato isomerasa (EC 5.3.1.6), transcetolasa (EC 2.2.1.1), transaldolasa (EC 2.2.1.2) y aldosa reductasa" (EC 1.1.1.21).

En una realización preferida, la célula eucariota de la invención que expresa araA, araB y araD, como se ha definido anteriormente en el presente documento, tiene la capacidad de isomerizar xilosa a xilulosa como se ha descrito, por ejemplo, en el documento WO 03/0624430 o en el documento WO 06/009434. La capacidad de isomerizar xilosa a xilulosa se confiere a la célula huésped por la transformación de la célula huésped con una construcción de ácidos nucleicos que comprende una secuencia de nucleótidos que codifica una xilosa isomerasa. La capacidad de la célula huésped transformada para isomerizar xilosa en xilulosa es la isomerización directa de xilosa a xilulosa. Esto se entiende que significa que la xilosa se isomeriza en xilulosa en una reacción única catalizada por una xilosa isomerasa, a diferencia de la conversión de dos etapas de xilosa en xilulosa mediante un producto intermedio de xilitol como se cataliza por xilosa reductasa y xilitol deshidrogenasa, respectivamente.

La secuencia de nucleótidos codifica una xilosa isomerasa que se expresa preferentemente en forma activa en la célula huésped transformada de la invención. Así, la expresión de la secuencia de nucleótidos en la célula huésped produce una xilosa isomerasa con una actividad específica de al menos 10 U de actividad de xilosa isomerasa por mg de proteína a 30 °C, preferentemente al menos 20, 25, 30, 50, 100, 200, 300 o 500 U por mg a 30 °C. La actividad específica de la xilosa isomerasa expresada en la célula huésped transformada se define en el presente documento como la cantidad de unidades de actividad de xilosa isomerasa por mg de proteína de lisado libre de células de la célula huésped, por ejemplo, un lisado libre de células de levadura. La determinación de la actividad de xilosa isomerasa ya se ha descrito anteriormente en el presente documento.

Preferentemente, la expresión de la secuencia de nucleótidos que codifica la xilosa isomerasa en la célula huésped produce una xilosa isomerasa con una K_m para xilosa que es inferior a 50, 40, 30 o 25 mM, más preferentemente, la K_m para xilosa es aproximadamente 20 mM o menos.

Una secuencia de nucleótidos preferida que codifica la xilosa isomerasa puede seleccionarse del grupo que consiste en:

(e) secuencias de nucleótidos que codifican un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 60, 65, 70, 75, 80, 85, 90, 95, 97, 98 o el 99 % de identidad de secuencias con la secuencia de aminoácidos de SEQ ID NO. 7 o SEQ ID NO: 15;

(f) secuencias de nucleótidos que comprenden una secuencia de nucleótidos que tiene al menos el 40, 50, 60, 70, 80, 90, 95, 97, 98 o el 99 % de identidad de secuencias con la secuencia de nucleótidos de SEQ ID NO. 8 o SEQ ID NO: 16;

(g) secuencias de nucleótidos cuya hebra complementaria se hibrida con una secuencia de la molécula de ácido nucleico de (a) o (b);

(h) secuencias de nucleótidos cuya secuencia se diferencia de la secuencia de una molécula de ácido nucleico de (c) debido a la degeneración del código genético.

La secuencia de nucleótidos que codifica la xilosa isomerasa puede codificar tanto una xilosa isomerasa procariota como una eucariota, es decir, una xilosa isomerasa con una secuencia de aminoácidos que es idéntica a la de una xilosa isomerasa que se produce naturalmente en el organismo procariota o eucariota. Los presentes inventores han encontrado que la capacidad de una xilosa isomerasa particular para conferir a una célula huésped eucariota la

capacidad de isomerizar xilosa en xilulosa no depende tanto de si la isomerasa es de origen procariota o eucariota. Más bien, esto depende de la relación de la secuencia de aminoácidos de la isomerasa con la de la secuencia de *Piromyces* (SEQ ID NO. 7). Sorprendentemente, la isomerasa de *Piromyces* eucariota está más relacionada con isomerasas procariotas que con otras isomerasas eucariotas conocidas. Por tanto, una secuencia de nucleótidos preferida codifica una xilosa isomerasa que tiene una secuencia de aminoácidos que está relacionada con la secuencia de *Piromyces* como se ha definido anteriormente. Una secuencia de nucleótidos preferida codifica una xilosa isomerasa fúngica (por ejemplo, de un *Basidiomycete*), más preferentemente una xilosa isomerasa de un hongo anaerobio, por ejemplo, una xilosa isomerasa de un hongo anaerobio que pertenece a las familias *Neocallimastix*, *Caecomyces*, *Piromyces*, *Orpinomyces* o *Ruminomyces*. Alternativamente, una secuencia de nucleótidos preferida codifica una xilosa isomerasa bacteriana, preferentemente una bacteria Gram-negativa, más preferentemente una isomerasa de la clase *Bacteroides*, o del género *Bacteroides*, lo más preferentemente de *B. thetaiotaomicron* (SEQ ID NO. 15).

Para aumentar la probabilidad de que la xilosa isomerasa se exprese en forma activa en una célula huésped eucariota tal como levadura, la secuencia de nucleótidos que codifica la xilosa isomerasa puede adaptarse para optimizar su uso de codones al de la célula huésped eucariota como se ha definido anteriormente en el presente documento.

Una célula huésped para transformación con la secuencia de nucleótidos que codifica la xilosa isomerasa, como se ha descrito anteriormente, es preferentemente un huésped capaz de transporte activo o pasivo de xilosa dentro de la célula. La célula huésped contiene preferentemente glicólisis activa. La célula huésped puede contener adicionalmente una vía de pentosa fosfato endógena y puede contener actividad de xilulosa cinasa endógena de manera que la xilulosa isomerizada a partir de xilosa pueda metabolizarse a piruvato. El huésped contiene preferentemente adicionalmente enzimas para la conversión de piruvato a un producto de fermentación deseado tal como etanol, ácido láctico, ácido 3-hidroxi-propiónico, ácido acrílico, ácido acético, ácido succínico, ácido cítrico, ácido málico, ácido fumárico, un aminoácido, 1,3-propano-diol, etileno, glicerol, butanol, un antibiótico β -lactámico o una cefalosporina. Una célula huésped preferida es una célula huésped que es naturalmente capaz de fermentación alcohólica, preferentemente fermentación alcohólica anaerobia. La célula huésped tiene preferentemente adicionalmente una alta tolerancia a etanol, una alta tolerancia a pH bajo (es decir, capaz de crecimiento a un pH inferior a 5, 4, 3 o 2,5) y hacia ácidos orgánicos como ácido láctico, ácido acético o ácido fórmico, y productos de degradación de azúcares tales como furfural e hidroxi-metilfurfural, y una alta tolerancia a temperaturas elevadas. Cualquiera de estas características o actividades de la célula huésped puede estar naturalmente presente en la célula huésped o puede introducirse o modificarse por modificación genética. Una célula adecuada es un microorganismo eucariota como, por ejemplo, un hongo, sin embargo, lo más adecuado como célula huésped son levaduras u hongos filamentosos. Levaduras y hongos filamentosos preferidos ya se han definido en el presente documento.

Como se usa en el presente documento, el término célula huésped tiene el mismo significado que célula.

La célula de la invención se transforma preferentemente con una construcción de ácidos nucleicos que comprende la secuencia de nucleótidos que codifica la xilosa isomerasa. La construcción de ácidos nucleicos que se usa preferentemente es la misma que la usada que comprende la secuencia de nucleótidos que codifica araA, araB o araD.

En otra realización preferida de la invención, la célula de la invención:

- que expresa araA, araB y araD, y que presenta la capacidad de isomerizar directamente xilosa en xilulosa, como se ha definido anteriormente en el presente documento

comprende además una modificación genética que aumenta el flujo de la vía de pentosa fosfato, como se describe en el documento WO 06/009434. En particular, la modificación genética produce un elevado flujo de la vía de pentosa fosfato de parte no oxidativa. Una modificación genética que produce un elevado flujo de la parte no oxidativa de la vía de pentosa fosfato se entiende en el presente documento que significa una modificación que aumenta el flujo al menos un factor de 1,1, 1,2, 1,5, 2, 5, 10 o 20 en comparación con el flujo en una cepa que es genéticamente idéntica, excepto por la modificación genética que causa el aumento de flujo. El flujo de la parte no oxidativa de la vía de pentosa fosfato puede medirse cultivando el huésped modificado sobre xilosa como la única fuente de carbono, determinándose la tasa específica de consumo de xilosa y restándose la tasa específica de producción de xilitol de la tasa específica de consumo de xilosa, si se produce cualquier xilitol. Sin embargo, el flujo de la parte no oxidativa de la vía de pentosa fosfato es proporcional a la velocidad de crecimiento sobre xilosa como única fuente de carbono, preferentemente con la velocidad de crecimiento anaerobia sobre xilosa como única fuente de carbono. Hay una relación lineal entre la velocidad de crecimiento sobre xilosa como única fuente de carbono ($\mu_{\text{máx}}$) y el flujo de la parte no oxidativa de la vía de pentosa fosfato. La tasa específica de consumo de xilosa (Q_s) es igual a la velocidad de crecimiento (μ) dividida entre el rendimiento de biomasa sobre azúcar (Y_{xs}) debido a que el rendimiento de biomasa sobre azúcar es constante (bajo un conjunto dado de condiciones: anaerobias, medio de crecimiento, pH, antecedentes genéticos de la cepa, etc.; es decir, $Q_s = \mu / Y_{\text{xs}}$). Por tanto, el elevado flujo de la parte no oxidativa de la vía de pentosa fosfato puede deducirse del aumento en la máxima velocidad de crecimiento bajo estas condiciones. En una realización preferida, la célula comprende una modificación genética que aumenta el flujo

de la vía de pentosa fosfato y tiene una tasa específica de consumo de xilosa de al menos 346 mg de xilosa/g de biomasa/h.

Las modificaciones genéticas que aumentan el flujo de la vía de pentosa fosfato pueden introducirse en la célula huésped de diversas formas. Incluyendo éstas, por ejemplo, lograr mayores niveles de actividad en estado estacionario de xilulosa cinasa y/o una o más de las enzimas de la vía de pentosa fosfato de parte no oxidativa y/o un nivel en estado estacionario reducido de actividad no específica de aldosa reductasa. Estos cambios en los niveles de actividad en estado estacionario pueden efectuarse por selección de mutantes (espontánea o inducida por productos químicos o radiación) y/o por tecnología de ADN recombinante, por ejemplo, por expresión en exceso o inactivación, respectivamente, de genes que codifican las enzimas o factores que regulan estos genes.

En una célula huésped más preferida, la modificación genética comprende la expresión en exceso de al menos una enzima de la vía de pentosa fosfato (de parte no oxidativa). Preferentemente, la enzima está seleccionada del grupo que consiste en las enzimas que codifican ribulosa-5-fosfato isomerasa, ribulosa-5-fosfato epimerasa, transcetolasa y transaldolasa, como se describen en el documento WO 06/009434.

Pueden expresarse en exceso diversas combinaciones de enzimas de la vía de pentosa fosfato (de parte no oxidativa). Por ejemplo, las enzimas que se expresan en exceso pueden ser al menos las enzimas ribulosa-5-fosfato isomerasa y ribulosa-5-fosfato epimerasa; o al menos las enzimas ribulosa-5-fosfato isomerasa y transcetolasa; o al menos las enzimas ribulosa-5-fosfato isomerasa y transaldolasa; o al menos las enzimas ribulosa-5-fosfato epimerasa y transcetolasa; o al menos las enzimas ribulosa-5-fosfato epimerasa y transaldolasa; o al menos las enzimas transcetolasa y transaldolasa; o al menos las enzimas ribulosa-5-fosfato epimerasa, transcetolasa y transaldolasa; o al menos las enzimas ribulosa-5-fosfato isomerasa, transcetolasa y transaldolasa; o al menos las enzimas ribulosa-5-fosfato isomerasa, ribulosa-5-fosfato epimerasa y transaldolasa; o al menos las enzimas ribulosa-5-fosfato isomerasa, ribulosa-5-fosfato epimerasa y transcetolasa. En una realización de la invención, cada una de las enzimas ribulosa-5-fosfato isomerasa, ribulosa-5-fosfato epimerasa, transcetolasa y transaldolasa se expresan en exceso en la célula huésped. Es más preferida una célula huésped en la que la modificación genética comprende al menos la expresión en exceso de tanto las enzimas transcetolasa como transaldolasa, ya que una célula huésped tal ya es capaz de crecimiento anaerobio sobre xilosa. En realidad, bajo algunas condiciones, los presentes inventores han encontrado que las células huésped que expresan en exceso solo la transcetolasa y la transaldolasa ya tienen la misma velocidad de crecimiento anaerobio sobre xilosa que las células huésped que expresan en exceso las cuatro las enzimas, es decir, la ribulosa-5-fosfato isomerasa, ribulosa-5-fosfato epimerasa, transcetolasa y transaldolasa. Además, se prefieren las células huésped que expresan en exceso ambas enzimas ribulosa-5-fosfato isomerasa y ribulosa-5-fosfato epimerasa con respecto a las células huésped que expresan en exceso solo la isomerasa o solo la epimerasa, ya que la expresión en exceso de solo una de estas enzimas puede producir desequilibrios metabólicos.

Hay diversos medios disponibles en la materia para la expresión en exceso de enzimas en las células de la invención. En particular, una enzima puede expresarse en exceso aumentando el número de copias del gen que codifica la enzima en la célula huésped, por ejemplo, integrando copias adicionales del gen en el genoma de la célula huésped, expresando el gen a partir de un vector de expresión de múltiples copias episómico o introduciendo un vector de expresión episómico que comprende múltiples copias del gen.

Alternativamente, la expresión en exceso de enzimas en las células huésped de la invención puede lograrse usando un promotor que no es nativo para la secuencia que codifica la enzima que va a expresarse en exceso, es decir, un promotor que es heterólogo a la secuencia codificante a la que está operativamente enlazado. Promotores adecuados para este fin ya se han definido en el presente documento.

La secuencia codificante usada para la expresión en exceso de las enzimas es preferentemente homóloga a la célula huésped de la invención. Sin embargo, asimismo pueden aplicarse secuencias codificantes que son heterólogas a la célula huésped de la invención, como se ha mencionado en el documento WO 06/009434.

Una secuencia de nucleótidos usada para la expresión en exceso de ribulosa-5-fosfato isomerasa en la célula huésped de la invención es una secuencia de nucleótidos que codifica un polipéptido con actividad de ribulosa-5-fosfato isomerasa, por lo que preferentemente el polipéptido tiene una secuencia de aminoácidos que tiene al menos el 50, 60, 70, 80, 90 o el 95 % de identidad con SEQ ID NO. 17 o por lo que la secuencia de nucleótidos es capaz de hibridarse con la secuencia de nucleótidos de SEQ ID NO. 18, bajo condiciones moderadas, preferentemente bajo condiciones rigurosas.

Una secuencia de nucleótidos usada para la expresión en exceso de ribulosa-5-fosfato epimerasa en la célula huésped de la invención es una secuencia de nucleótidos que codifica un polipéptido con actividad de ribulosa-5-fosfato epimerasa, por lo que preferentemente el polipéptido tiene una secuencia de aminoácidos que tiene al menos el 50, 60, 70, 80, 90 o el 95 % de identidad con SEQ ID NO. 19 o por lo que la secuencia de nucleótidos es capaz de hibridarse con la secuencia de nucleótidos de SEQ ID NO. 20, bajo condiciones moderadas, preferentemente bajo condiciones rigurosas.

Una secuencia de nucleótidos usada para la expresión en exceso de transcetolasa en la célula huésped de la invención es una secuencia de nucleótidos que codifica un polipéptido con actividad de transcetolasa, por lo que preferentemente el polipéptido tiene una secuencia de aminoácidos que tiene al menos el 50, 60, 70, 80, 90 o el 95 % de identidad con SEQ ID NO. 21 o por lo que la secuencia de nucleótidos es capaz de hibridarse con la secuencia de nucleótidos de SEQ ID NO. 22, bajo condiciones moderadas, preferentemente bajo condiciones rigurosas.

Una secuencia de nucleótidos usada para la expresión en exceso de transaldolasa en la célula huésped de la invención es una secuencia de nucleótidos que codifica un polipéptido con actividad de transaldolasa, por lo que preferentemente el polipéptido tiene una secuencia de aminoácidos que tiene al menos el 50, 60, 70, 80, 90 o el 95 % de identidad con SEQ ID NO. 23 o por lo que la secuencia de nucleótidos es capaz de hibridarse con la secuencia de nucleótidos de SEQ ID NO. 24, bajo condiciones moderadas, preferentemente bajo condiciones rigurosas.

Expresión en exceso de una enzima, cuando se refiere a la producción de la enzima en una célula huésped genéticamente modificada, significa que la enzima se produce a un nivel más alto de actividad enzimática específica en comparación con la célula huésped no modificada bajo condiciones idénticas. Normalmente, esto significa que la proteína enzimáticamente activa (o proteínas en el caso de enzimas de múltiples subunidades) se produce en mayores cantidades, o más bien a un mayor nivel en estado estacionario en comparación con la célula huésped sin modificar bajo condiciones idénticas. Similarmente, esto normalmente significa que el ARNm que codifica la proteína enzimáticamente activa se produce en cantidades mayores, o nuevamente más bien a un mayor nivel en estado estacionario en comparación con la célula huésped sin modificar bajo condiciones idénticas. La expresión en exceso de una enzima se determina así preferentemente midiendo el nivel de actividad específica de la enzima en la célula huésped usando ensayos enzimáticos apropiados como se describen en el presente documento. Alternativamente, la expresión en exceso de la enzima puede determinarse indirectamente cuantificando el nivel en estado estacionario específico de la proteína de enzima, por ejemplo, usando anticuerpos específicos para la enzima, o cuantificando el nivel estacionario específico del ARNm que codifica la enzima. Lo último puede ser particularmente adecuado para enzimas de la vía de pentosa fosfato para las que los ensayos enzimáticos no son fácilmente factibles, ya que los sustratos para las enzimas no están comercialmente disponibles. Preferentemente, en las células huésped de la invención, una enzima que va a expresarse en exceso se expresa en exceso al menos un factor de 1,1, 1,2, 1,5, 2, 5, 10 o 20 en comparación con una cepa que es genéticamente idéntica, excepto por la modificación genética que causa la expresión en exceso. Debe entenderse que estos niveles de expresión en exceso pueden aplicarse al nivel en estado estacionario de la actividad de la enzima, el nivel en estado estacionario de la proteína de la enzima, además de al nivel en estado estacionario del transcrito que codifica la enzima.

En otra realización preferida, la célula huésped de la invención:

- que expresa *araA*, *araB* y *araD*, y que presenta la capacidad de isomerizar directamente xilosa en xilulosa, y opcionalmente
- que comprende una modificación genética que aumenta el flujo de la vía de pentosa como se ha definido anteriormente en el presente documento

comprende además una modificación genética que aumenta la actividad específica de xilulosa cinasa. Preferentemente, la modificación genética produce la expresión en exceso de una xilulosa cinasa, por ejemplo, por expresión en exceso de una secuencia de nucleótidos que codifica una xilulosa cinasa. El gen que codifica la xilulosa cinasa puede ser endógeno a la célula huésped o puede ser una xilulosa cinasa que es heteróloga a la célula huésped. Una secuencia de nucleótidos usada para la expresión en exceso de xilulosa cinasa en la célula huésped de la invención es una secuencia de nucleótidos que codifica un polipéptido con actividad de xilulosa cinasa, por lo que preferentemente el polipéptido tiene una secuencia de aminoácidos que tiene al menos el 50, 60, 70, 80, 90 o el 95 % de identidad con SEQ ID NO. 25 o por lo que la secuencia de nucleótidos es capaz de hibridarse con la secuencia de nucleótidos de SEQ ID NO. 26, bajo condiciones moderadas, preferentemente bajo condiciones rigurosas.

Una xilulosa cinasa particularmente preferida es una xilosa cinasa que está relacionada con la xilulosa cinasa *xyIB* de *Piromyces* como se ha mencionado en el documento WO 03/0624430. Una secuencia de nucleótidos más preferida para su uso en la expresión en exceso de xilulosa cinasa en la célula huésped de la invención es una secuencia de nucleótidos que codifica un polipéptido con actividad de xilulosa cinasa, por lo que preferentemente el polipéptido tiene una secuencia de aminoácidos que tiene al menos el 45, 50, 55, 60, 65, 70, 80, 90 o el 95 % de identidad con SEQ ID NO. 27 o por lo que la secuencia de nucleótidos es capaz de hibridarse con la secuencia de nucleótidos de SEQ ID NO. 28, bajo condiciones moderadas, preferentemente bajo condiciones rigurosas.

En las células huésped de la invención, la modificación genética que aumenta la actividad específica de xilulosa cinasa puede combinarse con cualquiera de las modificaciones que aumentan el flujo de la vía de pentosa fosfato como se ha descrito anteriormente, pero esta combinación no es esencial para la invención. Así, una célula huésped de la invención que comprende una modificación genética que aumenta la actividad específica de xilulosa cinasa, además de la expresión de las enzimas *araA*, *araB* y *araD* como se define en el presente documento, se incluye

específicamente en la invención. Los diversos medios disponibles en la materia para alcanzar y analizar la expresión en exceso de una xilulosa cinasa en las células huésped de la invención son los mismos que se han descrito anteriormente para enzimas de la vía de pentosa fosfato. Preferentemente en las células huésped de la invención, una xilulosa cinasa que va a expresarse en exceso se expresa en exceso al menos un factor 1,1, 1,2, 1,5, 2, 5, 10 o 20 en comparación con una cepa que es genéticamente idéntica, excepto por la modificación genética que causa la expresión en exceso. Debe entenderse que estos niveles de expresión en exceso pueden aplicarse al nivel en estado estacionario de la actividad de la enzima, el nivel en estado estacionario de la proteína de la enzima, además de al nivel en estado estacionario del transcrito que codifica la enzima.

En otra realización preferida, la célula huésped de la invención:

- 10 - que expresa *araA*, *araB* y *araD*, y que presenta la capacidad de isomerizar directamente xilosa en xilulosa, y opcionalmente
- que comprende una modificación genética que aumenta el flujo de la vía de pentosa y/o
- que comprende además una modificación genética que aumenta la actividad específica de xilulosa cinasa, todas como se ha definido anteriormente en el presente documento

15 comprende además una modificación genética que reduce la actividad no específica de aldosa reductasa en la célula huésped. Preferentemente, la actividad no específica de aldosa reductasa se reduce en la célula huésped por una o más modificaciones genéticas que reducen la expresión de o inactivan un gen que codifica una aldosa reductasa no específica, como se describe en el documento WO 06/009434. Preferentemente, las modificaciones genéticas reducen o inactivan la expresión de cada copia endógena de un gen que codifica una aldosa reductasa no específica en la célula huésped. Las células huésped pueden comprender copias múltiples de genes que codifican aldosas reductasas no específicas como resultado de di-, poli- o aneu-ploidía, y/o la célula huésped puede contener varias (iso)enzimas diferentes con actividad de aldosa reductasa que se diferencian en la secuencia de aminoácidos y que están cada una codificada por un gen diferente. También en tales casos, preferentemente la expresión de cada gen que codifica una aldosa reductasa no específica se reduce o inactiva. Preferentemente, el gen se inactiva por la delección de al menos parte del gen o por la rotura del gen, por lo que en este contexto el término gen también incluye cualquier secuencia no codificante en la dirección 5' o en la dirección 3' de la secuencia codificante, produciendo la delección o inactivación (parcial) una reducción de la expresión de la actividad no específica de aldosa reductasa en la célula huésped. Una secuencia de nucleótidos que codifica una aldosa reductasa cuya actividad va a reducirse en la célula huésped de la invención es una secuencia de nucleótidos que codifica un polipéptido con la actividad de aldosa reductasa, por lo que preferentemente el polipéptido tiene una secuencia de aminoácidos que tiene al menos el 50, 60, 70, 80, 90 o el 95 % de identidad con SEQ ID NO. 29, o por lo que la secuencia de nucleótidos es capaz de hibridarse con la secuencia de nucleótidos de SEQ ID NO. 30 bajo condiciones moderadas, preferentemente bajo condiciones rigurosas.

En las células huésped de la invención, la expresión de las enzimas *araA*, *araB* y *araD*, como se define en el presente documento, se combina con la modificación genética que reduce la actividad no específica de aldosa reductasa. La modificación genética que conduce a la reducción de la actividad no específica de aldosa reductasa puede combinarse con cualquiera de las modificaciones que aumentan el flujo de la vía de pentosa fosfato y/o con cualquiera de las modificaciones que aumentan la actividad específica de xilulosa cinasa en las células huésped como se ha descrito anteriormente, pero estas combinaciones no son esenciales para la invención. Así, una célula huésped que expresa *araA*, *araB* y *araD*, que comprende una modificación genética adicional que reduce la actividad no específica de aldosa reductasa, está específicamente incluida en la invención.

En una realización preferida, la célula huésped es CBS 120327 depositada en el CBS Institute (Los Países Bajos) el 27 de septiembre de 2006.

En otra realización preferida, la invención se refiere a células huésped modificadas que se adaptan adicionalmente a L-arabinosa (usan L-arabinosa y/o la convierten en L-ribulosa y/o xilulosa-5-fosfato y/o en un producto de fermentación deseado y opcionalmente utilización de xilosa por selección de mutantes, tanto espontáneos como inducidos (por ejemplo, por radiación o productos químicos), para el crecimiento sobre L-arabinosa y opcionalmente xilosa, preferentemente sobre L-arabinosa y opcionalmente xilosa como única fuente de carbono, y más preferentemente en condiciones anaerobias. La selección de mutantes puede realizarse sometiendo a pases en serie cultivos como, por ejemplo, se describe por Kuyper et al. (2004, FEMS Yeast Res. 4: 655-664) y/o por cultivo bajo presión selectiva en un cultivo en quimiostato como se describe en el Ejemplo 4 del documento WO 06/009434. Este proceso de selección puede continuarse en tanto que sea necesario. Este proceso de selección se lleva a cabo preferentemente durante una semana hasta un año. Sin embargo, el proceso de selección puede llevarse a cabo durante un periodo de tiempo más largo, si fuera necesario. Durante el proceso de selección, las células se cultivan preferentemente en presencia de aproximadamente 20 g/l de L-arabinosa y/o aproximadamente 20 g/l de xilosa. Se espera que la célula obtenida al final de este proceso de selección mejore en cuanto a sus capacidades de uso de L-arabinosa y/o xilosa, y/o de conversión de L-arabinosa en L-ribulosa y/o xilulosa-5-fosfato y/o un producto de fermentación deseado tal como etanol. En este contexto "célula mejorada" puede significar que la célula obtenida es capaz de usar L-arabinosa y/o xilosa de una forma más eficiente que la célula de la que se deriva. Por ejemplo, se

espera que la célula obtenida crezca mejor: aumento de la velocidad de crecimiento específica de al menos el 2 % con respecto a la célula de la que se deriva bajo las mismas condiciones. Preferentemente, el aumento es de al menos el 4 %, 6 %, 8 %, 10 %, 15 %, 20 %, 25 % o más. La velocidad de crecimiento específica puede calcularse a partir de la DO_{660} como se conoce por el experto. Por tanto, monitorizando la DO_{660} , puede deducirse la velocidad de crecimiento específica. En este contexto, "célula mejorada" puede también significar que la célula obtenida convierte L-arabinosa en L-ribulosa y/o xilulosa-5-fosfato y/o un producto de fermentación deseado tal como etanol de una forma más eficiente que la célula de la que se deriva. Por ejemplo, se espera que la célula obtenida produzca mayores cantidades de L-ribulosa y/o xilulosa-5-fosfato y/o un producto de fermentación deseado tal como etanol: aumento de al menos uno de estos compuestos de al menos el 2 % con respecto a la célula de la que se deriva bajo las mismas condiciones. Preferentemente, el aumento es de al menos el 4 %, 6 %, 8 %, 10 %, 15 %, 20 %, 25 % o más. En este contexto, "célula mejorada" también puede significar que la célula obtenida convierte xilosa en xilulosa y/o un producto de fermentación deseado tal como etanol de una forma más eficiente que la célula de la que se deriva. Por ejemplo, se espera que la célula obtenida produzca mayores cantidades de xilulosa y/o un producto de fermentación deseado tal como etanol: aumento de al menos uno de estos compuestos de al menos el 2 % con respecto a la célula de la que se deriva bajo las mismas condiciones. Preferentemente, el aumento es de al menos el 4 %, 6 %, 8 %, 10 %, 15 %, 20 %, 25 % o más.

En una célula huésped preferida de la invención, al menos una de las modificaciones genéticas descritas anteriormente, que incluye modificaciones obtenidas por selección de mutantes, confiere a la célula huésped la capacidad de crecer sobre L-arabinosa y opcionalmente xilosa como fuente de carbono, preferentemente como la única fuente de carbono, y preferentemente en condiciones anaerobias. Preferentemente, la célula huésped modificada no produce esencialmente xilitol, por ejemplo, el xilitol producido está por debajo del límite de detección o, por ejemplo, inferior al 5, 2, 1, 0,5 o 0,3 % del carbono consumido en una base molar.

Preferentemente, la célula huésped modificada tiene la capacidad de crecer sobre L-arabinosa y opcionalmente xilosa como la única fuente de carbono a una tasa de al menos 0,001, 0,005, 0,01, 0,03, 0,05, 0,1, 0,2, 0,25 o 0,3 h^{-1} bajo condiciones aerobias, o, si es aplicable, a una tasa de al menos 0,001, 0,005, 0,01, 0,03, 0,05, 0,07, 0,08, 0,09, 0,1, 0,12, 0,15 o 0,2 h^{-1} en condiciones anaerobias. Preferentemente, la célula huésped modificada tiene la capacidad de crecer sobre una mezcla de glucosa y L-arabinosa y opcionalmente xilosa (en una relación 1:1 de peso) como única fuente de carbono a una tasa de al menos 0,001, 0,005, 0,01, 0,03, 0,05, 0,1, 0,2, 0,25 o 0,3 h^{-1} bajo condiciones aerobias, o, si es aplicable, a una tasa de al menos 0,001, 0,005, 0,01, 0,03, 0,05, 0,1, 0,12, 0,15 o 0,2 h^{-1} en condiciones anaerobias.

Preferentemente, la célula huésped modificada tiene una tasa de consumo de L-arabinosa y opcionalmente xilosa específica de al menos 346, 350, 400, 500, 600, 650, 700, 750, 800, 900 o 1000 mg/g de células/h. Preferentemente, la célula huésped modificada tiene un rendimiento de producto de fermentación (tal como etanol) sobre L-arabinosa y opcionalmente xilosa que es al menos el 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 85, 90, 95 o el 98 % del rendimiento de la célula huésped del producto de fermentación (tal como etanol) sobre glucosa. Más preferentemente, el rendimiento de producto de fermentación (tal como etanol) de la célula huésped modificada sobre L-arabinosa y opcionalmente xilosa es igual al rendimiento de producto de fermentación (tal como etanol) de la célula huésped sobre glucosa. Asimismo, el rendimiento de biomasa de la célula huésped modificada sobre L-arabinosa y opcionalmente xilosa es preferentemente al menos el 55, 60, 70, 80, 85, 90, 95 o 98 % del rendimiento de biomasa de la célula huésped sobre glucosa. Más preferentemente, el rendimiento de biomasa de la célula huésped modificada sobre L-arabinosa y opcionalmente xilosa es igual al rendimiento de biomasa de la célula huésped sobre glucosa. Se entiende que en la comparación de rendimiento sobre glucosa y L-arabinosa y opcionalmente xilosa ambos rendimientos se comparan en condiciones aerobias o ambos en condiciones anaerobias.

En una realización más preferida, la célula huésped es CBS 120328 depositada en el CBS Institute (Los Países bajos) el 27 de septiembre de 2006 o CBS 121879 depositada en el CBS Institute (Los Países bajos) el 20 de septiembre de 2007.

En una realización preferida, la célula expresa una o más enzimas que confieren a la célula la capacidad de producir al menos un producto de fermentación seleccionado del grupo que consiste en etanol, ácido láctico, ácido 3-hidroxi-propiónico, ácido acrílico, ácido acético, ácido succínico, ácido cítrico, ácido málico, ácido fumárico, un aminoácido, 1,3-propano-diol, etileno, glicerol, butanol, un antibiótico β -lactámico y una cefalosporina. En una realización más preferida, la célula huésped de la invención es una célula huésped para la producción de etanol. En otra realización preferida, la invención se refiere a una célula huésped transformada para la producción de productos de fermentación distintos de etanol. Tales productos de fermentación no etanólicos incluyen en principio cualquier producto químico a granel o fino que pueda producirse por un microorganismo eucariota tal como una levadura o un hongo filamentoso. Tales productos de fermentación incluyen, por ejemplo, ácido láctico, ácido 3-hidroxi-propiónico, ácido acrílico, ácido acético, ácido succínico, ácido cítrico, ácido málico, ácido fumárico, un aminoácido, 1,3-propano-diol, etileno, glicerol, butanol, un antibiótico β -lactámico y una cefalosporina. Una célula huésped preferida de la invención para la producción de productos de fermentación no etanólicos es una célula huésped que contiene una modificación genética que produce actividad de alcohol deshidrogenasa disminuida.

Método

En otro aspecto, la invención se refiere a procesos de fermentación en los que una célula huésped de la invención se usa para la fermentación de una fuente de carbono que comprende una fuente de L-arabinosa y opcionalmente una fuente de xilosa. Preferentemente, la fuente de L-arabinosa y la fuente de xilosa son L-arabinosa y xilosa. Además, la fuente de carbono en el medio de fermentación también puede comprender una fuente de glucosa. La fuente de L-arabinosa, xilosa o glucosa puede ser L-arabinosa, xilosa o glucosa como tales o puede ser cualquier oligo- o polímero de hidrato de carbono que comprende unidades de L-arabinosa, xilosa o glucosa, tales como, por ejemplo, lignocelulosa, xilanos, celulosa, almidón, arabinano y similares. Para la liberación de las unidades de xilosa o glucosa de tales hidratos de carbono, carbohidrasas apropiadas (tales como xilanasas, glucanasas, amilasas y similares) pueden añadirse al medio de fermentación o pueden producirse por la célula huésped modificada. En el último caso, la célula huésped modificada puede ser genéticamente manipulada para producir y excretar tales carbohidrasas. Una ventaja adicional de uso de fuentes de glucosa oligo- o poliméricas es que permite mantener una concentración (más) baja de glucosa libre durante la fermentación, por ejemplo, usando cantidades limitantes de la velocidad de las carbohidrasas. Esto, a su vez, prevendrá la represión de sistemas requeridos para el metabolismo y el transporte de azúcares no de glucosa tales como xilosa. En un proceso preferido, la célula huésped modificada fermenta tanto la L-arabinosa (opcionalmente xilosa) como la glucosa, preferentemente simultáneamente, en cuyo caso se usa preferentemente una célula huésped modificada que es insensible a la represión de glucosa para prevenir el crecimiento diaúxico. Además de una fuente de L-arabinosa, opcionalmente xilosa (y glucosa) como fuente de carbono, el medio de fermentación comprenderá además el componente apropiado requerido para el crecimiento de la célula huésped modificada. Composiciones de medios de fermentación para el crecimiento de microorganismos tales como levaduras u hongos filamentosos son muy conocidas en la técnica.

En un proceso preferido, se proporciona un proceso de producción de un producto de fermentación seleccionado del grupo que consiste en etanol, ácido láctico, ácido 3-hidroxi-propiónico, ácido acrílico, ácido acético, ácido succínico, ácido cítrico, ácido málico, ácido fumárico, un aminoácido, 1,3-propano-diol, etileno, glicerol, butanol, un antibiótico β -lactámico y una cefalosporina, por el cual el proceso comprende las etapas de:

(a) fermentar un medio que contiene una fuente de L-arabinosa y opcionalmente xilosa con una célula huésped modificada como se define en el presente documento, por el cual la célula huésped fermenta L-arabinosa y opcionalmente xilosa en el producto de fermentación, y opcionalmente,

(b) recuperar el producto de fermentación.

El proceso de fermentación es un proceso para la producción de un producto de fermentación tal como, por ejemplo, etanol, ácido láctico, ácido 3-hidroxi-propiónico, ácido acrílico, ácido acético, ácido succínico, ácido cítrico, ácido málico, ácido fumárico, un aminoácido, 1,3-propano-diol, etileno, glicerol, butanol, un antibiótico β -lactámico, tal como penicilina G o penicilina V y derivados fermentativos de los mismos, y/o una cefalosporina. El proceso de fermentación puede ser un proceso de fermentación aerobio o anaerobio. Un proceso de fermentación anaerobio se define en el presente documento como un proceso de fermentación realizado en ausencia de oxígeno o en el que sustancialmente no se consume oxígeno, preferentemente menos de 5, 2,5 o 1 mmol/l/h, más preferentemente se consumen 0 mmol/l/h (es decir, el consumo de oxígeno no es detectable), y en el que las moléculas orgánicas sirven de tanto donante de electrones como de aceptores de electrones. En ausencia de oxígeno, la NADH producida en la glicólisis y formación de biomasa no puede oxidarse por fosforilación oxidativa. Para resolver este problema, muchos microorganismos usan piruvato o uno de sus derivados como aceptor de electrones y de hidrógeno, regenerando así NAD^+ . Así, en un proceso de fermentación anaerobio preferido se usa piruvato como electrón (y aceptor de hidrógeno) y se reduce dando productos de fermentación tales como etanol, ácido láctico, ácido 3-hidroxi-propiónico, ácido acrílico, ácido acético, ácido succínico, ácido cítrico, ácido málico, ácido fumárico, un aminoácido, 1,3-propano-diol, etileno, glicerol, butanol, un antibiótico β -lactámicos y una cefalosporina. En una realización preferida, el proceso de fermentación es anaerobio. Un proceso anaerobio es ventajoso ya que es más barato que los procesos aerobios: se necesita equipo menos especial. Además, se espera que los procesos anaerobios den un mayor rendimiento de producto que los procesos aerobios. Bajo condiciones aerobias, normalmente el rendimiento de la biomasa es superior a en condiciones anaerobias. Como consecuencia, normalmente bajo condiciones aerobias, el rendimiento de producto esperado es más bajo que bajo condiciones anaerobias. Según los inventores, el proceso de la invención es el primer proceso de fermentación anaerobio con un medio que comprende una fuente de L-arabinosa que ha sido desarrollado hasta la fecha.

En otra realización preferida, el proceso de fermentación es bajo condiciones limitadas de oxígeno. Más preferentemente, el proceso de fermentación es aerobio y bajo condiciones limitadas de oxígeno. Un proceso de fermentación limitado en oxígeno es un proceso en el que el consumo de oxígeno está limitado por la transferencia de oxígeno del gas al líquido. El grado de limitación de oxígeno se determina por la cantidad y composición del flujo de gas entrante, además de las propiedades de mezcla/transferencia de masa reales del equipo de fermentación usado. Preferentemente, en un proceso bajo condiciones limitadas de oxígeno, la tasa de consumo de oxígeno es al menos 5,5, más preferentemente al menos 6, e incluso más preferentemente al menos 7 mmol/l/h.

El proceso de fermentación se realiza preferentemente a una temperatura que es óptima para la célula modificada. Así, para la mayoría de las levaduras o células fúngicas, el proceso de fermentación se realiza a una temperatura

que es inferior a 42 °C, preferentemente inferior a 38 °C. Para levadura o células huésped fúngicas filamentosas, el proceso de fermentación se realiza preferentemente a una temperatura que es inferior a 35, 33, 30 o 28 °C y a una temperatura que es superior a 20, 22 o 25 °C.

5 Un proceso preferido es un proceso para la producción de etanol, por el cual el proceso comprende las etapas de:
 (a) fermentar un medio que contiene una fuente de L-arabinosa y opcionalmente xilosa con una célula huésped modificada como se define en el presente documento, por el cual la célula huésped fermenta L-arabinosa y
 10 opcionalmente xilosa en etanol; y opcionalmente, (b) recuperación del etanol. El medio de fermentación también puede comprender una fuente de glucosa que también se fermenta en etanol. En una realización preferida, el proceso de fermentación para la producción de etanol es anaerobio. Anaerobio ya se ha definido anteriormente en el
 presente documento. En otra realización preferida, el proceso de fermentación para la producción de etanol es aerobio. En otra realización preferida, el proceso de fermentación para la producción de etanol es bajo condiciones
 15 limitadas de oxígeno, más preferentemente aerobio y bajo condiciones limitadas de oxígeno. Las condiciones limitadas de oxígeno ya se han definido anteriormente en el presente documento.

En el proceso, la productividad volumétrica del etanol es preferentemente al menos 0,5, 1,0, 1,5, 2,0, 2,5, 3,0, 5,0 o
 15 10,0 g de etanol por litro por hora. El rendimiento de etanol sobre L-arabinosa y opcionalmente xilosa y/o glucosa en el proceso es preferentemente al menos el 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 95 o el 98 %. El rendimiento de etanol se define en el presente documento como un porcentaje del máximo rendimiento teórico que, para glucosa y
 L-arabinosa y opcionalmente xilosa, es 0,51 g de etanol por g de glucosa o xilosa. En otra realización preferida, la
 20 invención se refiere a un proceso de producción de un producto de fermentación seleccionado del grupo que consiste en ácido láctico, ácido 3-hidroxi-propiónico, ácido acrílico, ácido acético, ácido succínico, ácido cítrico, ácido málico, ácido fumárico, un aminoácido, 1,3-propano-diol, etileno, glicerol, butanol, un antibiótico β-lactámico y una
 cefalosporina. El proceso comprende preferentemente las etapas de (a) fermentar un medio que contiene una fuente
 de L-arabinosa y opcionalmente xilosa con una célula huésped modificada como se define en el presente documento
 25 anteriormente, por el cual la célula huésped fermenta L-arabinosa y opcionalmente xilosa en el producto de fermentación, y opcionalmente, (b) recuperación del producto de fermentación. En un proceso preferido, el medio también contiene una fuente de glucosa.

En el proceso de fermentación de la invención que conduce a la producción de etanol, pueden citarse varias ventajas por comparación con procesos de fermentación de etanol conocidos:

- son posibles procesos anaerobios.
- 30 - también son posibles condiciones limitadas de oxígeno.
- pueden obtenerse mayores rendimientos de etanol y velocidades de producción de etanol.
- la cepa usada puede ser capaz de usar L-arabinosa y opcionalmente xilosa.

Alternativamente a los procesos de fermentación descritos anteriormente, otro proceso de fermentación se
 35 proporciona como otro aspecto de la invención en el que al menos dos células distintas se usan para la fermentación de una fuente de carbono que comprende al menos dos fuentes de carbono seleccionadas del grupo que consiste en, pero no se limita a: una fuente de L-arabinosa, una fuente de xilosa y una fuente de glucosa. En este proceso de fermentación, "al menos dos células distintas" significa que este proceso es preferentemente un proceso de co-
 fermentación. En una realización preferida, se usan dos células distintas: siendo una la de la invención como se ha
 40 definido anteriormente capaz de usar L-arabinosa, y/o de convertirla en L-ribulosa y/o xilulosa-5-fosfato y/o en un producto de fermentación deseado tal como etanol y siendo opcionalmente capaz de usar xilosa, siendo la otra, por ejemplo, una cepa que es capaz de usar xilosa y/o de convertirla en un producto de fermentación deseado tal como etanol como se define en los documentos WO 03/062430 y/o WO 06/009434. Una célula que es capaz de usar xilosa es preferentemente una cepa que presenta la capacidad de isomerizar directamente xilosa en xilulosa (en una
 45 etapa) como se ha definido anteriormente en el presente documento. Estas dos cepas distintas se cultivan preferentemente en presencia de una fuente de L-arabinosa, una fuente de xilosa y opcionalmente una fuente de glucosa. Pueden co-cultivarse tres células distintas o más y/o pueden usarse tres o más fuentes de carbono, siempre que al menos una célula sea capaz de usar al menos una fuente de carbono presente y/o convertirla en un
 producto de fermentación deseado tal como etanol. La expresión "uso de al menos una fuente de carbono" tiene el
 mismo significado que la expresión "uso de L-arabinosa". La expresión "convertirla (es decir, una fuente de carbono)
 50 en un producto de fermentación deseado tiene el mismo significado que la expresión "convierten L-arabinosa en un producto de fermentación deseado".

En una realización preferida, la invención se refiere a un proceso de producción de un producto de fermentación
 55 seleccionado del grupo que consiste en etanol, ácido láctico, ácido 3-hidroxi-propiónico, ácido acrílico, ácido acético, ácido succínico, ácido cítrico, ácido málico, ácido fumárico, aminoácidos, 1,3-propano-diol, etileno, glicerol, butanol, antibióticos β-lactámicos y cefalosporinas, por el cual el proceso comprende las etapas de:

- (a) fermentar un medio que contiene al menos una fuente de L-arabinosa y una fuente de xilosa con una célula de la invención como se ha definido anteriormente en el presente documento y una célula capaz de

usar xilosa y/o que presenta la capacidad de isomerizar directamente xilosa en xilulosa, por el cual cada célula fermenta L-arabinosa y/o xilosa en el producto de fermentación, y opcionalmente,

(b) recuperar el producto de fermentación.

5 Todas las realizaciones preferidas de los procesos de fermentación que se han descrito anteriormente también son realizaciones preferidas de estos procesos de fermentación adicionales: identidad del producto de fermentación, identidad de la fuente de L-arabinosa y fuente de xilosa, condiciones de fermentación (condiciones aerobias o anaerobias, condiciones limitadas de oxígeno, temperatura a la que se está llevando a cabo el proceso, productividad de etanol, rendimiento de etanol).

Modificaciones genéticas

10 Para la expresión en exceso de enzimas en las células huésped de la invención como se ha descrito anteriormente, además de para modificación genética adicional de células huésped, preferentemente levaduras, las células huésped se transforman con las diversas construcciones de ácidos nucleicos de la invención por métodos muy conocidos en la técnica. Tales métodos son conocidos, por ejemplo, de libros de texto estándar, tales como Sambrook y Russel (2001) "Molecular Cloning: A Laboratory Manual (3ª edición), Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, o F. Ausubel et al, eds., "Current protocols in molecular biology", Green Publishing and Wiley Interscience, New York (1987). Métodos de transformación y modificación genética de células huésped fúngicas se conocen de, por ejemplo, los documentos EP-A-0 635 574, WO 98/46772, WO 99/60102 y WO 00/37671.

20 Se han descrito anteriormente promotores para su uso en las construcciones de ácidos nucleicos para la expresión en exceso de enzimas en las células huésped de la invención. En las construcciones de ácidos nucleicos para la expresión en exceso, el extremo 3' de la secuencia de ácido nucleótido que codifica la(s) enzima(s) está preferentemente operativamente enlazado a una secuencia terminadora de la transcripción. Preferentemente, la secuencia terminadora es operable en una célula huésped de elección tal como, por ejemplo, la especie de levadura de elección. En cualquier caso, la elección del terminador no es crítica; puede, por ejemplo, ser de cualquier gen de levadura, aunque los terminadores pueden algunas veces funcionar si son de un gen eucariota no de levadura. La secuencia de terminación de la transcripción comprende preferentemente adicionalmente una señal de poliadenilación. Secuencias terminadoras preferidas son los terminadores de alcohol deshidrogenasa (ADH1) y de PGI1. Más preferentemente, los terminadores de ADH1 y de PGI1 son ambos de *S. cerevisiae* (SEQ ID NO: 50 y SEQ ID NO: 53, respectivamente).

30 Opcionalmente, un marcador de selección puede estar presente en la construcción de ácidos nucleicos. Como se usa en el presente documento, el término "marcador" se refiere a un gen que codifica un rasgo o un fenotipo que permite la selección de, o el cribado de, una célula huésped que contiene el marcador. El gen marcador puede ser un gen de resistencia a antibiótico, por el cual el antibiótico apropiado puede usarse para seleccionar células transformadas de entre células que no están transformadas. Preferentemente, sin embargo, se usan marcadores de resistencia no de antibiótico, tales como marcadores auxotróficos (URA3, TRP1, LEU2). En una realización preferida, las células huésped transformadas con las construcciones de ácidos nucleicos están libres de gen marcador. Métodos de construcción de células huésped microbianas libres de gen marcador recombinante se desvelan en el documento EPA-0 635 574 y se basan en el uso de marcadores bidireccionales. Alternativamente, un marcador que puede cribarse tal como la proteína verde fluorescente, *lacZ*, luciferasa, cloranfenicol acetiltransferasa, beta-glucuronidasa puede incorporarse en las construcciones de ácidos nucleicos de la invención permitiendo cribar células transformadas.

45 Elementos adicionales opcionales que pueden estar presentes en las construcciones de ácidos nucleicos de la invención incluyen, pero no se limitan a, una o más secuencias conductoras, potenciadores, factores de integración y/o genes indicadores, secuencias de intrones, centrómeros, telómeros y/o secuencias de unión a matriz (MAR). Las construcciones de ácidos nucleicos de la invención pueden comprender además una secuencia para la replicación autónoma, tal como una secuencia ARS. Construcciones de ácidos nucleicos episómicas adecuadas pueden, por ejemplo, basarse en los plásmidos 2 μ o pKD1 de levadura (Fleer et al., 1991, Biotechnology 9:968-975). Alternativamente, la construcción de ácidos nucleicos puede comprender secuencias para la integración, preferentemente por recombinación homóloga. Tales secuencias pueden así ser secuencias homólogas al sitio diana para la integración en el genoma de la célula huésped. Las construcciones de ácidos nucleicos de la invención pueden proporcionarse de un modo por sí conocido, que generalmente implica técnicas tales como restringir y enlazar ácidos nucleicos/secuencias de ácidos nucleicos, para lo que se hace referencia a los libros de texto estándar, tales como Sambrook y Russel (2001) "Molecular Cloning: A Laboratory Manual (3ª edición), Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press.

55 Métodos de inactivación y rotura de genes en levadura u hongos son muy conocidos en la técnica (véase, por ejemplo, Fincham, 1989, Microbiol Rev. 53(1):148-70 y el documento EP-A-0 635 574).

En este documento y en sus reivindicaciones, el verbo "comprender" y sus conjugaciones se usan en su sentido no limitante para significar que las cosas que siguen a la palabra están incluidas, pero las cosas no específicamente

mencionadas no se excluyen. Además, referencia a un elemento por el artículo indefinido "un" o "una" no excluye la posibilidad de que más de un elemento esté presente, a menos que el contexto requiera claramente que haya uno y solo uno de los elementos. El artículo indefinido "un" o "una" significa así normalmente "al menos uno".

5 La invención se describe adicionalmente por los siguientes ejemplos, que no deben interpretarse como limitantes del alcance de la invención.

Ejemplos

Construcción de plásmidos y de cepas

Cepas

10 La cepa de *Saccharomyces cerevisiae* que consume L-arabinosa descrita en este trabajo se basa en la cepa RWB220, que es ella misma un derivado de RWB217. RWB217 es una cepa CEN.PK en la que cuatro genes que codifican la expresión de enzimas en la vía de pentosa fosfato han sido expresados en exceso, *TAL1*, *TKL1*, *RPE1*, *RKI1* (Kuyper et al., 2005a). Además, se ha delecionado el gen que codifica una aldosa reductasa (*GRE3*). La cepa RWB217 también contiene dos plásmidos, una plásmido de copia única con un marcador *LEU2* para la expresión en exceso de la xilulocinasa (*XKS1*) y un plásmido episómico de copias múltiples con *URA3* como marcador para la

15 expresión de la xilosa isomerasa, *XylA*. Se sometió RWB217 a un procedimiento de selección para crecimiento mejorado sobre xilosa que se describe en Kuyper et al. (2005b). El procedimiento produjo dos cepas puras, RWB218 (Kuyper et al., 2005b) y RWB219. La diferencia entre RWB218 y RWB219 es que después del procedimiento de selección, RWB218 se obtuvo sembrando y volviendo a cultivar en línea sobre medio mineral con glucosa como fuente de carbono, mientras que para RWB219 se usó xilosa.

20 La cepa RWB219 se cultivó no selectivamente sobre YP con glucosa (YPD) como fuente de carbono con el fin de facilitar la pérdida de ambos plásmidos. Después de sembrar sobre YPD, se probaron colonias individuales para la pérdida de plásmidos estudiando cuidadosamente la auxotropía de uracilo y leucina. Una cepa que había perdido ambos plásmidos se transformó con pSH47, que contiene la recombinasa cre, con el fin de eliminar un casete KanMX (Guldener et al., 1996), todavía presente después de integrar la construcción de expresión en exceso *RKI1*.

25 Colonias con el plásmido se resuspendieron en medio de peptona de levadura (YP) (10 g/l de extracto de levadura y 20 g/l de peptona, ambos de BD Difco Belgium) con 1 % de galactosa y se incubaron durante 1 hora a 30 °C. Se sembraron aproximadamente 200 células sobre YPD. Las colonias resultantes se comprobaron para pérdida del marcador KanMX (resistencia a G418) y pSH47 (*URA3*). Una cepa que había perdido tanto el marcador KanMX como el plásmido pSH47 se llamó entonces RWB220. Para obtener la cepa probada en esta patente, RWB220 se transformó con pRW231 y pRW243 (Tabla 2), produciendo la cepa IMS0001.

30

Durante la construcción, las cepas se mantuvieron sobre YP complejo: 10 g l⁻¹ de extracto de levadura (BD Difco), 20 g l⁻¹ de peptona (BD Difco) o medio sintético (MY) (Verduyn et al., 1992) complementado con glucosa (2 %) como fuente de carbono (YPD o MYD) y 1,5 % de agar en el caso de placas. Después de la transformación con plásmidos, las cepas se sembraron sobre MYD.

35 Las transformaciones de levadura se hicieron según Gietz y Woods (2002). Los plásmidos se amplificaron en la cepa XL-1 blue de *Escherichia coli* (Stratagene, La Jolla, CA, EE.UU.). La transformación se realizó según Inoue et al. (1990). Se cultivó *E. coli* sobre placas de LB (Luria-Bertani) o en medio TB líquido (caldo Terrific) para el aislamiento de plásmidos (Sambrook et al, 1989).

Plásmidos

40 Con el fin de cultivar sobre L-arabinosa, la levadura necesita expresar tres genes diferentes, una L-arabinosa isomerasa (*AraA*), una L-ribulocinasa (*AraB*) y una L-ribulosa-5-P 4-epimerasa (*AraD*) (Becker y Boles, 2003). En este trabajo, los presentes inventores han elegido expresar *AraA*, *AraB* y *AraD* a partir de la bacteria ácido láctica *Lactobacillus plantarum* en *S. cerevisiae*. Debido a que el objetivo eventual es consumir L-arabinosa en combinación con otros azúcares, como D-xilosa, los genes que codifican la vía de L-arabinosa bacteriana se combinaron sobre el mismo plásmido con los genes que codifican el consumo de D-xilosa.

45

Con el fin de conseguir un alto nivel de expresión, los genes *AraA* y *AraD* de *L. plantarum* se ligaron dentro del plásmido pAKX002, el plásmido que lleva 2μ *XylA*.

Se construyó el casete de *AraA* amplificando una versión truncada del promotor de TDH3 con SpeI5'Ptdh3 y 5'AraAPtdh3 (SEQ ID NO: 49), el gen *AraA* con Ptdh5'AraA y Tadh3'AraA y el terminador ADH1 (SEQ ID NO: 50) con 3'AraATadh1 y 3'Tadh1-SpeI. Los tres fragmentos se extrajeron del gel y se mezclaron en cantidades aproximadamente equimolares. Sobre esta mezcla se realizó una PCR usando los oligonucleótidos SpeI-5'Ptdh3 y 3'Tadh1SpeI. El casete P_{TDH3}-*AraA*-T_{ADH1} resultante se purificó en gel, se cortó en los sitios SpeI de 5' y 3' y entonces se ligó en pAKX002 cortado con NheI, produciendo el plásmido pRW230.

50

La construcción de *AraD* se produjo amplificando primero una versión truncada del promotor de HXT7 (SEQ ID NO: 52) con los oligonucleótidos Sall5'Phxt7 y 5'AraDPhxt, el gen *AraD* con Phxt5' *AraD* y Tpgi3'AraD y la región

55

terminadora GPI1 (SEQ ID NO: 53) con los oligonucleótidos 3'AraDTpgi y 3'TpgiSall. Los fragmentos resultantes se extrajeron del gel y se mezclaron en cantidades aproximadamente equimolares, después de lo cual se realizó una PCR usando los oligonucleótidos Sall5'Phxt7 y 3'Tpgi1Sall. El casete P_{HXT7}-AraD-T_{PGI1} resultante se purificó en gel, se cortó en los sitios Sall de 5' y 3' y entonces se ligó en pRW230 cortado con XhoI, produciendo el plásmido pRW231 (Figura 1).

Debido a que una expresión demasiado alta de la L-ribulocinasa es perjudicial para el crecimiento (Becker y Boles, 2003), el gen *AraB* se combinó con el gen XKS1, que codifica xilulocinasa, sobre un plásmido de integración. Para esto, p415ADHXKS (Kuyper et al., 2005a) se cambió primero a pRW229, cortando tanto p415ADHXKS como pRS305 con PvuI y ligando el fragmento de PvuI que contiene ADHXKS de p415ADHXKS al esqueleto de vector de pRS305, produciendo pRW229.

Se produjo un casete, que contiene el gen *AraB* de *L. plantarum* entre el promotor de PGI1 (SEQ ID NO: 51) y el terminador de ADH1 (SEQ ID NO: 50) amplificando el promotor de PGI1 con los oligonucleótidos SacI5'Ppgi1 y 5'AraBPpgi1, el gen *AraB* con los oligonucleótidos Ppgi5'AraB y Tadh3'AraB y el terminador de ADH1 con los oligonucleótidos 3'AraBTadh1 y 3'Tadh1SacI. Los tres fragmentos se extrajeron del gel y se mezclaron en cantidades aproximadamente equimolares. Sobre esta mezcla se realizó una PCR usando los oligonucleótidos SacI-5'Ppgi1 y 3'Tadh1SacI. El casete P_{PGI1}-AraB-T_{ADH1} resultante se purificó en gel, se cortó en los sitios SacI de 5' y 3' y entonces se ligó en pRW229 cortado con SacI, produciendo el plásmido pRW243 (Figura 1).

La cepa RWB220 se transformó con pRW231 y pRW243 (Tabla 2), produciendo la cepa IMS0001.

Se usaron endonucleasas de restricción (New England Biolabs, Beverly, MA, EE.UU. y Roche, Basilea, Suiza) y ADN ligasa (Roche) según las especificaciones del fabricante. El aislamiento del plásmido de *E. coli* se realizó con el kit Qiaprep spin miniprep (Qiagen, Hilden, Alemania). Los fragmentos de ADN se separaron sobre un gel de 1 % de agarosa (Sigma, St. Louis, MO, EE.UU.) en 1×TBE (Sambrook et al, 1989). El aislamiento de fragmentos de gel se llevó a cabo con el kit de extracción en gel Qiaquick (Quiagen). La amplificación de los (elementos de los) casetes de AraA, AraB y AraD se hizo con ADN polimerasa Vent_R (New England Biolabs) según la especificación del fabricante. El molde fue ADN cromosómico de *S. cerevisiae* CEN.PK113-7D para los promotores y terminadores, o *Lactobacillus plantarum* DSM20205 para los genes Ara. La reacción en cadena de la polimerasa (PCR) se realizó en un ciclador térmico Biometra TGradient (Biometra, Göttingen, Alemania) con los siguientes parámetros: 30 ciclos de 1 min de hibridación a 55 °C, 60 °C o 65 °C, 1 a 3 min de extensión a 75 °C, dependiendo del tamaño de fragmento esperado, y 1 min de desnaturalización a 94 °C.

30 Cultivo y medios

Se realizaron cultivos en matraces con agitación a 30 °C en un medio sintético (Verduyn et al., 1992). El pH del medio se ajustó a 6,0 con KOH 2 M antes de la esterilización. Para medio sintético sólido se añadió 1,5 % de agar.

Se prepararon pre-cultivos inoculando 100 ml de medio que contenía el azúcar apropiado en un matraz oscilante de 500 ml con un cultivo madre congelado. Después de la incubación a 30 °C en un agitador orbital (200 rpm), este cultivo se usó para inocular tanto los cultivos de matraz agitado como los cultivos de fermentador. El medio sintético para el cultivo anaerobio se complementó con 0,01 g l⁻¹ de ergosterol y 0,42 g l⁻¹ de Tween 80 disuelto en etanol (Andreasen y Stier, 1953; Andreasen y Stier, 1954). El cultivo discontinuo anaerobio (secuenciación) se llevó a cabo a 30 °C en fermentadores de laboratorio de 2 l (Applikon, Schiedam, Los Países Bajos) con un volumen de trabajo de 1 l. El pH del cultivo se mantuvo a pH 5,0 por adición automática de KOH 2 M. Los cultivos se agitaron a 800 rpm y se burbujearon con 0,5 l min⁻¹ de gas nitrógeno (<10 ppm de oxígeno). Para minimizar la difusión de oxígeno, los fermentadores se equiparon con tubos de Norprene (Col Palmer Instrument company, Vernon Hills, EE.UU.). El oxígeno disuelto se monitorizó con un electrodo de oxígeno (Applisens, Schiedam, Los Países Bajos). Se lograron las condiciones limitadas de oxígeno en la misma configuración experimental por aireación del espacio de cabeza a aproximadamente 0,05 l min⁻¹.

45 Determinación del peso seco

Se filtraron muestras de cultivo (10,0 ml) sobre filtros de nitrocelulosa previamente pesados (tamaño de poro 0,45 μm; Gelman laboratory, Ann Arbor, EE.UU.). Después de eliminar el medio, los filtros se lavaron con agua desmineralizada y se secaron en un horno microondas (Bosch, Stuttgart, Alemania) durante 20 min a 360 W y se pesaron. Las determinaciones por duplicado variaron menos del 1 %.

50 Análisis de gases

Se enfrió gas de escape en un condensador (2 °C) y se secó con una secadora Permapure tipo MD-110-48P-4 (Permapure, Toms River, EE.UU.). Se determinaron las concentraciones de O₂ y CO₂ con un analizador NGA 2000 (Rosemount Analytical, Orrville, EE.UU.). Se determinaron el flujo del gas de escape y las tasas específicas de consumo de oxígeno y de producción de dióxido de carbono como se ha descrito previamente (Van Urk et al., 1988; Weusthuis et al., 1994). En el cálculo de estas tasas específicas de biomasa, se tuvieron en cuenta los cambios de volumen producidos por la extracción de muestras de cultivo.

Análisis de metabolitos

Se analizaron glucosa, xilosa, arabinosa, xilitol, ácidos orgánicos, glicerol y etanol por HPLC usando un Waters Alliance 2690 HPLC (Waters, Milford, EE.UU.) suministrado con una columna BioRad HPX 87H (BioRad, Hercules, EE.UU.), un detector del índice de refracción 2410 de Waters y un detector de UV 2487 de Waters. La columna se eluyó a 60 °C con 0,5 g l⁻¹ de ácido sulfúrico a un caudal de 0,6 ml min⁻¹.

Ensayo para xilulosa-5-fosfato (Zaldivar J., et al, Appl. Microbiol. Biotechnol., (2002), 59:436-442)

Para el análisis de metabolitos intracelulares tales como xilulosa-5-fosfato, se recogieron 5 ml de caldo por duplicado de los reactores, antes del agotamiento de la glucosa (a las 22 y 26 h de cultivo) y después del agotamiento de la glucosa (42, 79 y 131 h de cultivo). Los procedimientos para la parada metabólica, extracción de metabolitos en fase sólida y análisis se han descrito en detalle por Smits H.P. et al. (Anal. Biochem., 261:36-42, (1998)). Sin embargo, el análisis por cromatografía de intercambio iónico a alta presión acoplado a detección amperométrica pulsada usado para analizar extractos celulares se modificó ligeramente. Las disoluciones usadas fueron eluyente A, NaOH 75 mM, y eluyente B, NaAc 500 mM. Para prevenir la contaminación de carbonato en las disoluciones de eluyente, se usó una disolución al 50 % de NaOH con baja concentración de carbonato (Baker Analysed, Deventer, Los Países Bajos) en lugar de lentejas de NaOH. Los eluyentes se desgasificaron con helio (He) durante 30 min y a continuación se mantuvieron bajo una atmósfera de He. La bomba de gradiente se programó para generar los siguientes gradientes: 100 % de A y 0 % de B (0 min), una disminución lineal de A al 70 % y un aumento lineal de B al 30 % (0-30 min), una disminución lineal de A al 30 % y un aumento lineal de B al 70 % (30-70 min), una disminución lineal de A al 0 % y un aumento lineal de B al 100 % (70-75 min), 0 % de A y 100 % de B (75-85 min), un aumento lineal de A al 100 % y una disminución lineal de B al 0 % (85-95 min). La fase móvil se ejecutó a un caudal de 1 ml/min. Otras condiciones fueron según Smits et al. (1998).

Recuperación de carbono

Se calcularon las recuperaciones de carbono como carbono en los productos formados dividido entre la cantidad total de carbono de azúcar consumido, y se basaron en un contenido de carbono de la biomasa del 48 %. Para corregir para la evaporación de etanol durante las fermentaciones, la cantidad de etanol producida se supuso que era igual a la producción acumulada medida de CO₂ menos la producción de CO₂ que se produjo debido a la síntesis de biomasa (5,85 mmoles de CO₂ por gramo de biomasa (Verduyn et al., 1990)) y el CO₂ asociado a la formación de acetato.

Selección para el crecimiento sobre L-arabinosa

Se construyó la cepa IMS0001 (CBS 120327 depositada en CBS el 27/09/06), que contiene los genes que codifican las vías para tanto la metabolización de xilosa (XylA y XKS1) como de arabinosa (AraA, AraB, AraD), según el procedimiento descrito anteriormente. Aunque es capaz de crecer sobre xilosa (datos no mostrados), la cepa IMS0001 no pareció ser capaz de crecer sobre medio sintético sólido complementado con 2 % de L-arabinosa. Se seleccionaron mutantes de IMS0001 capaces de utilizar L-arabinosa como fuente de carbono para el crecimiento por transferencia en serie en matraces con agitación y por cultivo discontinuo secuencial en fermentadores (SBR).

Para los experimentos de transferencia en serie, un matraz oscilante de 500 ml que contiene 100 ml de medio sintético que contiene 0,5 % de galactosa se inoculó con cualquiera de la cepa IMS0001, o la cepa de referencia RWB219. Después de 72 horas, a una densidad óptica a 660 nm de 3,0, los cultivos se usaron para inocular un nuevo matraz oscilante que contenía 0,1 % de galactosa y 2 % de arabinosa. Basándose en la determinación por HPLC con D-ribulosa como patrón de calibración, se determinó que ya en los primeros cultivos de la cepa IMS0001, sobre medio que contiene una mezcla de galactosa/arabinosa, parte de la arabinosa se convirtió en ribulosa y posteriormente se secretó al sobrenadante. Estos análisis de HPLC se realizaron usando un Waters Alliance 2690 HPLC (Waters, Milford, EE.UU.) suministrado con una columna BioRad HPX 87H (BioRad, Hercules, EE.UU.), un detector del índice de refracción 2410 de Waters y un detector de UV 2487 de Waters. La columna se eluyó a 60 °C con 0,5 g l⁻¹ de ácido sulfúrico a un caudal de 0,6 ml min⁻¹. A diferencia de la cepa de referencia RWB219, la DO₆₆₀ del cultivo de la cepa IMS0001 aumentó después del agotamiento de la galactosa. Cuando después de aproximadamente 850 horas se observó el crecimiento sobre arabinosa por la cepa IMS0001 (Figura 2), este cultivo se transfirió a una DO₆₆₀ de 1,7 a un matraz oscilante que contenía 2 % de arabinosa. Entonces, los cultivos se transfirieron secuencialmente a medio fresco que contenía 2 % de arabinosa a una DO₆₆₀ de 2-3. La utilización de arabinosa se confirmó midiendo ocasionalmente concentraciones de arabinosa por HPLC (datos no mostrados). La velocidad de crecimiento de estos cultivos aumentó de 0 a 0,15 h⁻¹ en aproximadamente 3600 horas (Figura 3).

Se empezó una fermentación discontinua bajo condiciones limitadas de oxígeno inoculando 1 l de medio sintético complementado con 2 % de arabinosa con un matraz oscilante de 100 ml de cultivo de células IMS0001 cultivadas con arabinosa con una velocidad de crecimiento máxima sobre 2 % de L-arabinosa de aproximadamente 0,12 h⁻¹. Cuando se observó crecimiento sobre arabinosa, el cultivo se sometió a condiciones anaerobias burbujeando con gas nitrógeno. Los ciclos secuenciales de cultivo discontinuo anaerobio empezaron por sustitución tanto manual como automatizada del 90 % del cultivo con medio sintético con 20 g l⁻¹ de arabinosa. Para cada ciclo durante la fermentación en SBR, la velocidad de crecimiento exponencial se estimó a partir del perfil de CO₂ (Figura 4). En 13

ciclos, la velocidad de crecimiento exponencial aumentó de 0,025 a 0,08 h⁻¹. Después de 20 ciclos se tomó una muestra, y se sembró sobre medio sintético sólido complementado con 2 % de L-arabinosa y se incubó a 30 °C durante varios días. Volvieron a cultivarse en línea dos veces colonias separadas sobre medio sintético sólido con L-arabinosa. Finalmente, un matraz oscilante que contenía medio sintético con 2 % de L-arabinosa se inoculó con una única colonia, y se incubó durante 5 días a 30 °C. Este cultivo se designó la cepa IMS0002 (CBS 120328 depositada en Centraal Bureau voor Schimmelculturen (CBS) el 27/09/06). Se tomaron muestras de cultivo, se añadió 30 % de glicerol y las muestras se almacenaron a -80 °C.

Fermentación de cultivos mixtos

Los hidrolizados de biomasa, una materia prima deseada para la biotecnología industrial, contienen mezclas complejas que consisten en diversos azúcares entre los que la glucosa, xilosa y arabinosa están comúnmente presentes en fracciones significativas. Para realizar la fermentación etanólica de no solo glucosa y arabinosa, sino también xilosa, se realizó una fermentación discontinua anaerobia con un cultivo mixto de la cepa IMS0002 que fermenta arabinosa, y la cepa RWB218 que fermenta xilosa. Un fermentador discontinuo anaerobio que contiene 800 ml de medio sintético suministrado con 30 g l⁻¹ de D-glucosa, 15 g l⁻¹ de D-xilosa y 15 g l⁻¹ de L-arabinosa se inoculó con 100 ml de pre-cultivo de la cepa IMS0002. Después de 10 horas, se añadió un inóculo de 100 ml de RWB218. A diferencia de la fermentación de azúcares mixtos con solo la cepa IMS0002, tanto la xilosa como la arabinosa se consumieron después del agotamiento de la glucosa (Fig. 5D). El cultivo mixto consumió completamente todos los azúcares, y en el plazo de 80 horas se produjeron 564,0 ± 6,3 mmol l⁻¹ de etanol (calculado a partir de la producción de CO₂) con un alto rendimiento global de 0,42 g g⁻¹ de azúcar. El xilitol se produjo solo en pequeñas cantidades, a una concentración de 4,7 mmol l⁻¹.

Caracterización de la cepa IMS0002

El crecimiento y la formación de productos de la cepa IMS0002 se determinó durante las fermentaciones discontinuas anaerobias sobre medio sintético con tanto L-arabinosa como la única fuente de carbono, como una mezcla de glucosa, xilosa y L-arabinosa. Los pre-cultivos para estas fermentaciones discontinuas anaerobias se prepararon en matraces con agitación que contenían 100 ml de medio sintético con 2 % de L-arabinosa, inoculando con reservas congeladas a -80 °C de la cepa IMS0002, e incubando durante 48 horas a 30 °C.

La Figura 5A muestra que la cepa IMS0002 es capaz de fermentar 20 g l⁻¹ de L-arabinosa a etanol durante una fermentación discontinua anaerobia de aproximadamente 70 horas. La velocidad de crecimiento específica en condiciones anaerobias con L-arabinosa como única fuente de carbono fue 0,05 ± 0,001 h⁻¹. Teniendo en cuenta la evaporación de etanol durante la fermentación discontinua, el rendimiento de etanol a partir de 20 g l⁻¹ de arabinosa fue 0,43 ± 0,003 g g⁻¹. Sin corrección por la evaporación, el rendimiento del etanol fue 0,35 ± 0,01 g g⁻¹ de arabinosa. No se observó formación de arabinitol durante el crecimiento anaerobio sobre arabinosa.

En la Figura 5B se muestra la fermentación etanólica de una mezcla de 20 g l⁻¹ de glucosa y 20 g l⁻¹ de L-arabinosa por cepa IMS0002. El consumo de L-arabinosa empezó después del agotamiento de glucosa. En el plazo de 70 horas, tanto la glucosa como la L-arabinosa se habían consumido completamente. El rendimiento de etanol a partir del total de los azúcares fue 0,42 ± 0,003 g g⁻¹.

En la Figura 5C se muestra el perfil de fermentación de una mezcla de 30 g l⁻¹ de glucosa, 15 g l⁻¹ de D-xilosa y 15 g l⁻¹ de L-arabinosa por cepa IMS0002. El consumo de arabinosa empezó después del agotamiento de glucosa. En el plazo de 80 horas, tanto la glucosa como la arabinosa se habían consumido completamente. Solo se consumieron 20 mM de 100 mM de xilosa por la cepa IMS0002. Además, se observó la formación de 20 mM de xilitol. Aparentemente, la xilosa se convirtió en xilitol por la cepa IMS0002. Por lo tanto, el rendimiento de etanol del total de azúcares fue más bajo que para las fermentaciones anteriormente descritas: 0,38 ± 0,001 g g⁻¹. El rendimiento de etanol a partir del total de glucosa y arabinosa fue similar al de otras fermentaciones: 0,43 ± 0,001 g g⁻¹.

La Tabla 1 muestra las tasas de consumo de arabinosa y las tasas de producción de etanol observadas para la fermentación discontinua anaerobia de la cepa IMS0002. La arabinosa se consumió con una tasa de 0,23 - 0,75 g h⁻¹ g⁻¹ de peso seco de biomasa. La tasa de etanol producida a partir de la arabinosa varió de 0,08 - 0,31 g h⁻¹ g⁻¹ de peso seco de biomasa.

Inicialmente, la cepa IMS0001 construida fue capaz de fermentar xilosa (datos no mostrados). A diferencia de las expectativas de los presentes inventores, la cepa seleccionada IMS0002 no fue capaz de fermentar xilosa a etanol (Fig. 5C). Para recuperar la capacidad de fermentar xilosa, una colonia de la cepa IMS0002 se transfirió a medio sintético sólido con 2 % de D-xilosa, y se incubó en un recipiente anaerobio a 30 °C durante 25 días. Posteriormente, una colonia se transfirió otra vez a medio sintético sólido con 2 % de arabinosa. Después de 4 días de incubación a 30 °C, se transfirió una colonia a un matraz oscilante que contenía medio sintético con 2 % de arabinosa. Después de la incubación a 30 °C durante 6 días, se añadió 30 % de glicerol, se tomaron muestras y se guardaron a -80 °C. Se inoculó un matraz oscilante que contenía 100 ml de medio sintético con 2 % de arabinosa con una reserva congelada, y se usó como precultivo para una fermentación discontinua anaerobia sobre medio sintético con 20 g l⁻¹ de xilosa y 20 g l⁻¹ de arabinosa. En la Figura 6 se muestra el perfil de fermentación de esta fermentación discontinua. La xilosa y la arabinosa se consumieron simultáneamente. La arabinosa se completó en el plazo de 70

horas, mientras que la xilosa se consumió completamente en 120 horas. Se produjeron al menos 250 mM de etanol a partir del total de azúcares, no teniendo en cuenta la evaporación del etanol. Suponiendo un peso seco de biomasa final de 3,2 g l⁻¹ (suponiendo un rendimiento de la biomasa de 0,08 g g⁻¹ de azúcar), la concentración de etanol final estimada a partir de la producción de CO₂ acumulada (355 mmol l⁻¹) fue aproximadamente 330 mmol l⁻¹, correspondiente a un rendimiento de etanol de 0,41 g g⁻¹ de azúcar pentosa. Además de etanol, glicerol y ácidos orgánicos, se produjo una pequeña cantidad de xilitol (aproximadamente 5 mM).

Selección de la cepa IMS0003

Inicialmente, la cepa construida IMS0001 fue capaz de fermentar xilosa (datos no mostrados). A diferencia de las expectativas de los presentes inventores, la cepa seleccionada IMS0002 no fue capaz de fermentar xilosa a etanol (Fig. 5C). Para recuperar la capacidad de fermentar xilosa, una colonia de la cepa IMS0002 se transfirió a medio sintético sólido con 2 % de D-xilosa, y se incubó en un recipiente anaerobio a 30 °C durante 25 días. Posteriormente, una colonia se transfirió otra vez a medio sintético sólido con 2 % de arabinosa. Después de 4 días de incubación a 30 °C, se transfirió una colonia a un matraz oscilante que contenía medio sintético con 2 % de arabinosa. Después de la incubación a 30 °C durante 6 días, se añadió 30 % de glicerol, se tomaron muestras y se guardaron a -80 °C.

A partir de la reserva congelada, las muestras se extendieron sobre medio sintético sólido con 2 % de L-arabinosa y se incubaron a 30 °C durante varios días. Volvieron a cultivarse dos veces en línea colonias separadas sobre medio sintético sólido con L-arabinosa. Finalmente, un matraz oscilante que contenía medio sintético con 2 % de L-arabinosa se inoculó con una única colonia, y se incubó durante 4 días a 30 °C. Este cultivo se designó la cepa IMS0003 (CBS 121879 depositada en CBS el 20/09/07). Se tomaron muestras de cultivo, se añadió 30 % de glicerol y las muestras se almacenaron a -80 °C.

Caracterización de la cepa IMS0003

Se determinó el crecimiento y la formación de productos de la cepa IMS0003 durante una fermentación discontinua anaerobia sobre medio sintético con una mezcla de 30 g l⁻¹ de glucosa, 15 g l⁻¹ de D-xilosa y 15 g l⁻¹ de L-arabinosa. El pre-cultivo para esta fermentación discontinua anaerobia se preparó en un matraz con agitación que contenía 100 ml de medio sintético con 2 % de L-arabinosa, inoculando con una reserva congelada a -80 °C de la cepa IMS0003, y se incubó durante 48 horas a 30 °C.

En la Figura 7 se muestra el perfil de fermentación de una mezcla de 30 g l⁻¹ de glucosa, 15 g l⁻¹ de D-xilosa y 15 g l⁻¹ de L-arabinosa por cepa IMS0003. El consumo de arabinosa empezó después del agotamiento de la glucosa. En el plazo de 70 horas, la glucosa, la xilosa y la arabinosa se consumieron completamente. La xilosa y la arabinosa se consumieron simultáneamente. Se produjo al menos 406 mM de etanol a partir del total de azúcares, no teniendo en cuenta la evaporación del etanol. La concentración de etanol final calculada a partir de la producción de CO₂ acumulada fue 572 mmol l⁻¹, correspondiente a un rendimiento de etanol de 0,46 g g⁻¹ de azúcar total. A diferencia de la fermentación de una mezcla de glucosa, xilosa y arabinosa por cepa IMS0002 (Figura 5C) o un cultivo mixto de las cepas IMS0002 y RWB218 (Figura 5D), la cepa IMS0003 no produjo cantidades detectables de xilitol.

TABLAS

Tabla 1: Cepas de *S. cerevisiae* usadas.

Cepa	Características	Referencia
RWB217	<i>MATA ura3-52 leu2-112 loxP-P_{TPI}::(-266,-1)TAL1 gre3::hphMX pUGP_{TPI} TKL1 pUGP_{TPI}-RPE1 KanloxP-P_{TPI}::(-?-1)RKI1 {p415ADHXKS, pAKX002}</i>	Kuyper et al. 2005a
RWB218	<i>MATA ura3-52 leu2-112 loxP-P_{TPI}::(-266,-1)TAL1 gre3::hphMX pUGP_{TPI} TKL1 pUGP_{TPI}-RPE1 KanloxP-P_{TPI}::(-?-1)RKI1 {p415ADHXKS1, pAKX002}</i>	Kuyper et al. 2005b
RWB219	<i>MATA ura3-52 leu2-112 loxP-P_{TPI}::(-266,-1)TAL1 gre3::hphMX pUGP_{TPI} TKL1 pUGP_{TPI}-RPE1 KallloxP-P_{TPI}::(-?-1)RKI1 {p415ADHXKS1, pAKX002}</i>	Este trabajo
RWB220	<i>MATA ura3-52 leu2-112 loxP-P_{TPI}::(-266,-1)TAL1 gre3::hphMX pUGP_{TPI} TKL1 pUGP_{TPI}-RPE1 loxP-P_{TPI}::(-?-1)RKI1</i>	Este trabajo
IMS0001	<i>MATA ura3-52 leu2-112 loxP-P_{TPI}::(-266,-1)TAL1 gre3::hphMX pUGP_{TPI} TKLI pUGP_{TPI}-RPE1 loxP-P_{TPI}::(-?-1)RKI1 {pRW231, PRW243}</i>	Este trabajo
IMS0002	<i>MATA ura3-52 leu2-112 loxP-P_{TPI}::(-266,-1)TAL1 gre3:: hphMX pUGP_{TPI} TKL1 pUGP_{TPI}-RPE1 loxP-P_{TPI}::(-?-1)RKI1 {pRW231, PRW243}</i> seleccionada para crecimiento anaerobio sobre L-arabinosa	Este trabajo

Tabla 2: Plásmidos usados

plásmido	características	Referencia
pRS305	Integración, LEU2	Gietz and Sugino, 1988
pAKX002	2 μ , URA3, P _{TP11} -Piromyces xylA	Kuyper et al. 2003
p415ADHXKS1	CEN, LEU2, P _{ADH1} -S.cerXKS1	Kuyper et al., 2005a
pRW229	integración, LEU2, P _{ADH1} -S.cerXKS1	Este trabajo
pRW230	pAKX002 con P _{TDH3} -AraA	Este trabajo
pRW231	pAKX002 con P _{TDH3} -AraA y P _{HXT7} -AraD	Este trabajo
pRW243	LEU2, integración, P _{ADH1} -ScXKS1-T _{CYC} , P _{PGH1} -L. plantarum AraB-T _{ADH1}	Este trabajo

Tabla 3: oligonucleótidos usados en este trabajo

Oligonucleótido	Secuencia de ADN
Casete de expresión de AraA	
SpeI5'Ptdh3 SEQ ID NO:31	5'GACTAGTTCGAGTTTATCATTATCAATACTGC3'
5'AraAPtdh SEQ ID NO:32	5'CTCATAATCAGGTACTGATAACATTTTGTGGTTTATGTGTGTTTATTC3'
Ptdh5' AraA SEQ ID NO:33	5'GAATAAACACACATAAACAAACAAAATGTTATCAGTACCTGATTATGAG3
Tadh3'AraA SEQ ID NO:34	5'AATCATAAATCATAAGAAATTCGCTTACTTTAAGAATGCCTTAGTCAT3'
3'AraATadh1 SEQ ID NO:35	5'ATGACTAAGGCATTCTTAAAGTAAGCGAATTTCTTATGATTTATGATT3'
3'Tadh1SpeI SEQ ID NO:36	5'CACTAGTCTCGAGTGTGGAAGAACGATTACAACAGG3'
Casete de expresión de AraB	
SacI5'Ppgi1 SEQ ID NO:37	5'CGAGCTCGTGGGTGTATTGGATTATAGGAAG3'
5'AraBPpgi1 SEQ ID NO:38	5'TTGGGCTGTTTCAACTAAATTCATTTTTAGGCTGGTATCTTGATTCTA3'
Ppgi5' AraB SEQ ID NO:39	5'TAGAATCAAGATACCAGCCTAAAAATGAATTTAGTTGAAACAGCCCAA3'
Tadh3'AraB SEQ ID NO:40	5'AATCATAAATCATAAGAAATTCGCTCTAATATTTGATTGCTTGCCAG3'
3'AraBTadh1 SEQ ID NO:41	5'CTGGGCAAGCAATCAAATATTAGAGCGAATTTCTTATGATTTATGATT3'
3'Tadh1SacI SEQ ID NO:42	5'TGAGCTCGTGTGGAAGAACGATTACAACAGG3'
Casete de expresión de AraD	
Sall5'Phxt7 SEQ ID NO:43	5'ACGCGTCGACTCGTAGGAACAATTCGG3'
5'AraDPhxt SEQ ID NO:44	5'CTTCTTGTTTTAATGCTTCTAGCATTTTTTGATTAATAAATAAAAAACTT3'
Phxt5'AraD SEQ ID NO:45	5'AAGTTTTTTAATTTAATCAAAAAATGCTAGAAGCATTAAAAACAAGAAG3'
Tpgi3'AraD SEQ ID NO:46	5'GGTATATATTTAAGAGCGATTTGTTTACTTGCGAACTGCATGATCC3'
3'AraDTpgi SEQ ID NO:47	5'GGATCATGCAGTTCGCAAGTAAACAAATCGCTCTTAAATATATACC3'
3'TpgiSall SEQ ID NO:48	5'CGCAGTCGACCTTTTAAACAGTTGATGAGAACC3'

Tabla 4.

Tasas específicas de consumo de glucosa y arabinosa observadas máximas y tasas de producción de etanol durante las fermentaciones discontinuas anaerobias de IMS0002 de *S. cerevisiae*.

Q_{glu} : tasa específica de consumo de glucosa

5 Q_{ara} : tasa específica de consumo de arabinosa

$Q_{et,glu}$: tasa específica de producción de etanol durante el crecimiento sobre glucosa

$Q_{et,ara}$: tasa específica de producción de etanol durante el crecimiento sobre arabinosa

Fuente de C	Q_{glu} g h ⁻¹ g ⁻¹ de PS	Q_{ara} g h ⁻¹ g ⁻¹ de PS	$Q_{et,glu}$ g h ⁻¹ g ⁻¹ de PS	$Q_{et,ara}$ g h ⁻¹ g ⁻¹ de PS
20 g l ⁻¹ de arabinosa	-	0,75 ± 0,04	-	0,31 ± 0,02
20 g l ⁻¹ de glucosa 20 g l ⁻¹ de arabinosa	2,08 ± 0,09	0,41 ± 0,01	0,69 ± 0,00	0,19 ± 0,00
30 g l ⁻¹ de glucosa 15 g l ⁻¹ de xilosa 15 g l ⁻¹ de arabinosa	1,84 ± 0,04	0,23 ± 0,01	0,64 ± 0,03	0,08 ± 0,01

Lista de referencias

10 Andreasen AA, Stier TJ (1954) Anaerobic nutrition of *Saccharomyces cerevisiae*. II. Unsaturated fatty acid requirement for growth in a defined medium. *J Cell Physiol* 43:271-281

Andreasen AA, Stier TJ (1953) Anaerobic nutrition of *Saccharomyces cerevisiae*. I. Ergosterol requirement for growth in a defined medium. *J Cell Physiol* 41:23-36

15 Becker J, Boles E (2003) A modified *Saccharomyces cerevisiae* strain that consumes L-Arabinose and produces ethanol. *Appl Environ Microbiol* 69:4144-4150

Gietz R.D., Sugino A. (1988). New yeast-*Escherichia coli* shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites. *Gene* 74:527-534.

Gietz, R. D., and R. A. Woods. 2002. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* 350:87-96.

20 Guldener U, Heck S, Fielder T, Beinbauer J, Hegemann JH. (1996) A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res.* 1996 Jul 1;24(13):2519-24.

Hauf J, Zimmermann FK, Muller S. Simultaneous genomic overexpression of seven glycolytic enzymes in the yeast *Saccharomyces cerevisiae*. *Enzyme Microb Technol.* 2000 Jun 1;26(9-10):688-698.

25 Inoue H., H. Nojima and H. Okayama, High efficiency transformation of *Escherichia coli* with plasmids. *Gene* 96 (1990), pp. 23-28

Kuyper M, Hartog MMP, Toirkens MJ, Almering MJH, Winkler AA, Van Dijken JP, Pronk JT (2005a) Metabolic engineering of a xylose-isomerase-expressing *Saccharomyces cerevisiae* strain for rapid anaerobic xylose fermentation. *Fems Yeast Research* 5:399-409

30 Kuyper M, Toirkens MJ, Diderich JA, Winkler AA, Van Dijken JP, Pronk JT (2005b) Evolutionary engineering of mixed-sugar utilization by a xylose-fermenting *Saccharomyces cerevisiae* strain. *Fems Yeast Research* 5:925-934

Sambrook, K., Fritsch, E.F. and Maniatis, I. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

35 Van Urk H, Mak PR, Scheffers WA, Van Dijken JP (1988) Metabolic responses of *Saccharomyces cerevisiae* CBS 8066 and *Candida utilis* CBS 621 upon transition from glucose limitation to glucose excess. *Yeast* 4:283-291

ES 2 601 052 T3

Verduyn C, Postma E, Scheffers WA, Van Dijken JP (1990) Physiology of *Saccharomyces cerevisiae* in anaerobic glucose-limited chemostat cultures. *J Gen Microbiol* 136:395-403

Verduyn C, Postma E, Scheffers WA, Van Dijken JP (1992) Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* 8:501-517

5 Weusthuis RA, Visser W, Pronk JT, Scheffers WA, Van Dijken JP (1994) Effects of oxygen limitation on sugar metabolism in yeasts - a continuous-culture study of the Kluver effect. *Microbiology* 140:703-715

LISTADO DE SECUENCIAS

<110> TU Delft

AJA van Maris

10 JT Pronk

HW Wisselink

JP van Dijk

AA Winkler

JH de Winde

15

<120> Ingeniería metabólica de células eucariotas que fermentan arabinosa

<130> P6011342PCT

20 <150> EP06121633.9

<151> 10-02-2006

<150> US 60/848.357

<151> 10-02-2006

25

<160> 53

<170> PatentIn versión 3.3

30 <210> 1

<211> 474

<212> PRT

<213> *Lactobacillus plantarum*

35 <400> 1

ES 2 601 052 T3

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

ES 2 601 052 T3

		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										

<210> 2

<211> 1425

5

<212> ADN

<213> Lactobacillus plantarum

<400> 2

```

atgttatcag tacctgatta tgagttttgg tttgttaccg gttcacaaca cctttatggt      60
gaagaacaat tgaagtctgt tgctaaggat gcgcaagata ttgctgataa attgaatgca      120
agcggcaagt taccttataa agtagtcttt aaggatgtta tgacgacggc tgaagatc      180
accaacttta tgaagaagt taattacaat gataaggtag ccggtgttat tacttgatg      240
cacacattct caccagctaa gaactggatt cgtggaactg aactgttaca aaaaccatta      300
ttacacttag caacgcaata tttgaataat attccatattg cagacattga ctttgattac      360
atgaacctta accaaagtgc ccatggcgac cgcgagtatg cctacattaa cgcccggttg      420
cagaacata ataagattgt ttacggctat tggggcgatg aagatgtgca agagcagatt      480
gcacgttggg aagacgtcgc cgtagcgtac aatgagagct ttaaagttaa ggttgctcgc      540
tttggcgaca caatgcgtaa tgtggccggt actgaagggtg acaagggtga agctcaaatt      600
aagatgggct ggacagttag ctattatggt atcggtgact tagttgaaga gatcaataag      660
gtttcggatg ctgatgttga taaggaatac gctgacttgg agtctcggta tgaaatggtc      720
caagtgtgata acgatgcgga cacgtataaa cattcagttc gggttcaatt ggcacaatat      780
ctgggtatta agcggttctt agaaagaggc ggttacacag cctttaccac gaactttgaa      840
gatctttggg ggatggagca attacctggg ctagcttcac aattattaat tcgtgatggg      900
tatggttttg gtgctgaagg tgactggaag acggctgctt taggacgggt tatgaagatt      960
atgtctcaca acaagcaaac cgctttatg gaagactaca cgttagactt gcgtcatggt     1020
catgaagcga tcttaggttc acacatggtg gaagttgatc cgtctatcgc aagtgataaa     1080
ccacgggtcg aagttcatcc attggatatt gggggtaaag atgatcctgc tcgcctagta     1140
tttactggtt cagaagggtga agcaattgat gtcaccgttg ccgatttccg tgatgggttc     1200
aagatgatta gctacgcggg agatgcgaat aagccagaag ccgaaacacc taatttacca     1260
gttgctaagc aattatggac cccaaagatg ggcttgaaga aggggtgact agaatggatg     1320
caagctgggt gtggtcacca cagcatgctg tccttctcgt taactgaaga acaaatggaa     1380
gactatgcaa ccatggttgg catgactaag gcattcttaa agtaa                       1425

```

10

ES 2 601 052 T3

<210> 3

<211> 533

<212> PRT

<213> Lactobacillus plantarum

5

<400> 3

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

ES 2 601 052 T3

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 Arg 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 Gln 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 2 601 052 T3

<211> 1602

<212> ADN

<213> Lactobacillus plantarum

5

<400> 4

```

atgaatttag ttgaaacagc ccaagcgatt aaaactggca aagtttcttt aggaattgag      60
cttgctcaa ctcgaattaa agccgttttg atcacggacg attttaatac gattgcttcg      120
ggaagttacg tttgggaaaa ccaatttggt gatggtactt ggacttacgc acttgaagat      180
gtctggaccg gaattcaaca aagttatacg caattagcag cagatgtccg cagtaaatat      240
cacatgagtt tgaagcatat caatgctatt ggcattagtg ccatgatgca cggataccta      300
gcatttgatc aacaagcгаа attattagtt ccgtttcggg cttggcgtaa taacattacg      360
gggcaagcag cagatgaatt gaccgaatta tttgatttca acattccaca acggtggagt      420
atcgcgcact tataccaggc aatcttaaat aatgaagcgc acgttaaaca ggtggacttc      480
ataacaacgc tggctggcta tgtaacctgg aaattgtcgg gtgagaaagt tctaggaatc      540
ggtgatgctg ctggcgtttt cccaattgat gaaacgactg acacatacaa tcagacgatg      600
ttaaccaagt ttagccaact tgacaaagt aaaccgtatt catgggatat cgggcatatt      660
ttaccgctgg ttttaccagc gggagccatt gctggaaagt taacggctgc cggggcgagc      720
ttacttgatc agagcggcac gctcgacgct ggcagtgtta ttgcaccgcc agaaggggat      780
gctggaacag gaatggtcgg tacgaacagc gtcggtaaac gcacgggtaa catctcggtg      840
ggaacctcag cattttcgat gaacgttcta gataaacat tgtctaaagt ctatcgcgat      900
attgatattg ttatgacgcc agatgggtca ccagttgcaa tgggtgcatgt taataattgt      960
tcatcagata ttaatgctg ggcaacgatt tttcgtgagt ttgcagcccg gttgggaatg     1020
gaattgaaac cggatcgatt atatgaaacg ttattcttgg aatcaactcg cgctgatgcg     1080
gatgctggag ggttggctaa ttatagttat caatccggtg agaatattac taagattcaa     1140
gctggtcggc cgctatttgt acggacacca aacagtaaat ttagtttacc gaactttatg     1200
ttgaccaat tatatgcggc gttcgcaccc ctccaacttg gtatggatat tctgtttaac     1260
gaagaacatg ttcaaacgga cgttatgatt gcacaggggtg gattgttccg aacgccggtg     1320
attggccaac aagtattggc caacgcaactg aacattccga ttactgtaat gagtactgct     1380
ggtgaaggcg gcccatgggg gatggcagtg ttagccaact ttgcttgtcg gcaaactgca     1440
atgaacctag aagatttctt agatcaagaa gtctttaaag agccagaaag tatgacgttg     1500
agtccagaac cggaacgggt ggccggatat cgtgaattta ttcaacgtta tcaagctggc     1560
ttaccagttg aagcagcggc tgggcaagca atcaaatatt ag                               1602

```

<210> 5

10

<211> 242

<212> PRT

<213> Lactobacillus plantarum

ES 2 601 052 T3

<400> 5

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

5

<210> 6

<211> 729

<212> ADN

<213> Lactobacillus plantarum

10

ES 2 601 052 T3

<400> 6

```

atgctagaag cattaaaaca agaagtttat gaggctaaca tgcagcttcc aaagctgggc      60
ctggttactt ttacctgggg caatgtctcg ggcattgacc gggaaaaagg cctattcgtg      120
atcaagccat ctgggtgtga ttatggtgaa ttaaaaccaa gcgatttagt cgttgttaac      180
ttacaggggtg aagtgggtga aggtaaacta aatccgtcta gtgatacgcc gactcatacg      240
gtgttatata acgcttttcc taatattggc ggaattgtcc atactcattc gccatgggca      300
gttgccatag cagctgctca aatggatgtg ccagctatga acacgacca tgctgatagc      360
ttctatgggtg acgtgcccggc cgcggatgcg ctgactaagg aagaaattga agcagattat      420
gaaggcaaca cgggtaaaac cattgtgaag acgttccaag aacggggcct cgattatgaa      480
gctgtaccag cctcattagt cagccagcac ggcccatttg cttggggacc aacgccagct      540
aaagccgttt acaatgctaa agtgttggaa gtggttgccg aagaagatta tcatactgcg      600
caattgacct gtgcaagtag cgaattacca caatatttat tagataagca ttatttacgt      660
aagcatgggtg caagtgccta ttatggtcaa aataatgcbc attctaagga tcatgcagtt      720
cgcaagtaa                                          729
    
```

5 <210> 7

<211> 437

<212> PRT

<213> Especie de Piromyces

10 <400> 7

```

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
    
```

ES 2 601 052 T3

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 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

ES 2 601 052 T3

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

<210> 8

<211> 1669

5

<212> ADN

<213> Especie de Piromyces

<400> 8

gtaaatggct aaggaatatt tcccacaaat tcaaaagatt aagttcgaag gtaaggattc 60
 taagaatcca ttagccttcc actactacga tgctgaaaag gaagtcattg gtaagaaaa 120
 gaaggattgg ttacgtttcg ccatggcctg gtggcacact ctttgcgccg aagggtgctga 180
 ccaattcggg ggaggtacaa agtctttccc atggaacgaa ggtactgatg ctattgaaat 240
 tgccaagcaa aaggttgatg ctggtttcga aatcatgcaa aagcttggtg ttccatacta 300
 ctgtttccac gatgttgatc ttgtttccga aggtaactct attgaagaat acgaatccaa 360
 ccttaaggct gtcgttgctt acctcaagga aaagcaaaag gaaaccggta ttaagcttct 420
 ctggagtact gctaactgct tcggtcacia gcggttacatg aacgggtgct cactaacc 480
 agactttgat gttgtcgccc gtgctattgt tcaaattaag aacgccatag acgccggtat 540
 tgaacttggg gctgaaaact acgtcttctg ggggtggtcgt gaaggttaca tgagtctcct 600
 taacactgac caaaagcgtg aaaaggaaca catggccact atgcttacca tggctcgtga 660
 ctacgctcgt tccaagggat tcaagggtac tttcctcatt gaaccaaagc caatggaacc 720
 aaccaagcac caatacgaat ttgacactga aaccgctatt ggtttcctta aggcccacia 780
 cttagacaag gacttcaagg tcaacattga agttaaccac gctactcttg ctggtcacac 840
 tttcgaacac gaacttgctt gtgctgttga tgctgggatg ctcggttcca ttgatgctaa 900
 ccgtgggtgac taccaaaacg gttgggatac tgatcaattc ccaattgatc aatacgaact 960
 cgccaagct tggatggaaa tcatccgtgg tgggtggttc gttactggtg gtaccaactt 1020
 cgatgccaag actcgtcgta actctactga cctcgaagac atcatcattg cccacgtttc 1080
 tggatggat gctatggctc gtgctcttga aaacgctgcc aagctcctcc aagaatctcc 1140
 atacaccaag atgaagaagg aacgttacgc ttccttcgac agtggtattg gtaaggactt 1200
 tgaagatggg aagctcacc tcgaacaagt ttacgaatac ggtaagaaga acggtgaacc 1260

10

ES 2 601 052 T3

aaagcaaact tctggtaagc aagaactcta cgaagctatt gttgccatgt accaataagt 1320
 taatcgtagt taaattggta aaataattgt aaaatcaata aacttgtcaa tcctccaatc 1380
 aagtttaaaa gatcctatct ctgtactaat taaatatagt acaaaaaaaaa atgtataaac 1440
 aaaaaaaagt ctaaaagacg gaagaattta atttagggaa aaaataaaaa taataataaa 1500
 caatagataa atcctttata ttaggaaaat gtcccattgt attattttca tttctactaa 1560
 aaaagaaagt aaataaaaca caagaggaaa ttticccttt tttttttttt tgtaataaat 1620
 tttatgcaaa tataaatata aataaaataa taaaaaaaa aaaaaaaaa 1669

<210> 9

<211> 496

5 <212> PRT

<213> Bacillus subtilis

<400> 9

Met Leu Gln Thr Lys Asp Tyr Glu Phe Trp Phe Val Thr Gly Ser Gln
 1 5 10 15
 His Leu Tyr Gly Glu Glu Thr Leu Glu Leu Val Asp Gln His Ala Lys
 20 25 30
 Ser Ile Cys Glu Gly Leu Ser Gly Ile Ser Ser Arg Tyr Lys Ile Thr
 35 40 45
 His Lys Pro Val Val Thr Ser Pro Glu Thr Ile Arg Glu Leu Leu Arg
 50 55 60
 Glu Ala Glu Tyr Ser Glu Thr Cys Ala Gly Ile Ile Thr Trp Met His
 65 70 75 80
 Thr Phe Ser Pro Ala Lys Met Trp Ile Glu Gly Leu Ser Ser Tyr Gln
 85 90 95
 Lys Pro Leu Met His Leu His Thr Gln Tyr Asn Arg Asp Ile Pro Trp
 100 105 110
 Gly Thr Ile Asp Met Asp Phe Met Asn Ser Asn Gln Ser Ala His Gly
 115 120 125
 Asp Arg Glu Tyr Gly Tyr Ile Asn Ser Arg Met Gly Leu Ser Arg Lys
 130 135 140
 Val Ile Ala Gly Tyr Trp Asp Asp Glu Glu Val Lys Lys Glu Met Ser
 145 150 155 160
 Gln Trp Met Asp Thr Ala Ala Ala Leu Asn Glu Ser Arg His Ile Lys
 165 170 175
 Val Ala Arg Phe Gly Asp Asn Met Arg His Val Ala Val Thr Asp Gly

10

ES 2 601 052 T3

<210> 10
 <211> 1511
 <212> ADN
 <213> Bacillus subtilis

5

<400> 10

```

atgagaaagg ggcagtttac atgcttcaga caaaggatta tgaattctgg tttgtgacag    60
gaagccagca cctatacggg gaagagacgc tggaactcgt agatcagcat gctaaaagca    120
tttgtgaggg gctcagcggg atttcttcca gatataaaat cactcataag cccgtcgtca    180
cttcaccgga aaccattaga gagctgttaa gagaagcgga gtacagtgag acatgtgctg    240
gcatcattac atggatgcac acattttccc ctgcaaaaat gtggatagaa ggcctttcct    300
cttatcaaaa accgcttatg cattttgcata cccaatataa tcgcgatata ccgtggggta    360
cgattgacat ggattttatg aacagcaacc aatccgcgca tggcgatcga gagtacggtt    420
acatcaactc gagaatgggg cttagccgaa aagtcattgc cggctattgg gatgatgaag    480
aagtgaaaaa agaaatgtcc cagtggatgg atacggcggc tgcattaaat gaaagcagac    540
atattaaggt tgccagatth ggagataaca tgcgtcatgt cgcggtaacg gacggagaca    600
aggtgggagc gcatattcaa tttggctggc aggttgacgg atatggcatc ggggatctcg    660
ttgaagtgat ggatcgcatt acggacgacg aggttgacac gctttatgcc gagtatgaca    720
gactatatgt gatcagtgag gaaacaaaac gtgacgaagc aaaggtagcg tccattaaag    780
aacaggcgaa aattgaactt ggattaaccg cttttcttga gcaaggcgga tacacagcgt    840
ttacgacatc gtttgaagtg ctgcacggaa tgaaacagct gccgggactt gccgttcagc    900
gcctgatgga gaaaggctat gggtttgccg gtgaaggaga ttggaagaca gcggcccttg    960
tacggatgat gaaaatcatg gctaaaggaa aaagaacttc cttcatggaa gattacacgt   1020
accatthtga accgggaaat gaaatgattc tgggctctca catgcttgaa gtgtgtccga   1080
ctgtcgcttt ggatcagccg aaaatcgagg ttcattcgct ttcgattggc ggcaaagagg   1140
accctgcgcy tttggatatt aacggcatca gcggttctgc cattcaagct agcattgttg   1200
atattggcgg gcgthtccgc cttgtgctga atgaagtcaa cgccaggaa attgaaaag    1260
acatgccgaa tttaccggtt gcccggttc tctggaagcc ggagccgtca ttgaaaacag   1320
cagcggaggc atggatttta gccggcggty cacaccatac ctgcctgtct tatgaaactga   1380
cagcggagca aatgcttgat tggcgggaaa tggcgggaat cgaaagtgtt ctcatttccc   1440
gtgatacgac aattcataaa ctgaaacacg agttaaaatg gaacgaggcg ctttaccggc   1500
ttcaaaagta g
    
```

10 <210> 11
 <211> 566
 <212> PRT
 <213> E. coli

ES 2 601 052 T3

<400> 11

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

ES 2 601 052 T3

Asp Thr Trp Thr Ala Asp Ile Pro Val Gly Thr Leu Cys Pro Glu Trp
 245 250 255
 Ala Gln Arg Leu Gly Leu Pro Glu Ser Val Val Ile Ser Gly Gly Ala
 260 265 270
 Phe Asp Cys His Met Gly Ala Val Gly Ala Gly Ala Gln Pro Asn Ala
 275 280 285
 Leu Val Lys Val Ile Gly Thr Ser Thr Cys Asp Ile Leu Ile Ala Asp
 290 295 300
 Lys Gln Ser Val Gly Glu Arg Ala Val Lys Gly Ile Cys Gly Gln Val
 305 310 315 320
 Asp Gly Ser Val Val Pro Gly Phe Ile Gly Leu Glu Ala Gly Gln Ser
 325 330 335
 Ala Phe Gly Asp Ile Tyr Ala Trp Phe Gly Arg Val Leu Ser Trp Pro
 340 345 350
 Leu Glu Gln Leu Ala Ala Gln His Pro Glu Leu Lys Ala Gln Ile Asn
 355 360 365
 Ala Ser Gln Lys Gln Leu Leu Pro Ala Leu Thr Glu Ala Trp Ala Lys
 370 375 380
 Asn Pro Ser Leu Asp His Leu Pro Val Val Leu Asp Trp Phe Asn Gly
 385 390 395 400
 Arg Arg Ser Pro Asn Ala Asn Gln Arg Leu Lys Gly Val Ile Thr Asp
 405 410 415
 Leu Asn Leu Ala Thr Asp Ala Pro Leu Leu Phe Gly Gly Leu Ile Ala
 420 425 430
 Ala Thr Ala Phe Gly Ala Arg Ala Ile Met Glu Cys Phe Thr Asp Gln
 435 440 445
 Gly Ile Ala Val Asn Asn Val Met Ala Leu Gly Gly Ile Ala Arg Lys
 450 455 460
 Asn Gln Val Ile Met Gln Ala Cys Cys Asp Val Leu Asn Arg Pro Leu
 465 470 475 480
 Gln Ile Val Ala Ser Asp Gln Cys Cys Ala Leu Gly Ala Ala Ile Phe
 485 490 495
 Ala Ala Val Ala Ala Lys Val His Ala Asp Ile Pro Ser Ala Gln Gln
 500 505 510

ES 2 601 052 T3

Lys Met Ala Ser Ala Val Glu Lys Thr Leu Gln Pro Arg Ser Glu Gln
 515 520 525
 Ala Gln Arg Phe Glu Gln Leu Tyr Arg Arg Tyr Gln Gln Trp Ala Met
 530 535 540
 Ser Ala Glu Gln His Tyr Leu Pro Thr Ser Ala Pro Ala Gln Ala Ala
 545 550 555 560
 Gln Ala Val Ala Thr Leu
 565

<210> 12

<211> 1453

5 <212> ADN

<213> E. coli

<400> 12

atggcgattg caattggcct cgattttggc agtgattctg tgcgagcttt ggcggtggac 60
 tgcgccagcg gtgaagagat cgccaccagc gtagagtggg atccccgttg gcaaaaaggg 120
 caattttgtg atgccccgaa taaccagttc cgatcatcatc cgcgtgacta cattgagtca 180
 atggaagcgg cactgaaaac cgtgcttgca gagcttagcg tcgaacagcg cgcagctgtg 240
 gtcgggattg gcgttgacag taccggctcg acgcccgcac cgattgatgc cgacggtaac 300
 gtgctggcgc tgcgcccgga gtttgccgaa aaccgaaacg cgatgttcgt attgtggaaa 360
 gaccacactg cgggtgaaag aagcgaagag attaccggtt tgtgccacgc gccgggcaat 420
 gttgactact cccgctatat tggcgggtatt tattccagcg aatggttctg ggcaaaaatc 480
 ctgcatgtga ctgcccagga cagcgcctg gcgcaatctg ccgcatcgtg gattgagctg 540
 tgcgactggg tgccagctct gctttccggt accaccgcc cgaggatat tcgtcgcgga 600
 cgttgcagcg cggggcataa atctctgtgg cacgaaagct gggcgggctt gccgccagcc 660
 agtttctttg atgagctgga cccgatcctc aatcgccatt tgccttcccc gctgttcaact 720
 gacacctgga ctgccgatat tccgggtggc accttatgcc cggaatgggc gcagcgtctc 780
 ggcctgcctg aaagcgtggg gatttccggc ggcgcgtttg actgccatat gggcgcagtt 840
 ggcgcaggcg cacagcctaa cgcactggta aaagttatcg gtacttccac ctgcgacatt 900
 ctgattgccg acaaacagag cgttggcgag cgggcagtta aaggattttg cggtcaggtt 960
 gatggcagcg tgggtgcctgg atttatcggg ctggaagcag gccaatcggc gtttggtgat 1020
 atctacgcct ggttcggctc cgtactcagc tggccgctgg aacagcttgc cgcccagcat 1080
 ccggaactga aagcgcgaaat caacgccagc cagaaacaac tgcttccggc gctgaccgaa 1140
 gcatgggcca aaaatccgtc tctggatcac ctgccggtgg tgctcgactg gtttaacggt 1200
 cgctcgtcgc caaacgctaa ccaacgcctg aaaggggtga ttaccgatct taacctcgct 1260
 accgacgctc cgctgctgtt cggcggtttg attgctgcca ccgcctttgg cgcacgcgca 1320
 atcatggaggt gctttaccga tcaggggatc gccgtcaata acgtgatggc gctgggaggc 1380

10

ES 2 601 052 T3

atcgcgcgga aaaaccaagt cattatgcag gcctgctgcg acgtgctgaa tcgcccgctg 1440
 caaattgttg cct 1453

<210> 13

<211> 231

5 <212> PRT

<213> E. coli

<400> 13

Met Leu Glu Asp Leu Lys Arg Gln Val Leu Glu Ala Asn Leu Ala Leu
 1 5 10 15
 Pro Lys His Asn Leu Val Thr Leu Thr Trp Gly Asn Val Ser Ala Val
 20 25 30
 Asp Arg Glu Arg Gly Val Phe Val Ile Lys Pro Ser Gly Val Asp Tyr
 35 40 45
 Ser Ile Met Thr Ala Asp Asp Met Val Val Val Ser Ile Glu Thr Gly
 50 55 60
 Glu Val Val Glu Gly Ala Lys Lys Pro Ser Ser Asp Thr Pro Thr His
 65 70 75 80
 Arg Leu Leu Tyr Gln Ala Phe Pro Ser Ile Gly Gly Ile Val His Thr
 85 90 95
 His Ser Arg His Ala Thr Ile Trp Ala Gln Ala Gly Gln Ser Ile Pro
 100 105 110
 Ala Thr Gly Thr Thr His Ala Asp Tyr Phe Tyr Gly Thr Ile Pro Cys
 115 120 125
 Thr Arg Lys Met Thr Asp Ala Glu Ile Asn Gly Glu Tyr Glu Trp Glu
 130 135 140
 Thr Gly Asn Val Ile Val Glu Thr Phe Glu Lys Gln Gly Ile Asp Ala
 145 150 155 160
 Ala Gln Met Pro Gly Val Leu Val His Ser His Gly Pro Phe Ala Trp
 165 170 175
 Gly Lys Asn Ala Glu Asp Ala Val His Asn Ala Ile Val Leu Glu Glu
 180 185 190
 Val Ala Tyr Met Gly Ile Phe Cys Arg Gln Leu Ala Pro Gln Leu Pro
 195 200 205
 Asp Met Gln Gln Thr Leu Leu Asn Lys His Tyr Leu Arg Lys His Gly
 210 215 220
 Ala Lys Ala Tyr Tyr Gly Gln
 225 230

10

ES 2 601 052 T3

<210> 14

<211> 696

<212> ADN

<213> E. coli

5

<400> 14

```

atgttagaag atctcaaacg ccaggattta gaggccaacc tggcgctgcc aaaacataac    60
ctggtcacgc tcacatgggg caacgtcagc gccgttgatc gcgagcgcgg cgtctttgtg   120
atcaaacctt ccggcgtcga ttacagcatc atgaccgctg acgatatggt cgtggttagc   180
atcgaaaccg gtgaagtggg tgaagggtcg aaaaagccct cctccgatac gccaaactcac   240
cgactgctct atcaggcatt cccgtccatt ggcggcattg tgcacacaca ctcgcgccac   300
gccactatct gggcgcaggc gggccagtcg attccagcaa ccggcaccac ccacgccgac   360
tatttctacg gcaccattcc ctgcacccgc aaaatgaccg acgcagaaat caacggtgaa   420
tatgagtggg aaaccggtaa cgtcatcgta gaaaccttcg aaaaacaggg tatcgatgca   480
gcgcaaatgc ccggcgtcct ggtccattct cacggcccat ttgcatgggg caaaaatgcc   540
gaagatgcgg tgcataacgc catcgtgctg gaagaggtcg cttatatggg gatattctgc   600
cgtcagttag cgccgcagtt accggatatg cagcaaacgc tgctgaataa acactatctg   660
cgtaagcatg gcgcgaaggc atattacggg cagtaa                                696
    
```

10

<210> 15

<211> 438

<212> PRT

<213> Bacteroides thetaiotaomicron

15

<400> 15

ES 2 601 052 T3

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

ES 2 601 052 T3

100					105					110					
Ser	Val	Glu	Glu	Tyr	Glu	Ala	Asn	Leu	Lys	Glu	Ile	Val	Ala	Tyr	Ala
		115					120					125			
Lys	Gln	Lys	Gln	Ala	Glu	Thr	Gly	Ile	Lys	Leu	Leu	Trp	Gly	Thr	Ala
		130					135					140			
Asn	Val	Phe	Gly	His	Ala	Arg	Tyr	Met	Asn	Gly	Ala	Ala	Thr	Asn	Pro
145					150					155					160
Asp	Phe	Asp	Val	Val	Ala	Arg	Ala	Ala	Val	Gln	Ile	Lys	Asn	Ala	Ile
				165					170					175	
Asp	Ala	Thr	Ile	Glu	Leu	Gly	Gly	Glu	Asn	Tyr	Val	Phe	Trp	Gly	Gly
			180					185					190		
Arg	Glu	Gly	Tyr	Met	Ser	Leu	Leu	Asn	Thr	Asp	Gln	Lys	Arg	Glu	Lys
		195					200					205			
Glu	His	Leu	Ala	Gln	Met	Leu	Thr	Ile	Ala	Arg	Asp	Tyr	Ala	Arg	Ala
	210					215					220				
Arg	Gly	Phe	Lys	Gly	Thr	Phe	Leu	Ile	Glu	Pro	Lys	Pro	Met	Glu	Pro
225						230					235				240
Thr	Lys	His	Gln	Tyr	Asp	Val	Asp	Thr	Glu	Thr	Val	Ile	Gly	Phe	Leu
				245					250					255	
Lys	Ala	His	Gly	Leu	Asp	Lys	Asp	Phe	Lys	Val	Asn	Ile	Glu	Val	Asn
			260					265					270		
His	Ala	Thr	Leu	Ala	Gly	His	Thr	Phe	Glu	His	Glu	Leu	Ala	Val	Ala
		275					280					285			
Val	Asp	Asn	Gly	Met	Leu	Gly	Ser	Ile	Asp	Ala	Asn	Arg	Gly	Asp	Tyr
	290					295					300				
Gln	Asn	Gly	Trp	Asp	Thr	Asp	Gln	Phe	Pro	Ile	Asp	Asn	Tyr	Glu	Leu
305						310					315				320
Thr	Gln	Ala	Met	Met	Gln	Ile	Ile	Arg	Asn	Gly	Gly	Leu	Gly	Thr	Gly
				325					330					335	
Gly	Thr	Asn	Phe	Asp	Ala	Lys	Thr	Arg	Arg	Asn	Ser	Thr	Asp	Leu	Glu
			340					345					350		
Asp	Ile	Phe	Ile	Ala	His	Ile	Ala	Gly	Met	Asp	Ala	Met	Ala	Arg	Ala
		355					360					365			
Leu	Glu	Ser	Ala	Ala	Ala	Leu	Leu	Asp	Glu	Ser	Pro	Tyr	Lys	Lys	Met

ES 2 601 052 T3

370	375	380
Leu Ala Asp Arg Tyr Ala Ser Phe Asp Gly Gly Lys Gly Lys Glu Phe 385 390 400		
Glu Asp Gly Lys Leu Thr Leu Glu Asp Val Val Ala Tyr Ala Lys Thr 405 410 415		
Lys Gly Glu Pro Lys Gln Thr Ser Gly Lys Gln Glu Leu Tyr Glu Ala 420 425 430		
Ile Leu Asn Met Tyr Cys 435		

<210> 16

<211> 1317

5 <212> ADN

<213> Bacteroides thetaiotaomicron

<400> 16

```

atggcaaca aagaatttt tccgggaatt gaaaagatta aatttgaagg taaagatagt 60
aagaaccga tggcattccg ttattacgat gcagagaagg tgattaatgg taaaaagatg 120
aaggattgac tgagattcgc tatggcatgg tggcacacat tgtgcgctga aggtggtgat 180
cagttcggtg gcggaacaaa gcaattccca tgggaatggta atgcagatgc taticagga 240
gcaaaagata agatggatgc aggatttgaa ttcatgcaga agatgggtat cgaatactat 300
tgcttccatg acgtagactt ggtttcggaa ggtgccagtg tagaagaata cgaagctaac 360
ctgaaagaaa tcgtagctta tgcaaacag aaacaggcag aaaccggtat caaactactg 420
tggggtactg ctaatgtatt cggtcacgcc cgctatatga acggtgcagc taccaatcct 480
gacttcgatg tagtagctcg tctgctgtt cagatcaaaa atgcgattga tgcaacgatt 540
gaacttggcg gagagaatta tigtttttgg ggtggctcgtg aaggctatat gtctcttctg 600
aacacagatc agaaacgtga aaaagaacac cttgcacaga tgttgacgat tgctcgtgac 660
tatgcccgtg cccgtggttt caaaggtact ttctgatcg aaccgaaacc gatggaaccg 720
actaaacatc aatatgacgt agatacggaa actgtaactg gcttctgaa agctcatggt 780
ctggataagg atttcaaagt aaatatcgag gtgaatcacg caactttggc aggtcacact 840
ttcgagcatg aattggctgt agctgtagac aatggtatgt tgggctcaat tgacccaat 900
cgtggtgact atcagaatgg ctgggataca gaccaattcc cgatcgacaa ttatgaactg 960
actcaggcta tgatgcagat tatccgtaat ggtggtctcg gtaccggtgg tacgaacttt 1020
gatgctaaaa cccgtcgtaa ttctactgat ctggaagata tctttattgc tcacatcgca 1080
gggatggacg ctatggcccg tgcaactcga agtgcagcgg ctctgctcga cgaatctccc 1140
tataagaaga tgctggctga ccgttatgct tcatttgatg ggggcaaagg taaagaattt 1200
gaagacggca agctgactct ggaggatgtg gttgcttatg caaaaacaaa aggcgaaccg 1260
aaacagacta gcggcaagca agaactttat gaggcaattc tgaatatgta ttgctaa 1317

```

10

ES 2 601 052 T3

<210> 17

<211> 258

<212> PRT

<213> Saccharomyces cerevisiae

5

<400> 17

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

ES 2 601 052 T3

Ala Ser Lys Ala Tyr Phe Gly Asn Ser Asp Gly Ser Val Glu Val Thr
 245 250 255

Glu Lys

<210> 18

<211> 2467

5

<212> ADN

<213> *Saccharomyces cerevisiae*

<400> 18

```

ggatccaaga ccattattcc atcagaatgg aaaaaagttt aaaagatcac ggagatthttg      60
ttcttctgag cttctgctgt ccttgaaaac aaattattcc gctggccgcc ccaaacaaaa      120
acaacccccg ttttaataaca ttgtcacagt attagaaatt ttctttttac aaattaccat      180
ttccagctta ctacttccta taatcctcaa tcttcagcaa gcgacgcagg gaatagccgc      240
tgagggtgcat aactgtcact tttcaattcg gccaatgcaa tctcaggcgg acgaataagg      300
gggccctctc gagaaaaaca aaaggaggat gagattagta ctttaatggt gtgttcagta      360
attcagagac agacaagaga ggttccaac acaatgtctt tagactcata ctatcttggg      420
tttgatcttt cgaccaaca actgaaatgt ctcgccatta accaggacct aaaaattgtc      480
cattcagaaa cagtggaatt tgaaaaggat cttccgcatt atcacacaaa gaaggggtgtc      540
tatatacacg gcgacactat cgaatgtccc gtagccatgt ggtaggggc tctagatctg      600
gttctctcga aatatcgcga ggctaaatth ccattgaaca aagttatggc cgtctcaggg      660
tcctgccagc agcacgggtc tgtctactgg tcctcccaag ccgaatctct gtagagcaa      720
ttgaataaga aaccggaaaa agatttattg cactacgtga gctctgtagc atttgcaagg      780
caaaccgccc ccaattggca agaccacagt actgcaaagc aatgtcaaga gtttgaagag      840
tgcataggtg ggccctgaaa aatggctcaa ttaacagggt ccagagccca ttttagatth      900
actggtcctc aaattctgaa aattgcacaa ttagaaccag aagcttacga aaaaacaaag      960
accatttctt tagtgtctaa ttttttgact tctatcttag tgggccatct tgttgaatta     1020
gaggaggcag atgcctgtgg tatgaacctt tatgatatac gtgaaagaaa attcatgtat     1080
gagctactac atctaattga tagttcttct aaggataaaa ctatcagaca aaaattaatg     1140
agagcaccca tgaaaaatth gatagcgggt accatctgta aatatttht tgagaagtac     1200
ggtttcaata caaactgcaa ggtctctccc atgactgggg ataatttagc cactatatgt     1260
tctttacccc tgcggaagaa tgacgttctc gtttccttag gaacaagtac tacagttctt     1320
ctggtcaccg ataagtatca cccctctccg aactatcatc ttttcattca tccaactctg     1380
ccaaaccatt atatgggtat gatttgttat tgtaatgggt ctttgcaag ggagaggata     1440
agagacgagt taaacaaaga acgggaaaat aattatgaga agactaacga ttggactctt     1500
tttaatacaag ctgtgctaga tgactcagaa agtagtgaaa atgaattagg tgtatattth     1560
cctctggggg agatcgttcc tagcgtaaaa gccataaaca aaagggttat cttcaatcca     1620
    
```

10

ES 2 601 052 T3

aaaacgggta tgattgaaag agaggtggcc aagttcaaag acaagaggca cgatgccaaa 1680
 aatattgtag aatcacaggc ttaagttgc agggtaagaa tatctcccct gctttcggat 1740
 tcaaacgcaa gtcacaaca gagactgaac gaagatacaa tcgtgaagtt tgattacgat 1800
 gaatctccgc tgcgggacta cctaaataaa aggccagaaa ggactttttt tgtaggtggg 1860
 gcttctaaaa acgatgctat tgtgaagaag tttgctcaag tcattgggtgc taaaagggt 1920
 aattttaggc tagaaacacc aaactcatgt gcccttgggtg gttgttataa ggccatgtgg 1980
 tcattgttat atgactctaa taaaattgca gttccttttg ataaatttct gaatgacaat 2040
 tttccatggc atgtaatgga aagcatatcc gatgtggata atgaaaattg gatcgctata 2100
 attccaagat tgtcccctta agcgaactgg aaaagactct catctaaaat atgtttgaat 2160
 aatttatcat gccctgacaa gtacacacaa acacagacac ataatataca tacatatata 2220
 tatatcaccg ttattatgcg tgcacatgac aatgcccttg tatgtttcgt atactgtagc 2280
 aagtagtcat cttttgttc cccgttcgga aatgacaaa aagtaaaatc aataaatgaa 2340
 gagtaaaaaa caatttatga aagggtgagc gaccagcaac gagagagaca aatcaaatta 2400
 gcgctttcca gtgagaatat aagagagcat tgaagagct aggttattgt taaatcatct 2460
 cgagctc 2467

<210> 19

<211> 238

5 <212> PRT

<213> Saccharomyces cerevisiae

<400> 19

Met Val Lys Pro Ile Ile Ala Pro Ser Ile Leu Ala Ser Asp Phe Ala
 1 5 10 15

Asn Leu Gly Cys Glu Cys His Lys Val Ile Asn Ala Gly Ala Asp Trp
 20 25 30

Leu His Ile Asp Val Met Asp Gly His Phe Val Pro Asn Ile Thr Leu
 35 40 45

Gly Gln Pro Ile Val Thr Ser Leu Arg Arg Ser Val Pro Arg Pro Gly
 50 55 60

Asp Ala Ser Asn Thr Glu Lys Lys Pro Thr Ala Phe Phe Asp Cys His
 65 70 75 80

Met Met Val Glu Asn Pro Glu Lys Trp Val Asp Asp Phe Ala Lys Cys
 85 90 95

Gly Ala Asp Gln Phe Thr Phe His Tyr Glu Ala Thr Gln Asp Pro Leu
 100 105 110

His Leu Val Lys Leu Ile Lys Ser Lys Gly Ile Lys Ala Ala Cys Ala

10

ES 2 601 052 T3

115	120	125
Ile Lys Pro Gly Thr Ser Val Asp Val Leu Phe Glu Leu Ala Pro His 130 135 140		
Leu Asp Met Ala Leu Val Met Thr Val Glu Pro Gly Phe Gly Gly Gln 145 150 155 160		
Lys Phe Met Glu Asp Met Met Pro Lys Val Glu Thr Leu Arg Ala Lys 165 170 175		
Phe Pro His Leu Asn Ile Gln Val Asp Gly Gly Leu Gly Lys Glu Thr 180 185 190		
Ile Pro Lys Ala Ala Lys Ala Gly Ala Asn Val Ile Val Ala Gly Thr 195 200 205		
Ser Val Phe Thr Ala Ala Asp Pro His Asp Val Ile Ser Phe Met Lys 210 215 220		
Glu Glu Val Ser Lys Glu Leu Arg Ser Arg Asp Leu Leu Asp 225 230 235		

<210> 20

<211> 1328

5

<212> ADN

<213> Saccharomyces cerevisiae

<400> 20

gtaggact tacgtatcct gtatagtagg aatggctcgg tttatgtata ttaggagatc	60
aaaacgagaa aaaaatacca tatcgtatag tatagagagt ataaatataa gaaatgccgc	120
atatgtacaa ctaatctagc aaatctctag aacgcaattc cttcgagact tcttctttca	180
tgaaggagat aacatcgtgc gggtcagctg cagtgaaaac actggtacca gcgacaataa	240
cgttggcacc ggctttggcg gctttcggga tggctctcct gcccaaacca ccatcgactt	300
ggatattcaa atgggggaac ttggctctca aagtttccac ttttggcatc atgtctttca	360
tgaatttttg gcctccaaac ccaggttcca cagtcataac aagagccata tccaaatgag	420
gagctagttc aaataaaacg tcaacagaag taccaggttt gatggcgcac gcagctttga	480
tgcccttaga ctaaatcaac ttaactaaat gcaaagggtc ttgtgtggcc tcgtagtgga	540
acgtaaattg gtcagcacca catttagcaa aatcgtcgac ccatttttca ggattttcaa	600
ccatcatgtg acaatcgaag aacgcagtgg gcttcttttc tgtgttgcta gcatcgccag	660
ggcgtggcac agaacgacgt agggaggtaa caattggttg gccagagta atgtttggaa	720
caaatggcc gtccatgaca tcgatatgta accaatctgc gccggcgttg atgaccttat	780
gacattcgca acccaagttg gcgaagtcag aagcaaggat actgggagct ataattggtt	840
tgaccatttt ttcttgtgtg ttacctcgc tcttggatt agcaaatggc cttcttgcac	900
gaaattgtat cgagtttgct ttatttttct tttaacgggc ggattctttc tattctggct	960

10

ES 2 601 052 T3

ttcctataac agagatcatg aaagaagttc cagcttacgg atcaagaaag tacctataca 1020
 tatacaaaaa tctgattact ttccagctc gacttggata gctgttcttg ttttctcttg 1080
 gcgacacatt ttttgtttct gaagccacgt cctgctttat aagaggacat ttaaagttgc 1140
 aggacttgaa tgcaattacc ggaagaagca accaaccggc atggttcagc atacaatata 1200
 catttgatta gaaaagcaga gaataaatag acatgatacc tctcttttta tcctctgcag 1260
 cgtattattg tttattccac gcaggcatcg gtcgttggct gttgttatgt ctcagataag 1320
 cgcgtttg 1328

<210> 21

<211> 680

5 <212> PRT

<213> *Saccharomyces cerevisiae*

<400> 21

Met Thr Gln Phe Thr Asp Ile Asp Lys Leu Ala Val Ser Thr Ile Arg
 1 5 10 15
 Ile Leu Ala Val Asp Thr Val Ser Lys Ala Asn Ser Gly His Pro Gly
 20 25 30
 Ala Pro Leu Gly Met Ala Pro Ala Ala His Val Leu Trp Ser Gln Met
 35 40 45
 Arg Met Asn Pro Thr Asn Pro Asp Trp Ile Asn Arg Asp Arg Phe Val
 50 55 60
 Leu Ser Asn Gly His Ala Val Ala Leu Leu Tyr Ser Met Leu His Leu
 65 70 75 80
 Thr Gly Tyr Asp Leu Ser Ile Glu Asp Leu Lys Gln Phe Arg Gln Leu
 85 90 95
 Gly Ser Arg Thr Pro Gly His Pro Glu Phe Glu Leu Pro Gly Val Glu
 100 105 110
 Val Thr Thr Gly Pro Leu Gly Gln Gly Ile Ser Asn Ala Val Gly Met
 115 120 125
 Ala Met Ala Gln Ala Asn Leu Ala Ala Thr Tyr Asn Lys Pro Gly Phe
 130 135 140
 Thr Leu Ser Asp Asn Tyr Thr Tyr Val Phe Leu Gly Asp Gly Cys Leu
 145 150 155 160
 Gln Glu Gly Ile Ser Ser Glu Ala Ser Ser Leu Ala Gly His Leu Lys
 165 170 175
 Leu Gly Asn Leu Ile Ala Ile Tyr Asp Asp Asn Lys Ile Thr Ile Asp

10

ES 2 601 052 T3

180 185 190

Gly Ala Thr Ser Ile Ser Phe Asp Glu Asp Val Ala Lys Arg Tyr Glu
195 200 205

Ala Tyr Gly Trp Glu Val Leu Tyr Val Glu Asn Gly Asn Glu Asp Leu
210 215 220

Ala Gly Ile Ala Lys Ala Ile Ala Gln Ala Lys Leu Ser Lys Asp Lys
225 230 235 240

Pro Thr Leu Ile Lys Met Thr Thr Thr Ile Gly Tyr Gly Ser Leu His
245 250 255

Ala Gly Ser His Ser Val His Gly Ala Pro Leu Lys Ala Asp Asp Val
260 265 270

Lys Gln Leu Lys Ser Lys Phe Gly Phe Asn Pro Asp Lys Ser Phe Val
275 280 285

Val Pro Gln Glu Val Tyr Asp His Tyr Gln Lys Thr Ile Leu Lys Pro
290 295 300

Gly Val Glu Ala Asn Asn Lys Trp Asn Lys Leu Phe Ser Glu Tyr Gln
305 310 315 320

Lys Lys Phe Pro Glu Leu Gly Ala Glu Leu Ala Arg Arg Leu Ser Gly
325 330 335

Gln Leu Pro Ala Asn Trp Glu Ser Lys Leu Pro Thr Tyr Thr Ala Lys
340 345 350

Asp Ser Ala Val Ala Thr Arg Lys Leu Ser Glu Thr Val Leu Glu Asp
355 360 365

Val Tyr Asn Gln Leu Pro Glu Leu Ile Gly Gly Ser Ala Asp Leu Thr
370 375 380

Pro Ser Asn Leu Thr Arg Trp Lys Glu Ala Leu Asp Phe Gln Pro Pro
385 390 395 400

Ser Ser Gly Ser Gly Asn Tyr Ser Gly Arg Tyr Ile Arg Tyr Gly Ile
405 410 415

Arg Glu His Ala Met Gly Ala Ile Met Asn Gly Ile Ser Ala Phe Gly
420 425 430

Ala Asn Tyr Lys Pro Tyr Gly Gly Thr Phe Leu Asn Phe Val Ser Tyr
435 440 445

Ala Ala Gly Ala Val Arg Leu Ser Ala Leu Ser Gly His Pro Val Ile

ES 2 601 052 T3

450 455 460

Trp Val Ala Thr His Asp Ser Ile Gly Val Gly Glu Asp Gly Pro Thr
465 470 475 480

His Gln Pro Ile Glu Thr Leu Ala His Phe Arg Ser Leu Pro Asn Ile
485 490 495

Gln Val Trp Arg Pro Ala Asp Gly Asn Glu Val Ser Ala Ala Tyr Lys
500 505 510

Asn Ser Leu Glu Ser Lys His Thr Pro Ser Ile Ile Ala Leu Ser Arg
515 520 525

Gln Asn Leu Pro Gln Leu Glu Gly Ser Ser Ile Glu Ser Ala Ser Lys
530 535 540

Gly Gly Tyr Val Leu Gln Asp Val Ala Asn Pro Asp Ile Ile Leu Val
545 550 555 560

Ala Thr Gly Ser Glu Val Ser Leu Ser Val Glu Ala Ala Lys Thr Leu
565 570 575

Ala Ala Lys Asn Ile Lys Ala Arg Val Val Ser Leu Pro Asp Phe Phe
580 585 590

Thr Phe Asp Lys Gln Pro Leu Glu Tyr Arg Leu Ser Val Leu Pro Asp
595 600 605

Asn Val Pro Ile Met Ser Val Glu Val Leu Ala Thr Thr Cys Trp Gly
610 615 620

Lys Tyr Ala His Gln Ser Phe Gly Ile Asp Arg Phe Gly Ala Ser Gly
625 630 635 640

Lys Ala Pro Glu Val Phe Lys Phe Phe Gly Phe Thr Pro Glu Gly Val
645 650 655

Ala Glu Arg Ala Gln Lys Thr Ile Ala Phe Tyr Lys Gly Asp Lys Leu
660 665 670

Ile Ser Pro Leu Lys Lys Ala Phe
675 680

<210> 22

<211> 2046

5 <212> ADN

<213> Saccharomyces cerevisiae

<400> 22

atggcacagt tctccgacat tgataaactt gcggtttcca cttaagatt actttccggt 60
gaccaggtgg aaagcgaca atctggccac ccaggtgcac cactaggatt ggcaccagtt 120

ES 2 601 052 T3

gccccatgtaa ttttcaagca actgcgctgt aaccctaaca atgaacattg gatcaataga 180
gacaggtttg ttctgtcgaa cggctactca tgcgctcttc tgtactcaat gctccatcta 240
ttaggatacg attactctat cgaggacttg agacaattta gacaagtaaa ctcaaggaca 300
ccgggtcatc cagaattcca ctacagcggga gtggaaatca cttccgggcc gctaggccag 360
ggtatctcaa atgctgttgg tatggcaata ggcgaggcca accttgccgc cacttataac 420
gaggatggct ttcccatttc cgactcatat acgtttgcta ttgtagggga tggttgctta 480
caagaggggtg tttcttcgga gacctcttcc ttagcgggac atctgcaatt gggtaacttg 540
attacgtttt atgacagtaa tagcatttcc attgacggta aaacctcgta ctcgttcgac 600
gaagatgttt tgaagcgata cgaggcatat ggttgggaag tcatggaagt cgataaagga 660
gacgacgata tggaaatccat ttctagcgtt ttggaaaagg caaaactatc gaaggacaag 720
ccaaccataa tcaaggaac tactacaatt ggatttgggt ccctacaaca gggactgct 780
ggtgttcacg ggtccgcttt gaaggcagat gatgttaaac agttgaagaa gaggtggggg 840
tttgacccaa ataatcatt tgtagtacct caagaggtgt acgattatta taagaagact 900
gttgtggaac ccggtcaaaa acttaatgag gaatgggata ggtgtttga agaatacaaa 960
accaaatttc ccgagaaggg taaagaattg caaagaagat tgaatggtga gttaccggaa 1020
ggttgggaaa agcatttacc gaagtttact ccggacgacg atgctctggc aacaagaaag 1080
acatcccagc aggtgctgac gaacatggtc caagttttgc ctgaattgat cggtggttct 1140
gccgatttga caccttcgaa tctgacaagg tgggaaggcg cggtagattt ccaacctccc 1200
attacccaac taggtaacta tgcaggaagg tacattagat acggtgtgag ggaacacgga 1260
atgggtgccca ttatgaacgg tatctctgcc tttggtgcaa actacaagcc ttacggtggt 1320
acctttttga acttcgtctc ttatgctgca ggagccgta ggttagccgc cttgtctggt 1380
aatccagtca tttgggttgc aacacatgac tctatcgggc ttggtgagga tggccaacg 1440
caccaaccta ttgaaactct ggctcacttg agggctattc caaacatgca tgtatggaga 1500
cctgctgatg gtaacgaaac ttctgctgcg tattattctg ctatcaaatc tggtcgaaca 1560
ccatctgttg tggctttatc acgacagaat cttcctcaat tggagcattc ctcttttgaa 1620
aaagccttga aggggtggcta tgtgatccat gacgtggaga atcctgatat taccctggtg 1680
tcaacaggat cagaagtctc catttctata gatgcagcca aaaaattgta cgatactaaa 1740
aaaatcaaag caagagttgt ttccctgcc aacttttata cttttgacag gcaaagtgaa 1800
gaatacagat tctctgttct accagacggt gttccgatca tgtccttga agtattggct 1860
acttcaagct ggggtaagta tgctcatcaa tcgttcggac tcgacgaatt tggctgttca 1920
ggcaaggggc ctgaaattta caaattgttc gatttcacag cggacggtgt tgcgtcaagg 1980
gctgaaaaga caatcaatta ctacaaagga aagcagttgc tttctcctat ggggaagagct 2040
ttctaa 2046

<210> 23

<211> 335

5 <212> PRT

<213> *Saccharomyces cerevisiae*

ES 2 601 052 T3

<400> 23

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

ES 2 601 052 T3

	245		250		255										
Ser	Pro	Ala	Leu	Leu	Asp	Lys	Leu	Met	Asn	Ser	Thr	Glu	Pro	Phe	Pro
			260					265					270		
Arg	Val	Leu	Asp	Pro	Val	Ser	Ala	Lys	Lys	Glu	Ala	Gly	Asp	Lys	Ile
		275					280					285			
Ser	Tyr	Ile	Ser	Asp	Glu	Ser	Lys	Phe	Arg	Phe	Asp	Leu	Asn	Glu	Asp
	290					295					300				
Ala	Met	Ala	Thr	Glu	Lys	Leu	Ser	Glu	Gly	Ile	Arg	Lys	Phe	Ser	Ala
305					310					315					320
Asp	Ile	Val	Thr	Leu	Phe	Asp	Leu	Ile	Glu	Lys	Lys	Val	Thr	Ala	
				325					330					335	

<210> 24

<211> 2046

5

<212> ADN

<213> *Saccharomyces cerevisiae*

<400> 24

```

atggcacagt tctccgacat tgataaactt gcggtttcca cttaagatt actttccggtt    60
gaccaggtgg aaagcgcaca atctggccac ccagggtgcac cactaggatt ggcaccagtt    120
gcccatgtaa tttcaagca actgcgctgt aaccctaaca atgaacattg gatcaataga    180
gacaggtttg ttctgtcgaa cggtcactca tgcgctcttc tgtactcaat gctccatcta    240
ttaggatacg attactctat cgaggacttg agacaattta gacaagtaaa ctcaaggaca    300
ccgggtcatc cagaattcca ctacgcggga gtggaaatca cttccggtcc gctaggccag    360
ggtatctcaa atgctggttg tatggcaata gcgcaggcca actttgccgc cacttataac    420
gaggatggct ttcccatttc cgactcatat acgtttgcta ttgtagggga tggttgctta    480
caagagggtg tttcttcgga gacctcttc ttagcgggac atctgcaatt gggtaacttg    540
attacgtttt atgacagtaa tagcatttcc attgacggta aaacctcgta ctcgttcgac    600
gaagatgttt tgaagcgata cgaggcatat ggttgggaag tcatggaagt cgataaagga    660
gacgacgata tggaatccat ttctagcgtt ttggaaaagg caaaactatc gaaggacaag    720
ccaaccataa tcaaggtaac tactacaatt ggatttgggt ccctacaaca gggtagtctg    780
ggtgttcatg ggtccgcttt gaaggcagat gatgttaaac agtgaagaa gaggtggggg    840
tttgaccaa ataaatcatt tgtagtacct caagaggtgt acgattatta taagaagact    900
gttgtggaac ccggtcaaaa acttaatgag gaatgggata ggatgttga agaatacaaa    960
accaaatttc ccgagaaggg taaagaattg caaagaagat tgaatgggta gttaccggaa   1020
ggttgggaaa agcatttacc gaagtttact ccggacgacg atgctctggc aacaagaaag   1080
acatcccagc aggtgctgac gaacatgggc caagttttgc ctgaattgat cggtggttct   1140
gccgatttga caccttcgaa tctgacaagg tgggaaggcg cggtagattt ccaacctccc   1200

```

10

ES 2 601 052 T3

attacccaac taggtaacta tgcaggaagg tacattagat acggtgtgag ggaacacgga 1260
atgggtgccca ttatgaacgg tatctctgcc tttggtgcaa actacaagcc ttacggtggt 1320
acctttttga acctcgtctc ttatgctgca ggagccgtta ggttagccgc cttgtctggt 1380
aatccagtca tttgggttgc aacacatgac tctatcgggc ttggtgagga tggccaacg 1440
caccaacctt ttgaaactct ggctcacttg agggctattc caaacatgca tgtatggaga 1500
cctgctgatg gtaacgaaac ttctgctgcg tattattctg ctatcaaadc tggtcgaaca 1560
ccatctggtt tggctttatc acgacagaat cttcctcaat tggagcattc ctcttttgaa 1620
aaagccttga aggggtggcta tgtgatccat gacgtggaga atcctgatat tatcctggtg 1680
tcaacaggat cagaagtctc catttctata gatgcagcca aaaaattgta cgatactaaa 1740
aaaatcaaag caagagttgt ttccctgccca gacttttata cttttgacag gcaaagtgaa 1800
gaatacagat tctctgttct accagacggg gttccgatca tgtcctttga agtattggct 1860
acttcaagct ggggtaagta tgctcatcaa tcggttcggac tgcacgaatt tggtcgttca 1920
ggcaaggggc ctgaaattta caaattgttc gatttcacag cggacgggtg tgcgtcaagg 1980
gctgaaaaga caatcaatta ctacaaagga aagcagttgc tttctctat ggaagagct 2040
ttctaa 2046

<210> 25

<211> 600

5 <212> PRT

<213> Saccharomyces cerevisiae

<400> 25

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

10 Val Tyr Trp Ser Ser Gln Ala Glu Ser Leu Leu Glu Gln Leu Asn Lys

ES 2 601 052 T3

```

385                390                395                400
Leu Gly Val Tyr Phe Pro Leu Gly Glu Ile Val Pro Ser Val Lys Ala
      405
Ile Asn Lys Arg Val Ile Phe Asn Pro Lys Thr Gly Met Ile Glu Arg
      420      425
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
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
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
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> 26

<211> 2467

5 <212> ADN

<213> Saccharomyces cerevisiae

<400> 26

10

```

ggatccaaga ccattattcc atcagaatgg aaaaaagttt aaaagatcac ggagattttg      60
ttcttctgag cttctgctgt ccttgaaaac aaattattcc gctggccgcc ccaaacaaaa      120
acaacccccga tttataaca ttgtcacagt attagaaatt ttctttttac aaattacat      180
ttccagctta ctacttccta taatcctcaa tcttcagcaa gcgaçgcagg gaatagccgc      240

```

ES 2 601 052 T3

tgagggtgcat aactgtcact tttcaattcg gccaatgcaa tctcaggcgg acgaataagg 300
 gggccctctc gagaaaaaca aaaggaggat gagattagta ctttaatggt gtgttcagta 360
 attcagagac agacaagaga ggtttccaac acaatgtctt tagactcata ctatcttggg 420
 tttgatcttt cgaccecaaca actgaaatgt ctcgccatta accaggacct aaaaattgtc 480
 cattcagaaa cagtggaatt tgaaaaggat cttccgcatt atcacacaaa gaagggtgtc 540
 tatatacacg gcgacactat cgaatgtccc gttagccatgt ggtagggggc tctagatctg 600
 gttctctcga aatatcgca ggctaaatct cattgaaca aagttatggc cgtctcaggg 660
 tcctgccagc agcacgggtc tgtctactgg tcctcccaag ccgaatctct gttagagcaa 720
 ttgaataaga aaccggaaaa agatttattg cactacgtga gctctgtagc atttgcaagg 780
 caaacgcccc ccaattggca agaccacagt actgcaaagc aatgtcaaga gtttgaagag 840
 tgcataagtg ggcctgaaaa aatggctcaa ttaacagggt ccagagccca ttttagattt 900
 actggtcctc aaattctgaa aattgcacaa ttgaaccag aagcttacga aaaaacaaa 960
 accatttctt tagtgtctaa tttttgact tctatcttag tgggccatct tgttgaatta 1020
 gaggaggcag atgcctgtgg tatgaacctt tatgatatac gtgaaagaaa attcatgtat 1080
 gagctactac atctaattga tagttcttct aaggataaaa ctatcagaca aaaattaatg 1140
 agagcaccca tgaaaaatct gatagcgggt accatctgta aatattttat tgagaagtac 1200
 ggtttcaata caaactgcaa ggtctctccc atgactgggg ataatttagc cactatatgt 1260
 tctttacccc tgcggaagaa tgacgttctc gtttccctag gaacaagtac tacagttctt 1320
 ctggtcaccg ataagtatca cccctctccg aactatcatc ttttcattca tccaactctg 1380
 ccaaaccatt atatgggtat gatttgttat tgtaatgggt ctttggcaag ggagaggata 1440
 agagacgagt taaacaaaga acgggaaaat aattatgaga agactaacga ttggactctt 1500
 tttaatcaag ctgtgctaga tgactcagaa agtagtgaaa atgaattagg tgtatatttt 1560
 cctctggggg agatcgttcc tagcgtaaaa gccataaaca aaagggttat cttcaatcca 1620
 aaaacgggta tgattgaaag agaggtggcc aagttcaaag acaagaggca cgatgccaaa 1680
 aatattgtag aatcacaggc ttttaagtgc agggtaagaa tatctcccct gctttcggat 1740
 tcaaacgcaa gctcacaaca gagactgaac gaagatacaa tcgtgaagtt tgattacgat 1800
 gaatctccgc tgcgggacta cctaaataaa aggccagaaa ggactttttt ttaggtggg 1860
 gcttctaaaa acgatgctat tgtgaagaag tttgctcaag tcattggtgc tacaagggt 1920
 aattttaggc tagaaacacc aaactcatgt gcccttgggt gttgttataa ggccatgtgg 1980
 tcattgttat atgactctaa taaaattgca gttccttttg ataaatttct gaatgacaat 2040
 tttccatggc atgtaatgga aagcatatcc gatgtggata atgaaaattg gatcgctata 2100
 attccaagat tgtcccctta agcgaactgg aaaagactct catctaaaat atgtttgaat 2160
 aatttatcat gccctgacaa gtacacacaa acacagacac ataataatac tacatatata 2220
 tatatcaccg ttattatgca tgcacatgac aatgcccttg tatgtttcgt atactgtagc 2280
 aagtagtcat cattttgttc cccgttcgga aaatgacaaa aagtaaaatc aataaatgaa 2340
 gagtaaaaaa caatttatga aagggtgagc gaccagcaac gagagagaca aatcaaatta 2400
 gcgctttcca gtgagaatat aagagagcat tgaaagagct aggttattgt taaatcatct 2460
 cgagctc 2467

ES 2 601 052 T3

<211> 494

<212> PRT

<213> Especie de Piromyces

5

<400> 27

Met Lys Thr Val Ala Gly Ile Asp Leu Gly Thr Gln Ser Met Lys Val
 1 5 10 15

Val Ile Tyr Asp Tyr Glu Lys Lys Glu Ile Ile Glu Ser Ala Ser Cys
 20 25 30

Pro Met Glu Leu Ile Ser Glu Ser Asp Gly Thr Arg Glu Gln Thr Thr
 35 40 45

Glu Trp Phe Asp Lys Gly Leu Glu Val Cys Phe Gly Lys Leu Ser Ala
 50 55 60

Asp Asn Lys Lys Thr Ile Glu Ala Ile Gly Ile Ser Gly Gln Leu His
 65 70 75 80

Gly Phe Val Pro Leu Asp Ala Asn Gly Lys Ala Leu Tyr Asn Ile Lys
 85 90 95

Leu Trp Cys Asp Thr Ala Thr Val Glu Glu Cys Lys Ile Ile Thr Asp
 100 105 110

Ala Ala Gly Gly Asp Lys Ala Val Ile Asp Ala Leu Gly Asn Leu Met
 115 120 125

Leu Thr Gly Phe Thr Ala Pro Lys Ile Leu Trp Leu Lys Arg Asn Lys
 130 135 140

Pro Glu Ala Phe Ala Asn Leu Lys Tyr Ile Met Leu Pro His Asp Tyr
 145 150 155 160

Leu Asn Trp Lys Leu Thr Gly Asp Tyr Val Met Glu Tyr Gly Asp Ala
 165 170 175

Ser Gly Thr Ala Leu Phe Asp Ser Lys Asn Arg Cys Trp Ser Lys Lys
 180 185 190

Ile Cys Asp Ile Ile Asp Pro Lys Leu Leu Asp Leu Leu Pro Lys Leu
 195 200 205

ES 2 601 052 T3

Ile Glu Pro Ser Ala Pro Ala Gly Lys Val Asn Asp Glu Ala Ala Lys
 210 215 220

Ala Tyr Gly Ile Pro Ala Gly Ile Pro Val Ser Ala Gly Gly Gly Asp
 225 230 235 240

Asn Met Met Gly Ala Val Gly Thr Gly Thr Val Ala Asp Gly Phe Leu
 245 250 255

Thr Met Ser Met Gly Thr Ser Gly Thr Leu Tyr Gly Tyr Ser Asp Lys
 260 265 270

Pro Ile Ser Asp Pro Ala Asn Gly Leu Ser Gly Phe Cys Ser Ser Thr
 275 280 285

Gly Gly Trp Leu Pro Leu Leu Cys Thr Met Asn Cys Thr Val Ala Thr
 290 295 300

Glu Phe Val Arg Asn Leu Phe Gln Met Asp Ile Lys Glu Leu Asn Val
 305 310 315 320

Glu Ala Ala Lys Ser Pro Cys Gly Ser Glu Gly Val Leu Val Ile Pro
 325 330 335

Phe Phe Asn Gly Glu Arg Thr Pro Asn Leu Pro Asn Gly Arg Ala Ser
 340 345 350

Ile Thr Gly Leu Thr Ser Ala Asn Thr Ser Arg Ala Asn Ile Ala Arg
 355 360 365

Ala Ser Phe Glu Ser Ala Val Phe Ala Met Arg Gly Gly Leu Asp Ala
 370 375 380

Phe Arg Lys Leu Gly Phe Gln Pro Lys Glu Ile Arg Leu Ile Gly Gly
 385 390 395 400

Gly Ser Lys Ser Asp Leu Trp Arg Gln Ile Ala Ala Asp Ile Met Asn
 405 410 415

Leu Pro Ile Arg Val Pro Leu Leu Glu Glu Ala Ala Ala Leu Gly Gly
 420 425 430

Ala Val Gln Ala Leu Trp Cys Leu Lys Asn Gln Ser Gly Lys Cys Asp
 435 440 445

Ile Val Glu Leu Cys Lys Glu His Ile Lys Ile Asp Glu Ser Lys Asn
 450 455 460

Ala Asn Pro Ile Ala Glu Asn Val Ala Val Tyr Asp Lys Ala Tyr Asp
 465 470 475 480

Glu Tyr Cys Lys Val Val Asn Thr Leu Ser Pro Leu Tyr Ala
 485 490

<210> 28

5 <211> 2041

<212> ADN

ES 2 601 052 T3

<213> Especie de Piromyces

<400> 28

```

attatataaa ataactttaa ataaaacaat ttttatttgt ttatttaatt attcaaaaaa    60
aattaaagta aaagaaaaat aatacagtag aacaatagta ataatatcaa aatgaagact    120
gttgctggta ttgatcttgg aactcaaagt atgaaagtcg ttatttacga ctatgaaaag    180
aaagaaatta ttgaaagtgc tagctgtcca atggaattga tttccgaaag tgacgggtacc    240
cgtgaacaaa ccactgaatg gtttgacaag ggtcttgaag tttgttttgg taagcttagt    300
gctgataaca aaaagactat tgaagctatt ggtatttctg gtcaattaca cggttttggt    360
cctcttgatg ctaacggtaa ggctttatac aacatcaaac tttgggtgta tactgctacc    420
gttgaagaat gtaagattat cactgatgct gccgggtggg acaaggctgt tattgatgcc    480
cttggttaacc ttatgctcac cggtttcacc gctccaaaga tcctctggct caagcgcaac    540
aagccagaag ctttcgctaa cttaaagtac attatgcttc cacacgatta cttaaactgg    600
aagcttactg gtgattacgt tatggaatac ggtgatgcct ctggtaccgc tctcttcgat    660
tctaagaacc gttgctggtc taagaagatt tgcgatatca ttgacccaaa acttttagat    720
ttacttccaa agttaattga accaagcgct ccagctggta aggttaatga tgaagccgct    780
aaggcttacg gtattccagc cggattcca gtttccgctg gtgggtggta taacatgatg    840
ggtgctgttg gtactggtag tgttgctgat ggtttcctta ccatgtctat gggacttct    900
ggtactcttt acggttacag tgacaagcca attagtacc cagctaattgg ttaagtgggt    960
ttctgttctt ctactggtgg atggcttcca ttactttgta ctatgaactg tactgttgcc   1020
actgaattcg ttcgtaacct cttccaaatg gatattaagg aacttaatgt tgaagctgcc   1080
aagcttccat gtggttagtga aggtgtttta gttattccat tcttcaatgg tgaagaact   1140
ccaaacttac caaacggtcg tgctagtatt actggtctta cttctgctaa caccagccgt   1200
gctaacattg ctgctgctag tttcgaatcc gccgttttcg ctatgcgtgg tggtttgat   1260
gctttccgta agttaggttt ccaaccaaag gaaattcgtc ttattggtgg tggttctaag   1320
tctgatctct ggagacaaat tgccgctgat atcatgaacc ttccaatcag agttccactt   1380
ttagaagaag ctgctgctct tgggtggtgct gttcaagctt tatggtgtct taagaaccaa   1440
tctggtaagt gtgatattgt tgaactttgc aaagaacaca ttaagattga tgaatctaag   1500
aatgctaacc caattgccga aaatgttgct gtttacgaca aggcttacga tgaatactgc   1560
aaggttgtaa atactctttc tccattatat gcttaaattg ccaatgtaaa aaaaaatata   1620
atgccatata atgaccttgt caatacactg ttcattgttca tataatcata ggacattgaa   1680
5 tttacaaggt ttatacaatt aatatctatt atcatattat tatacagcat ttcattttct   1740
aagattagac gaaacaattc ttggttcctt gcaatataca aaatttacat gaatttttag   1800
aatagtctcg tatttatgcc caataatcag gaaaattacc taatgctgga ttcttggttaa   1860
taaaaacaaa ataaataaat taaataaaca aataaaaatt ataagtaaat ataaatata   1920
aagtaatata aaaaaaaagt aaataaataa ataaataaat aaaaattttt tgcaaatata   1980
taaataaata aataaaatat aaaaataatt tagcaataaa attaaaaaaa aaaaaaaaaa   2040
a                                                                                   2041

```

ES 2 601 052 T3

<210> 29

<211> 327

<212> PRT

<213> Saccharomyces cerevisiae

5

<400> 29

```

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

```


ES 2 601 052 T3

Ile Glu His His Pro Tyr Leu Thr Gln Glu His Leu Val Glu Phe Cys
 195 200 205

Lys Leu His Asp Ile Gln Val Val Ala Tyr Ser Ser Phe Gly Pro Gln
 210 215 220

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

Phe Glu Asn Asp Val Ile Lys Lys Val Ser Gln Asn His Pro Gly Ser
 245 250 255

Thr Thr Ser Gln Val Leu Leu Arg Trp Ala Thr Gln Arg Gly Ile Ala
 260 265 270

Val Ile Pro Lys Ser Ser Lys Lys Glu Arg Leu Leu Gly Asn Leu Glu
 275 280 285

Ile Glu Lys Lys Phe Thr Leu Thr Glu Gln Glu Leu Lys Asp Ile Ser
 290 295 300

Ala Leu Asn Ala Asn Ile Arg Phe Asn Asp Pro Trp Thr Trp Leu Asp
 305 310 315 320

Gly Lys Phe Pro Thr Phe Ala
 325

<210> 30

<211> 984

5 <212> ADN

<213> Saccharomyces cerevisiae

<400> 30

atgtcttcac tggttactct taataacggg ctgaaaatgc ccctagtcgg cttaggggtgc 60
 tggaaaattg acaaaaaagt ctgtgcgaat caaatttatg aagctatcaa attaggtctac 120
 cgtttattcg atgggtgctg cgactacggc aacgaaaagg aagttggtga aggtatcagg 180
 aaagccatct ccgaaggtct tgtttctaga aaggatataat ttgttgtttc aaagtatgg 240
 aacaattttc accatcctga tcatgtaaaa ttagctttaa agaagacctt aagcgatatg 300
 ggacttgatt atttagacct gtattatatt cacttcccaa tcgccttcaa atatgttcca 360
 tttgaagaga aataccctcc aggattctat acgggcgcag atgacgagaa gaaaggtcac 420
 atcaccgaag cacatgtacc aatcatagat acgtaccggg ctctggaaga atgtgttgat 480
 gaaggcttga ttaagtctat tgggttttcc aactttcagg gaagcttgat tcaagattta 540
 ttacgtggtt gtagaatcaa gcccggtggct ttgcaaattg aacaccatcc ttatttgact 600
 caagaacacc tagttgagtt ttgtaaatta cacgatatcc aagtagttgc ttactcctcc 660
 ttcggtcctc aatcattcat tgagatggac ttacagttgg caaaaaccac gccaaactctg 720
 ttcgagaatg atgtaatcaa gaaggcttca caaaaccatc caggcagtac cacttcccaa 780

10

ES 2 601 052 T3

gtattgctta gatgggcaac tcagagagggc attgccgtca ttccaaaatc ttccaagaag 840
gaaaggttac ttggcaacct agaaatcgaa aaaaagttca ctttaacgga gcaagaattg 900
aaggatattt ctgcactaaa tgccaacatc agatttaatg atccatggac ctggttggat 960
ggtaaattcc ccacttttgc ctga 984

<210> 31

<211> 31

5 <212> ADN

<213> ARTIFICIAL

<220>

<223> cebador

10

<400> 31

gactagtcga gttatcatt atcaatactg c 31

<210> 32

15 <211> 49

<212> ADN

<213> Artificial

<220>

20 <223> cebador

<400> 32

ctcataatca ggtactgata acattttggt tgttatgtg tgtttattc 49

25 <210> 33

<211> 49

<212> ADN

<213> artificial

30 <220>

<223> cebador

<400> 33

gaataaacac acataaacaa acaaaatggt atcagttacct gattatgag 49

ES 2 601 052 T3

<210> 34
<211> 48
<212> ADN
<213> artificial
5
<220>
<223> cebador

<400> 34
10 aatcataaat cataagaaat tcgcttactt taagaatgcc ttatgcat 48

<210> 35
<211> 48
<212> ADN
15 <213> artificial

<220>
<223> cebador

<400> 35
20 atgactaagg cattcttaaa gtaagcgaat ttcttatgat ttatgatt 48

<210> 36
<211> 36
25 <212> ADN
<213> artificial

<220>
<223> cebador
30

<400> 36
cactagtctc gagtgtggaa gaacgattac aacagg 36

<210> 37
35 <211> 31
<212> ADN
<213> artificial

ES 2 601 052 T3

<220>

<223> cebador

<400> 37

5 cgagctcgtg ggtgtattgg attataggaa g 31

<210> 38

<211> 48

<212> ADN

10 <213> artificial

<220>

<223> cebador

15 <400> 38

ttgggctgt tcaactaaat tcatttttag gctggtatct tgattcta 48

<210> 39

<211> 48

20 <212> ADN

<213> artificial

<220>

<223> cebador

25

<400> 39

tagaatcaag ataccagcct aaaaatgaat ttagttgaaa cagcccaa 48

<210> 40

30 <211> 48

<212> ADN

<213> artificial

<220>

35 <223> cebador

<400> 40

ES 2 601 052 T3

aatcataaat cataagaaat tcgctctaat atttgattgc ttgccag 48

<210> 41
<211> 48

5 <212> ADN
<213> artificial

<220>
<223> cebador

10 <400> 41
ctgggcaagc aatcaaatat tagagcgaat ttcttatgat ttatgatt 48

<210> 42

15 <211> 31
<212> ADN
<213> artificial

<220>

20 <223> cebador

<400> 42
tgagctcgtg tggaagaacg attacaacag g 31

25 <210> 43
<211> 28
<212> ADN
<213> artificial

30 <220>
<223> cebador

<400> 43
acgcgctcgac tcgtaggaac aatttcgg 28

35 <210> 44
<211> 50

ES 2 601 052 T3

<212> ADN

<213> artificial

<220>

5 <223> cebador

<400> 44

cttcttggtt taatgcttct agcattttt gattaaaatt aaaaaaactt 50

10 <210> 45

<211> 50

<212> ADN

<213> artificial

15 <220>

<223> cebador

<400> 45

aagtttttt aatttaatc aaaaaatgct agaagcatta aaacaagaag 50

20

<210> 46

<211> 46

<212> ADN

<213> artificial

25

<220>

<223> cebador

<400> 46

30 ggtatatatt taagagcgat ttgttactt gcgaactgca tgatcc 46

<210> 47

<211> 46

<212> ADN

35 <213> artificial

<220>

ES 2 601 052 T3

<223> cebador

<400> 47

ggatcatgca gttcgcaagt aaacaaatcg ctcttaaata tataacc 46

5

<210> 48

<211> 33

<212> ADN

<213> artificial

10

<220>

<223> cebador

<400> 48

15 cgcagtcgac ctttaaaca gttgatgaga acc 33

<210> 49

<211> 676

<212> ADN

20 <213> artificial

<220>

<223> promotor

25 <400> 49

tcgagtttat cattatcaat actgccatth caaagaatac gtaaataatt aatagtagtg 60
atthttcctaa cthttatthtag tcaaaaaatt agcctththaa thctgctgta acccgtacat 120
gccccaaaata gggggcgggt tacacagaat atataacatc gtaggtgtct gggggaacag 180
thttatthctg gcatccacta aatataatgg agcccgttht ttaagctggc atccagaaaa 240
aaaaagaatc ccagcaccaa aatattgtht thttcaccaa ccatcagthc ataggtccat 300
thctthtagcg caactacaga gaacaggggc acaaacaggc aaaaaacggg cacaacctca 360
atggagtgat gcaacctgcc tggagthaat gatgacacaa ggcaattgac ccacgcatgt 420
atctatctca thttcttaca cthtctatta cthtctgctc thctctgattt ggaaaaagct 480
gaaaaaaaag gthgaaacca gthccctgaa atthattcccc tacttgacta ataagtatat 540
aaagacggta ggtattgatt gthattctgt aatctattht cthtaacttc thaaattcta 600
ctthttatagt tagthcttht thtagththt aaacaccaag aacttagtht cgaataaaca 660
cacataaaca aacaaa 676

ES 2 601 052 T3

<210> 50

<211> 326

<212> ADN

<213> artificial

5

<220>

<223> terminador

<400> 50

10

```

gcgaatttct tatgatttat gatTTTTtatt attaaataag ttataaaaaa aataagtgta      60
tacaaatttt aaagtgactc ttaggtttta aaacgaaaat tcttattctt gagtaactct      120
ttcctgtagg tcaggttgct ttctcaggta tagcatgagg tcgctcttat tgaccacacc      180
tctaccggca tgccgagcaa atgcctgcaa atcgctcccc atttcaccca attgtagata      240
tgctaactcc agcaatgagt tgatgaatct cggtgtgtat tttatgtcct cagaggacaa      300
cacctgttgt aatcgttctt ccacac                                             326
    
```

<210> 51

15

<211> 374

<212> ADN

<213> artificial

<220>

20

<223> promotor

<400> 51

```

gtgggtgtat tggattatag gaagccacgc gctcaacctg gaattacagg aagctggtaa      60
ttttttgggt ttgcaatcat caccatctgc acgttggtat aatgtcccgt gtctatatat      120
atccattgac ggtattctat ttttttgcta ttgaaatgag cgttttttgt tactacaatt      180
ggttttacag acggaatttt ccctatttgt ttcgtcccat ttttcctttt ctattgttc      240
tcatactta aaaaggctct ttcttcataa tcaatgcttt cttttactta atattttact      300
tgcattcagt gaattttaat acatattcct ctagtcttgc aaaatcgatt tagaatcaag      360
ataccagcct aaaa                                                         374
    
```

25

<210> 52

<211> 390

ES 2 601 052 T3

<212> ADN

<213> artificial

<220>

5 <223> promotor

<400> 52

```

ctcgtaggaa caatttcggg cccctgcgtg ttcttctgag gttcatcttt tacatttgct    60
tctgctggat aattttcaga ggcaacaagg aaaaattaga tggcaaaaag tcgtctttca    120
aggaaaaatc cccaccatct ttcgagatcc cctgtaactt attggcaact gaaagaatga    180
aaaggaggaa aatacaaaat atactagaac tgaaaaaaaaa aaagtataaa tagagacgat    240
atatgccaat acttcacaat gttcgaatct attcttcatt tgcagctatt gtaaaataat    300
aaaacatcaa gaacaacaa gctcaacttg tcttttctaa gaacaagaa taaacacaaa    360
aacaaaaagt tttttaatt ttaatcaaaa    390

```

10

<210> 53

<211> 302

<212> ADN

<213> artificial

15

<220>

<223> terminador

<400> 53

20

```

acaaatcgct cttaaatata tacctaaaga acattaaagc tatattataa gcaagatac    60
gtaaattttg cttatattat tatacacata tcatatttct atatttttaa gatttggtta    120
tataatgtac gtaatgcaa ggaaataaat ttatacatt attgaacagc gtccaagtaa    180
ctacattatg tgcactaata gtttagcgtc gtgaagactt tattgtgtcg cgaaaagtaa    240
aaattttaaa aattagagca ccttgaactt gcgaaaaagg ttctcatcaa ctgttttaaa    300
gg    302

```

REIVINDICACIONES

1. Una célula de levadura capaz de expresar las siguientes secuencias de nucleótidos, en la que la expresión de estas secuencias de nucleótidos confiere a la célula de levadura la capacidad de usar L-arabinosa y/o de convertir L-arabinosa en L-ribulosa y/o xilulosa-5-fosfato y/o en etanol, en cuya célula de levadura se expresan:

5 (a) una secuencia de nucleótidos que codifica una arabinosa isomerasa (araA), en la que dicha secuencia de nucleótidos está seleccionada del grupo que consiste en:

i. secuencias de nucleótidos que codifican una araA, comprendiendo dicha araA una secuencia de aminoácidos que tiene al menos el 80 % de identidad de secuencias con la secuencia de aminoácidos de SEQ ID NO: 1.

10 ii. secuencias de nucleótidos que comprenden una secuencia de nucleótidos que tiene al menos 80 % de identidad de secuencias con la secuencia de nucleótidos de SEQ ID NO: 2;

iii. secuencias de nucleótidos cuya hebra complementaria se hibrida con una molécula de ácido nucleico de secuencia de (i) o (ii);

15 (b) una secuencia de nucleótidos que codifica una L-ribulocinasa (araB), en la que dicha secuencia de nucleótidos está seleccionada del grupo que consiste en:

i. secuencias de nucleótidos que codifican una araB, comprendiendo dicha araB una secuencia de aminoácidos que tiene al menos el 80 % de identidad de secuencias con la secuencia de aminoácidos de SEQ ID NO: 3.

20 ii. secuencias de nucleótidos que comprenden una secuencia de nucleótidos que tiene al menos el 80 % de identidad de secuencias con la secuencia de nucleótidos de SEQ ID NO: 4.

iii. secuencias de nucleótidos cuya hebra complementaria se hibrida con una molécula de ácido nucleico de secuencia de (i) o (ii); y

(c) una secuencia de nucleótidos que codifica una L-ribulosa-5-P-4-epimerasa (araD), en la que dicha secuencia de nucleótidos está seleccionada del grupo que consiste en:

25 i. secuencias de nucleótidos que codifican una araD, comprendiendo dicha araD una secuencia de aminoácidos que tiene al menos el 80 % de identidad de secuencias con la secuencia de aminoácidos de SEQ ID NO: 5.

ii. secuencias de nucleótidos que comprenden una secuencia de nucleótidos que tiene al menos el 80 % de identidad de secuencias con la secuencia de nucleótidos de SEQ ID NO: 6;

30 iii. secuencias de nucleótidos cuya hebra complementaria se hibrida con una molécula de ácido nucleico de secuencia de (i) o (ii),

en la que en los puntos iii. la hibridación se determina bajo condiciones de hibridación que permiten que una secuencia de ácidos nucleicos de 200 nucleótidos se hibride a una temperatura de 65 °C en una disolución que comprende sal aproximadamente 1 M, y lavar a 65 °C en una disolución de aproximadamente 0,1 M, en la que la hibridación se realiza durante 10 horas y el lavado se realiza una hora con dos cambios de la disolución de lavado.

35 2. Una célula de levadura según la reivindicación 1, en la que una, dos o tres de las secuencias de nucleótidos de araA, araB y araD se originan a partir de un género de *Lactobacillus*, preferentemente una especie de *Lactobacillus plantarum*.

40 3. Una célula de levadura según la reivindicación 1 o 2, en la que la célula es una célula de levadura, preferentemente que pertenece a uno de los géneros: *Saccharomyces*, *Kluyveromyces*, *Candida*, *Pichia*, *Schizosaccharomyces*, *Hansenula*, *Kloeckera*, *Schwanniomyces* o *Yarrowia*.

4. Una célula de levadura según la reivindicación 3, en la que la célula de levadura pertenece a una de las especies: *S. cerevisiae*, *S. bulderi*, *S. barnetti*, *S. exiguus*, *S. uvarum*, *S. diastaticus*, *K. lactis*, *K. marxianus* o *K. fragilis*.

45 5. Una célula de levadura según una cualquiera de las reivindicaciones 1 a 4, en la que las secuencias de nucleótidos que codifican araA, araB y/o araD están operativamente enlazadas a un promotor que produce expresión suficiente de las secuencias de nucleótidos correspondientes en la célula para conferir a la célula la capacidad de usar L-arabinosa y/o de convertir L-arabinosa en L-ribulosa y/o xilulosa-5-fosfato y/o en etanol.

6. Una célula de levadura según una cualquiera de las reivindicaciones 1 a 5, en la que la célula de levadura presenta la capacidad de isomerizar directamente xilosa en xilulosa.

7. Una célula de levadura según la reivindicación 6, en la que la célula de levadura comprende una modificación genética que aumenta el flujo de la vía de pentosa fosfato.
8. Una célula de levadura según la reivindicación 6 o 7, en la que la modificación genética comprende la expresión en exceso de al menos un gen de la parte no oxidativa de la vía de pentosa fosfato.
- 5 9. Una célula de levadura según la reivindicación 8, en la que el gen está seleccionado del grupo que consiste en los genes que codifican ribulosa-5-fosfato isomerasa, ribulosa-5-fosfato epimerasa, transcetolasa y transaldolasa.
10. Una célula de levadura según la reivindicación 8, en la que la modificación genética comprende la expresión en exceso de al menos los genes que codifican una transcetolasa y una transaldolasa.
- 10 11. Una célula de levadura según una cualquiera de las reivindicaciones 8 a 10, en la que la célula de levadura comprende además una modificación genética que aumenta la actividad específica de xilulosa cinasa.
12. Una célula de levadura según la reivindicación 11, en la que la modificación genética comprende expresión en exceso de un gen que codifica una xilulosa cinasa.
13. Una célula de levadura según una cualquiera de las reivindicaciones 8 a 12, en la que el gen que se expresa en exceso es endógeno a la célula.
- 15 14. Una célula de levadura según una cualquiera de las reivindicaciones 5 a 13, en la que la célula de levadura comprende una modificación genética que reduce la actividad no específica de aldosa reductasa en la célula de levadura.
15. Una célula de levadura según la reivindicación 14, en la que la modificación genética reduce la expresión de, o inactiva un gen que codifica una aldosa reductasa no específica.
- 20 16. Una célula de levadura según la reivindicación 15, en la que el gen se inactiva por delección de al menos parte del gen o por rotura del gen.
17. Una célula de levadura según las reivindicaciones 14 o 15, en la que la expresión de cada gen en la célula de levadura que codifica una aldosa reductasa no específica se reduce o inactiva.
- 25 18. Una célula de levadura según una cualquiera de las reivindicaciones precedentes, en la que el producto de fermentación es etanol.
19. Una construcción de ácidos nucleicos que comprende una secuencia de ácidos nucleicos que codifica una araA, una secuencia de ácidos nucleicos que codifica una araB y/o una secuencia de ácidos nucleicos que codifica una araD, todas como se define en la reivindicación 1 o 2.
20. Un proceso de producción de etanol, por el cual el proceso comprende:
- 30 a. fermentar un medio que contiene una fuente de arabinosa y opcionalmente xilosa con una célula de levadura modificada como se define en una cualquiera de las reivindicaciones 1 a 18, por el cual la célula de levadura fermenta arabinosa y opcionalmente xilosa a etanol; y opcionalmente,
- b. recuperar el etanol.
21. Un proceso producción de etanol, en el que el proceso comprende:
- 35 a. fermentar un medio que contiene al menos una fuente de L-arabinosa y una fuente de xilosa con una célula de levadura como se define en una cualquiera de las reivindicaciones 1 a 18 y una célula de levadura capaz de usar xilosa y/o que presenta la capacidad de isomerizar directamente xilosa en xilulosa, por el cual cada célula de levadura fermenta L-arabinosa y/o xilosa en etanol; y opcionalmente,
- b. recuperar el producto de fermentación.
- 40 22. Un proceso según la reivindicación 20 o 21, en el que el medio también contiene una fuente de glucosa.
23. Un proceso según cualquiera de las reivindicaciones 20-22, en el que la productividad volumétrica del etanol es al menos 0,5 g de etanol por litro por hora.
24. Un proceso según la reivindicación 23, en el que el rendimiento del etanol es al menos el 30 %.
25. Un proceso según una cualquiera de las reivindicaciones 20 a 24, en el que el proceso es anaerobio.
- 45 26. Un proceso según una cualquiera de las reivindicaciones 20 a 24, en la que el proceso es aerobio, realizado preferentemente bajo condiciones limitadas de oxígeno.

Fig 1

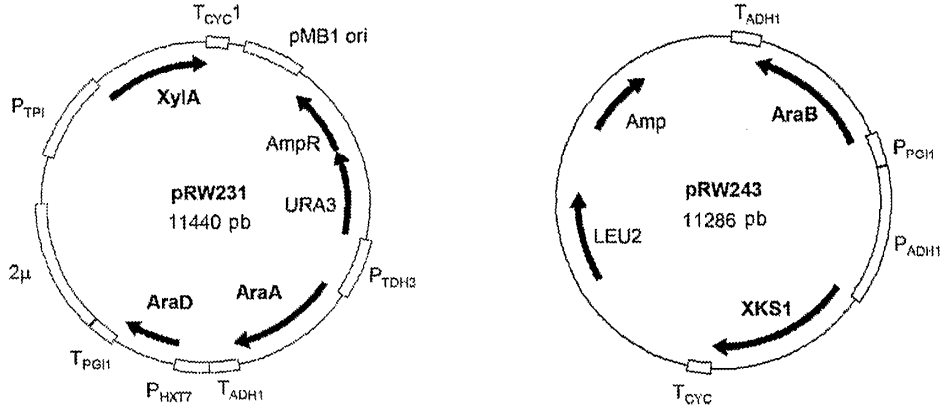


Fig 2

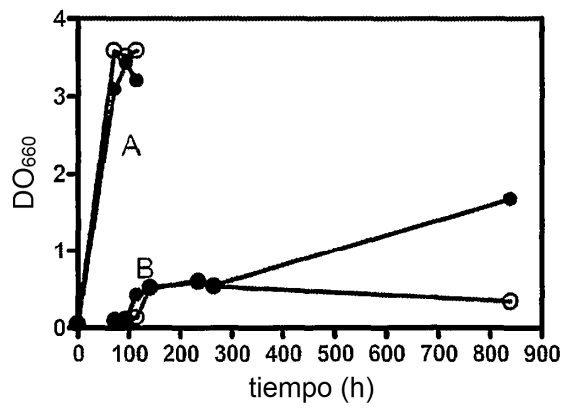


Fig 3

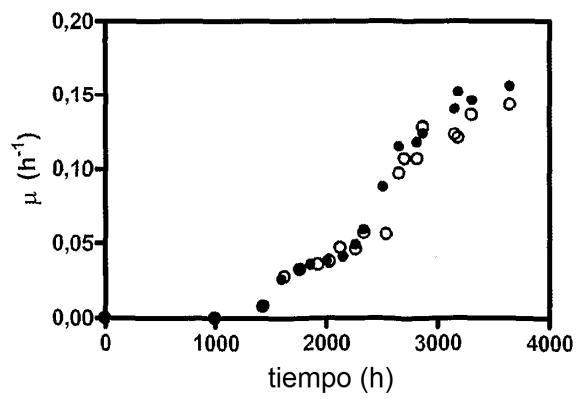


Fig 4

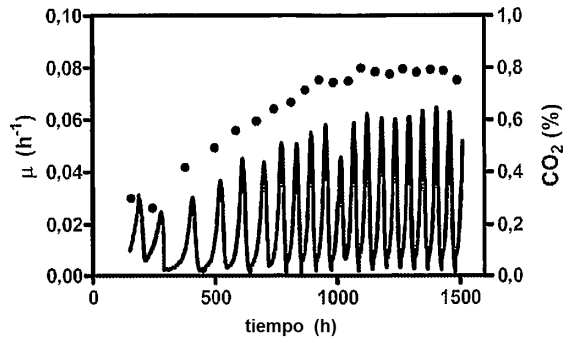


Fig 5a

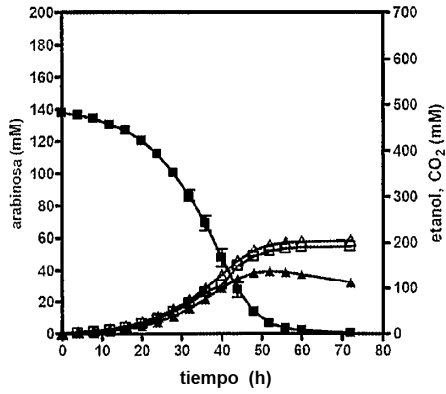


Fig 5b

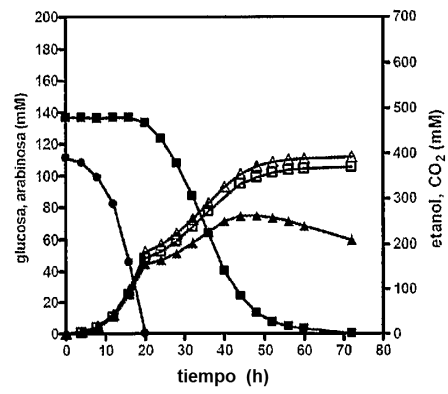


Fig 5c

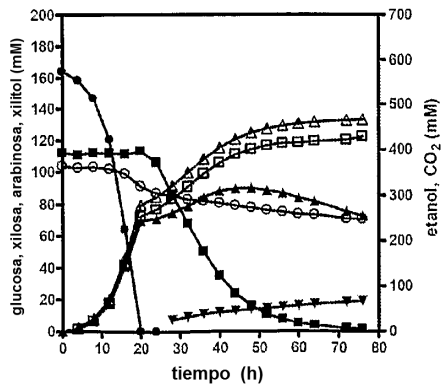


Fig 5d

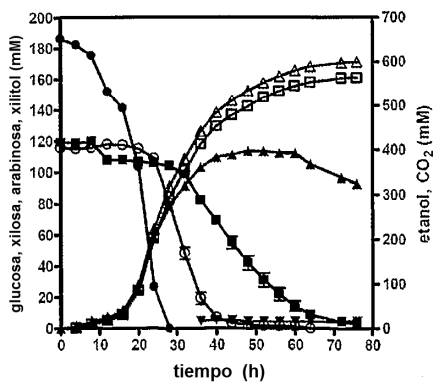


Fig 6

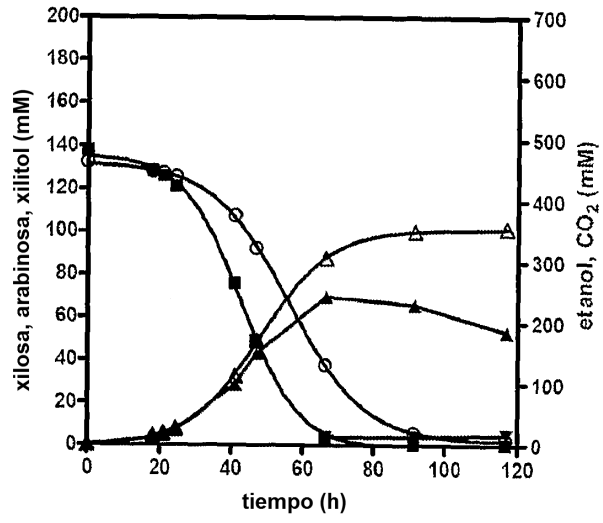


Fig 7

