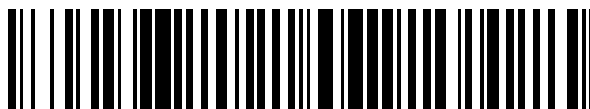


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 602 808**

51 Int. Cl.:

C12N 1/19

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **11.04.2012** **PCT/US2012/033030**

87 Fecha y número de publicación internacional: **18.10.2012** **WO12142094**

96 Fecha de presentación y número de la solicitud europea: **11.04.2012** **E 12770605 (9)**

97 Fecha y número de publicación de la concesión europea: **10.08.2016** **EP 2697252**

54 Título: **Composiciones y métodos para la elevada producción de etanol a partir de biomasa**

30 Prioridad:

11.04.2011 US 201161474035 P

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

22.02.2017

73 Titular/es:

CARGILL, INCORPORATED (100.0%)
15407 McGinty Road West
Wayzata, MN 55391, US

72 Inventor/es:

JESSEN, HOLLY J.;
YI, JIAN;
LUNDORFF, JOSHUA;
LIAO, HANS;
NEGRETE-RAYMOND, ANA;
SUOMINEN, PIRKKO y
ARISTIDOU, ARISTOS

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 602 808 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

Composiciones y métodos para la elevada producción de etanol a partir de biomasa

Antecedentes

Se ha realizado un gran trabajo en los últimos años para desarrollar métodos rentables para generar etanol a partir de biomasa. El uso de biomasa para generar etanol para combustible presenta varias ventajas con respecto al uso de fuentes de materia prima más tradicionales. Los posibles materiales de partida son abundantes y diversos, el uso de estas materias primas no se desvía del suministro de alimentos, y posiblemente presentan una presencia de carbono más pequeña.

Aunque la biomasa proporciona un sustrato atractivo para la producción de etanol, también presenta varios retos. Primero, la biomasa contiene tanto celulosa, que puede degradarse en azúcares de hexosa tales como glucosa, como hemicelulosa, que puede degradarse en azúcares de pentosa tales como arabinosa y xilosa. Muchos de los microorganismos tradicionalmente usados en la fermentación del etanol son incapaces de fermentar tanto azúcares de hexosa como de pentosa en etanol. Segundo, a diferencia de fuentes de materia prima más tradicionales de etanol (p. ej., maíz, azúcar de caña), la biomasa incluye componentes estructurales de fuentes de planta. Debido a que el material de la fuente es estructural y más difícil de degradar, la biomasa requiere más procesamiento para generar los monómeros de azúcar que funcionan como sustrato de fermentación. Tercero, el hidrolizado resultante del pre-tratamiento de la biomasa presenta un ambiente agresivo para los microorganismos fermentadores.

Varias especies bacterianas son capaces de fermentar azúcares de pentosa a etanol, pero estas especies generalmente producen una mezcla de productos en vez de un único producto. Frecuentemente, uno o más de estos productos son perjudiciales para las bacterias. Además, las bacterias pueden presentar tasas de fermentación drásticamente reducidas en el ambiente agresivo del hidrolizado de materia de planta.

Generalmente se considera que las levaduras son candidatos más atractivos para la fermentación del etanol a escala industrial que las bacterias. Sin embargo, muy pocas levaduras son capaces de fermentar azúcares de pentosa a etanol. Se han introducido diversas modificaciones genéticas en diferentes especies de levadura en un intento por vencer este problema, pero ninguna de estas cepas modificadas previamente desarrolladas ha demostrado ser completamente satisfactoria para la producción de etanol a gran escala a partir de biomasa. Por tanto, hay una necesidad en la materia de nuevas cepas de levadura genéticamente modificadas capaces de fermentar biomasa a etanol.

Compendio

La invención se define en las reivindicaciones.

En la presente memoria se proporcionan en ciertas realizaciones polinucleótidos KHT105 y RAG4 aislados. En ciertas de estas realizaciones, los polinucleótidos codifican un polipéptido con la secuencia de aminoácidos de SEQ ID NOs:2 (KHT105) o 4 (RAG4), o un polipéptido que comprende una secuencia de aminoácidos con al menos el 90 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NOs:2 o 4. En ciertas realizaciones, los polinucleótidos comprenden la secuencia de nucleótidos de SEQ ID NOs:1 o 3, o una secuencia de nucleótidos con al menos el 90 % de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NOs:1 o 3. En otras realizaciones, los polinucleótidos codifican un polipéptido que comprende una secuencia de aminoácidos con al menos el 90 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NOs:2 o 4, donde el polipéptido codificado es capaz de transportar xilosa en una célula de levadura. En ciertas de estas realizaciones, una célula de levadura que expresa en exceso el polinucleótido consume una mayor cantidad de xilosa con respecto a glucosa que una célula de levadura idéntica que no expresa en exceso el polinucleótido. En ciertas realizaciones, los polinucleótidos comprenden una secuencia de nucleótidos con al menos el 90 % de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NOs:1 o 3.

En la presente memoria se proporcionan en ciertas realizaciones polipéptidos KHT105 y RAG4 aislados. En ciertas de estas realizaciones, los polipéptidos comprenden la secuencia de aminoácidos de SEQ ID NOs:2 (KHT105) o 4 (RAG4), o una secuencia de aminoácidos con al menos el 90 % de identidad de secuencia con SEQ ID NOs:2 o 4. En otras realizaciones, los polipéptidos comprenden una secuencia de aminoácidos con al menos el 90 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NOs:2 o 4 y son capaces de transportar xilosa en una célula de levadura. En ciertas de estas realizaciones, una célula de levadura que expresa en exceso el polipéptido consume una mayor cantidad de xilosa con respecto a glucosa que una célula de levadura idéntica que no expresa en exceso el polipéptido.

En la presente memoria se proporcionan en ciertas realizaciones polinucleótidos RPE, RKI, TKL y TAL aislados. En ciertas de estas realizaciones, los polinucleótidos codifican un polipéptido con la secuencia de aminoácidos de SEQ ID NOs:34 (RPE), 40 (RKI), 46 (TKL) o 52 (TAL), o un polipéptido que comprende una secuencia de aminoácidos con al menos el 80 % de identidad de secuencia con SEQ ID NOs:34, 40, 46 o 52. En ciertas realizaciones, los polinucleótidos comprenden la secuencia de nucleótidos de SEQ ID NOs:33, 39, 45 o 51, o una secuencia de nucleótidos con al menos el 80 % de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NOs:33, 39,

45 o 51.

En la presente memoria se proporcionan en ciertas realizaciones polipéptidos RPE, RKI, TKL y TAL aislados. En ciertas de estas realizaciones, los polipéptidos comprenden la secuencia de aminoácidos de SEQ ID NOs:34 (RPE), 40 (RKI), 46 (TKL) o 52 (TAL), o una secuencia de aminoácidos con al menos el 80 % de identidad de secuencia con SEQ ID NOs:34, 40, 46 o 52.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que expresan en exceso uno o más genes transportadores de xilosa. En ciertas de estas realizaciones, el gen transportador de xilosa expresado en exceso es un gen KHT105 o RAG4 que codifica un polipéptido con al menos el 90 % de identidad de secuencia con SEQ ID NO:2 o SEQ ID NO:4, respectivamente. En ciertas realizaciones, las células de levadura genéticamente modificadas pertenecen al clado de *I. orientalis*/*P. fermentans*, y en ciertas de estas realizaciones las células son *I. orientalis*.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que comprenden una vía de arabinosa activa para convertir arabinosa en xilulosa 5-fosfato, en donde las células comprenden uno o más genes exógenos de la vía de fermentación de arabinosa seleccionados del grupo que consiste en genes AI, RK y RE. En ciertas realizaciones, las células de levadura genéticamente modificadas pertenecen al clado de *I. orientalis*/*P. fermentans*, y en ciertas de estas realizaciones las células son *I. orientalis*.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que expresan en exceso uno o más genes transportadores de xilosa y comprenden una vía de arabinosa activa para convertir arabinosa en xilulosa 5-fosfato, en donde las células comprenden uno o más genes exógenos de la vía de fermentación de arabinosa seleccionados del grupo que consiste en genes AI, RK y RE. En ciertas realizaciones, las células de levadura genéticamente modificadas pertenecen al clado de *I. orientalis*/*P. fermentans*, y en ciertas de estas realizaciones las células son *I. orientalis*.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que expresan en exceso uno o más genes transportadores de xilosa y comprenden una vía de fermentación de xilosa activa para convertir xilosa en xilulosa 5-fosfato, en donde las células comprenden uno o más genes exógenos de la vía de fermentación de xilosa seleccionados del grupo que consiste en genes XR, XDH y XK. En ciertas realizaciones, las células de levadura genéticamente modificadas pertenecen al clado de *I. orientalis*/*P. fermentans*, y en ciertas de estas realizaciones las células son *I. orientalis*.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que comprenden una vía de arabinosa activa para convertir arabinosa en xilulosa 5-fosfato y una vía de fermentación de xilosa activa para convertir xilosa en xilulosa 5-fosfato, en donde las células comprenden uno o más genes exógenos de la vía de fermentación de arabinosa seleccionados del grupo que consiste en genes AI, RK y RE y uno o más genes exógenos de la vía de fermentación de xilosa seleccionados del grupo que consiste en genes XR, XDH y XK. En ciertas realizaciones, las células de levadura genéticamente modificadas pertenecen al clado de *I. orientalis*/*P. fermentans*, y en ciertas de estas realizaciones las células son *I. orientalis*.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que expresan en exceso uno o más genes transportadores de xilosa y comprenden una vía de arabinosa activa para convertir arabinosa en xilulosa 5-fosfato y una vía de fermentación de xilosa activa para convertir xilosa en xilulosa 5-fosfato, en donde las células comprenden uno o más genes exógenos de la vía de fermentación de arabinosa seleccionados del grupo que consiste en genes AI, RK y RE y uno o más genes exógenos de la vía de fermentación de xilosa seleccionados del grupo que consiste en genes XR, XDH y XK. En ciertas realizaciones, las células de levadura genéticamente modificadas pertenecen al clado de *I. orientalis*/*P. fermentans*, y en ciertas de estas realizaciones las células son *I. orientalis*.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que expresan en exceso uno o más genes transportadores de xilosa y comprenden una vía de fermentación de xilosa activa para convertir xilosa en xilulosa 5-fosfato, en donde las células comprenden uno o más genes exógenos de la vía de fermentación de xilosa seleccionados del grupo que consiste en genes XI y XK. En ciertas realizaciones, las células de levadura genéticamente modificadas pertenecen al clado de *I. orientalis*/*P. fermentans*, y en ciertas de estas realizaciones las células son *I. orientalis*.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que comprenden una vía de arabinosa activa para convertir arabinosa en xilulosa 5-fosfato y una vía de fermentación de xilosa activa para convertir xilosa en xilulosa 5-fosfato, en donde las células comprenden uno o más genes exógenos de la vía de fermentación de arabinosa seleccionados del grupo que consiste en genes AI, RK y RE y uno o más genes exógenos de la vía de fermentación de xilosa seleccionados del grupo que consiste en genes XI y XK. En ciertas realizaciones, las células de levadura genéticamente modificadas pertenecen al clado de *I. orientalis*/*P. fermentans*, y en ciertas de estas realizaciones las células son *I. orientalis*.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que expresan en exceso uno o más genes transportadores de xilosa y comprenden una vía de arabinosa activa para

activa para convertir xilosa en xilulosa 5-fosfato, y comprenden una vía de pentosa fosfato no oxidativa activa para convertir xilulosa 5-fosfato más ribosa 5-fosfato en F6P y G3P, en donde las células comprenden uno o más genes exógenos de la vía de fermentación de arabinosa seleccionados del grupo que consiste en genes AI, RK o RE, uno o más genes exógenos de la vía de fermentación de xilosa seleccionados del grupo que consiste en genes XI y XK, y uno o más genes exógenos de la vía de pentosa fosfato no oxidativa seleccionados del grupo que consiste en genes TKL y TAL. En ciertas realizaciones, las células comprenden además uno o más genes exógenos de la vía de pentosa fosfato no oxidativa seleccionados del grupo que consiste en genes RPE y RKI. En ciertas realizaciones, las células de levadura genéticamente modificadas pertenecen al clado de *I. orientalis*/*P. fermentans*, y en ciertas de estas realizaciones las células son *I. orientalis*.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que comprenden una vía de arabinosa activa para convertir arabinosa en xilulosa 5-fosfato, una vía de fermentación de xilosa activa para convertir xilosa en xilulosa 5-fosfato, y una vía de pentosa fosfato no oxidativa activa para convertir xilulosa 5-fosfato más ribosa 5-fosfato en F6P y G3P, en donde las células comprenden uno o más genes exógenos de la vía de fermentación de arabinosa seleccionados del grupo que consiste en genes AI, RK o RE, uno o más genes exógenos de la vía de fermentación de xilosa seleccionados del grupo que consiste en genes XI y XK, y uno o más genes exógenos de la vía de pentosa fosfato no oxidativa seleccionados del grupo que consiste en genes TKL y TAL. En ciertas realizaciones, las células comprenden además uno o más genes exógenos de la vía de pentosa fosfato no oxidativa seleccionados del grupo que consiste en genes RPE y RKI. En ciertas realizaciones, las células de levadura genéticamente modificadas pertenecen al clado de *I. orientalis*/*P. fermentans*, y en ciertas de estas realizaciones las células son *I. orientalis*.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que expresan en exceso uno o más genes transportadores de xilosa, comprenden una vía de arabinosa activa para convertir arabinosa en xilulosa 5-fosfato, comprenden una vía de fermentación de xilosa activa para convertir xilosa en xilulosa 5-fosfato, y comprenden una vía de pentosa fosfato no oxidativa activa para convertir xilulosa 5-fosfato más ribosa 5-fosfato en F6P y G3P, en donde las células comprenden además una delección o alteración de uno o más genes exógenos de la vía de fermentación de arabinosa seleccionados del grupo que consiste en genes AI, RK o RE, uno o más genes exógenos de la vía de fermentación de xilosa seleccionados del grupo que consiste en genes XI y XK, y uno o más genes exógenos de la vía de pentosa fosfato no oxidativa seleccionados del grupo que consiste en genes TKL y TAL. En ciertas realizaciones, las células comprenden además uno o más genes exógenos de la vía de pentosa fosfato no oxidativa seleccionados del grupo que consiste en genes RPE y RKI. En ciertas realizaciones, las células de levadura genéticamente modificadas pertenecen al clado de *I. orientalis*/*P. fermentans*, y en ciertas de estas realizaciones las células son *I. orientalis*.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que expresan en exceso uno o más genes transportadores de xilosa, comprenden una vía de arabinosa activa para convertir arabinosa en xilulosa 5-fosfato, comprenden una vía de fermentación de xilosa activa para convertir xilosa en xilulosa 5-fosfato, y/o comprenden una vía de pentosa fosfato no oxidativa activa para convertir xilulosa 5-fosfato más ribosa 5-fosfato en F6P y G3P, en donde las células comprenden además una delección o alteración de uno o más genes que codifican enzimas involucradas en una vía de fermentación de xilosa activa que convierte xilosa en xilulosa 5-fosfato mediante productos intermedios de xilitol y de D-xilulosa. En ciertas realizaciones, las células comprenden una delección o alteración de uno o más de los genes AR/XR, arabitol 4-deshidrogenasa, xilulosa reductasa o XDH. En ciertas realizaciones, el gen AR/XR delecionado o alterado codifica un polipéptido con al menos el 50 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NOs:64, 66, 68, 69 o 71, y/o comprende una secuencia de nucleótidos con al menos el 50 % de identidad de secuencia con la región codificante de los nucleótidos secuencia de SEQ ID NOs:63, 65, 67 o 70. En ciertas realizaciones, el gen de xilulosa reductasa delecionado o alterado codifica un polipéptido con al menos el 50 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO:58 y/o comprende una secuencia de nucleótidos con al menos el 50 % de identidad de secuencia con la región codificante de la secuencia de nucleótidos de SEQ ID NO:57. En ciertas realizaciones, el gen XDH delecionado o alterado codifica un polipéptido con al menos el 50 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NOs:60 o 62 y/o comprende una secuencia de nucleótidos con al menos el 50 % de identidad de secuencia con la región codificante de la secuencia de nucleótidos expuesta en SEQ ID NOs:59 o 61.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que expresan en exceso uno o más genes transportadores de xilosa, comprenden una vía de arabinosa activa para convertir arabinosa en xilulosa 5-fosfato, comprenden una vía de fermentación de xilosa activa para convertir xilosa en xilulosa 5-fosfato, y/o comprenden una vía de pentosa fosfato no oxidativa activa para convertir xilulosa 5-fosfato más ribosa 5-fosfato en F6P y G3P, en donde las células comprenden además una delección o alteración de uno o más genes ALD o ADH. En ciertas realizaciones, el gen ALD delecionado o alterado codifica un polipéptido con al menos el 50 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO:73 y/o comprende una secuencia de nucleótidos con al menos el 50 % de identidad de secuencia con la región codificante de la secuencia de nucleótidos de SEQ ID NO:72. En ciertas realizaciones, el gen ADH delecionado o alterado codifica un polipéptido con al menos el 50 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NOs:75 o 85 y/o comprende una secuencia de nucleótidos con al menos el 50 % de identidad de secuencia con la región codificante de la secuencia de nucleótidos expuesta en SEQ ID NOs:74 o 84.

En la presente memoria se proporcionan en ciertas realizaciones métodos de fermentación que utilizan una o más de las células de levadura genéticamente modificadas proporcionadas en la presente memoria. En ciertas realizaciones, el medio de fermentación contiene xilosa. En ciertas de estas realizaciones, el medio contiene al menos 10 g/l de xilosa de un hidrolizado de biomasa vegetal, y en ciertas realizaciones la xilosa es el azúcar más abundante en el medio.

En la presente memoria se proporcionan en ciertas realizaciones métodos de producción de etanol usando una o más de las células de levadura genéticamente modificadas proporcionadas en la presente memoria. En ciertas realizaciones, las células se cultivan en un medio que contiene xilosa. En ciertas de estas realizaciones, el medio contiene al menos 10 g/l de xilosa de un hidrolizado de biomasa vegetal, y en ciertas realizaciones la xilosa es el azúcar más abundante en el medio.

Breve descripción de dibujos

Figura 1: Vías de levadura para el metabolismo de xilosa y arabinosa.

Figura 2: Fermentación de L-arabinosa a etanol por las cepas 1822, JY30.1 y JY30.2.

Figura 3: Fermentación de arabinosa a etanol por las cepas 3922, 3936, 3937 y 3408 en medio YP con 20 g/l de arabinosa.

Figura 4: Fermentación de xilosa a etanol por las cepas 3922, 3936, 3937 y 3408 en medio YP con 10 g/l de dextrosa, 40 g/l de xilosa y 10 g/l de arabinosa.

Figura 5: Fermentación de xilosa a etanol por las cepas 3922, 3936, 3937 y 3408 en medio YP con 20 g/l de dextrosa, 80 g/l de xilosa y 10 g/l de arabinosa.

Figura 6: Fermentación de arabinosa por las cepas 3922, 3936, 3937 y 3408 en medio YP con 10 g/l de dextrosa, 40 g/l de xilosa y 10 g/l de arabinosa.

Figura 7: Fermentación de xilosa a etanol por las cepa yJY19 en medio YP con 20 g/l de glucosa y 55 g/l de xilosa.

Figura 8: Fermentación de xilosa a etanol por las cepa yJY20 en medio YP con 20 g/l de glucosa y 55 g/l de xilosa.

Figura 9: Fermentación de xilosa a etanol por las cepa yJLJ77 en medio YP con 20 g/l de glucosa y 55 g/l de xilosa.

Figura 10: Fermentación de xilosa por las cepas 2973, 3097 y yJY28.

Figura 11: Fermentación de xilosa por las cepas 2973, 3097 y yJY28.

Figura 12: La producción de etanol por las cepas 2973, 3097 y yJY28.

Figura 13: Fermentación de xilosa y dextrosa a etanol por las cepas 3415 y 3416 en 30 % de medio DMDX de CSH.

Figura 14: Fermentación de arabinosa a etanol por las cepas 3408 y 3812 en medio YP con 40 g/l de arabinosa y 10 g/l de dextrosa.

Figura 15: Fermentación de xilosa a etanol por las cepas 3937 y 12038 en medio YP con 20 g/l de dextrosa, 80 g/l de xilosa y 10 g/l de arabinosa.

Figura 16: Fermentación de arabinosa por las cepas 3937 y 12038 en medio YP con 20 g/l de dextrosa, 80 g/l de xilosa y 10 g/l de arabinosa.

Figura 17: Fermentación de xilosa a etanol por las cepas 3937 y 12038 en medio YP con 10 g/l de dextrosa, 40 g/l de xilosa y 10 g/l de arabinosa.

Figura 18: Fermentación de arabinosa por las cepas 3937 y 12038 en medio YP con 10 g/l de dextrosa, 40 g/l de xilosa y 10 g/l de arabinosa.

Figura 19: Fermentación de xilosa a etanol por las cepas 3489 y 3861 en medio YP con 20 g/l de dextrosa y 80 g/l de xilosa.

Figura 20: Producción de acetato por las cepas 3489 y 3861 en medio YP con 20 g/l de dextrosa y 80 g/l de xilosa.

Figura 21: Fermentación de xilosa a etanol por las cepas 3489 y 3861 en medio YP con 20 g/l de dextrosa y 80 g/l de xilosa.

Figura 22: Producción de acetato por las cepas 3489 y 3861 en medio YP con 20 g/l de dextrosa y 80 g/l de xilosa.

Figura 23: Fermentación de xilosa a etanol por las cepas 3489 y 3861 en medio YP con 20 g/l de dextrosa y 80 g/l

de xilosa.

Figura 24: Fermentación de xilosa a etanol por las cepas 4084 y 4085 en 50 % de medio CSH.

Figura 25: Fermentación de xilosa a etanol por las cepas 4083, 4085 y 4086 en 50 % de medio CSH.

5 Figura 26: La producción de etanol por las cepas 12053, 12124 y 12125 en 50 % de medio CSH con tanto 20 g/l de dextrosa y 80 g/l de xilosa como 70 g/l de dextrosa y 40 g/l de xilosa.

Figura 27: Fermentación de xilosa por las cepas 12053, 12124 y 12125 en 50 % de medio CSH con tanto 20 g/l de dextrosa y 80 g/l de xilosa como 70 g/l de dextrosa y 40 g/l de xilosa.

Figura 28: Consumo de arabinosa por las cepas 12038, yACN168, yACN170 y yACN172 en medio DM con 20 g/l de dextrosa, 35 g/l de xilosa y 35 g/l de arabinosa.

10 Figura 29: Consumo de arabinosa por las cepas 12038, yACN174, yACN176 y yACN178 en medio DM con 20 g/l de dextrosa, 35 g/l de xilosa y 35 g/l de arabinosa.

Figura 30: Consumo de xilosa por las cepas 3937, 12215, y 12216 en medio YP con 20 g/l de dextrosa, 80 g/l de xilosa y 10 g/l de arabinosa.

15 Figura 31: Producción de etanol por las cepas 3937, 12215, y 12216 en medio YP con 20 g/l de dextrosa, 80 g/l de xilosa y 10 g/l de arabinosa.

Figura 32: Consumo de arabinosa por las cepas 3937, 12215 y 12216 en medio YP con 50 g/l de arabinosa.

Figura 33: Fermentación de glucosa y xilosa a etanol por las cepas 3118, 3082 y 3862 en medio YP con 20 g/l de glucosa y 80 g/l de xilosa.

20 Figura 34: Fermentación de glucosa y xilosa a etanol por las cepas 3083 y 3352 en medio YP con 20 g/l de dextrosa y 80 g/l de xilosa.

Figura 35: Fermentación de glucosa y xilosa a etanol por las cepas 3356 y 12293 en medio YP con 20 g/l de dextrosa, 80 g/l de xilosa y 10 g/l de arabinosa.

Descripción detallada

25 La siguiente descripción de la invención se prevé simplemente para ilustrar diversas realizaciones de la invención. Como tales, las modificaciones específicas tratadas no deben interpretarse como limitaciones al alcance de la invención. Será evidente para un experto en la materia que pueden hacerse diversos equivalentes, cambios y modificaciones sin apartarse del alcance de la invención reivindicada, y se entiende que tales realizaciones equivalentes deben incluirse en la presente memoria.

30 A menos que se indique lo contrario, todos los números que expresan concentraciones de componentes, condiciones de fermentación, realización de la fermentación, etc., usados en la memoria descriptiva deben entenderse como que están modificados en todos los casos por el término "aproximadamente". Por consiguiente, a menos que se indique lo contrario, los parámetros numéricos expuestos en la memoria descriptiva son aproximaciones que pueden variar dependiendo al menos de la técnica analítica específica. Cualquier valor numérico contiene inherentemente ciertos errores resultantes necesariamente de la desviación estándar encontrada en sus mediciones de prueba respectivas.

Abreviaturas:

ADH, alcohol deshidrogenasa; Al/araA, arabinosa isomerasa; ALD, aldehído deshidrogenasa; AR, aldosa reductasa; CYB2, L-(+)-lactato-citocromo c oxidoreductasa; CYC, iso-2-citocromo c; DHAP, dihidroxiacetona P; ENO1, enolasa 1; E4P, eritrosa 4-fosfato; F6P, fructosa 6-fosfato; GAL6, cisteína aminopeptidasa; GAPDH, gliceraldehído-3-fosfato deshidrogenasa 3; G3P, gliceraldehído 3-fosfato; G3PDH, glicerol-3-fosfato deshidrogenasa; PDC1, piruvato descarboxilasa 1; PGK, fosfoglicerato cinasa; PPP, vía de pentosa fosfato; RE/araD, ribulosa 5-fosfato 4-epimerasa; RK/araB, ribulocinasa; RKL, ribosa 5-fosfato cetol-isomerasa; RPE, ribulosa 5-fosfato 3-epimerasa; S7P, sedoheptulosa 7-fosfato; TAL, transaldolasa; TDH3, gliceraldehído-3-fosfato deshidrogenasa; TEF1, factor-1 de elongación de la traducción; TEF2, factor-2 de elongación de la traducción; TKL, transcetolasa; TPI, triosafosfato isomerasa; URA3, orotidina 5'-fosfato descarboxilasa; XDH, xilitol deshidrogenasa; XI, xilosa isomerasa; XK, xilulocinasa; XR, xilosa reductasa.

En la presente memoria se proporcionan células de levadura genéticamente modificadas para la producción de etanol, métodos de preparación de estas células de levadura, y métodos de uso de estas células para producir etanol.

La especie de levadura ideal para la producción de etanol a escala industrial a partir de biomasa debe presentar resistencia a ambientes de pH bajo, la capacidad de fermentar tanto azúcares de hexosa como de pentosa a etanol, y resistencia a compuestos inhibidores presentes en hidrolizado de materia de planta y que surgen de la fermentación, que incluyen acetato, HMF, furfural, fenólicos, aldehídos, cetonas, y el propio etanol.

5 *Saccharomyces cerevisiae* y la mayoría de las otras especies de levadura son capaces de fermentar azúcares de hexosa a etanol. Sin embargo, la mayoría de las especies de levadura son incapaces de fermentar azúcares de pentosa tales como arabinosa y xilosa. Aquellas especies de levadura que son capaces de metabolizar azúcares de pentosa lo hacen mediante una vía compleja. Las vías convencionales de la levadura para el metabolismo de la xilosa y la arabinosa (los dos azúcares de pentosa más comunes en biomasa celulósica) utilizan un producto intermedio de xilitol. La D-xilosa se reduce a xilitol por xilosa reductasa (XR). La arabinosa se convierte en xilitol mediante un proceso de tres etapas. La L-arabinosa se reduce a L-arabitol por aldosa reductasa (AR), el L-arabitol se convierte en L-xilulosa por L-arabitol 4-deshidrogenasa, y la L-xilulosa se convierte en xilitol por L-xilulosa reductasa. En ambas vías, el xilitol se oxida a D-xilulosa por xilitol deshidrogenasa (XDH), y la D-xilulosa se fosforila por xilulocinasa (XK) para producir D-xilulosa 5-P. La D-xilulosa 5-P resultante entra en la vía de pentosa fosfato (PPP), que genera fructosa 6-fosfato (F6P) y gliceraldehído 3-fosfato (G3P), ambos de los cuales entran en el ciclo glicolítico. Esta vía se ilustra en la Figura 1. El piruvato que surge de la glicólisis se convierte en acetaldehído y CO₂ por piruvato descarboxilasa, y el acetaldehído se reduce a etanol por la alcohol deshidrogenasa (ADH).

Como las reductasas de la vía de utilización de arabinosa fúngica utilizan NADPH como reductor y las deshidrogenasas son específicas para NAD⁺, un desequilibrio del cofactor produce un crecimiento anaeróbico lento sobre la L-arabinosa y bajos niveles de producción de etanol aún cuando el proceso es redox neutro.

A diferencia de la levadura, las bacterias no utilizan un producto intermedio de xilitol cuando se metaboliza arabinosa. En bacterias, la L-arabinosa se convierte en L-ribulosa por L-arabinosa isomerasa (AI). La L-ribulosa se convierte en L-ribulosa 5-fosfato por L-ribulocinasa (RK), que entonces se convierte en D-xilulosa 5-fosfato por L-ribulosa-fosfato 4-epimerasa (RE). Ninguna de estas etapas enzimáticas requiere un cofactor de NAD o NADH, que significa que la vía de arabinosa bacteriana no tiene problemas que compliquen el desequilibrio de factores. Se han hecho intentos previos para utilizar la vía de arabinosa bacteriana en levadura. Se incorporaron genes AI, RK y RE de fuentes bacterianas en *S. cerevisiae*, y la cepa de levadura genéticamente modificada resultante presentó la capacidad de fermentar arabinosa a etanol (Becker y Boles Appl. Environ Microbiol 69:4144 (2003)). Sin embargo, *S. cerevisiae* tiene tolerancia limitada al acetato libre y otros inhibidores comunes en los hidrolizados.

Se han hecho intentos previos para generar especies de levadura adicionales que son capaces de fermentar azúcares de pentosa y tolerantes a inhibidores del hidrolizado. Se generó una cepa de *I. orientalis* que contenía una inactivación de los supuestos genes de ADH ADHa y ADHb y también expresó en exceso un supuesto gen ADH1. La cepa de levadura resultante mostró una elevada capacidad para fermentar xilosa a etanol. Sin embargo, fue incapaz de fermentar arabinosa.

Como se ha descrito en la presente memoria, los genes AI (araA), RK (araB) y RE (araD) de la vía de arabinosa bacteriana de *Bacteroides thetaiotaomicron*, *Escherichia coli*, *Lactobacillus plantarum* y *Bacillus licheniformis* se incorporaron en una cepa de *Issatchenkia orientalis* en diversas combinaciones (Ejemplo 1). Los genes bacterianos fueron normalmente, pero no siempre, de codón optimizado para *I. orientalis*. Cada una de las cepas resultantes presentó actividad de AI, RK y/o RE apropiada (Ejemplo 2). Se probaron varias cepas que contienen un conjunto completo de genes de la vía de arabinosa bacteriana (es decir, al menos una copia de cada uno de los genes AI, RK y RE) para su capacidad para fermentar arabinosa. Estas cepas presentaron tanto consumo de arabinosa como producción de etanol a partir de arabinosa (Ejemplo 3). Los resultados descritos en la presente memoria confirman que los genes de la vía de arabinosa bacteriana son activos cuando se expresan en *I. orientalis*.

Se incorporó un conjunto completo de genes de la vía de arabinosa de *B. thetaiotaomicron* en una cepa de *I. orientalis* que previamente había sido manipulada para fermentar xilosa a etanol con el fin de crear una cepa de vía doble capaz de fermentar tanto xilosa como arabinosa (Ejemplo 4). Las cepas de vía doble resultantes presentaron la capacidad de fermentar tanto arabinosa como xilosa a etanol, y ambas produjeron más etanol que las cepas de control que contienen solo xilosa o solo genes de la vía de arabinosa (Ejemplo 5). Sin embargo, la utilización de xilosa disminuyó en las cepas de la vía doble frente a la cepa de solo xilosa, incluso en medios que carecían de arabinosa. Además, el consumo de arabinosa no empezó hasta que tanto la dextrosa como la xilosa estuvieron prácticamente agotadas. Se generaron cepas de *I. orientalis* adicionales que contenían genes araB de *B. thetaiotaomicron* y *L. citreum* no de codón optimizado (Ejemplo 6). Estas cepas presentaron utilización de xilosa y producción de etanol mejoradas frente a una cepa que contiene el gen de *B. thetaiotaomicron* de codón optimizado.

Como se ha descrito en la presente memoria, se cribó el genoma de *K. marxianus* para identificar posibles transportadores de azúcar (Ejemplo 7). Se caracterizaron dos posibles genes transportadores de azúcar de *K. marxianus*, Kit105 y RAG4. Ambos genes se integraron en cepas de *I. orientalis* que habían sido previamente manipuladas para contener una vía de xilosa básica (XI, XK) con el fin de evaluar el efecto de la posible expresión del transportador en la utilización de xilosa (Ejemplo 8). Las cepas resultantes presentaron elevado co-consumo de glucosa y xilosa, así se integró una segunda copia de cada gen transportador en las células. Las células que contienen dos copias del gen KHT105 presentaron mayor utilización de xilosa y producción de etanol que la cepa

parental o cepas que contienen dos copias del gen RAG4.

Los efectos de la expresión de KHT105 se probaron adicionalmente integrando dos copias del gen en una cepa de *I. orientalis* que contenía ingeniería de xilosa más avanzada, que incluye expresión en exceso de los genes de la vía de pentosa no oxidativa transaldolasa (TAL), ribosa 5-fosfato cetol-isomerasa (RKI) y ribulosa 5-fosfato 3-epimerasa (RPE) (Ejemplo 9). En fermentadores con medio de hidrolizado, la cepa que expresa KHT105 presentó un aumento del 80 % del consumo de xilosa y la producción de etanol frente a una cepa de control.

Para evaluar el efecto de la expresión de KHT105 sobre el consumo de arabinosa, se integró una única copia del gen en el locus S141G4546 de una cepa de *I. orientalis* que contenía los genes de la vía de arabinosa (Ejemplo 10). S141G4546 es un homólogo de la butanodiol deshidrogenasa y la xilitol deshidrogenasa. La cepa resultante presentó un ligero aumento en el consumo de arabinosa y la producción de etanol frente a una cepa principal.

Basándose en los datos que muestran que el transportador KHT105 aumentó tanto el consumo de xilosa como de arabinosa, se integraron dos copias del gen KHT105 en el locus S141G4546 de las cepas de vía doble de *I. orientalis* descritas anteriormente (Ejemplo 11). Las cepas que contienen el transportador KHT105 presentaron mayor producción de etanol y consumo de xilosa y de arabinosa que la cepa principal (Ejemplo 12). Los beneficios de la expresión de KHT105 fueron particularmente evidentes en medio que contenía niveles más altos de azúcar.

Para evaluar métodos adicionales para mejorar la producción de etanol en *I. orientalis*, se desarrolló una cepa de inactivación de aldehído deshidrogenasa (ALD). *I. orientalis* tiene tres homólogos principales a los genes ALD4, ALD5 y ALD6 de *S. cerevisiae*: S141G5680 ("ALD5680"), S141G9161 ("ALD9161") y S141G6502 ("ALD6502"). Las inactivaciones se dirigieron a ALD5680, que presenta elevada expresión cuando las células se cultivan sobre xilosa. Ambas copias de ALD5680 se inactivaron en una cepa de *I. orientalis* que había sido previamente manipulada para fermentar xilosa a etanol (Ejemplo 13). La cepa inactivada en ALD5680 presentó consumo de xilosa y producción de etanol elevados y producción de acetato reducida bajo ciertas condiciones, pero los resultados fueron parcialmente dependientes de las condiciones de fermentación precisas usadas (Ejemplo 14).

Se integraron copias adicionales del gen KHT105 de *K. marxianus* en una cepa de *I. orientalis* que había sido previamente manipulada para contener dos copias de KHT105 en el locus S141G9091 (homólogo de ADH) (Ejemplo 15). Se integraron copias adicionales de KHT105 en los loci S141G4546 o ALD5680, y se evaluó el efecto del elevado número de copias de KHT105 y la inactivación de S141G4546/ALD5680 sobre el consumo de azúcar y la producción de etanol en el medio de hidrolizado. Entre ambas cepas inactivadas en S141G4546 y ALD5680, la presencia de una cuarta copia del gen KHT105 aumentó el consumo de xilosa y la producción de etanol frente a cepas que solo contenían tres copias del gen, presentando las cepas inactivadas en ALD resultados ligeramente mejores que las cepas inactivadas en S141G4546.

Se evaluaron a continuación los efectos de la expresión en exceso de KHT105 y/o la inactivación de ALD5680 en una cepa de *I. orientalis* resistente a etanol. La expresión en exceso de KHT105 produjo un aumento significativo en la producción de etanol y el consumo de xilosa en medio definido bajo en dextrosa, pero solo tuvo un ligero efecto en medio alto en dextrosa (Ejemplo 16).

Se integraron genes *araD* de *Bifidobacterium animalis* y *Lactococcus lactis* (Ejemplo 17) y genes *araA* de *Lactobacillus sakei* y *B. thetaiotaomicron* alternativos (Ejemplo 18) en cepas de *I. orientalis* de la vía doble que expresan en exceso KHT105 para evaluar su efecto sobre la fermentación de arabinosa. Estas cepas presentaron un elevado consumo de arabinosa frente a las cepas principales.

Como se ha descrito en la presente memoria, se identificaron secuencias novedosas de los genes TAL, RKI, TKL y RPE de *I. orientalis*. Se integraron copias exógenas de estos genes en *I. orientalis* para evaluar el efecto de su expresión en exceso sobre el consumo de xilosa y la producción de etanol (Ejemplos 19-21). Las cepas resultantes presentaron utilización de xilosa y producción de etanol elevadas frente a cepas principales.

En la presente memoria se proporcionan en ciertas realizaciones se aíslan polinucleótidos transportadores KHT105 y RAG4. En ciertas realizaciones, estos polinucleótidos aislados comprenden una región codificante que codifica un polipéptido que tiene la secuencia de aminoácidos expuesta en SEQ ID NOs:2 o 4, respectivamente. En ciertas de estas realizaciones, los polinucleótidos comprenden la región codificante de la secuencia de nucleótidos expuesta en SEQ ID NOs:1 o 3. En otras realizaciones, los polinucleótidos comprenden una secuencia de nucleótidos con al menos el 90 % de identidad de secuencia con la región codificante de la secuencia de nucleótidos expuesta en SEQ ID NOs:1 o 3. En ciertas de estas realizaciones, los polinucleótidos comprenden una secuencia de nucleótidos que tiene al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, al menos el 99 %, o al menos el 99,5 % de identidad de secuencia con la región codificante de la secuencia de nucleótidos expuesta en SEQ ID NOs:1 o 3.

En ciertas realizaciones, los polinucleótidos KHT105 y RAG4 aislados proporcionados en la presente memoria comprenden una región codificante que codifica un polipéptido que comprende una secuencia de aminoácidos con al menos el 90 % de identidad de secuencia con la secuencia de aminoácidos expuesta en SEQ ID NOs:2 o 4, respectivamente. En ciertas de estas realizaciones, el polipéptido codificado comprende una secuencia de aminoácidos con al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, al menos el 99 %, o al

menos el 99,5 % de identidad de secuencia con la secuencia de aminoácidos expuesta en SEQ ID NOs:2 o 4. En ciertas realizaciones, los polinucleótidos aislados comprenden una secuencia de nucleótidos con al menos el 90 % de identidad de secuencia con la región codificante de la secuencia de nucleótidos expuesta en SEQ ID NOs:1 o 3. En ciertas de estas realizaciones, los polinucleótidos aislados comprenden una secuencia de nucleótidos que tiene al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, al menos el 99 %, o al menos el 99,5 % de identidad de secuencia con la región codificante de la secuencia de nucleótidos expuesta en SEQ ID NOs:1 o 3.

En ciertas realizaciones, los polinucleótidos KHT105 y RAG4 aislados proporcionados en la presente memoria comprenden una región codificante que codifica un polipéptido con el 70 % o mayor identidad de secuencia con la secuencia de aminoácidos expuesta en SEQ ID NOs:2 o 4, en donde una célula de levadura que expresa en exceso el polinucleótido consume una mayor cantidad de xilosa con respecto a glucosa que una célula idéntica que no expresa en exceso el polinucleótido. Similarmente, en ciertas realizaciones, los polinucleótidos proporcionados en la presente memoria comprenden una región codificante que codifica un polipéptido con el 70 % o mayor identidad de secuencia con la secuencia de aminoácidos expuesta en SEQ ID NOs:2 o 4, en donde el polipéptido codificado es capaz de transportar xilosa en una célula de levadura. En ciertas de estas realizaciones, los polinucleótidos comprenden una región codificante que codifica un polipéptido con al menos el 70 %, al menos el 75 %, al menos el 80 %, al menos el 85 %, o al menos el 90 % de identidad de secuencia con la secuencia de aminoácidos expuesta en SEQ ID NOs:2 o 4. En ciertas de estas realizaciones, los polinucleótidos comprenden una secuencia de nucleótidos con al menos el 70 %, al menos el 75 %, al menos el 80 %, al menos el 85 %, o al menos el 90 % de identidad de secuencia con la región codificante de la secuencia de nucleótidos expuesta en SEQ ID NOs:1 o 3.

En la presente memoria se proporcionan en ciertas realizaciones polinucleótidos RPE, RKI, TKL y TAL de *I. orientalis* aislados. En ciertas realizaciones, estos polinucleótidos aislados comprenden una región codificante que codifica un polipéptido que tiene la secuencia de aminoácidos expuesta en SEQ ID NOs:34, 40, 46 o 52, respectivamente. En ciertas de estas realizaciones, los polinucleótidos comprenden la región codificante de la secuencia de nucleótidos expuesta en SEQ ID NOs:33, 39, 45 o 51. En otras realizaciones, los polinucleótidos comprenden una secuencia de nucleótidos con al menos el 80 % de identidad de secuencia con la región codificante de la secuencia de nucleótidos expuesta en SEQ ID NOs:33, 39, 45 o 51. En ciertas de estas realizaciones, los polinucleótidos comprenden una secuencia de nucleótidos que tiene al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, al menos el 99 %, o al menos el 99,5 % de identidad de secuencia con la región codificante de la secuencia de nucleótidos expuesta en SEQ ID NOs:33, 39, 45 o 51.

En ciertas realizaciones, los polinucleótidos de RKI, TKL y TAL *I. orientalis* aislados proporcionados en la presente memoria comprenden una región codificante que codifica un polipéptido que comprende una secuencia de aminoácidos con al menos el 80 % de identidad de secuencia con la secuencia de aminoácidos expuesta en SEQ ID NOs:34, 40, 46 o 52, respectivamente. En ciertas de estas realizaciones, el polipéptido codificado comprende una secuencia de aminoácidos con al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, al menos el 99 %, o al menos el 99,5 % de identidad de secuencia con la secuencia de aminoácidos expuesta en SEQ ID NOs:34, 40, 46 o 52. En ciertas realizaciones, los polinucleótidos aislados comprenden una secuencia de nucleótidos con al menos el 80 % de identidad de secuencia con la región codificante de la secuencia de nucleótidos expuesta en SEQ ID NOs:33, 39, 45 o 51. En ciertas de estas realizaciones, los polinucleótidos aislados comprenden una secuencia de nucleótidos que tiene al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, al menos el 99 %, o al menos el 99,5 % de identidad de secuencia con la región codificante de la secuencia de nucleótidos expuesta en SEQ ID NOs:33, 39, 45 o 51.

En la presente memoria se proporcionan en ciertas realizaciones construcciones que comprenden uno o más de los polinucleótidos aislados Kit105, RAG4, RKI de *I. orientalis*, TKL de *I. orientalis* y/o TAL de *I. orientalis* proporcionados en la presente memoria. El término “construcción”, como se emplea en la presente memoria, se refiere a una secuencia de ADN que se usa para transformar una célula. La construcción puede ser, por ejemplo, un plásmido o vector circular, una porción de un plásmido o vector circular (tal como un producto de digestión de enzima de restricción), un plásmido o vector linealizado, o un producto de PCR preparado usando un plásmido o ADN genómico como molde. Además del uno o más de los polinucleótidos proporcionados en la presente memoria, una construcción puede comprender uno o más elementos reguladores (p. ej., promotores, terminadores) operativamente unidos a la secuencia de polinucleótidos. Como se emplea en la presente memoria, el término “promotor” se refiere a una secuencia no traducida localizada aguas arriba (es decir, 5') con respecto al codón de iniciación de la traducción de un gen (generalmente dentro de aproximadamente 1 a 1000 pares de bases (pb), preferiblemente dentro de aproximadamente 1 a 500 pb) que controla el inicio de la transcripción del gen. El término “terminador”, como se emplea en la presente memoria, se refiere a una secuencia no traducida localizada aguas abajo (es decir, 3') con respecto al codón de terminación de la traducción de un gen (generalmente dentro de aproximadamente 1 a 1000 pb, preferiblemente dentro de aproximadamente 1 a 500 pb, y especialmente dentro de aproximadamente 1 a 100 pb) que controla el extremo de transcripción del gen. Un promotor o terminador está “operativamente unido” a un gen si su posición en el genoma con respecto a la del gen es tal que el promotor o terminador, según pueda ser el caso, realiza su función de control transcripcional. Promotores y terminadores adecuados se describen, por ejemplo, en los documentos WO99/14335, WO00/71738, WO02/42471, WO03/102201, WO03/102152 y WO03/049525. Una construcción puede comprender además uno o más componentes adicionales, que incluyen, por ejemplo, uno o más sitios de restricción y/o uno o más genes de marcadores de selección,

opcionalmente unidos a uno o más elementos reguladores. Un “gen de marcador de selección” es un gen que codifica una proteína necesaria para la supervivencia y/o crecimiento de la célula transformada en un medio de cultivo selectivo, y, por tanto, puede usarse para aplicar presión de selección a la célula.

En la presente memoria se proporcionan en ciertas realizaciones polipéptidos KHT105 y RAG4 aislados. En ciertas realizaciones, estos polipéptidos comprenden la secuencia de aminoácidos expuesta en SEQ ID NOs:2 o 4. En otras realizaciones, los polipéptidos comprenden una secuencia de aminoácidos con al menos el 90 % de identidad de secuencia con la secuencia de aminoácidos expuesta en SEQ ID NOs:2 o 4. En ciertas de estas realizaciones, los polipéptidos comprenden una secuencia de aminoácidos con al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, al menos el 99 %, o al menos el 99,5 % de identidad de secuencia con la secuencia de aminoácidos expuesta en SEQ ID NOs:2 o 4. En otras realizaciones más, los polipéptidos comprenden una secuencia de aminoácidos con al menos el 70 % de identidad de secuencia con la secuencia de aminoácidos expuesta en SEQ ID NOs:2 o 4 y son capaces de transportar xilosa en una célula de levadura. Similarmente, en ciertas realizaciones, los polipéptidos proporcionados en la presente memoria comprenden una secuencia de aminoácidos con al menos el 70 % de identidad de secuencia con la secuencia de aminoácidos expuesta en SEQ ID NOs:2 y 4, y una célula de levadura que expresa en exceso el polipéptido consume una mayor cantidad de xilosa con respecto a glucosa que una célula idéntica que no expresa en exceso el polipéptido. En ciertas de estas realizaciones, los polipéptidos comprenden una secuencia de aminoácidos con al menos el 75 %, al menos el 80 %, al menos el 85 %, o al menos el 90 % de identidad de secuencia con la secuencia de aminoácidos expuesta en SEQ ID NOs:2 o 4.

En la presente memoria se proporcionan en ciertas realizaciones polipéptidos RPE, RKI, TKL, y TAL de *I. orientalis* aislados. En ciertas realizaciones, estos polipéptidos comprenden la secuencia de aminoácidos expuesta en SEQ ID NOs:34, 40, 46 o 52, respectivamente. En otras realizaciones, los polipéptidos comprenden una secuencia de aminoácidos con al menos el 80 % de identidad de secuencia con la secuencia de aminoácidos expuesta en SEQ ID NOs:34, 40, 46 o 52. En ciertas de estas realizaciones, los polipéptidos comprenden una secuencia de aminoácidos con al menos el 80 %, al menos 85%, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, al menos el 99 %, o al menos el 99,5 % de identidad de secuencia con la secuencia de aminoácidos expuesta en SEQ ID NOs:34, 40, 46 o 52.

Los porcentajes de identidad de secuencia para las secuencias de nucleótidos o de aminoácidos pueden calcularse por métodos conocidos en la técnica, tales como, por ejemplo, usando el software BLAST (Basic Local Alignment Search Tool del National Center for Biological Information (NCBI)) versión 2.2.1 con parámetros por defecto. Se considera que las secuencias que tienen una puntuación de identidad de al menos el 90 %, usando el algoritmo BLAST versión 2.2.1 con parámetros por defecto, tienen al menos el 90 % de identidad de secuencia. El software BLAST está disponible del NCBI, Bethesda, Maryland.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que comprenden uno o más genes KHT105 y/o RAG4. En ciertas realizaciones, estos genes comprenden la secuencia de nucleótidos de los polinucleótidos KHT105 y/o RAG4 descritos en la presente memoria. Similarmente, en ciertas realizaciones, los genes transportadores de xilosa codifican un transportador de xilosa que comprende la secuencia de aminoácidos de los polipéptidos KHT105 y/o RAG4 descritos en la presente memoria. En ciertas realizaciones, las células genéticamente modificadas presentan un grado más alto de transporte de xilosa que las células no mutantes correspondientes.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que tienen al menos una vía de fermentación de arabinosa activa para convertir arabinosa en xilulosa 5-fosfato. Una célula de levadura que tiene una “vía de fermentación de arabinosa activa”, como se emplea en la presente memoria, produce enzimas activas necesarias para catalizar cada reacción en una vía de fermentación de arabinosa, y por tanto es capaz de convertir arabinosa en xilulosa 5-fosfato cuando se cultiva en condiciones de fermentación en presencia de arabinosa. Una célula de levadura que tiene una vía de fermentación de arabinosa activa comprende uno o más genes de la vía de fermentación de arabinosa. Un “gen de la vía de fermentación de arabinosa”, como se emplea en la presente memoria, se refiere a la región codificante de una secuencia de nucleótidos que codifica una enzima involucrada en una vía de fermentación de arabinosa activa. En ciertas realizaciones, las células de levadura proporcionadas en la presente memoria tienen una vía de fermentación de arabinosa activa que convierte arabinosa en xilulosa 5-fosfato sin pasar por un producto intermedio de arabitol, xilulosa, xilitol o xilulosa. En ciertas de estas realizaciones, las células de levadura tienen una vía de fermentación de arabinosa activa que convierte arabinosa en xilulosa 5-fosfato mediante productos intermedios de ribulosa y ribulosa 5-fosfato. En estas realizaciones, las células de levadura comprenden al menos una copia de cada uno de los genes de la vía de fermentación de arabinosa AI, RK y RE.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que tienen al menos una vía de fermentación de arabinosa activa para convertir arabinosa en xilulosa 5-fosfato, y que comprenden además uno o más genes transportadores de xilosa. En ciertas realizaciones, la vía de fermentación de arabinosa activa convierte arabinosa en xilulosa 5-fosfato mediante productos intermedios de ribulosa y de ribulosa 5-fosfato, y en estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de arabinosa AI, RK y RE. En ciertas realizaciones, los genes transportadores

de xilosa comprenden la secuencia de nucleótidos de uno o más de los polinucleótidos KHT105 y/o RAG4 descritos en la presente memoria. Similarmente, en ciertas realizaciones, los genes transportadores de xilosa codifican un transportador de xilosa que comprende la secuencia de aminoácidos de uno o más de los polipéptidos KHT105 y/o RAG4 descritos en la presente memoria.

- 5 En ciertas realizaciones, las células de levadura genéticamente modificadas proporcionadas en la presente memoria comprenden una vía de fermentación de xilosa activa para convertir xilosa en xilulosa 5-fosfato. Una célula de levadura que tiene una “vía de fermentación de xilosa activa”, como se emplea en la presente memoria, produce enzimas activas necesarias para catalizar cada reacción en una vía de fermentación de xilosa, y por tanto es capaz de convertir xilosa en xilulosa 5-fosfato cuando se cultiva en condiciones de fermentación en presencia de xilosa.
- 10 Una célula de levadura que tiene una vía de fermentación de xilosa activa comprende uno o más genes de la vía de fermentación de xilosa. Un “gen de la vía de fermentación de xilosa”, como se emplea en la presente memoria, se refiere a la región codificante de una secuencia de nucleótidos que codifica una enzima involucrada en una vía de fermentación de xilosa activa. En ciertas realizaciones, una vía de fermentación de xilosa activa convierte xilosa en xilulosa 5-fosfato mediante productos intermedios de xilitol y de xilulosa. En estas realizaciones, las células de levadura comprenden al menos una copia de cada uno de los genes de la vía de fermentación de xilosa XR, XDH y XK. En otras realizaciones, una vía de fermentación de xilosa activa convierte xilosa en xilulosa 5-fosfato mediante un producto intermedio de xilulosa solo. En estas realizaciones, las células de levadura comprenden al menos una copia de cada uno de los genes de la vía de fermentación de xilosa xilosa isomerasa (XI) y XK.

- 20 En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que tienen al menos una vía de fermentación de xilosa activa para convertir xilosa en xilulosa 5-fosfato, y que comprenden además uno o más genes transportadores de xilosa. En ciertas realizaciones, la vía de fermentación de xilosa activa convierte xilosa en xilulosa 5-fosfato mediante productos intermedios de xilitol y de xilulosa, y en ciertas de estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de xilosa XR, XDH y XK. En otras realizaciones, la vía de fermentación de xilosa activa convierte xilosa en xilulosa 5-fosfato mediante un producto intermedio de xilulosa solo, y en ciertas de estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de xilosa XI y XK. En ciertas realizaciones, los genes transportadores de xilosa comprenden la secuencia de nucleótidos de uno o más de los polinucleótidos kit105 y/o RAG4 descritos en la presente memoria. Similarmente, en ciertas realizaciones, los genes transportadores de xilosa codifican un transportador de xilosa que comprende la secuencia de aminoácidos de uno o más de los polipéptidos KHT105 y/o RAG4 descritos en la presente memoria.

- 35 En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que tienen al menos una vía de fermentación de xilosa activa para convertir xilosa en xilulosa 5-fosfato, y que comprenden además una vía de fermentación de arabinosa activa para convertir arabinosa en xilulosa 5-fosfato. En ciertas realizaciones, la vía de fermentación de xilosa activa convierte xilosa en xilulosa 5-fosfato mediante productos intermedios de xilitol y de xilulosa, y en ciertas de estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de xilosa XR, XDH y XK. En otras realizaciones, la vía de fermentación de xilosa activa convierte xilosa en xilulosa 5-fosfato mediante un producto intermedio de xilulosa solo, y en ciertas de estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de xilosa XI y XK. En ciertas realizaciones, la vía de fermentación de arabinosa activa convierte arabinosa en xilulosa 5-fosfato mediante productos intermedios de ribulosa y de ribulosa 5-fosfato, y en estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de arabinosa AI, RK y RE.

- 45 En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que tienen al menos una vía de fermentación de xilosa activa para convertir xilosa en xilulosa 5-fosfato, y que comprenden además uno o más genes transportadores de xilosa y una vía de fermentación de arabinosa activa para convertir arabinosa en xilulosa 5-fosfato. En ciertas realizaciones, la vía de fermentación de xilosa activa convierte xilosa en xilulosa 5-fosfato mediante productos intermedios de xilitol y de xilulosa, y en ciertas de estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de xilosa XR, XDH y XK. En otras realizaciones, la vía de fermentación de xilosa activa convierte xilosa en xilulosa 5-fosfato mediante un producto intermedio de xilulosa solo, y en ciertas de estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de xilosa XI y XK. En ciertas realizaciones, los genes transportadores de xilosa comprenden la secuencia de nucleótidos de uno o más de los polinucleótidos KHT105 y/o RAG4 descritos en la presente memoria. Similarmente, en ciertas realizaciones, los genes transportadores de xilosa codifican un transportador de xilosa que comprende la secuencia de aminoácidos de uno o más de los polipéptidos KHT105 y/o RAG4 descritos en la presente memoria. En ciertas realizaciones, la vía de fermentación de arabinosa activa convierte arabinosa en xilulosa 5-fosfato mediante productos intermedios de ribulosa y de ribulosa 5-fosfato, y en estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de arabinosa AI, RK y RE.

- 60 En ciertas realizaciones, las células de levadura genéticamente modificadas proporcionadas en la presente memoria comprenden una vía de pentosa fosfato no oxidativa activa. Una célula de levadura que tiene una “vía de pentosa fosfato no oxidativa activa”, como se emplea en la presente memoria, produce enzimas activas necesarias para convertir xilulosa 5-fosfato más ribosa 5-fosfato en F6P y G3P. Una célula de levadura que tiene una vía de pentosa

fosfato no oxidativa activa comprende uno o más genes de la vía de pentosa fosfato no oxidativa. Un “gen de la vía de pentosa fosfato no oxidativa”, como se emplea en la presente memoria, se refiere a la región codificante de una secuencia de nucleótidos que codifica una enzima involucrada en una vía de pentosa fosfato no oxidativa activa. En ciertas realizaciones, una célula de levadura que tiene una vía de pentosa fosfato no oxidativa activa comprende al menos una copia de cada uno de los genes de la vía de pentosa fosfato no oxidativa TKL y TAL. En ciertas de estas realizaciones, la célula de levadura comprende además unas o más copias de los genes de la vía de pentosa fosfato no oxidativa RPE y RKI. En ciertas realizaciones, una célula de levadura que tiene una vía de pentosa fosfato no oxidativa activa comprende al menos una copia de un gen RPE, RKI, TKL y/o TAL de *I. orientalis*, y en ciertas realizaciones estos genes comprenden la secuencia de ADN de los polinucleótidos RPE, RKI, TKL y/o TAL descritos en la presente memoria y/o codifican un polipéptido que comprende la secuencia de aminoácidos de los polipéptidos RPE, RKI, TKL y/o TAL descritos en la presente memoria.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que tienen al menos una vía de pentosa fosfato no oxidativa activa para convertir xilulosa 5-fosfato más ribosa 5-fosfato en F6P y G3P, y que comprenden además una vía de fermentación de arabinosa activa para convertir arabinosa en xilulosa 5-fosfato. En ciertas realizaciones, las células comprenden al menos una copia de los genes de la vía de pentosa fosfato no oxidativa TKL y TAL, y en ciertas realizaciones las células comprenden además al menos una copia de los genes de la vía de pentosa fosfato no oxidativa RPE y RKI. En ciertas realizaciones, los genes TKL, TAL, RPE y RKI comprenden la secuencia de nucleótidos de uno o más de los polinucleótidos TKL, TAL, RPE y/o RKI descritos en la presente memoria. Similarmente, en ciertas realizaciones, los genes TKL, TAL, RPE y RKI codifican polipéptidos que comprenden la secuencia de aminoácidos de uno o más de los polipéptidos TKL, TAL, RPE y/o RKI descritos en la presente memoria. En ciertas realizaciones, la vía de fermentación de arabinosa activa convierte arabinosa en xilulosa 5-fosfato mediante productos intermedios de ribulosa y de ribulosa 5-fosfato, y en estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de arabinosa AI, RK, y RE.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que tienen al menos una vía de pentosa fosfato no oxidativa activa para convertir xilulosa 5-fosfato más ribosa 5-fosfato en F6P y G3P, y que comprenden además uno o más genes transportadores de xilosa y una vía de fermentación de arabinosa activa para convertir arabinosa en xilulosa 5-fosfato. En ciertas realizaciones, las células comprenden al menos una copia de los genes de la vía de pentosa fosfato no oxidativa TKL y TAL, y en ciertas realizaciones las células comprenden además al menos una copia de los genes de la vía de pentosa fosfato no oxidativa RPE y RKI. En ciertas realizaciones, los genes TKL, TAL, RPE y RKI comprenden la secuencia de nucleótidos de uno o más de los polinucleótidos TKL, TAL, RPE y/o RKI descritos en la presente memoria. Similarmente, en ciertas realizaciones, los genes TKL, TAL, RPE y RKI codifican polipéptidos que comprenden la secuencia de aminoácidos de uno o más de los polipéptidos TKL, TAL, RPE y/o RKI descritos en la presente memoria. En ciertas realizaciones, los genes transportadores de xilosa comprenden la secuencia de nucleótidos de uno o más de los polinucleótidos KHT105 y/o RAG4 descritos en la presente memoria. Similarmente, en ciertas realizaciones, los genes transportadores de xilosa codifican un transportador de xilosa que comprende la secuencia de aminoácidos de uno o más de los polipéptidos KHT105 y/o RAG4 descritos en la presente memoria. En ciertas realizaciones, la vía de fermentación de arabinosa activa convierte arabinosa en xilulosa 5-fosfato mediante productos intermedios de ribulosa y de ribulosa 5-fosfato, y en estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de arabinosa AI, RK y RE.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que tienen al menos una vía de pentosa fosfato no oxidativa activa para convertir xilulosa 5-fosfato más ribosa 5-fosfato en F6P y G3P, y que comprenden además uno o más genes transportadores de xilosa y una vía de fermentación de xilosa activa para convertir xilosa en xilulosa 5-fosfato. En ciertas realizaciones, las células comprenden al menos una copia de los genes de la vía de pentosa fosfato no oxidativa TKL y TAL, y en ciertas realizaciones las células comprenden además al menos una copia de los genes de la vía de pentosa fosfato no oxidativa RPE y RKI. En ciertas realizaciones, los genes TKL, TAL, RPE y RKI comprenden la secuencia de nucleótidos de uno o más de los polinucleótidos TKL, TAL, RPE y/o RKI descritos en la presente memoria. Similarmente, en ciertas realizaciones, los genes TKL, TAL, RPE y RKI codifican polipéptidos que comprenden la secuencia de aminoácidos de uno o más de los polipéptidos TKL, TAL, RPE y/o RKI descritos en la presente memoria. En ciertas realizaciones, la vía de fermentación de xilosa activa convierte xilosa en xilulosa 5-fosfato mediante productos intermedios de xilitol y de xilulosa, y en ciertas de estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de xilosa XR, XDH y XK. En otras realizaciones, la vía de fermentación de xilosa activa convierte xilosa en xilulosa 5-fosfato mediante un producto intermedio de xilulosa solo, y en ciertas de estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de xilosa XI y XK. En ciertas realizaciones, los genes transportadores de xilosa comprenden la secuencia de nucleótidos de uno o más de los polinucleótidos KHT105 y/o RAG4 descritos en la presente memoria. Similarmente, en ciertas realizaciones, los genes transportadores de xilosa codifican un transportador de xilosa que comprende la secuencia de aminoácidos de uno o más de los polipéptidos KHT105 y/o RAG4 descritos en la presente memoria.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que tienen al menos una vía de pentosa fosfato no oxidativa activa para convertir xilulosa 5-fosfato más ribosa 5-

fosfato en F6P y G3P, y que comprenden además una vía de fermentación de arabinosa activa para convertir arabinosa en xilulosa 5-fosfato y una vía de fermentación de xilosa activa para convertir xilosa en xilulosa 5-fosfato. En ciertas realizaciones, las células comprenden al menos una copia de los genes de la vía de pentosa fosfato no oxidativa TKL y TAL, y en ciertas realizaciones las células comprenden además al menos una copia de los genes de la vía de pentosa fosfato no oxidativa RPE y RKI. En ciertas realizaciones, los genes TKL, TAL, RPE y RKI comprenden la secuencia de nucleótidos de uno o más de los polinucleótidos TKL, TAL, RPE y/o RKI descritos en la presente memoria. Similarmente, en ciertas realizaciones, los genes TKL, TAL, RPE y RKI codifican polipéptidos que comprenden la secuencia de aminoácidos de uno o más de los polipéptidos TKL, TAL, RPE y/o RKI descritos en la presente memoria. En ciertas realizaciones, la vía de fermentación de arabinosa activa convierte arabinosa en xilulosa 5-fosfato mediante productos intermedios de ribulosa y de ribulosa 5-fosfato, y en estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de arabinosa AI, RK y RE. En ciertas realizaciones, la vía de fermentación de xilosa activa convierte xilosa en xilulosa 5-fosfato mediante productos intermedios de xilitol y de xilulosa, y en ciertas de estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de xilosa XR, XDH y XK. En otras realizaciones, la vía de fermentación de xilosa activa convierte xilosa en xilulosa 5-fosfato mediante un producto intermedio de xilulosa solo, y en ciertas de estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de xilosa XI y XK.

En la presente memoria se proporcionan en ciertas realizaciones células de levadura genéticamente modificadas que tienen al menos una vía de pentosa fosfato no oxidativa activa para convertir xilulosa 5-fosfato más ribosa 5-fosfato en F6P y G3P, y que comprenden además uno o más genes transportadores de xilosa, una vía de fermentación de arabinosa activa para convertir arabinosa en xilulosa 5-fosfato, y una vía de fermentación de xilosa activa para convertir xilosa en xilulosa 5-fosfato. En ciertas realizaciones, las células comprenden al menos una copia de los genes de la vía de pentosa fosfato no oxidativa TKL y TAL, y en ciertas realizaciones las células comprenden además al menos una copia de los genes de la vía de pentosa fosfato no oxidativa RPE y RKI. En ciertas realizaciones, los genes TKL, TAL, RPE y RKI comprenden la secuencia de nucleótidos de uno o más de los polinucleótidos TKL, TAL, RPE y/o RKI descritos en la presente memoria. Similarmente, en ciertas realizaciones, los genes TKL, TAL, RPE y RKI codifican polipéptidos que comprenden la secuencia de aminoácidos de uno o más de los polipéptidos TKL, TAL, RPE y/o RKI descritos en la presente memoria. En ciertas realizaciones, los genes transportadores de xilosa comprenden la secuencia de nucleótidos de uno o más de los polinucleótidos KHT105 y/o RAG4 descritos en la presente memoria. Similarmente, en ciertas realizaciones, los genes transportadores de xilosa codifican un transportador de xilosa que comprende la secuencia de aminoácidos de uno o más de los polipéptidos KHT105 y/o RAG4 descritos en la presente memoria. En ciertas realizaciones, la vía de fermentación de arabinosa activa convierte arabinosa en xilulosa 5-fosfato mediante productos intermedios de ribulosa y de ribulosa 5-fosfato, y en estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de arabinosa AI, RK y RE. En ciertas realizaciones, la vía de fermentación de xilosa activa convierte xilosa en xilulosa 5-fosfato mediante productos intermedios de xilitol y de xilulosa, y en ciertas de estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de xilosa XR, XDH y XK. En otras realizaciones, la vía de fermentación de xilosa activa convierte xilosa en xilulosa 5-fosfato mediante un producto intermedio de xilulosa solo, y en ciertas de estas realizaciones las células comprenden al menos una copia de cada uno de los genes de la vía de fermentación de xilosa XI y XK.

Los genes de la vía de fermentación de arabinosa, transportadores de xilosa, de la vía de fermentación de xilosa y de la vía de pentosa fosfato no oxidativa en las células de levadura genéticamente modificadas proporcionadas en la presente memoria pueden ser endógenos o exógenos. "Endógeno", como se emplea en la presente memoria, con respecto a componentes genéticos tales como genes, promotores y secuencias terminadoras significa que el componente genético está presente en una localización particular en el genoma de una forma nativa de una célula de levadura particular. "Exógeno", como se emplea en la presente memoria, con respecto a componentes genéticos significa que el componente genético no está presente en una localización particular en el genoma de una forma nativa de una célula de levadura particular. "Nativo", como se emplea en la presente memoria, con respecto a una célula de levadura se refiere a una célula de levadura no mutante de una especie de levadura particular. "Nativo", como se emplea en la presente memoria, con respecto a una vía metabólica se refiere a una vía metabólica que existe y es activa en una célula de levadura nativa.

Un componente genético exógeno puede tener tanto una secuencia nativa como no nativa. Un componente genético exógeno con una secuencia nativa comprende una secuencia idéntica (aparte de mutaciones de individuo a individuo que no afectan la función) a un componente genético que está presente en el genoma de una célula nativa (es decir, el componente genético exógeno es idéntico a un componente genético endógeno). Sin embargo, el componente exógeno está presente en una localización diferente en el genoma de la célula huésped que en el componente endógeno. Por ejemplo, un gen XI exógeno que es idéntico a un gen XI endógeno puede insertarse en una célula de levadura, produciendo una célula modificada con un número (elevado) de copias del gen XI no nativo. Un componente genético exógeno con una secuencia no nativa comprende una secuencia que no se encuentra en el genoma de una célula nativa. Por ejemplo, un gen XI exógeno de una especie particular puede insertarse en una célula de levadura de otra especie. Un gen exógeno se integra preferiblemente en el genoma de la célula huésped de una manera funcional, que significa que es capaz de producir una proteína activa en la célula huésped. Sin embargo, en ciertas realizaciones, el gen exógeno puede introducirse en la célula como parte de un vector que se

mantiene establemente en el citoplasma del huésped.

En ciertas realizaciones, las células de levadura proporcionadas en la presente memoria comprenden uno o más genes exógenos de la vía de fermentación de arabinosa, transportadores de xilosa, de la vía de fermentación de xilosa o de la vía de pentosa fosfato no oxidativa. En ciertas realizaciones, las células de levadura genéticamente modificadas descritas en la presente memoria comprenden un único gen exógeno. En otras realizaciones, las células de levadura comprenden múltiples genes exógenos. En estas realizaciones, las células de levadura pueden comprender múltiples copias de un único gen exógeno y/o copias de dos o más genes exógenos diferentes. Células de levadura que comprenden múltiples genes exógenos pueden comprender cualquier número de genes exógenos. Por ejemplo, estas células de levadura pueden comprender 1 a 20 genes exógenos, y en ciertas realizaciones preferidas pueden comprender 1 a 7 genes exógenos. Pueden integrarse múltiples copias de un gen exógeno en un único locus de forma que sean adyacentes entre sí. Alternativamente, pueden integrarse en varios loci dentro del genoma de la célula huésped. Una célula de levadura como se proporciona en la presente memoria puede comprender solo un tipo de gen exógeno o genes exógenos de solo una vía. Por ejemplo, los genes exógenos en una célula de levadura pueden limitarse a genes de la vía de fermentación de arabinosa o a genes transportadores de xilosa. Alternativamente, una célula de levadura puede comprender genes exógenos de dos o más vías o de una o más vías en combinación con un gen exógeno transportador de xilosa. Por ejemplo, una célula de levadura puede comprender uno o más genes exógenos de la vía de fermentación de arabinosa y uno o más genes exógenos transportadores de xilosa.

En ciertas realizaciones, las células de levadura proporcionadas en la presente memoria comprenden uno o más genes endógenos de la vía de fermentación de arabinosa, transportadores de xilosa, de la vía de fermentación de xilosa y de la vía de pentosa fosfato no oxidativa. En ciertas de estas realizaciones, las células pueden manipularse para expresar en exceso uno o más de estos genes endógenos, que significa que las células modificadas expresan el gen endógeno a un nivel más alto que una célula nativa bajo al menos algunas condiciones. En ciertas de estas realizaciones, el gen endógeno que se expresa en exceso puede estar operativamente unido a uno o más elementos reguladores exógenos. Por ejemplo, pueden introducirse uno o más promotores fuertes exógenos nativos o no nativos en una célula de forma que están operativamente unidos a uno o más genes endógenos.

Los genes de la vía de fermentación de arabinosa, transportadores de xilosa, de la vía de fermentación de xilosa y/o de la vía de pentosa fosfato no oxidativa en las células de levadura genéticamente modificadas proporcionadas en la presente memoria pueden estar operativamente unidos a uno o más elementos reguladores tales como un promotor o terminador. Como se emplea en la presente memoria, el término "promotor" se refiere a una secuencia no traducida localizada aguas arriba (es decir, 5') con respecto al codón de iniciación de la traducción de un gen (generalmente dentro de aproximadamente 1 a 1000 pares de bases (pb), preferiblemente dentro de aproximadamente 1 a 500 pb) que controla el inicio de la transcripción del gen. El término "terminador", como se emplea en la presente memoria, se refiere a una secuencia no traducida localizada aguas abajo (es decir, 3') con respecto al codón de terminación de la traducción de un gen (generalmente dentro de aproximadamente 1 a 1000 pb, preferiblemente dentro de aproximadamente 1 a 500 pb, y especialmente dentro de aproximadamente 1 a 100 pb) que controla el fin de la transcripción del gen. Un promotor o terminador está "operativamente unido" a un gen si su posición en el genoma con respecto a la del gen es tal que el promotor o terminador, según sea el caso, realiza su función de control transcripcional. Promotores y terminadores adecuados se describen, por ejemplo, en los documentos WO99/14335, WO00/71738, WO02/42471, WO03/102201, WO03/102152 y WO03/049525.

Los elementos reguladores unidos a genes de la vía de fermentación de arabinosa, transportadores de xilosa, de la vía de fermentación de xilosa o de la vía de pentosa fosfato no oxidativa en las células de levadura proporcionadas en la presente memoria pueden ser endógenos o exógenos. Por ejemplo, puede insertarse un gen exógeno de la vía de fermentación de arabinosa o transportador de xilosa en una célula de levadura de forma que esté bajo el control transcripcional de un promotor y/o terminador endógeno. Alternativamente, el gen exógeno de la vía de fermentación de arabinosa o transportador de xilosa puede unirse a uno o más elementos reguladores exógenos. Por ejemplo, puede introducirse un gen exógeno en la célula como parte de una construcción de expresión de gen que comprende uno o más elementos reguladores exógenos. En ciertas realizaciones, los elementos reguladores exógenos, o al menos las porciones funcionales de elementos reguladores exógenos, pueden comprender secuencias nativas. En otras realizaciones, los elementos reguladores exógenos pueden comprender secuencias no nativas. En estas realizaciones, los elementos reguladores exógenos pueden comprender una secuencia con un grado de identidad de secuencia relativamente alto con un elemento regulador nativo. Por ejemplo, un gen exógeno puede unirse a un promotor o terminador exógeno que tiene al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, o al menos el 90 % de identidad de secuencia con un promotor o terminador nativo. Los porcentajes de identidad de secuencia para las secuencias de nucleótidos o de aminoácidos pueden calcularse por métodos conocidos en la técnica, tales como, por ejemplo, usando el software BLAST (Basic Local Alignment Search Tool del National Center for Biological Information (NCBI)) versión 2.2.1 con parámetros por defecto. Por ejemplo, una secuencia que tiene una puntuación de identidad de al menos el 90 % usando el algoritmo BLAST versión 2.2.1 con parámetros por defecto se considera que tiene al menos el 90 % de identidad de secuencia. El software BLAST está disponible de NCBI, Bethesda, Maryland. En aquellas realizaciones en donde múltiples genes exógenos se insertan en una célula huésped, cada gen exógeno puede estar bajo el control de un elemento regulador diferente, o dos o más genes exógenos pueden estar bajo el control de los mismos elementos reguladores. Por ejemplo, donde un primer gen exógeno está unido a un primer elemento regulador, un segundo gen

exógeno también puede unirse al primer elemento regulador, o puede unirse a un segundo elemento regulador. El primer y segundo elementos reguladores pueden ser idénticos o compartir un alto grado de identidad de secuencia, o estar completamente sin relacionar.

Ejemplos de promotores que pueden unirse a uno o más genes de la vía de fermentación de arabinosa, transportadores de xilosa, de la vía de fermentación de xilosa o de la vía de pentosa fosfato no oxidativa en las células de levadura proporcionadas en la presente memoria incluyen, pero no se limitan a, promotores para los genes piruvato descarboxilasa 1 (PDC1), enolasa 1 (ENO1), factor-1 o -2 de elongación de la traducción (TEF1, TEF2), fosfoglicerato cinasa (PGK), XR, XDH, L-(+)-lactato-citocromo c oxidoreductasa (CYB2), gliceraldehído-3-fosfato deshidrogenasa 3 (GAPDH/TDH3) y orotidina 5'-fosfato descarboxilasa (URA3). En estos ejemplos, los genes pueden unirse a promotores endógenos o exógenos para los genes PDC1, PGK, XR, XDH, CYB2, TEF1, TEF2, ENO1, TDH3 o URA3. Cuando los promotores son exógenos, pueden ser idénticos o compartir un alto grado de identidad de secuencia (es decir, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, o al menos aproximadamente el 99 %) con promotores nativos para los genes PDC1, ENO1, TEF1, TEF2, PGK, XR, XDH, CYB2, TDH3 o URA3.

Ejemplos de terminadores que pueden unirse a uno o más genes de la vía de fermentación de arabinosa, transportadores de xilosa, de la vía de fermentación de xilosa o de la vía de pentosa fosfato no oxidativa en las células de levadura proporcionadas en la presente memoria incluyen, pero no se limitan a, terminadores para los genes PDC1, XR, XDH, TAL, TKL, RKI, CYB2 o iso-2-citocromo c (CYC) o la familia galactosa de genes (especialmente el terminador GAL10). En estos ejemplos, los genes pueden unirse a terminadores endógenos o exógenos para los genes PDC1, XR, XDH, TAL, TKL, RKI, CYB2 o CYC o genes de la familia galactosa. Cuando los terminadores son exógenos, pueden ser idénticos o compartir un alto grado de identidad de secuencia (es decir, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, o al menos aproximadamente el 99 %) con terminadores nativos para los genes PDC1, XR, XDH, TAL, TKL, RKI, CYB2 o CYC o genes de la familia galactosa. En ciertas realizaciones, los genes se unen a un terminador que comprende una porción funcional de un gen nativo GAL10 nativo a la célula huésped o una secuencia que comparte al menos el 80 %, al menos el 85 %, al menos el 90 %, o al menos el 95 % de identidad de secuencia con un terminador GAL10 nativo.

Pueden insertarse genes exógenos en una célula huésped de levadura mediante cualquier método conocido en la técnica. En realizaciones preferidas, los genes se integran en el genoma de la célula huésped. Los genes exógenos pueden integrarse en el genoma en una forma dirigida o al azar. En aquellas realizaciones donde el gen se integra de una forma dirigida, pueden integrarse en los loci para un gen particular, de forma que la integración del gen exógeno se acople a delección o alteración de un gen nativo. Por ejemplo, la introducción de un gen exógeno de la vía de fermentación de arabinosa, de transporte de xilosa, de la vía de fermentación de xilosa o de la vía de pentosa fosfato no oxidativa puede acoplarse a delección o alteración de uno o más genes que codifican enzimas involucradas en otras vías de productos de fermentación. Alternativamente, el gen exógeno puede integrarse en una porción del genoma que no se corresponde con un gen.

La integración y/o delección dirigida puede utilizar una construcción de integración. El término “construcción”, como se emplea en la presente memoria, se refiere a una secuencia de ADN que se usa para transformar una célula. La construcción puede ser, por ejemplo, un plásmido o vector circular, una porción de un plásmido o vector circular (tal como un producto de digestión de enzima de restricción), un plásmido o vector linealizado, o un producto de PCR preparado usando un plásmido o ADN genómico como molde. Métodos de transformación de una célula de levadura con una construcción exógena se describen en, por ejemplo., los documentos WO99/14335, WO00/71738, WO02/42471, WO03/102201, WO03/102152 y WO03/049525. Una construcción de integración puede ensamblarse usando dos secuencias de ADN diana clonadas a partir de un diana del sitio de inserción. Las dos secuencias de ADN diana pueden estar contiguas o no contiguas en el genoma huésped nativo. En este contexto, “no contiguo” significa que las secuencias de ADN no están inmediatamente adyacentes la una a la otra en el genoma nativo, pero en su lugar están separadas por una región que va a delecionarse. Las secuencias “contiguas”, como se emplea en la presente memoria, son directamente adyacentes la una a la otra en el genoma nativo. Si la integración dirigida va a acoplarse a delección o alteración de un gen diana, la construcción de integración también puede denominarse una construcción de delección. En una construcción de delección, una de las secuencias diana pueden incluir una región 5' con respecto al promotor del gen diana, toda o una porción de la región promotora, toda o una porción de la secuencia codificante del gen diana, o alguna combinación de los mismos. La otra secuencia diana puede incluir una región 3' con respecto al terminador del gen diana, toda o una porción de la región de terminador, y/o toda o una porción de la secuencia codificante del gen diana. Si la integración dirigida no debe acoplarse a delección o alteración de un gen nativo, las secuencias diana se seleccionan de forma que la inserción de una secuencia intercalada no altere la expresión génica nativa. Se prepara una construcción de integración o delección de forma que las dos secuencias diana estén orientadas en la misma dirección en relación la una con la otra como aparecen nativamente en el genoma de la célula huésped. Si se usa una construcción de integración o delección para introducir un gen exógeno en una célula huésped, se clona un casete de expresión del gen en la construcción entre las dos secuencias del gen diana para permitir la expresión del gen exógeno. El casete de expresión del gen contiene el gen exógeno, y puede incluir además una o más secuencias reguladoras tales como promotores o terminadores operativamente unidos al gen exógeno. También pueden construirse construcciones de delección que no contienen un casete de expresión del gen. Tales construcciones se diseñan para deleccionar o alterar una secuencia del gen sin

la inserción de un gen exógeno.

Una construcción de integración o delección puede comprender uno o más casetes de marcador de selección clonados en la construcción entre las dos secuencias del gen diana. El casete del marcador de selección contiene al menos un gen de marcador de selección que permite la selección de transformantes. Transformantes satisfactorios contendrán el gen de marcador de selección, que confiere a la célula satisfactoriamente transformada al menos una característica que proporciona una base para la selección. Genes de marcadores de selección típicos codifican proteínas que (a) confieren resistencia a antibióticos u otras toxinas (p. ej., resistencia a bleomicina o zeomicina (p. ej., gen *ble* de *Streptoalloteichus hindustanus*), aminoglucósidos tales como G418 o kanamicina (p. ej., gen de resistencia a kanamicina del transposón Tn903), o higromicina (p. ej., gen de resistencia a antibióticos de aminoglucósido de *E. coli*)), (b) deficiencias auxotróficas del complemento de la célula (p. ej., deficiencias en leucina (p. ej., gen *LEU2* de *K. marxianus*), uracilo (p. ej., gen *URA3* de *K. marxianus*, *S. cerevisiae* o *I. orientalis*), o triptófano (p. ej., gen *TRP* de *K. marxianus*, *S. cerevisiae* o *I. orientalis*)), (c) permiten que la célula sintetice nutrientes críticos no disponibles de medios simples, o (d) confieren la capacidad de que la célula crezca sobre una fuente de carbono particular (p. ej., gen *MEL5* de *S. cerevisiae*, que codifica la enzima alfa-galactosidasa (melibiosa) y confiere la capacidad de crecer sobre melibiosa como única fuente de carbono). Marcadores de selección preferidos incluyen el gen *URA3*, gen de resistencia a zeocina, gen de resistencia a G418, gen *MEL5* y gen de resistencia a higromicina. Otro marcador de selección preferido es un casete del gen *CYB2*, a condición de que la célula huésped tanto carezca nativamente de un gen tal como su(s) gen(es) *CYB2* nativo(s) se delecionen o alteren primero. Un gen de marcador de selección está operativamente unido a una o más secuencias de promotor y/o terminador que están operativas en la célula huésped. En ciertas realizaciones, estas secuencias de promotor y/o terminador son secuencias de promotor y/o terminador exógenas que están incluidas en el casete de marcador de selección. Promotores y terminadores adecuados son como se han descrito anteriormente.

En otras realizaciones, una construcción de integración o delección puede no contener un casete de marcador de selección, pero puede, sin embargo, permitir la selección de transformantes basándose en la expresión en exceso de un gen exógeno (en el caso de construcciones de inserción) o delección de un gen endógeno (en el caso de construcciones de delección). Por ejemplo, donde una construcción de integración comprende uno o más genes exógenos de la vía de fermentación de arabinosa, pueden seleccionarse transformantes basándose en su capacidad para cultivar sobre arabinosa.

Se usa una construcción de integración o delección para transformar la célula huésped. La transformación puede llevarse a cabo usando, por ejemplo, métodos de electroporación y/o transformación química (p. ej., cloruro de calcio, basada en acetato de litio, etc.). La selección o cribado basado en la presencia o ausencia del marcador de selección puede realizarse para identificar transformantes satisfactorios. En transformantes satisfactorios, un evento de recombinación homóloga en el locus del sitio diana produce la alteración o la delección de la secuencia del sitio diana. Cuando la construcción se dirige a un gen nativo para la delección o alteración, todo o una porción del gen nativo diana, su promotor, y/o su terminador, pueden ser delecionados durante este evento de recombinación. El casete de expresión, casete de marcador de selección, y cualquier otro material genético entre las secuencias diana en la construcción de integración, se inserta en el genoma del huésped en el locus correspondiente a las secuencias diana. Puede realizarse análisis por PCR o análisis de Southern para confirmar que ha tenido lugar la inserción/delección deseada.

En algunas realizaciones, la transformación de las células puede realizarse usando ADN de dos o más construcciones, productos de PCR, o una combinación de los mismos, en vez de una única construcción o producto de PCR. En estas realizaciones, el extremo 3' de un fragmento de integración se solapa con el extremo 5' del otro fragmento de integración. En un ejemplo, una construcción contendrá la primera secuencia del locus de la secuencia diana y una parte no funcional del casete del gen marcador, mientras que la otra contendrá la segunda secuencia del locus de la secuencia diana y una segunda parte no funcional del casete del gen marcador. Las partes del casete del gen marcador están seleccionadas de forma que puedan combinarse para formar un casete completo. La célula se transforma con estos trozos simultáneamente, produciendo la formación de un marcador funcional completo o casete de gen estructural. Pueden seleccionarse transformantes satisfactorios basándose en las características conferidas por el marcador de selección. En otro ejemplo, el marcador de selección reside en un fragmento, pero las secuencias diana están en fragmentos separados, de manera que los fragmentos de integración tienen una alta probabilidad de integrarse en el sitio de interés. En otras realizaciones, puede usarse la transformación de tres ADN lineales para integrar el material genético exógeno. En estas realizaciones, un fragmento se solapa sobre el extremo 5' con un segundo fragmento y sobre el extremo 3' con un tercer fragmento.

Puede diseñarse una construcción de integración o de delección de forma que el gen de marcador de selección y algunos o todos de sus elementos reguladores puedan llegar a delecionarse espontáneamente como resultado de un evento de recombinación homóloga posterior. Una forma conveniente de realizar esto es diseñar la construcción de forma que el gen de marcador de selección y/o elementos reguladores estén flanqueados por secuencias de repetición. Las secuencias de repetición son secuencias de ADN idénticas, nativas o no nativas para la célula huésped, y orientadas en la construcción en la misma dirección la una con respecto a la otra. Las secuencias de repetición tienen ventajosamente aproximadamente 25 a 1500 pb de longitud, y no tienen que codificar nada. La inclusión de las secuencias de repetición permite que se produzca un evento de recombinación homóloga, que produce la delección del gen de marcador de selección y una de las secuencias de repetición. Como la

recombinación homóloga se produce con frecuencia relativamente baja, puede ser necesario cultivar transformantes durante varias rondas sobre medios no selectivos para permitir que se produzca la recombinación homóloga espontánea en algunas de las células. Pueden seleccionarse o cribarse células en las que el gen de marcador de selección se ha delecionado espontáneamente basándose en su pérdida de la característica de selección conferida por el gen de marcador de selección. En ciertos casos, la expresión de una enzima recombinasa puede potenciar la recombinación entre los sitios repetidos.

Un gen exógeno de la vía de fermentación de arabinosa, transportador de xilosa, de la vía de fermentación de xilosa o de la vía de pentosa fosfato no oxidativa en las células de levadura modificadas proporcionadas en la presente memoria puede derivarse de un gen fuente de cualquier organismo fuente adecuado. Por ejemplo, un gen exógeno puede derivarse de una fuente de levadura, fúngica, bacteriana, de planta, insecto, o de mamífero. Como se emplea en la presente memoria, un gen exógeno que se "deriva de" un gen de fuente nativa codifica un polipéptido que 1) es idéntico a un polipéptido codificado por el gen nativo, 2) comparte al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con un polipéptido codificado por el gen nativo, y/o 3) tiene la misma función en una vía de fermentación de arabinosa, vía de fermentación de xilosa, o vía de pentosa fosfato no oxidativa o en el transporte de xilosa que el polipéptido codificado por el gen nativo. Por ejemplo, un gen transportador de xilosa que se deriva de un gen KHT105 de *K. marxianus* puede codificar un polipéptido que comprende la secuencia de aminoácidos de SEQ ID NO:2, un polipéptido con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO:2, y/o un polipéptido que tiene la capacidad de transportar xilosa en una célula de levadura. Un gen derivado de un gen nativo puede comprender una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la región codificante del gen nativo. En ciertas realizaciones, un gen derivado de un gen nativo puede comprender una secuencia de nucleótidos que es idéntica a la región codificante del gen fuente. Por ejemplo, un gen transportador de xilosa que se deriva de un gen KHT105 de *K. marxianus* puede comprender la secuencia de nucleótidos de SEQ ID NO:1 o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO:1.

En ciertas realizaciones de las células de levadura modificadas proporcionadas en la presente memoria, un gen exógeno de la vía de fermentación de arabinosa, transportador de xilosa, de la vía de fermentación de xilosa o de la vía de pentosa fosfato no oxidativa puede derivarse de las especies de levadura huésped. Por ejemplo, donde la célula huésped es *I. orientalis*, un gen exógeno puede derivarse de un gen de *I. orientalis*. En estas realizaciones, el gen exógeno puede comprender una secuencia de nucleótidos idéntica a la región codificante del gen nativo, de forma que la incorporación del gen exógeno en la célula huésped aumenta el número de copias de una secuencia de genes nativos y/o cambia la regulación o nivel de expresión del gen si está bajo el control de un promotor que es diferente del promotor que conduce la expresión del gen en una célula no mutante. En otras realizaciones, el gen exógeno puede comprender una secuencia de nucleótidos que se diferencia de la región codificante de un gen nativo, pero sin embargo codifica un polipéptido que es idéntico al polipéptido codificado por el gen nativo. En otras realizaciones más, el gen exógeno puede comprender una secuencia de nucleótidos que codifica un polipéptido con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con un polipéptido codificado por uno o más genes nativos. En ciertas de estas realizaciones, el gen exógeno comprende una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la región codificante de uno o más genes nativos. En otras realizaciones más, el gen exógeno de la vía de fermentación de arabinosa, transportador de xilosa, de la vía de fermentación de xilosa o de la vía de pentosa fosfato no oxidativa puede codificar un polipéptido que tiene menos del 50 % de identidad de secuencia con un polipéptido codificado por un gen nativo de la vía de fermentación de arabinosa, transportador de xilosa, de la vía de fermentación de xilosa o de la vía de pentosa fosfato no oxidativa, pero que sin embargo tiene la misma función que el polipéptido nativo en una vía de fermentación de arabinosa, fermentación de xilosa o de pentosa fosfato no oxidativa (es decir, la capacidad para catalizar la misma reacción entre productos intermedios de reacción) o en el transporte de xilosa (es decir, la capacidad para transportar xilosa en una célula).

En otras realizaciones, un gen exógeno de la vía de fermentación de arabinosa, transportador de xilosa, de la vía de fermentación de xilosa o de la vía de pentosa fosfato no oxidativa puede derivarse de una especie que es diferente de la de la célula de levadura huésped. En ciertas de estas realizaciones, el gen exógeno puede derivarse de una especie de levadura diferente de la célula huésped. Por ejemplo, donde la célula huésped es *I. orientalis*, el gen exógeno puede derivarse de *S. cerevisiae*. En otras realizaciones, el gen exógeno puede derivarse de una fuente fúngica, bacteriana, de planta, insecto, o de mamífero. Por ejemplo, donde la célula huésped es *I. orientalis*, el gen exógeno puede derivarse de una fuente bacteriana tal como *E. coli*. En aquellas realizaciones donde el gen exógeno se deriva de una fuente no de levadura, la secuencia del gen exógeno puede ser de codón optimizado para la expresión en una célula huésped de levadura.

En aquellas realizaciones donde el gen exógeno de la vía de fermentación de arabinosa, transportador de xilosa, de la vía de fermentación de xilosa o de la vía de pentosa fosfato no oxidativa se deriva de una especie distinta de la especie de la célula huésped, el gen exógeno puede codificar un polipéptido idéntico a un polipéptido codificado por un gen nativo del organismo fuente. En ciertas de estas realizaciones, el gen exógeno puede ser idéntico a un gen nativo del organismo fuente. En otras realizaciones, el gen exógeno puede compartir al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la región codificante de un gen nativo del organismo fuente. En otras realizaciones, el gen exógeno puede codificar un polipéptido que comparte al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con un polipéptido codificado por un gen nativo del organismo fuente. En ciertas de estas realizaciones, el gen exógeno puede comprender una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la región codificante de uno o más genes nativos del organismo fuente. En otras realizaciones más, el gen exógeno puede codificar un polipéptido que tiene menos del 50 % de identidad de secuencia con un polipéptido codificado por un gen nativo del organismo fuente, pero que sin embargo tiene la misma función que el polipéptido nativo del organismo fuente en una vía de fermentación de arabinosa nativa, vía de fermentación de xilosa o de la vía de pentosa fosfato no oxidativa o en el transporte de xilosa. Un gen exógeno fuente puede someterse a mutagénesis si fuera necesario para proporcionar una secuencia codificante que empieza con el codón de iniciación eucariota usual (ATG), o para otros fines.

Un “gen de arabinosa isomerasa”, “gen AI” o “gen araA”, como se emplea en la presente memoria, se refiere a cualquier gen que codifica un polipéptido con actividad de arabinosa isomerasa, que significa la capacidad para catalizar la conversión de arabinosa en ribulosa. En ciertas realizaciones, un gen AI puede derivarse de una fuente bacteriana. Por ejemplo, un gen AI puede derivarse de un gen araA1 de *B. thetaiotaomicron* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:6, un gen araA2 de *B. thetaiotaomicron* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:8, un gen AI de *L. sakei* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:10, un gen AI de *L. plantarum* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:81, o un gen AI de *B. licheniformis* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:83. En otras realizaciones, el gen puede codificar una secuencia de aminoácidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NOS:6, 8, 10, 81 u 83. En ciertas realizaciones, un gen AI derivado de *B. thetaiotaomicron*, *L. sakei*, *L. plantarum* o *B. licheniformis* puede comprender la secuencia de nucleótidos expuesta en SEQ ID NOS:5, 7, 9, 80 u 82, o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de nucleótidos expuesta en SEQ ID NOS:5, 7, 9, 80 u 82.

Un “gen de ribulocinasa”, “gen RK” o “gen araB”, como se emplea en la presente memoria, se refiere a cualquier gen que codifica un polipéptido con actividad de ribulocinasa, que significa la capacidad para catalizar la conversión de ribulosa en ribulosa 5-fosfato. En ciertas realizaciones, un gen RK puede derivarse de una fuente bacteriana. Por ejemplo, un gen RK puede derivarse de un gen RK de *B. thetaiotaomicron* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:12 o un gen RK de *Leuconostoc citreum* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:14. En otras realizaciones, el gen puede codificar una secuencia de aminoácidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NOS:12 o 14. En ciertas realizaciones, un gen RK derivado de *B. thetaiotaomicron* o *L. citreum* puede comprender la secuencia de nucleótidos expuesta en SEQ ID NOS :11, 86 o 13, o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de nucleótidos expuesta en SEQ ID NOS :11, 86 o 13.

Una “ribulosa-fosfato 4-epimerasa”, “gen RE” o “gen araD”, como se emplea en la presente memoria, se refiere a cualquier gen que codifica un polipéptido con actividad de ribulosa-fosfato 4-epimerasa, que significa la capacidad para catalizar la conversión de ribulosa 5-fosfato en xilulosa 5-fosfato. En ciertas realizaciones, un gen RE puede derivarse de una fuente bacteriana. Por ejemplo, un gen RE puede derivarse de un gen RE de *B. thetaiotaomicron* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:16, un gen RE de *B. animalis* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:18, un gen RE de *L. lactis* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:20, un gen RE de *E. coli* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:77, o un gen RE de *L. plantarum* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:79. En otras realizaciones, el gen puede codificar una secuencia de aminoácidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NOS:16, 18, 20, 77 o 79. En ciertas realizaciones, un gen RE derivados de *B. thetaiotaomicron*, *B. animalis*, *L. lactis*, *E. coli*, o *L. plantarum* puede comprender la secuencia de nucleótidos expuesta en SEQ ID NOS:15, 17, 19, 76 o 78, o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la

secuencia de nucleótidos expuesta en SEQ ID NOs:15, 17, 19, 76, o 78.

Un "gen xilosa isomerasa" o "gen XI", como se emplea en la presente memoria, se refiere a cualquier gen que codifica un polipéptido con actividad de xilosa isomerasa, que significa la capacidad para catalizar la conversión de xilosa en xilulosa. En ciertas realizaciones, un gen XI puede derivarse de una fuente bacteriana. Por ejemplo, un gen XI puede derivarse de un gen XI de *B. thetaiotaomicron* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:22. En otras realizaciones, el gen puede codificar una secuencia de aminoácidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO:22. En ciertas realizaciones, un gen XI derivado de *B. thetaiotaomicron* puede comprender la secuencia de nucleótidos expuesta en SEQ ID NO:21, o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de nucleótidos expuesta en SEQ ID NO:21.

Un "gen xilulocinasa" o "gen XK", como se emplea en la presente memoria, se refiere a cualquier gen que codifica un polipéptido con actividad de xilulocinasa, que significa la capacidad para catalizar la conversión de xilulosa en xilulosa 5-fosfato. En ciertas realizaciones, un gen XK puede derivarse de una fuente de levadura. Por ejemplo, el gen XK puede derivarse de un gen XK de *I. orientalis* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:24, un gen XK de *S. cerevisiae* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:26, o un gen XK de *K. marxianus* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:28. En otras realizaciones, el gen puede codificar una secuencia de aminoácidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NOs:24, 26 o 28. En ciertas realizaciones, un gen XK derivado de *I. orientalis*, *S. cerevisiae* o *K. marxianus* puede comprender la secuencia de nucleótidos expuesta en SEQ ID NOs:23, 25 o 27 o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de nucleótidos expuesta en SEQ ID NOs:23, 25 o 27.

Un "gen xilosa reductasa" o "gen XR", como se emplea en la presente memoria, se refiere a cualquier gen que codifica un polipéptido con actividad de xilosa reductasa, que significa la capacidad para catalizar la conversión de xilosa en xilitol. En ciertas realizaciones, un gen XR puede derivarse de una fuente de levadura. Por ejemplo, el gen XR puede derivarse de un homólogo de XR/AR de *I. orientalis* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:71 o un gen XR de *Pichia stipitis* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:30. En otras realizaciones, el gen puede codificar una secuencia de aminoácidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NOs:71 o 30. En ciertas realizaciones, un gen XR de *I. orientalis* o derivado de *P. stipitis* puede comprender la secuencia de nucleótidos expuesta en SEQ ID NOs:70 o 29 o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de nucleótidos expuesta en SEQ ID NOs:70 o 29.

Un "gen xilitol deshidrogenasa" o "gen XDH", como se emplea en la presente memoria, se refiere a cualquier gen que codifica un polipéptido con actividad de xilitol deshidrogenasa, que significa la capacidad para catalizar la conversión de xilitol en xilulosa. En ciertas realizaciones, un gen XDH puede derivarse de una fuente de levadura. Por ejemplo, el gen XDH puede derivarse de un homólogo de XDH de *I. orientalis* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:60 o un gen XDH de *P. stipitis* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:32. En otras realizaciones, el gen puede codificar una secuencia de aminoácidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NOs:60 o 32. En ciertas realizaciones, un gen XDH de *I. orientalis* o derivado de *P. stipitis* puede comprender la secuencia de nucleótidos expuesta en SEQ ID NOs:59 o 31 o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de nucleótidos expuesta en SEQ ID NOs:59 o 31.

Un "gen ribulosa 5-fosfato 3-epimerasa" o "gen RPE", como se emplea en la presente memoria, se refiere a cualquier gen que codifica un polipéptido con actividad de ribulosa 5-fosfato 3-epimerasa, que significa la capacidad para catalizar la conversión de xilulosa 5-fosfato en ribulosa 5-fosfato. En ciertas realizaciones, un gen RPE puede derivarse de una fuente de levadura. Por ejemplo, el gen RPE puede derivarse de un gen RPE de *I. orientalis* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:34, un gen RPE de *S. cerevisiae* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:36, o un gen RPE de *K. marxianus* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:38. En otras realizaciones, el gen puede codificar una secuencia de aminoácidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NOs:34, 36, o 38. En ciertas realizaciones, un gen RPE derivado de *I. orientalis*, *S. cerevisiae* o *K. marxianus* puede comprender la secuencia de nucleótidos expuesta en SEQ ID NOs:33, 35 o 37 o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al

menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de nucleótidos expuesta en SEQ ID NOs:33, 35, o 37.

Un “gen 5-fosfato cetol-isomerasa” o “gen RKI”, como se emplea en la presente memoria, se refiere a cualquier gen que codifica un polipéptido con actividad de ribosa 5-fosfato cetol-isomerasa, que significa la capacidad para catalizar la conversión de ribulosa 5-fosfato en ribosa 5-fosfato. En ciertas realizaciones, un gen RKI puede derivarse de una fuente de levadura. Por ejemplo, el gen RKI puede derivarse de un gen RKI de *I. orientalis* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:40, un gen RKI de *S. cerevisiae* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:42, o un gen RKI de *K. marxianus* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:44. En otras realizaciones, el gen puede codificar una secuencia de aminoácidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NOs:40, 42, o 44. En ciertas realizaciones, un gen RKI derivado de *I. orientalis*, *S. cerevisiae* o *K. marxianus* puede comprender la secuencia de nucleótidos expuesta en SEQ ID NOs:39, 41 o 43 o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de nucleótidos expuesta en SEQ ID NOs:39, 41 o 43.

Un “gen transcetolasa” o “gen TKL”, como se emplea en la presente memoria, se refiere a cualquier gen que codifica un polipéptido con actividad de transcetolasa, que significa la capacidad para catalizar la conversión de xilulosa 5-fosfato y ribosa 5-fosfato en G3P y sedoheptulosa 7-fosfato (S7P) y la conversión de xilulosa 5-fosfato y eritrosa 4-fosfato en F6P y G3P. En ciertas realizaciones, un gen TKL puede derivarse de una fuente de levadura. Por ejemplo, el gen TKL puede derivarse de un gen TKL de *I. orientalis* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:46, un gen TKL de *S. cerevisiae* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:48, o un gen TKL de *K. marxianus* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:50. En otras realizaciones, el gen puede codificar una secuencia de aminoácidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NOs:46, 48, o 50. En ciertas realizaciones, un gen TKL derivado de *I. orientalis*, *S. cerevisiae* o *K. marxianus* puede comprender la secuencia de nucleótidos expuesta en SEQ ID NOs:45, 47 o 49 o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de nucleótidos expuesta en SEQ ID NOs:45, 47 o 49.

Un “gen transaldolasa” o “gen TAL”, como se emplea en la presente memoria, se refiere a cualquier gen que codifica un polipéptido con transaldolasa actividad, que significa la capacidad para catalizar la conversión de G3P y S7P en eritrosa 4-fosfato (E4P) y F6P. En ciertas realizaciones, un gen TAL puede derivarse de una fuente de levadura. Por ejemplo, el gen TAL puede derivarse de un gen TAL de *I. orientalis* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:52, un gen TAL de *S. cerevisiae* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:54 o un gen TAL de *K. marxianus* que codifica la secuencia de aminoácidos expuesta en SEQ ID NO:56. En otras realizaciones, el gen puede codificar una secuencia de aminoácidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NOs:52, 54 o 56. En ciertas realizaciones, un gen TAL derivado de *I. orientalis*, *S. cerevisiae* o *K. marxianus* puede comprender la secuencia de nucleótidos expuesta en SEQ ID NOs:51, 53, o 55 o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con la secuencia de nucleótidos expuesta en SEQ ID NOs:51, 53, o 55.

En ciertas realizaciones, las células de levadura genéticamente modificadas proporcionadas en la presente memoria comprenden además una delección o alteración de uno o más genes nativos. “Delección o alteración” con respecto a un gen nativo significa que tanto la región codificante entera del gen se elimina (delección) como la región codificante del gen, su promotor, y/o su región terminadora se modifica (tal como por delección, inserción o mutación) de forma que el gen ya no produce una enzima activa, produce una cantidad seriamente reducida (al menos el 75 % de reducción, preferiblemente al menos el 90 % de reducción) de una enzima activa, o produce una enzima con actividad seriamente reducida (al menos el 75 % reducida, preferiblemente al menos el 90 % reducida).

En ciertas realizaciones, la delección o alteración de uno o más genes nativos produce una delección o alteración de una o más vías metabólicas nativas. “Delección o alteración” con respecto a una vía metabólica significa que la vía es tanto inoperativa como incluso presenta actividad que se reducen al menos el 75 %, al menos el 85 %, o al menos el 95 % con respecto a la vía nativa.

En ciertas realizaciones, la delección o alteración del gen nativo puede llevarse a cabo por evolución forzada, mutagénesis, o métodos de ingeniería genética, seguido de selección o cribado apropiado para identificar los mutantes deseados. En ciertas realizaciones, la delección o alteración de un gen de células huésped nativa puede acoplarse a la incorporación de uno o más genes exógenos en la célula huésped, es decir, los genes exógenos pueden incorporarse usando una construcción de integración de la expresión génica que también es una

construcción de delección. En otras realizaciones, la delección o alteración puede llevarse a cabo usando una construcción de delección que no contiene un gen exógeno o por otros métodos conocidos en la técnica.

En ciertas realizaciones, las células de levadura genéticamente modificadas proporcionadas en la presente memoria comprenden una delección o alteración de uno o más genes nativos que codifican una enzima involucrada en una vía de fermentación de arabinosa activa que convierte arabinosa en xilulosa 5-fosfato mediante productos intermedios de arabitol, xilulosa, xilitol y xilulosa. En estas realizaciones, las células pueden comprender una delección o alteración de uno o más genes nativos AR, arabitol 4-deshidrogenasa, xilulosa reductasa o XDH. En aquellas realizaciones en donde las células tienen una vía de fermentación de arabinosa activa que convierte arabinosa en xilulosa 5-fosfato mediante productos intermedios de ribulosa y de ribulosa 5-fosfato, la delección o alteración de uno o más genes AR, arabitol 4-deshidrogenasa, xilulosa reductasa o XDH produce un aumento en la cantidad de arabinosa que entra en la vía de productos intermedios de ribulosa/ribulosa 5-fosfato. En ciertas realizaciones en donde la célula de levadura modificada es *I. orientalis*, las células pueden comprender una delección o alteración de un homólogo del gen de xilulosa reductasa que codifica la secuencia de aminoácidos de SEQ ID NO:58, un homólogo del gen XDH que codifica la secuencia de aminoácidos de SEQ ID NOs:60 o 62, y/o un homólogo del gen XR/AR que codifica la secuencia de aminoácidos de SEQ ID NOs:64, 66, 68, 69 o 71. En ciertas realizaciones en donde las células comprenden una delección o alteración de un homólogo del gen de xilulosa reductasa, el gen está localizado en el locus S141G8160 y/o comprende la secuencia de nucleótidos de SEQ ID NO:57 o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con SEQ ID NO:57. En ciertas realizaciones en donde las células comprenden una delección o alteración de un homólogo del gen XDH, el gen está localizado en el locus S141G4546 o S141G7675 y/o comprende la secuencia de nucleótidos de SEQ ID NOs:59 o 61 o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con SEQ ID NOs:59 o 61. En ciertas realizaciones en donde las células comprenden una delección o alteración de un homólogo del gen AR/XR, el gen está localizado en el locus S141G725, S141G4738, o S141G1158-1159, o S141G8885 y/o comprende la secuencia de nucleótidos de SEQ ID NOs:63, 65, 67 o 70 o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con SEQ ID NOs: 63, 65, 67 o 70.

En ciertas realizaciones, las células de levadura genéticamente modificadas proporcionadas en la presente memoria comprenden una delección o alteración de uno o más genes nativos que codifican una enzima involucrada en una vía de fermentación de xilosa activa que convierte xilosa en xilulosa 5-fosfato mediante productos intermedios de xilitol y de D-xilulosa. En estas realizaciones, las células pueden comprender una delección o alteración de uno o más genes nativos XDH o XR. En aquellas realizaciones en donde las células tienen una vía de fermentación de xilosa activa que convierte xilosa en xilulosa 5-fosfato sin un producto intermedio de xilitol (es decir, convirtiendo xilosa directamente en xilulosa), la delección o alteración de uno o más genes XDH o XR produce un aumento en la cantidad de xilosa que entra en la vía de producto intermedio de sólo xilulosa. En ciertas realizaciones en donde la célula de levadura modificada es *I. orientalis*, las células comprenden una delección o alteración de un homólogo del gen XDH que codifica la secuencia de aminoácidos de SEQ ID NOs:60 o 62 y/o un homólogo del gen AR/XR que codifica la secuencia de aminoácidos de SEQ ID NO:64, 66, 68, 69 o 71. En ciertas realizaciones en donde las células comprenden una delección o alteración de un homólogo del gen XDH, el gen está localizado en el locus S141G7675 o S141G4546 y/o comprende la secuencia de nucleótidos de SEQ ID NOs:59 o 61 o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con SEQ ID NOs:59 o 61. En ciertas realizaciones en donde las células comprenden una delección o alteración de un homólogo del gen AR/XR, el gen está localizado en el locus S141G725, S141G4738, S141G1158-1159, o S141G8885 y/o comprende la secuencia de nucleótidos de SEQ ID NOs:63, 65, 67 o 70 o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con SEQ ID NOs:63, 65, 67 o 70.

En ciertas realizaciones, las células de levadura genéticamente modificadas proporcionadas en la presente memoria comprenden una delección o alteración de uno o más genes nativos que codifican una enzima que desvía el carbono de la producción de etanol. En estas realizaciones, las células pueden comprender una delección o alteración de uno o más genes ALD o ADH. En ciertas realizaciones en donde la célula de levadura modificada es *I. orientalis*, las células comprenden una delección o alteración de un gen ALD que codifica la secuencia de aminoácidos de SEQ ID NO:73 (ALD5680) y/o un gen ADH que codifica la secuencia de aminoácidos de SEQ ID NOs:75 o 85. En ciertas realizaciones en donde las células comprenden una delección o alteración de un gen ALD, el gen ALD está localizado en el locus S141G5680 y/o comprende la secuencia de nucleótidos de SEQ ID NO:72 o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de secuencia con SEQ ID NO:72. En ciertas realizaciones en donde las células comprenden una delección o alteración de un gen ADH, el gen ADH está localizado en el locus S141G9091 o S141G1202 y/o comprende la secuencia de nucleótidos de SEQ ID NOs:74 o 84 o una secuencia de nucleótidos con al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 85 %, al menos el 90 %, al menos el 95 %, al menos el 97 %, o al menos el 99 % de identidad de

secuencia con SEQ ID NOs:74 o 84.

Las células de levadura genéticamente modificadas proporcionadas en la presente memoria pueden seleccionarse de una variedad de especies de levadura. En ciertas realizaciones, las células de levadura genéticamente modificadas proporcionadas en la presente memoria son células de levadura no de *Saccharomyces*. En ciertas de estas realizaciones, las células de levadura son células de levadura negativas para Crabtree, y en ciertas de estas realizaciones las células de levadura pertenecen al clado de *I. orientalis/Pichia fermentans*. El clado de *I. orientalis/P. fermentans* es el clado más terminal que contiene al menos las especies *I. orientalis*, *Pichia galeiformis*, *Pichia sp. YB-4149* (designación NRRL), *Candida ethanolica*, *Pichia deserticola*, *Pichia membranifaciens* y *Pichia fermentans*. Miembros del clado de *I. orientalis/P. fermentans* se identifican por análisis del dominio variable D1/D2 del ADN ribosómico 26S de especies de levadura, usando el método descrito por Kurtzman y Robnett en "Identification and Phylogeny of Ascomycetous Yeasts from Analysis of Nuclear Large Subunit (26S) Ribosomal DNA Partial Sequences", Antonie van Leeuwenhoek 73:331-371, 1998 (véase especialmente p. 349). El análisis del dominio variable D1/D2 del ADN ribosómico 26S de cientos de ascomicetos ha revelado que el clado de *I. orientalis/P. fermentans* contiene especies muy estrechamente relacionadas. Miembros del clado de *I. orientalis/P. fermentans* presentan mayor similitud en el dominio variable D1/D2 del ADN ribosómico 26S a otros miembros del clado que a especies de levadura fuera del clado. Por tanto, pueden identificarse otros miembros del clado de *I. orientalis/P. fermentans* por comparación de los dominios D1/D2 de su ADN ribosómico respectivo y comparando con el de los otros miembros del clado y especies estrechamente relacionadas fuera del clado, usando métodos de Kurtzman y Robnett. En ciertas realizaciones, las células de levadura genéticamente modificadas proporcionadas en la presente memoria pertenecen al género *Issatchenkia*, y en ciertas de estas realizaciones las células de levadura son *I. orientalis*. Cuando se caracterizaron por primera vez, a la especie *I. orientalis* se le asignó el nombre *Pichia kudriavzevii*. El anamorfo (forma asexual) de *I. orientalis* se conoce como *Candida krusei*. Se han enumerado numerosos sinónimos adicionales para la especie *I. orientalis* en cualquier parte (Kurtzman y Fell, The Yeasts, a Taxonomic Study. Section 35. *Issatchenkia* Kudryavtsev, pp 222-223 (1998)). *I. orientalis* y otros miembros del clado de *I. orientalis/P. fermentans* presentan ciertas características que los hacen ideales para la fermentación del etanol a partir de la biomasa, que incluyen tolerancia a pH bajo, etanol, temperatura alta (40 °C o superior), y diversos inhibidores presentes en el hidrolizado.

En ciertas realizaciones, se proporcionan procesos de fermentación en donde una célula de levadura genéticamente modificada como se proporciona en la presente memoria se cultiva bajo condiciones de fermentación. En ciertas de estas realizaciones, el proceso de fermentación produce la producción de etanol. Por consiguiente, en la presente memoria en ciertas realizaciones se proporcionan métodos de producción de etanol cultivando una célula de levadura genéticamente modificada como se proporciona en la presente memoria con uno o más azúcares de pentosa y/o hexosa.

En ciertas realizaciones de los procesos y métodos proporcionados en la presente memoria, el medio usado para cultivar las células de levadura genéticamente modificadas proporcionadas en la presente memoria comprende uno o más azúcares no de glucosa que son fermentables por las células. En ciertas de estas realizaciones, los azúcares no de glucosa pueden ser xilosa, xilano, otro oligómero de xilosa, y/o arabinosa. Estos azúcares no de glucosa pueden ser hidrolizados de una biomasa que contiene hemicelulosa tal como un hidrolizado de biomasa vegetal. El medio puede comprender además glucosa y/o oligómeros o polímeros de glucosa. Si están presentes azúcares multiméricos, puede ser necesario añadir enzimas al caldo de fermentación para digerir estos azúcares al azúcar monomérico correspondiente.

En ciertas realizaciones del proceso y métodos proporcionados en la presente memoria, el medio usado para cultivar las células de levadura genéticamente modificadas proporcionadas en la presente memoria es un medio que contiene xilosa, y en ciertas de estas realizaciones la xilosa se deriva de un hidrolizado de biomasa vegetal. En ciertas realizaciones, la xilosa puede estar presente en el medio a una concentración de aproximadamente 0 a aproximadamente 150 g/l al comienzo de la fermentación (es decir, en o antes del momento en el que las células se añaden al medio) y/o en diversos momentos de tiempo durante el proceso de fermentación. En ciertas de estas realizaciones, la xilosa puede estar presente en el medio a una concentración de al menos aproximadamente 10 g/l, 20 g/l, 30 g/l, 40 g/l, 50 g/l, 75 g/l, 100 g/l o 125 g/l. En ciertas realizaciones, el medio puede comprender uno o más azúcares, además de xilosa, que incluyen uno o más azúcares de pentosa y/o hexosa. En ciertas de estas realizaciones, la xilosa puede constituir aproximadamente del 10 a aproximadamente al 95 % del contenido de azúcar total del medio al comienzo de la fermentación y/o en diversos momentos de tiempo durante el proceso de fermentación. En ciertas de estas realizaciones, la xilosa puede constituir aproximadamente el 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, o el 90 % del contenido de azúcar total del medio. En ciertas realizaciones, las células de levadura genéticamente modificadas pueden fermentar uno o más de los azúcares adicionales presentes en el medio en etanol.

En ciertas realizaciones del proceso y métodos proporcionados en la presente memoria, el medio es un medio sintético tales como un medio de extracto de levadura/peptona, y en ciertas de estas realizaciones el medio puede contener acetato. En otras realizaciones, el medio es un medio sintético definido y en ciertas de estas realizaciones el medio puede contener acetato. En ciertas realizaciones, el medio comprende algún porcentaje de hidrolizado de biomasa, tal como hidrolizado de hojas y tallos de maíz. En estas realizaciones, el hidrolizado puede estar presente en el medio en cualquier parte de aproximadamente el 10 % al 100% del volumen de medio total. En ciertas de estas

realizaciones, el hidrolizado puede haber sido pretratado. Por ejemplo, el hidrolizado puede haber sido pretratado con uno o más ácidos o enzimas con el fin de degradar parcialmente la materia prima. En ciertas realizaciones, el hidrolizado es hidrolizado sin purificar. En aquellas realizaciones en donde el medio comprende hidrolizado a menos del 100 %, el resto del medio puede comprender uno o más agentes de dilución que incluyen medio sintético o agua.

- 5 En ciertas realizaciones, el cultivar las células proporcionadas en la presente memoria para producir etanol puede dividirse en fases. Por ejemplo, el proceso de cultivo celular puede dividirse en una fase de cultivo, una fase de producción y una fase de recuperación. Un experto habitual en la materia reconocerá que estas condiciones pueden variarse basándose en factores tales como la especie de levadura que se usa, la vía de fermentación específica utilizada por la levadura, el rendimiento deseado, u otros factores.
- 10 En ciertas realizaciones de los procesos y métodos proporcionados en la presente memoria, las células se cultivan a una temperatura de aproximadamente 20 °C a aproximadamente 60 °C. En ciertas de estas realizaciones, la fermentación tiene lugar a una temperatura que oscila de aproximadamente 30 °C a aproximadamente 50 °C, y en ciertas de estas realizaciones la fermentación tiene lugar a una temperatura de aproximadamente 35 °C a aproximadamente 45 °C. La temperatura puede variarse en todo el proceso de fermentación.
- 15 La fermentación puede realizarse aeróbicamente, microaeróbicamente, sustancialmente anaeróbicamente, o anaeróbicamente. Si se desea, puede variarse la velocidad de captación de oxígeno en toda la fermentación como un control de proceso (véase, p. ej., el documento WO03/102200). En ciertas realizaciones preferidas, la fermentación puede tener lugar bajo condiciones microaerobias, que se caracterizan por una velocidad de captación de oxígeno de aproximadamente 2 a aproximadamente 25 mmol/l/h.
- 20 Los siguientes ejemplos se proporcionan para ilustrar mejor la invención reivindicada y no deben interpretarse como limitantes del alcance de la invención. Hasta el punto que se mencionen materiales específicos, es simplemente para fines de ilustración y no pretende limitar la invención. Un experto en la materia puede desarrollar medios o reactantes equivalentes sin el ejercicio de capacidad inventiva y sin apartarse del alcance de la invención. Se entenderá que pueden hacerse muchas variaciones en los procedimientos descritos en la presente memoria mientras que todavía estén dentro de los límites de la presente invención. Es la intención de los inventores que tales variaciones estén incluidas dentro del alcance de la invención.
- 25

Ejemplos

Ejemplo 1: Integración de genes de la vía de arabinosa de *B. thetaiotaomicron*, *L. plantarum*, *E. coli* y *B. licheniformis* en *I. orientalis*:

- 30 Se incorporaron genes Al (araA), RK (araB) y RE (araD) no mutantes o de codón optimizado de *B. thetaiotaomicron*, *L. plantarum*, *E. coli* y *B. licheniformis* en la cepa 1822 de *I. orientalis* (una cepa resistente al ácido láctico) para determinar si confirieron la capacidad de utilizar arabinosa.

Ejemplo 1A: Integración de araB de *B. thetaiotaomicron* en un locus XR de *I. orientalis*:

- 35 El gen araB de *B. thetaiotaomicron* se optimizó en el codón para la expresión en *I. orientalis* usando una tabla de uso de codones de *I. orientalis* para la traducción inversa (SEQ ID NO:12). El gen araB de codón optimizado se sintetizó de manera que contuviera un sitio de restricción XbaI en el extremo 5' y un sitio de restricción PaeI en el extremo 3'. El producto de PCR se purificó en gel y se clonó en el vector TOPO PCR2.1. La secuenciación de insertos para clones múltiples produjo la identificación de un clon con la secuencia de ADN deseada.

- 40 Se clonó el gen araB de *B. thetaiotaomicron* bajo el control del promotor ENO1 de *I. orientalis* en un plásmido que contenía un terminador PDC de *I. orientalis*, un primer casete de marcador de selección URA3 (promotor/gen/terminador URA3) y una segunda copia del promotor URA3 en la dirección 3' del terminador para generar el plásmido pHJJ2.

- 45 Se clonaron las regiones en la dirección 5' y en la dirección 3' del locus del gen XYL1 de *I. orientalis* (XR) de forma contigua, separadas por un sitio de restricción NotI, en un vector de clonación para formar el plásmido pHJJ1. Se unió un fragmento NotI de pHJJ2 que contiene el promotor ENO1, gen araB y el casete de selección URA3 en pHJJ1 para formar pHJJ3 (orientación 1) y pHJJ18 (orientación 2).

- 50 Se linealizaron pHJJ3 y pHJJ18 por digestión secuencial con ApaI y SacI. El ADN linealizado se transformó en la cepa 2762 de *I. orientalis* (*ura3Δ ura3Δ*), y las células se sembraron en medio ScD-ura. Se purificaron las colonias transformadas en medio ScD-ura, y la integración en la localización XYL1 se confirmó por PCR. La cepa 2762 transformada con pHJJ3 formó la cepa yHJJ1 (2903), mientras que la cepa 2762 transformada con pHJJ18 formó la cepa yHJJ2 (2902).

- 55 Las cepas 2902 y 2903 se cultivaron durante la noche en medio YPD y se sembraron en medio ScD-FOA para seleccionar cepas en las que el marcador URA3 había salido del bucle mediante recombinación entre las regiones del promotor URA3. Las colonias resultantes se purificaron en medio YPD y se probaron en medio ScD-ura para confirmar la pérdida de URA3. Las colonias también se confirmaron por PCR de colonias. El derivado ura- de la

cepa 2902 se llamó la cepa yHJJ3 (2904) y el derivado ura- de la cepa 2903 se llamó la cepa yHJJ4 (2905).

La expresión del gen araB se confirmó usando qPCR. Se purificó ARN de la cepa principal y de la cepa 2902 usando un kit de ARN de ZymoResearch. Se usó un kit de RT-PCR Epicentre MasterAmp con cebadores de araB y de actina para la amplificación a partir de ARN. Los integrantes de araB mostraron Cts de aproximadamente 14 frente a aproximadamente 34 para la cepa 1822 y 18 para actina.

Ejemplo 1B: Integración de araD de *B. thetaiotaomicron* en un locus AR de *I. orientalis*:

El gen araD de *B. thetaiotaomicron* se optimizó en el codón para la expresión en *I. orientalis* (SEQ ID NO:15) y se clonó en PCR2.1-TOPO como se ha descrito anteriormente en el Ejemplo 1A, excepto que se usó el sistema de PCR ClonTech Genome Advantage2 en vez de ADN polimerasa rTth. Debido a que los seis clones secuenciados tuvieron al menos un error de nucleótido, el extremo de 5' libre de error (fragmento XbaI/PstI) de un clon se unió con el extremo 3' libre de error (fragmento PstI/PacI) de un segundo clon mediante digestión y ligación. El gen resultante se digirió con XbaI y PacI y se ligó en pHJJ3 similarmente cortado, creando el vector pHJJ5. pHJJ5 contuvo el promotor ENO1, gen araD y terminador PDC. Se unió un fragmento NotI que contenía el promotor ENO1, gen araD, terminador PDC y locus URA3 de pHJJ5 en el vector pHJJ4 para generar los vectores pHJJ9 (orientación 1) y pHJJ10 (orientación 2). pHJJ4 contuvo regiones en la dirección 5' y en la dirección 3' del locus S141G725 de *I. orientalis* (AR, "AXR1").

Se linealizó pHJJ9 por digestión secuencial con SacI y ApaI, liberando un fragmento que contenía el promotor ENO1, gen araD, terminador PDC, casete URA3 y secuencias que se dirigen a AXR1. Los fragmentos de integración se transformaron en la cepa 2904 de *I. orientalis* como se ha descrito anteriormente en el Ejemplo 1A. Se purificaron las colonias transformadas en medio ScD-ura, y se realizó PCR para confirmar la integración en el locus AXR1. La cepa 2904 transformada con pHJJ9 linealizado produjo las cepas yHJJ7 (2908) y yHJJ8 (2909), teniendo cada una una copia de araB y una copia de araD de *B. thetaiotaomicron*.

La cepa 2908 se cultivó durante la noche en medio YPD y se sembró en medio ScD-FOA para seleccionar cepas en las que el marcador URA3 había salido del bucle. Las colonias resultantes se purificaron en medio YPD y se probaron en medio ScD-ura para confirmar la auxotrofia del uracilo. Las colonias también se confirmaron por PCR de colonias. Los derivados ura- de la cepa 2908 fueron las cepas yHJJ13 (3009) y yHJJ14 (3010).

Se linealizó pHJJ10 por digestión secuencial con SacI y ApaI, liberando un fragmento que contenía el promotor ENO1, gen araD, terminador PDC, casete URA3, y secuencias que se dirigen a AXR1. Los fragmentos de integración se transformaron en la cepa 3009 de *I. orientalis*. Se purificaron colonias transformadas en medio ScD-ura, y se realizó PCR para confirmar la integración en el locus AXR1. La cepa 3009 transformada con pHJJ10 linealizado produjo la cepa yHJJ15 (3011), que tenía una copia de araB y dos copias de araD de *B. thetaiotaomicron*.

La expresión de araD se confirmó usando qPCR. Se purificó ARN de las cepas 2908 (araB/araD) y 2904 (araB) usando una extracción con fenol ácido. Se eliminó el ADN genómico usando un kit de ARN libre de ADN de ZymoResearch, y se preparó ADNc a partir de 4 µg de ARN usando transcriptasa inversa de Promega. Se aisló ADN genómico de la cepa 2908 para su uso como un patrón usando un kit de ADN genómico YeaStar de ZymoResearch. Se realizó qPCR usando mezcla maestra de PCR SYBR Green de Applied Biosystems y cebadores de araD y actina. Los integrantes de araD mostraron aproximadamente seis veces más expresión de araD que la expresión de actina, frente a ninguna expresión en la cepa de control de araB.

El marcador URA3 de yHJJ15 (3011) salió del bucle cultivando las células durante la noche en medio YPD y sembrando en placas de ScD-FOA. Las colonias se cribaron por PCR de colonias para identificar colonias que perdieron el marcador de selección, pero retuvieron el resto de la inserción de araD, y una colonia tal se llamó yJY21. Después se confirmó que la copia de araB de *B. thetaiotaomicron* se perdió durante el evento de salida del bucle, de manera que la cepa yJY21 solo tuvo las dos copias de araD.

Ejemplo 1C: Integración de araA de *B. thetaiotaomicron* en un locus XDH de *I. orientalis*:

El gen araA *B. thetaiotaomicron* se optimizó en el codón para la expresión en *I. orientalis* como se ha descrito anteriormente en el Ejemplo 1A (SEQ ID NO:5) y se sintetizó. Se usó mutagénesis dirigida al sitio para corregir los errores de nucleótidos en el gen ensamblado. Un clon que lleva el vector con la secuencia de genes deseada se llamó pJY13.

Se realizó una ligación de tres trozos usando un fragmento XbaI/PacI que contenía el gen araA de *B. thetaiotaomicron*, un fragmento XhoI/PacI de un vector de clonación que contenía secuencias que se dirigían a XYL2 (XDH), un terminador PDC y un casete de selección URA3, y un fragmento XhoI/XbaI que contenía el promotor TDH3 de *I. orientalis*. El plásmido pJY15 resultante contuvo el promotor TDH3, el gen araA de *B. thetaiotaomicron*, terminador PDC y el casete de marcador URA3 flanqueado por secuencias que se dirigen a XYL2.

El plásmido pJY15 se digirió con ApaI y KpnI para liberar el fragmento de integración, y se transformó ADN linealizado en la cepa 2904 del Ejemplo 1A (contiene el gen araB de *B. thetaiotaomicron* en el locus XYL1). Se

cribaron por PCR de colonias *ura+* para identificar colonias con integración en el locus deseado, y una de tales cepas se llamó yJY16. La cepa yJY16 contuvo una copia de cada uno de los genes *araB* y *araA* de *B. thetaiotaomicron*, y se usó para probar la actividad de AI con respecto a otras fuentes del gen *araA*.

También se transformó el fragmento de integración linealizado de pJY15 en la cepa yJY21 del Ejemplo 1B (contiene dos copias del gen *araD* de *B. thetaiotaomicron* en el locus S141G725). Se cribaron colonias *ura+* por PCR de colonias para identificar colonias con integración en el locus *XYL2*, y una de tales cepas se llamó yJY22. La cepa yJY22 contuvo una copia del gen *araA* de *B. thetaiotaomicron* y dos copias del gen *araD* de *B. thetaiotaomicron*.

El marcador *URA3* en yJY22 salió del bucle sembrando en placas de ScD-FOA. Las colonias se cribaron por PCR para identificar colonias que perdieron el marcador de selección pero retuvieron el resto de la inserción de *araA*, y una colonia tal se llamó yJY23.

La cepa yJY23 se transformó con el fragmento de integración *Apal/KpnI* de pJY15, y se cribaron por PCR colonias *ura+* para identificar colonias con integración en el locus deseado. Una cepa tal se llamó yJY24, que tuvo dos copias de cada uno de los genes *araA* y *araD* de *B. thetaiotaomicron*.

El marcador *URA3* de la cepa yJY24 salió del bucle sembrando células en placas de ScD-FOA. Las colonias se cribaron por PCR para identificar colonias que perdieron el marcador de selección pero retuvieron el resto de la inserción, y una colonia tal se llamó yJY29.

Se digirió el plásmido pHJJ3 (Ejemplo 1A) con *Apal* y *SacI* para liberar el fragmento de integración que contenía el gen *araB* de *B. thetaiotaomicron*, y se transformó ADN linealizado en la cepa yJY29. Se cribaron por PCR colonias *ura+* para identificar colonias con integración en el sitio *XYL1*, y una de tales cepas se llamó yJY30 (3409). La cepa 3409 tuvo dos copias de cada uno de los genes *araA* y *araD* de *B. thetaiotaomicron* y una copia del gen *araB*.

El marcador *URA3* de la cepa 3409 salió del bucle sembrando células en placas de ScD-FOA. Las colonias se cribaron por PCR para identificar colonias que perdieron el marcador de selección pero retuvieron el resto de la inserción, y una colonia tal se llamó yJY31. Se transformó el fragmento de integración linealizado a partir de pHJJ3 en la cepa yJY31 con el fin de insertar una segunda copia del gen *araB* en el sitio *XYL1*. Se cribaron por PCR colonias *ura+* para identificar colonias con integración en el locus deseado, y una de tales cepas se llamó la cepa yJY33 (3410).

Se amplificó una región de ADN que contenía el promotor *TEF1* a partir de ADN genómico de *I. orientalis* de manera que el extremo 5' contuviera un sitio de restricción *XhoI* y el extremo 3' contuviera un sitio *XbaI*. El producto de PCR cortado con *XhoI/XbaI* se unió en los plásmidos pHJJ3 y pHJJ18 (Ejemplo 1A) que habían sido similarmente digeridos para liberar el promotor *ENO1*. Las colonias transformadas con la ligación se cribaron por PCR para el inserto deseado y se confirmó por secuenciación. Estos vectores, que contuvieron el gen *araB* de *B. thetaiotaomicron* bajo el control del promotor *ENO1*, se llamaron pHJJ33 (derivado de pHJJ3) y pHJJ35 (derivado de pHJJ18).

El plásmido pHJJ33 se digirió con *Apal* y *SacI* para liberar el fragmento de integración, y el ADN linealizado se transformó en la cepa yJY29. Se cribaron por PCR colonias *ura+* para identificar colonias con integración en el sitio *XYL1*, y una de tales cepas se llamó yHJJ40 (3406). La cepa 3406 contuvo dos copias de los genes *araA* y *araD* de *B. thetaiotaomicron* y una copia de *araB* bajo el control del promotor *TEF1*.

El marcador *URA3* de 3406 salió del bucle cultivando células durante la noche en YPD y sembrando en placas de ScD-FOA. Las colonias se cribaron por PCR para identificar colonias que perdieron el marcador de selección pero retuvieron el resto de la inserción. Una colonia tal se llamó yHJJ44. El plásmido pHJJ35 se digirió con *Apal* y *SacI* para liberar el fragmento de integración, y el ADN linealizado se transformó en la cepa yHJJ44 para insertar *TEF1:araB* en un segundo locus *XYL1*. Se cribaron por PCR colonias *ura+* para identificar colonias con integración correcta, y una de tales cepas se llamó la cepa yHJJ47 (3408). La cepa 3408 contuvo dos copias de cada uno de los genes *araA*, *araD* y *araB* de *B. thetaiotaomicron*, con *araB* bajo el control del promotor *TEF1*.

Ejemplo 1D: Integración de *araD* de *E. coli* en el locus *AR* de la cepa de *I. orientalis* que contiene *araB* de *B. thetaiotaomicron*:

Se amplificó el gen *araD* de *E. coli* (SEQ ID NO:76) a partir de ADN genómico de la cepa MG1655 de manera que el extremo 5' del gen contuvo un sitio de restricción *XbaI* y el extremo 3' contuvo un sitio de restricción *PacI*. El producto de PCR se purificó en gel y se digirió con *XbaI* y *PacI*. El fragmento resultante se ligó en pHJJ18 (Ejemplo 1A) de los que habían sido digeridos el gen *araB* de *B. thetaiotaomicron* con *XbaI* y *PacI*. Las colonias que tenían el inserto de *araD* de *E. coli* se confirmaron por PCR, y se aisló ADN de plásmido (pHJJ12). El fragmento que contenía el promotor *ENO1*, gen *araD* de *E. coli*, terminador *PDC* y el casete de marcador *URA3* se digirió a partir de pHJJ12 con *NotI* y se ligó en pHJJ4 digerido con *NotI* (las secuencias que se dirigen a *AXR1* separadas por un sitio *NotI*) para obtener vectores pHJJ14 (orientación 1) y pHJJ19 (orientación 2).

El plásmido pHJJ14 se digirió con *Apal* y *SacI* para liberar el fragmento de integración, y el ADN linealizado se transformó en la cepa 2904 (Ejemplo 1A). Se cribaron por PCR colonias *ura+* para identificar colonias con

integración en el locus deseado, y una de tales cepas se llamó yHJJ9 (3005).

Ejemplo 1E: Integración de araD de *L. plantarum* en el locus AR de la cepa de *I. orientalis* que contiene araB de *B. thetaiotaomicron*:

El gen araD de *L. plantarum* se optimizó en el codón para la expresión en *I. orientalis* usando una tabla de uso de codones de *I. orientalis* para la traducción inversa y se sintetizó de manera que contuvo un sitio de restricción XbaI en el extremo 5' y un sitio de restricción PstI en el extremo 3' (SEQ ID NO:78). El producto de PCR de araD de *L. plantarum* se purificó en gel y se digirió con XbaI y PstI. Los fragmentos resultantes se unieron en pHJJ18 (Ejemplo 1A) de los que habían sido digeridos el gen araB de *B. thetaiotaomicron* con XbaI y PstI. Las colonias que tenían el inserto de araD de *L. plantarum* deseado se confirmaron por PCR, y se aisló ADN de plásmido (pHJJ13). El fragmento que contenía el promotor ENO1, araD de *L. plantarum*, terminador PDC y el casete de marcador URA3 se digirió a partir de pHJJ13 con NotI y se ligó en pHJJ4 digerido con NotI (las secuencias que se dirigen a AXR1 se separaron por un sitio NotI) para obtener vectores pHJJ15 (orientación 1) y pHJJ20 (orientación 2).

El plásmido pHJJ15 se digirió con ApaI y SacI para liberar el fragmento de integración, y el ADN linealizado se transformó en la cepa 2904 (Ejemplo 1 A). Se cribaron por PCR colonias ura⁺ para identificar colonias con integración en el locus AXR1, y una de tales cepas se llamó yHJJ11 (3007).

Ejemplo 1F: Integración de araA de *L. plantarum* en el locus XDH de la cepa de *I. orientalis* que contiene araB de *B. thetaiotaomicron*:

El gen araA de *L. plantarum* se optimizó en el codón para la expresión en *I. orientalis* usando una tabla de uso de codones de *I. orientalis* para la traducción inversa y se sintetizó de manera que contuvo un sitio de restricción XbaI en el extremo 5' y un sitio de restricción PstI en el extremo 3' (SEQ ID NO:80). El ADN se clonó en TOPO y el plásmido con la secuencia deseada se llamó pJY14.

Se realizó una ligación de tres trozos usando un fragmento XbaI/PstI de pJY14 que contenía el gen araA de *L. plantarum*, un fragmento XhoI/PstI que contenía secuencias que se dirigían a XYL2 (XDH), un terminador PDC de *I. orientalis* y un casete de selección URA3, y un fragmento XhoI/XbaI que contenía el promotor TDH3 de *I. orientalis*. El plásmido pJY17 resultante contuvo el promotor TDH3, gen araA de *L. plantarum*, terminador PDC y el casete de marcador URA3 flanqueado por secuencias que se dirigen a XYL2.

El plásmido pJY17 se digirió con ApaI y KpnI para liberar el fragmento de integración, y el ADN linealizado se transformó en la cepa 2904 (Ejemplo 1A). Se cribaron por PCR colonias ura⁺ para identificar colonias con integración en el locus XYL2, y una de tales cepas se llamó yJY17.

Ejemplo 1G: Integración de araA de *B. licheniformis* en el locus XDH de la cepa de *I. orientalis* que contiene araB de *B. thetaiotaomicron*:

El gen araA de *B. licheniformis* de 1,5 Kb se optimizó en el codón para la expresión en *I. orientalis* usando una tabla de uso de codones de *I. orientalis* para la traducción inversa y se construyó de manera que contuvo un sitio de restricción XbaI en el extremo 5' y un sitio de restricción PstI en el extremo 3' (SEQ ID NO:82). El producto de PCR se clonó en un vector TOPO, y se usó mutagénesis dirigida para corregir tres errores de nucleótidos. El plásmido resultante pJY23 contuvo el gen araA de *B. licheniformis* de codón optimizado correcto.

Se realizó una ligación de tres trozos usando un fragmento XbaI/PstI de pJY23 que contenía el gen araA de *B. licheniformis*, un fragmento XhoI/PstI de un vector de clonación que contenía secuencias que se dirigían a XYL2 (XDH), un terminador PDC y un casete de selección URA3, y un fragmento XhoI/XbaI que contenía el promotor TDH3 de *I. orientalis*. El plásmido resultante pJY24 contuvo el promotor TDH3, gen araA de *B. licheniformis*, terminador PDC y el casete de marcador URA3 flanqueado por secuencias que se dirigen a XYL2.

El plásmido pJY24 se digirió con ApaI y KpnI para liberar el fragmento de integración, y el ADN linealizado se transformó en la cepa 2904 (Ejemplo 1A). Se cribaron por PCR colonias ura⁺ para identificar colonias con integración en el locus XYL2, y una de tales cepas se llamó yJY18.

Las cepas de *I. orientalis* genéticamente modificadas generadas en los Ejemplos 1A a 1G se resumen en la Tabla 1.

Tabla 1:

Nombre de la cepa		araA	araB	araD
2762 (cepa principal)	Fuente	--	--	--
	N.º de copias	0	0	0
	Promotor	--	--	--
	Localización	--	--	--

Nombre de la cepa		araA	araB	araD
yHJJ2/2902 (ura+), yHJJ3/2904 (ura-)	Fuente	--	<i>B. thetaiotaomicron</i>	--
	N.º de copias	--	1	--
	Promotor	--	ENO1	--
	Localización	--	locus XYL1	--
yHJJ1/2903 (ura+), yHJJ4/2905 (ura-)	Fuente	--	<i>B. thetaiotaomicron</i>	--
	N.º de copias	0	1	0
	Promotor	--	ENO1	--
	Localización	--	locus XYL1	--
yHJJ7/2908 (ura+), yHJJ8/2909 (ura+), yHJJ13/3009 (ura-), yHJJ14/3010 (ura-)	Fuente	--	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>
	N.º de copias	0	1	1
	Promotor	--	ENO1	ENO1
	Localización	--	locus XYL1	locus AXR1
yHJJ9/3005	Fuente	--	<i>B. thetaiotaomicron</i>	<i>E. coli</i>
	N.º de copias	0	1	1
	Promotor	--	ENO1	ENO1
	Localización	--	locus XYL1	locus AXR1
yHJJ11/3007	Fuente	--	<i>B. thetaiotaomicron</i>	<i>L. plantarum</i>
	N.º de copias	0	1	1
	Promotor	--	ENO1	ENO1
	Localización	--	locus XYL1	locus AXR1
yHJJ15/3011	Fuente	--	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>
	N.º de copias	0	1	2
	Promotor	--	ENO1	ENO1
	Localización	--	locus XYL1	locus AXR1
yJY16	Fuente	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>	--
	N.º de copias	1	1	0
	Promotor	TDH3	ENO1	--
	Localización	locus XYL2	locus XYL1	--
yJY17	Fuente	<i>L. plantarum</i>	<i>B. thetaiotaomicron</i>	--
	N.º de copias	1	1	0
	Promotor	TDH3	ENO1	--
	Localización	locus XYL2	locus XYL1	--
yJY18	Fuente	<i>B. licheniformis</i>	<i>B. thetaiotaomicron</i>	--
	N.º de copias	1	1	0
	Promotor	TDH3	ENO1	--
	Localización	locus XYL2	locus XYL1	--
yJY21	Fuente	--	--	<i>B. thetaiotaomicron</i>
	N.º de copias	0	0	2
	Promotor	--	--	ENO1
	Localización	--	--	locus AXR1
yJY22 (ura+), yJY23 (ura-)	Fuente	<i>B. thetaiotaomicron</i>	--	<i>B. thetaiotaomicron</i>
	N.º de copias	1	0	2

Nombre de la cepa		araA	araB	araD
	Promotor	TDH3	--	ENO1
	Localización	locus XYL2	--	locus AXR1
yJY24 (ura+), yJY29 (ura-)	Fuente	<i>B. thetaiotaomicron</i>	--	<i>B. thetaiotaomicron</i>
	N.º de copias	2	0	2
	Promotor	TDH3	--	ENO1
	Localización	locus XYL2	--	locus AXR1
yHJJ40/3406 (ura+), yHJJ44 (ura-)	Fuente	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>
	N.º de copias	2	1	2
	Promotor	TDH3	TEF1	ENO1
	Localización	locus XYL2	locus XYL1	locus AXR1
yHJJ47/3408	Fuente	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>
	N.º de copias	2	2	2
	Promotor	TDH3	TEF1	ENO1
	Localización	locus XYL2	locus XYL1	locus AXR1
yJY30/3409 (ura+), yJY31 (ura-)	Fuente	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>
	N.º de copias	2	1	2
	Promotor	TDH3	ENO1	ENO1
	Localización	locus XYL2	locus XYL1	locus AXR1
yJY33/3410	Fuente	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>
	N.º de copias	2	2	2
	Promotor	TDH3	ENO1	ENO1
	Localización	locus XYL2	locus XYL1	locus AXR1

Ejemplo 2: Análisis de la actividad de RK, RE y AI en cepas de *I. orientalis* que contienen genes araA, araB y/o araD bacterianos:

Se probaron las cepas generadas en el Ejemplo 1 para actividad de RK, RE y AI.

5 Ejemplo 2A: Análisis de actividad de RK:

RK cataliza la conversión dependiente de ATP de L-ribulosa en L-ribulosa 5-fosfato, produciendo ADP. La actividad de RK va seguida de la regeneración de ATP con PEP catalizado por piruvato cinasa. Esta reacción produce piruvato, que se reduce a lactato con NADH y lactato deshidrogenasa.

10 Los ensayos contuvieron Tris HCl 30 mM, pH 7,5, MgCl₂ 3,3 mM, EDTA 0,3 mM, PEP 1,7 mM, ATP 0,7 mM, >4 U/ml de cada piruvato cinasa y lactato deshidrogenasa (premezclados PK+LDH de Sigma), ribulosa 2 mM, NADH 0,5 mM y extracto de células. En ensayos iniciales, se usó D-ribulosa como sustrato. En ensayos posteriores, se usó L-ribulosa (ZuChem). Debido al nivel de expresión generalmente alto de esta enzima, los extractos se diluyeron 10 veces en NaTES 50 mM, pH 7,0, NaCl 100 mM, MnCl₂ 0,1 mM, 0,01 % (v/v) de Tween 20. La reacción se llevó a cabo a temperatura ambiente, y el cambio en la absorbancia a 340 nm se monitorizó durante 10 minutos a intervalos de 15 segundos. Los ensayos se llevaron a cabo en pocillos de microtítulo con un volumen de ensayo final de 200 µl. La reacción se inició mediante la adición de NADH solo o con L-ribulosa. La ΔA₃₄₀ medida se convirtió en mM usando una longitud de trayectoria eficaz de 0,576 cm (determinada midiendo la absorbancia de una solución de NADH bajo estas condiciones frente a la medida en una cubeta de 1 cm, y aplicando la ley de Beer).

20 En ensayos con D-ribulosa como sustrato, se midió una actividad específica neta de 1,0 unidad/mg de proteína en extractos en bruto de la cepa 2902 (1 copia de araB de *B. thetaiotaomicron*). En ensayos con L-ribulosa como sustrato, la actividad específica de RK en extractos de la cepa 3409 (1 copia de araB de *B. thetaiotaomicron*, 2 copias de cada uno de los genes araA y araD de *B. thetaiotaomicron*) fue 1,4 unidades/mg de proteína.

25 Debido a que el ensayo de RK mide la producción de ADP que puede surgir de cualquier actividad de cinasa, tiene una alta actividad de fondo en ausencia de L-ribulosa (aproximadamente 1/3 como mucho en presencia de L-ribulosa). Esta actividad de fondo está presente en la cepa principal, y no aumenta cuando se añade L-ribulosa a

ensayos con extractos de estas células. La actividad de fondo no disminuye sustancialmente en extractos dializados, sugiriendo que las cinasas utilizan sustratos macromoleculares tales como proteínas o ácidos nucleicos. En las actividades específicas enumerada anteriormente, la actividad de fondo en ausencia de sustrato se resta de la actividad medida en presencia de L-ribulosa.

5 Ejemplo 2B: Análisis de la actividad de RE:

RE interconvierte L-ribulosa 5-P y D-xilulosa 5-P. L-ribulosa 5-P no está comercialmente disponible, y así necesita prepararse tanto en una reacción separada como en una reacción acoplada por RK. Como RK de *B. thetaiotaomicron* se expresa altamente y/o es activo en *I. orientalis*, los extractos de células con RK generalmente tienen actividad de RK en exceso con respecto a RE, que significa que producen un exceso de L-ribulosa 5-P a partir de L-ribulosa y ATP.

Se detecta D-xilulosa 5-P en un esquema de reacción acoplado añadiendo D-ribosa 5-P y TKL más tiamina PP para generar S7P más G3P; convirtiendo G3P en dihidroxiacetona P (DHAP) con triosafosfato isomerasa (TPI); y reduciendo DHAP a glicerol 3-P con glicerol 3-P deshidrogenasa (G3PDH) dependiente de NADH. La actividad de RE se acopla así a la oxidación de NADH. Las enzimas de acoplamiento se añaden exógenamente, pero es probable que también estén presentes en el extracto, que no debe afectar los resultados ya que todos deberían estar en exceso con respecto a la actividad de RE.

Los ensayos contuvieron Tris HCl 50 mM, pH 7,5, MgCl₂ 3,3 mM, ATP 2 mM, TPP 0,1 mM, D-ribosa 5-P 1 mM, transcetolasa 0,05 U/ml, 4,5 U/ml de TPI, 1,5 U/ml de G3PDH, L-ribulosa 4 mM y NADH 0,5 mM. La reacción se llevó a cabo a temperatura ambiente, y el cambio en la absorbancia a 340 nm se monitorizó durante 10 minutos a intervalos de 15 segundos. Los ensayos se llevaron a cabo en pocillos de microtítulo con un volumen de ensayo final de 200 µl. La reacción se inició mediante la adición de NADH solo o con L-ribulosa. La ΔA₃₄₀ medida se convirtió en mM usando una longitud de trayectoria eficaz de 0,576 cm (determinada midiendo la absorbancia de una solución de NADH bajo estas condiciones frente a la medida en una cubeta de 1 cm, y aplicando la ley de Beer).

Después de restarse la actividad de fondo, la cepa 2908 de *I. orientalis*, que contuvo copias únicas de los genes araB y araB de *B. thetaiotaomicron*, tuvo 0,022 U/mg de actividad de RE. La cepa 3005, que contuvo araD de *E. coli* y araB de *B. thetaiotaomicron*, tuvo actividad similar a la cepa 2908. La cepa 3007, que contuvo araD de *L. plantarum* y araB de *B. thetaiotaomicron*, tuvo aproximadamente la mitad de la actividad específica de las otras dos cepas.

Ejemplo 2C: Análisis de actividad de AI:

Los ensayos de AI contuvieron Na-TES 100 mM, pH 7,0, MnCl₂ 0,3 mM, 37,5 unidades/ml de sorbitol deshidrogenasa (SIGMA S3764), NADH 0,5 mM, L-arabinosa 66,7 mM y extracto de células. La reacción se llevó a cabo a temperatura ambiente, y el cambio en la absorbancia a 340 nm se monitorizó durante 10 minutos a intervalos de 15 segundos. Los ensayos se llevaron a cabo en pocillos de microtítulo a un volumen de ensayo final de 200 µl. La reacción se inició mediante la adición de NADH solo o con L-arabinosa. La ΔA₃₄₀ medida se convirtió en mM usando una longitud de trayectoria eficaz de 0,576 cm (determinada midiendo la absorbancia de una solución de NADH bajo estas condiciones frente a la medida en una cubeta de 1 cm, y aplicando la ley de Beer).

La cepa yJY16 (1 copia de cada uno de araA y araB de *B. thetaiotaomicron*) presentó una actividad específica de 0,045 unidades/mg, mientras que yJY17 (una copia de cada uno de araA de *L. plantarum* y araB de *B. thetaiotaomicron*) y yJY18 (una copia de cada uno de araA de *B. licheniformis* y araB de *B. thetaiotaomicron*) presentó actividades específicas de 0,012 y 0,010 unidades/mg, respectivamente. Se midieron actividades específicas significativamente más altas en extractos de células que llevan dos copias de araA de *B. thetaiotaomicron* (yJY24 y 3409), que puede ser un reflejo de la inestabilidad de heterocromosomas en *I. orientalis*.

La actividad de AI en extractos de la cepa 3409 que lleva la vía de arabinosa completa fue 0,24 U/mg. Este número fue superior al medido en el experimento preliminar por varios motivos: esta cepa llevó dos copias del gen araA de *B. thetaiotaomicron* integrado en lugar de una; se optimizó la cantidad de enzima de acoplamiento sorbitol deshidrogenasa; y mejoró el crecimiento celular y la extracción de proteína.

Ejemplo 3: Caracterización de cepas de *I. orientalis* que contienen genes de la vía de arabinosa de *B. thetaiotaomicron*:

Se probaron dos cepas hermanas de la cepa 3409 (yJY30.1 y yJY30.2) y 1822 (cepa principal) para su capacidad para fermentar arabinosa a etanol. Las cepas yJY30.1 y yJY30.2 contuvieron copias de los tres genes de *B. thetaiotaomicron* de la vía de arabinosa (dos copias de cada uno de araA y araD, una copia de araB). Las células se cultivaron en YP más 40 g/l de arabinosa a 37 °C y 100 rpm. Después de 132 horas, ambas cepas de la vía de arabinosa de *B. thetaiotaomicron* consumieron aproximadamente 22 g/l de arabinosa mientras que producían aproximadamente 4 g/l de etanol (Figura 2).

Ejemplo 4: Integración de genes de *B. thetaiotaomicron* de la vía de arabinosa en la cepa de *I. orientalis* manipulada para utilizar xilosa:

La cepa 3489 de *I. orientalis* se había manipulado previamente para fermentar xilosa a etanol. La cepa 3489 contuvo cuatro copias de un gen exógeno que codificaba XI de *B. thetaiotaomicron*, dos copias de un gen exógeno nativo que codifica XK, dos copias de cada uno de los genes nativos exógenos de la vía de pentosa fosfato no oxidativa TAL, RKI y RPE, y dos copias de un gen ADH1 exógeno nativo, todos bajo el control de promotores glicolíticos exógenos nativos fuertes. La construcción de esta cepa también incluyó delecciones de genes para XR, XDH, S141G725, S141G4738, S141G1158-1159, S141G8160 y GAL6. La cepa 3489 fue incapaz de fermentar arabinosa a etanol. Por tanto, los genes *araA* y *araD* de *B. thetaiotaomicron* se insertaron en la cepa 3489 en el sitio de un supuesto homólogo de ADH de *I. orientalis* (SEQ ID NO:74, locus S141G9091), y el gen *araB* de *B. thetaiotaomicron* se insertó en el sitio de un segundo homólogo de ADH (SEQ ID NO:84, locus S141G1202).

Ejemplo 4A: Construcción del vector de inserción de *araA* y *araD* de *B. thetaiotaomicron*:

Para generar un vector de inserción con el gen *araA* de *B. thetaiotaomicron* unido al promotor TDH3 de *I. orientalis* y el terminador TAL, se unió un fragmento XhoI/PacI del vector pJY39 que contenía el terminador TAL, sitios que se dirigen a XYL1, y un casete de selección URA3 a un fragmento XhoI/PacI de pJY15 (Ejemplo 1C) que contenía el promotor TDH3 de *I. orientalis* y gen *araA* de *B. thetaiotaomicron*. Los plásmidos de las colonias transformadas con la mezcla de ligación se cribaron por digestos de enzima de restricción con HindIII y SphI, y el plásmido correcto se llamó pLUN111.

Para combinar los genes *araA* y *araD* en un único plásmido, un fragmento AscI/ApaI de pLUN111 que contenía el promotor TDH3 de *I. orientalis*, gen *araA* de *B. thetaiotaomicron*, terminador TAL de *I. orientalis*, casete de selección URA3, y secuencia que se dirige en la dirección 3' se unió a un fragmento AscI/ApaI de pJY33, que contuvo el esqueleto del vector, en la secuencia que se dirige en la dirección 5', promotor ENO1 de *I. orientalis*, gen *araD* de *B. thetaiotaomicron* y terminador PDC de *I. orientalis*. Los plásmidos de las colonias transformadas con la ligación se cribaron para la inserción deseada por digestión con restricción. El plásmido resultante, pLUN112, contuvo el gen *araA* de *B. thetaiotaomicron* bajo el control del promotor TDH3 y el gen *araD* de *B. thetaiotaomicron* bajo el control del promotor ENO1.

Un fragmento NotI de pLUN112 que contenía los genes *araA* y *araD* y sus elementos reguladores, además del casete de selección URA3, se unió en el vector pHJJ22 cortado con NotI. pHJJ22 contuvo las regiones en la dirección 5' y en la dirección 3' del gen 9091 separadas por un sitio NotI. Así, la ligación insertó *araA/araD* entre las secuencias flanqueantes e 9091. Se cribaron las colonias resultantes de la transformación de la ligación para la presencia del inserto deseado por PCR de colonias. Los plásmidos se aislaron para clones que tenían los productos de PCR esperados, y los plásmidos aislados se cribaron para orientación del inserto de *araA/araD* usando una digestión con restricción SphI. Los plásmidos se llamaron pLUN113 (orientación 1) y pLUN114 (orientación 2).

Ejemplo 4B: Construcción del vector de inserción de *araB* de *B. thetaiotaomicron*:

Para generar un vector de inserción con el gen *araB* de *B. thetaiotaomicron* unido al promotor ENO1 entre las regiones flanqueantes del gen 1202, pHJJ74 cortado con NotI, un vector que contenía secuencias diana en la dirección 5' y en la dirección 3' separadas por un sitio NotI, se unió al inserto de NotI de pHJJ2 (Ejemplo 1 A), que contuvo el promotor ENO1 de *I. orientalis*, gen *araB* de *B. thetaiotaomicron*, terminador PDC de *I. orientalis*, y casete de selección URA3. Se aisló ADN de plásmido de las colonias transformadas con la ligación y se cribó por digestión con SphI y XhoI. Los plásmidos se llamaron pLUN125 (orientación 1) y pLUN126 (orientación 2).

Para generar un vector de inserción con el gen *araB* unido al promotor TEF1 entre las regiones flanqueantes del gen 1202, pHJJ74 cortado con NotI se unió al inserto de NotI de pHJJ33 (Ejemplo 1 C), que contuvo el promotor TEF1 de *I. orientalis*, gen *araB* de *B. thetaiotaomicron*, terminador PDC de *I. orientalis* y casete de selección URA3. Se aisló ADN de plásmido de las colonias transformadas con la ligación y se cribó por digestión con SphI y XhoI. Los plásmidos se llamaron pLUN127 (orientación 1) y pLUN128 (orientación 2).

Ejemplo 4C: Integración de *araA* y *araD* de *B. thetaiotaomicron* en la cepa 3514 de *I. orientalis*

Se transformó la cepa de *I. orientalis* yHJJ84 (3514), un derivado ura- de la cepa 3489, con pLUN113 linealizado con ApaI/SacI (Ejemplo 4A) y se sembró en medio ScD-ura. El ADN genómico de los transformantes purificados se cribó por PCR y los clones identificados por tener un casete de *araA/araD* correctamente insertado se llamaron yARA21.

El gen de marcador URA3 de yARA21 salió del bucle cultivando células durante la noche en YPD y sembrando en placas de ScD-FOA. Se cribó por PCR el ADN genómico preparado a partir de colonias salidas del bucle a través de ambos empalmes por integración y se identificó una colonia (yARA22) por haber retenido la integración, pero perdió el gen URA3. Esta colonia se sembró sobre placas de ScD-ura para verificar que carecían de crecimiento sin suplementación con uracilo.

Se añadió el segundo casete de integración de *araA/araD* a yARA22 transformando con pLUN114 linealizado con ApaI/SacI (Ejemplo 4A). Los transformantes se seleccionaron sobre placas de ScD-ura y se purificó una colonia

individual. Los dos empalmes para cada locus se cribaron en reacciones de PCR separadas. Los clones identificados como que tenían ambas copias de araA/araD insertadas correctamente se llamaron yARA25.

El gen de marcador URA3 de yARA25 salió del bucle cultivando células durante la noche en YPD y sembrando en placas de ScD-FOA. Se cribaron colonias salidas del bucle en dos reacciones de PCR separadas para identificar colonias que retuvieron la integración deseada. Estos clones se sembraron en placa sobre ScD-ura para verificar que carecían de crecimiento sin suplementación de uracilo. Los clones salidos del bucle correctos se llamaron yARA26.

Ejemplo 4D: Integración de araB de *B. thetaiotaomicron* en la cepa yARA26 de *I. orientalis*:

Para integrar la primera copia del gen araB de *B. thetaiotaomicron* unido al promotor ENO1, se transformó la cepa yARA26 de *I. orientalis* (Ejemplo 4C) con pLUN125 linealizado con Apal/SacI (Ejemplo 4B). Los transformantes se seleccionaron en placas de ScD-ura y se cribaron por PCR a través de ambos empalmes de integración. Los clones identificados por tener el fragmento de ENO1:araB insertado en el locus 1202 se llamaron yARA29.

Para integrar la primera copia del gen araB de *B. thetaiotaomicron* unido al promotor TEF1, se transformó la cepa yARA26 de *I. orientalis* (Ejemplo 4C) con pLUN127 linealizado con Apal/SacI (Ejemplo 4B). Los transformantes se seleccionaron en placas de ScD-ura y se cribaron por PCR a través de ambos empalmes de integración. Los clones identificados por tener el fragmento de TEF1:araB insertado en el locus 1202 se llamaron yARA30.

Para sacar del bucle el gen de marcador URA3 de yARA29 y yARA30, ambas cepas se cultivaron en YPD durante la noche y se sembraron en medio ScD-FOA. Se lisaron colonias individuales de yARA29 y se cribaron en dos reacciones de PCR separadas. Para las salidas de bucle de yARA30, se preparó ADN genómico y se cribó en dos reacciones de PCR separadas. Las colonias sacadas del bucle se cribaron por PCR para identificar aquellas que habían perdido el gen URA3, pero retuvieron la integración de araB. Estas cepas se sembraron reproducidas en ScD-ura para verificar que carecían de crecimiento sin uracilo. Las salidas de bucle correctas de yARA29 se llamaron yARA33, y las salidas de bucle correctas de yARA30 se llamaron yARA34.

Para integrar la segunda copia de araB unido al promotor ENO1, la cepa yARA33 se transformó con pLUN126 linealizado con Apal/SacI (Ejemplo 4B). Los transformantes se seleccionaron en ScD-ura y se cribaron en reacciones de PCR separadas para verificar ambos empalmes en el evento de integración. Los clones identificados por tener la segunda copia de ENO1:araB correctamente integrada se designaron la cepa yARA36 (3936). Estos clones contuvieron dos copias de cada uno de los genes araA, araB y araD de *B. thetaiotaomicron*, con los genes araB bajo el control del promotor ENO1.

Para integrar la segunda copia de araB unido al promotor TEF1, la cepa yARA34 se transformó con el fragmento de integración linealizado de pLUN128 digerido con Apal/SacI (Ejemplo 4B). Se purificó el ADN genómico de transformantes ura⁺ y se cribó en reacciones de PCR separadas para verificar ambos empalmes del evento de integración. Los clones identificados por tener la segunda copia de TEF1:gen araB correctamente integrada se designaron cepa yARA38 (3937). Estos clones contuvieron dos copias de cada uno de los genes araA, araB y araD de *B. thetaiotaomicron*, con los genes araB bajo el control del promotor TEF1.

Las cepas de *I. orientalis* genéticamente modificadas generadas en los Ejemplos 4C y 4D se resumen en la Tabla 2.

Tabla 2:

Nombre de la cepa		araA	araB	araD
3489 (ura⁺), 3514/yHJJ84 (ura⁻) (cepa principal fermentadora de xilosa)	Fuente	--	--	--
	N.º de copias	0	0	0
	Promotor	--	--	--
	Localización	--	--	--
yARA21 (ura⁺), yARA22 (ura⁻)	Fuente	<i>B. thetaiotaomicron</i>	--	<i>B. thetaiotaomicron</i>
	N.º de copias	1	0	1
	Promotor	TDH3	--	ENO1
	Localización	9091	--	9091
yARA25 (ura⁺), yARA26 (ura⁻)	Fuente	<i>B. thetaiotaomicron</i>	--	<i>B. thetaiotaomicron</i>
	N.º de copias	2	0	2
	Promotor	TDH3	--	ENO1
	Localización	9091	--	9091
yARA29 (ura⁺), yARA33	Fuente	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>

Nombre de la cepa		araA	araB	araD
(ura-)	N.º de copias	2	1	2
	Promotor	TDH3	ENO1	ENO1
	Localización	9091	1202	9091
yARA30 (ura+), yARA34 (ura-)	Fuente	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>
	N.º de copias	2	1	2
	Promotor	TDH3	TEF1	ENO1
	Localización	9091	1202	9091
3936/yARA36	Fuente	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>
	N.º de copias	2	2	2
	Promotor	TDH3	ENO1	ENO1
	Localización	9091	1202	9091
3937/yARA38(ura+), yLUN011 (ura-)	Fuente	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>	<i>B. thetaiotaomicron</i>
	N.º de copias	2	2	2
	Promotor	TDH3	TEF1	ENO1
	Localización	9091	1202	9091

Ejemplo 5: Caracterización de cepas de *I. orientalis* que utilizan xilosa manipuladas para contener genes de la vía de arabinosa de *B. thetaiotaomicron*:

Se caracterizaron cepas de la vía doble 3936 de *I. orientalis* (Ejemplo 4D; dos copias cada una de TDH3:araA, ENO1:araB y ENO1:araD) y 3937 (Ejemplo 4D; dos copias cada una de TDH3:araA, TEF1:araB y ENO1:araD) usando un experimento en matraz oscilante. Las cepas de control para estos experimentos fueron la cepa de la vía de arabinosa 3408 (Ejemplo 1C) y la cepa de la vía de xilosa yHJJ169 (3922). La cepa 3922 contiene los mismos antecedentes genéticos que la cepa 3489 (cepa que utiliza xilosa de las que se derivaron las cepas 3936 y 3937), junto con deleciones en los sitios 9091 y 1202 que sirvieron de sitios de integración para genes de la vía de arabinosa en las cepas de la vía doble. Así, la única diferencia genética entre la cepa 3922 y las cepas de la vía doble 3936/3937 es la presencia de los genes de la vía de arabinosa en la última.

Todas las cepas se cultivaron aeróbicamente durante la noche en YP con 20 g/l de arabinosa, y se calculó la cantidad de cultivo necesaria para inocular a una $DO_{600}=0,8$. El volumen calculado de cultivo se centrifugó a 4000 rpm durante cuatro minutos y el sedimento de células se resuspendió en 500 µl de YP + 20 g/l de arabinosa. Éste se usó para inocular matraces con agitación fermentativos a $DO_{600}=0,8$. Debido al crecimiento residual en YP, este protocolo fue suficiente para recoger biomasa suficiente para inocular la cepa 3922.

Las cepas 3936 y 3937 se comportan similarmente a o ligeramente mejor que la cepa 3408 con respecto al consumo de arabinosa y la producción de etanol (Figura 3), consumiendo cada cepa aproximadamente 12-14 g de arabinosa en 145 horas y produciendo aproximadamente 3-4 g/l de etanol. Como era de esperar, la cepa 3922 no consumió arabinosa ni produjo etanol. Estos resultados confirmaron que las vías de arabinosa exógenas en las cepas 3936 y 3937 fueron completas y confirieron a estas cepas la capacidad de fermentar arabinosa a etanol.

Las cuatro cepas se caracterizaron a continuación en medio YP que contenía tanto 20 g/l de dextrosa, 80 g/l de xilosa y 10 g/l de arabinosa como 10 g/l de dextrosa, 40 g/l de xilosa y 10 g/l de arabinosa. Las cepas 3936 y 3937 presentaron la capacidad de fermentar xilosa a etanol y actuaron similarmente a la cepa de control 3922 en el medio de azúcar más bajo (Figura 4). En el medio de azúcar más alto, sin embargo, la utilización de xilosa fue reducida en las cepas de la vía doble en comparación con la cepa 3922 de la vía de xilosa (Figura 5). Esta disminución en la utilización de xilosa se observó incluso en medio que carecía de arabinosa, que indica que una de las enzimas de la vía de arabinosa es responsable de la reducida utilización de xilosa.

El consumo de arabinosa en las cepas de la vía doble pareció empezar solo después de que se agotaran la dextrosa y la xilosa. En el medio de xilosa más bajo, las cepas de la vía doble usaron aproximadamente 5 g/l de arabinosa, pero este nivel de consumo requirió aproximadamente 160 horas, ya que la arabinosa solo se consumió después de que se agotara la xilosa (Figura 6). En el medio de xilosa más alto, no se consumieron los últimos 5 g de xilosa, y así no se utilizó arabinosa.

Ejemplo 6: Utilización del gen araB de *L. citreum* y de *B. thetaiotaomicron* de codón no optimizado:

La evaluación de las cepas que tienen vías de arabinosa parciales mostró que el efecto inhibitorio de la vía en la utilización de xilosa se produjo principalmente a partir de la acción del gen araB. Se identificaron araB alternativos

que tienen actividad más baja que el araB de *B. thetaiotaomicron* de codón optimizado. Un araB de *B. thetaiotaomicron* de codón no optimizado y un araB de *L. citreum* tuvo aproximadamente 1/100 y 1/33 de la actividad de araB de *B. thetaiotaomicron* de codón optimizado, respectivamente. Se integraron dos copias de los araB de *B. thetaiotaomicron* de codón no optimizado o *L. citreum* en la cepa yARA26 (Ejemplo 4C; contiene dos copias de araA y araD de *B. thetaiotaomicron*), que dan lugar a las cepas 12216 y 12215, respectivamente. Estas cepas se probaron, junto con la cepa de control 3937, en medio YP con azúcares mixtos (20 g/l de dextrosa, 80 g/l de xilosa y 10 g/l de arabinosa) y en medio YP con 50 g/l de arabinosa. En el medio de azúcares mixtos, las cepas 12215 y 12216 mostraron mejor utilización de xilosa y producción de etanol que la cepa 3937 (Figuras 30 y 31). En el medio de arabinosa solo, la cepa 3937 tuvo uso de arabinosa ligeramente más rápido que las cepas 12215 o 12216 (Figura 32).

Ejemplo 7: Identificación de genes transportadores de xilosa de *K. marxianus*:

Cepas de *I. orientalis* manipuladas para la utilización de xilosa no utilizan xilosa como fuente de carbono hasta que se ha utilizado la gran mayoría de la glucosa en el medio. Esto podría ser debido a la baja captación de xilosa en la célula con respecto a la captación de glucosa. Si éste es el caso, se esperaría que las modificaciones que aumentan la captación de xilosa en células de levadura también aumentarían la utilización de xilosa.

Se cribó el genoma de *K. marxianus* para los transportadores de azúcar sin caracterizar con el fin de evaluar el impacto de estos transportadores sobre el consumo de xilosa. Se seleccionaron dos de los posibles genes transportadores identificados en este cribado, KHT105 y RAG4, para estudio adicional. Las coincidencias de BLAST más próximas para ambos de estos genes fueron los transportadores de hexosa. La secuencia de nucleótidos de la región codificante del gen KHT105 se expone en SEQ ID NO:1, y la secuencia de aminoácidos codificada por el gen se expone en SEQ ID NO:2. La secuencia de nucleótidos de la región codificante del gen RAG4 se expone en SEQ ID NO:3, y la secuencia de aminoácidos codificada por el gen se expone en SEQ ID NO:4.

Ejemplo 8: Caracterización de genes transportadores de xilosa de *K. marxianus*:

El análisis de BLAST de los posibles genes transportadores de azúcar de *K. marxianus* del Ejemplo 7 indicaron que ambos genes compartieron su mayor grado de homología con los transportadores de hexosa. Para determinar si los transportadores de azúcar codificados por estos genes también fueron capaces de transportar azúcares de pentosa tales como xilosa, los genes se clonaron y se caracterizaron por diversos ensayos de utilización de xilosa.

Ambos genes se amplificaron a partir ADN genómico de *K. marxianus* usando cebadores que contenían sitios de restricción XbaI y PacI, y los genes se clonaron en TOPO y se secuenciaron. Los genes transportadores se digirieron a partir de los vectores TOPO con XbaI y PacI y se unieron en el vector pHJJ16 similarmente cortado, generando los plásmidos pJY20 (KHT105) y pJY21 (RAG4). El vector pHJJ16 contenía una secuencia de ARS de *I. orientalis*, que permite el mantenimiento del plásmido en el citoplasma del huésped, un promotor PDC en la dirección 5' del sitio de clonación XbaI-PacI y un marcador de selección URA3.

Se transformaron cepas de *I. orientalis* que fermentan xilosa yJY15 (3250) y yJLJ70 (3099) con los plásmidos pJY20, pJY21 y pHJJ26 (control). Antes de la transformación, la cepa 3250 contuvo dos copias cada una de un gen XI exógeno de *B. thetaiotaomicron*, un gen XK endógeno nativo y un gen XK exógeno de la secuencia nativa. La cepa 3099 tuvo los mismos cambios genéticos que 3250, con dos copias adicionales del gen XI de *B. thetaiotaomicron*. Se había mostrado previamente que las cepas de *I. orientalis* que contenían los genes XI y XK presentaban utilización de xilosa y producción de etanol (véase, p. ej., el documento WO04/099381). Las células transformadas se sembraron con xilosa como la única fuente de carbono, y se evaluó el crecimiento. Los transformantes presentaron elevado crecimiento sobre placas de xilosa a las 48 horas frente a la cepa de control, que indica que ambos genes funcionaron en el transporte de xilosa y que el transporte de xilosa era un factor limitante en la utilización de xilosa en *I. orientalis*.

Se generó el plásmido pJY27 uniendo un fragmento XhoI/PacI de un vector de integración AXR1 que contenía el terminador PDC de *I. orientalis* y el casete de selección URA3 y un fragmento XhoI/PacI que contenía el promotor PDC de *I. orientalis* y el gen transportador KHT105 de *K. marxianus*. El plásmido resultante se digirió con SacI y ApaI, y los fragmentos de integración linealizados se transformaron en la cepa 3099 de *I. orientalis*. Se cribaron por PCR de colonias colonias ura⁺ para identificar células con la integridad deseada usando cebadores fuera de AXR1 de 5' y 3' de cebadores en combinación con un homólogo de cebador al promotor PDC o casete URA3. Una de las colonias positivas se llamó yJY19.

Se generó el plásmido pJY28 uniendo un fragmento XbaI/PacI de un vector de integración AXR1 que contenía el promotor PDC de *I. orientalis*, terminador, casete de selección URA3 y secuencias que se dirigen a AXR1 y un fragmento XbaI/PacI que contenía el gen RAG4 de *K. marxianus*. El fragmento de integración se liberó por digestión de pJY28 con ApaI y KpnI, y los fragmentos de integración linealizados se transformaron en la cepa 3099 de *I. orientalis*. Se cribaron por PCR de colonias colonias ura⁺ para identificar células con la integridad deseada en el locus AXR1; una colonia tal se llamó yJY20.

Se cultivaron yJY19, yJY20, y la cepa 2973 de *I. orientalis* (versión ura⁺ de la cepa 3099) en un medio que contenía 20 g/l de glucosa y 55 g/l de xilosa a pH 4,8. Ambas cepas de transportador mostraron co-consumo de glucosa y

xilosa, mientras que la cepa 2973 solo consumió xilosa después de agotarse la glucosa (Figuras 7-9).

Basándose en estos resultados, se integró una segunda copia de cada gen transportador en el genoma. El marcador URA3 en las células yJY19 y yJY20 salió del bucle sembrando estas cepas en placas de ScD-FOA. Se cribaron por PCR de colonias las colonias para identificar colonias que retuvieron la integración, pero perdieron el gen URA3. Una de las cepas positivas que surge de yJY19 se llamó yJY25, y una de las cepas positivas que surge de yJY20 se llamó yJY26.

Se integró una segunda copia del casete de expresión de KHT105 de pJY27, como se ha descrito anteriormente, en la cepa yJY25. Se cribaron por PCR de colonias colonias ura+ para identificar células con la integridad deseada en el sitio AXR1, y un clon tal se llamó la cepa yJY27 (3097). El marcador URA3 en la cepa 3097 salió del bucle sembrando en placas de ScD-FOA después de crecimiento durante la noche. Las colonias se cribaron por PCR de colonias para confirmar la retención de la integración de KHT105, y una de las cepas ura- resultantes se llamó yJY32. yJY32 se transformó con ADN linealizado que llevaba el locus URA3 no mutante, y se cribaron por PCR de colonias colonias ura+ para identificar colonias con la integración correcta. Una de estas cepas que tiene URA3 en su locus original se llamó yJY34 (3081). Así, hubo tres cepas separadas que contenían dos copias del gen KHT105: 3097 (ura+), yJY32 (ura-) y 3081 (ura+).

Se integró una segunda copia del casete de expresión RAG4 de pJY28, como se ha descrito anteriormente, en la cepa yJY26. Se cribaron por PCR de colonias colonias ura+ para identificar células con la inserción deseada en el sitio AXR1, y un clon tal se llamó la cepa yJY28.

Las cepas 3097 (dos copias de KHT105) yJY28 (dos copias de RAG4) y 2973 (principal) se cultivaron durante la noche en YPD a 37 °C y 250 rpm. Los cultivos durante la noche se recogieron y se resuspendieron a una DO₆₀₀ objetivo de 3,0 en medio YP + 40 g/l de glucosa + 40 g/l de xilosa (pH 4,8, 37 °C, 100 rpm).

La cepa 3097 presentó mayor co-consumo de glucosa/xilosa que la cepa de control 2973 cuando la concentración de glucosa fue inferior a 20 g/l (Figura 10). Toda la glucosa fue consumida en aproximadamente cinco horas por la cepa 2973, frente a aproximadamente ocho horas para las cepas 3097 y yJY28. La xilosa se utilizó a una tasa más rápida por las cepas 3097 y yJY28 frente a la principal después de que se consumiera toda la glucosa (Figura 11). La combinación de mayor co-consumo y velocidades de utilización de xilosa más rápidas condujeron a mayor producción de etanol en la cepa 3097 (Figura 12). La cepa 3097 produjo 29 g/l de etanol en 25 horas quedando 7 g/l de xilosa. La cepa de control 2973 produjo 24 g/l de etanol en 25 horas quedando 13 g/l de xilosa. La cepa yJY28 produjo 22 g/l de etanol quedando 16 g/l de xilosa. Estos resultados mostraron que la incorporación del gen transportador KHT105 aumentó la productividad de etanol a partir de una mezcla de sustrato de glucosa/ xilosa.

Cepas de *I. orientalis* que contienen genes transportadores de *K. marxianus* se resumen en la Tabla 3.

Tabla 3:

Nombre de la cepa	Cepa principal	Gen transportador	N.º de copias	Localización de la inserción
2973 (ura+), 3099/yJLJ70 (ura-) (cepa principal fermentadora de xilosa)	--	--	0	--
yACN55 (ura-) (cepa principal fermentadora de xilosa con genes TAL, RKI, y RPE)	--	--	0	--
3408/yHJJ47 (ura+), yJY39 (ura-) (cepa principal con la vía de arabinosa completa de <i>B. thetaiotaomicron</i> , delección de XYL1, XYL2 y AXR1)	--	--	0	--
3937/yARA38 (ura+), yLUN011 (ura-) (cepa principal con la vía de arabinosa completa de <i>B. thetaiotaomicron</i> , delección de 9091 y 1202)	--	--	0	--
12053/yGP44 (ura+), yLUN027 (ura-) (cepa principal tolerante a etanol)	--	--	0	--
yJY19 (ura+), yJY25 (ura-)	3099	KHT105 de <i>K. marxianus</i>	1	AXR1 (1)
yJY20 (ura+), yJY26 (ura-)	3099	RAG4 de <i>K. marxianus</i>	1	AXR1 (1)

Nombre de la cepa	Cepa principal	Gen transportador	N.º de copias	Localización de la inserción
3097/yJY27 (ura+), yJY32 (ura-), 3081/yJY34 (URA3 reintegrado en el locus original)	yJY25	KHT105 de <i>K. marxianus</i>	2	AXR1 (2)
yJY28	yJY26	RAG4 de <i>K. marxianus</i>	2	AXR1 (2)
yACN59 (ura+), yACN60 (ura+), yACN67 (ura-), yACN68 (ura-)	yACN55	KHT105 de <i>K. marxianus</i>	1	9091 (1)
3415/yACN71 (ura+), yACN72 (ura+), yACN74 (ura-), yACN75 (ura-), 4141	yACN67	KHT105 de <i>K. marxianus</i>	2	9091 (2)
3849 (ura+), yHJJ172 (ura-) (2X ADH1)	3415	KHT105 de <i>K. marxianus</i>	2	9091 (2)
4014 (ura+), yHJJ182 (ura-), 4084	yHJJ172	KHT105 de <i>K. marxianus</i>	3	9091 (2), S141G4546 (1)
4083 (ura+), yLUN005 (ura-)	yHJJ172	KHT105 de <i>K. marxianus</i>	3	9091 (2), ALD5680 (1)
4085	yHJJ182	KHT105 de <i>K. marxianus</i>	4	9091 (2), S141G4546 (2)
4086/yLUN007 (ura+), 4117 (ura-)	yLUN005	KHT105 de <i>K. marxianus</i>	4	9091 (2), ALD5680 (2)
12037/yLUN013	4117	KHT105 de <i>K. marxianus</i>	6	9091 (2), ALD5680 (2), S141G4546 (2)
3812/yARA19	yJY39	KHT105 de <i>K. marxianus</i>	1	S141G4546 (1)
yLUN031 (ura+), yLUN033 (ura-)	yLUN027	KHT105 de <i>K. marxianus</i>	1	ALD5680 (1)
12125/yLUN036	yLUN033	KHT105 de <i>K. marxianus</i>	2	ALD5680 (2)
yLUN015 (ura+), yLUN016 (ura-)	yLUN011	KHT105 de <i>K. marxianus</i>	1	S141G4546 (1)
12038/yLUN018	yLUN016	KHT105 de <i>K. marxianus</i>	2	S141G4546 (2)

Ejemplo 9: Integración del gen transportador KHT105 de *K. marxianus* en una cepa de la vía de xilosa de *I. orientalis* más avanzada:

- 5 Se analizó una cepa de *I. orientalis* modificada que contenía el transportador KHT105 de *K. marxianus* en combinación con la vía de utilización de xilosa XI/XK, expresión en exceso de los genes de pentosa fosfato no oxidativa, e inactivación del gen 9091 para su capacidad para fermentar xilosa y glucosa a etanol con respecto a una cepa comparable sin el transportador.

Se insertó un fragmento NotI que lleva el casete URA3 en el sitio NotI de pHJJ22 (Ejemplo 4A) para crear los plásmidos de delección de 9091 pHJJ27 (orientación 1) y pHJJ28 (orientación 2).

- 10 Se clonó un fragmento NotI del vector pJY27 (Ejemplo 8) que lleva el promotor PDC de *I. orientalis*, gen transportador KHT105 de *K. marxianus*, terminador PDC de *I. orientalis* y casete de selección URA3 en pHJJ22 (Ejemplo 4A) para crear los vectores de expresión KHT105 pHJJ23 (orientación 1) y pHJJ24 (orientación 2).

- 15 Se digirió pHJJ23 con Apal y KpnI para liberar el fragmento de integración, y se transformó ADN linealizado en células yACN55. yACN55 es una cepa ura- que contiene cuatro copias de un gen XI exógeno de *B. thetaiotaomicron*, dos copias de un gen XK exógeno de la secuencia nativa y dos copias de cada uno de los genes exógenos de la vía de pentosa fosfato de la secuencia nativa (TAL, RKI, RPE), además de copias endógenas de genes XK, TAL, TKL, RPE y RKI. La principal de ura+ de yACN55 es la cepa 3356/yACN53.

Se seleccionaron transformantes y se purificaron sobre placas de ScD-ura. Se cribaron por PCR de colonias colonias ura+ para la integración correcta en el locus 9091. Dos cepas aisladas se llamaron yACN59 y yACN60. La

cepa yACN59 se cultivó durante la noche en YPD y se sembró en placas de ScD-FOA para sacar del bucle el gen URA3. Se usó PCR de colonias para confirmar la retención de la integración, y dos cepas aisladas se llamaron yACN67 y yACN68.

5 Se digirió pHJJ24 con Apal y KpnI para liberar el fragmento de integración, y se transformó ADN linealizado en células yACN67. Se seleccionaron transformantes y se purificaron en placas de ScD-ura. Se cribaron por PCR de colonias colonias ura⁺ para la integración correcta. Dos cepas aisladas se llamaron cepas yACN71 (3415) y yACN72. La cepa 3415 se cultivó durante la noche en YPD y se sembró en placas de ScD-FOA para sacar del bucle el gen URA3. Se usó PCR de colonias para confirmar la integración correcta en el locus 9091, y dos de tales cepas aisladas se llamaron yACN74 y yACN75.

10 Se digirió pHJJ28 con Apal y SacI para liberar el fragmento de integración, y se transformó ADN linealizado en células yACN55. Se seleccionaron transformantes y se purificaron en placas de ScD-ura. Se cribaron por PCR de colonias colonias ura⁺ para la integración correcta en el locus 9091, y dos de tales cepas aisladas se llamaron yACN61 y yACN62. La cepa yACN61 se cultivó durante la noche en medio YPD y se sembró en placas de ScD-FOA para sacar del bucle el gen URA3. Se usó PCR de colonias para confirmar la retención de la integración; dos de
15 tales cepas aisladas se llamaron yACN69 y yACN70.

Se digirió pHJJ27 con Apal y SacI para liberar el fragmento de integración, y se transformó ADN linealizado en células yACN69. Se seleccionaron transformantes y se purificaron en placas de ScD-ura. Se cribaron por PCR de colonias colonias ura⁺ para la integración correcta en el locus 9091, y una cepa aislada tal se llamó la cepa yACN73 (3416).

20 Se caracterizaron cepas 3415 (2 copias de KHT105, ambas copias delecionadas de 9091) y 3416 (ambas copias delecionadas de 9091) en fermentadores para el rendimiento en medio de hidrolizado. Se usaron bucles de biomasa de placas YPD para inocular matraces con deflectores de 250 ml que contenían 100 ml de medio definido (DMDX) que tenían 20 g/l de dextrosa y 80 g/l de xilosa y el pH se ajustó a aproximadamente 5,0. El medio definido contuvo urea como fuente de nitrógeno y tampón MES 0,2 M. Las células se incubaron a 250 rpm y 37 °C durante 15-24
25 horas, y se recogieron en fase de crecimiento exponencial media-tardía. Los cultivos se mezclaron con el 80 % de solución madre de glicerol y se separaron en alícuotas de 1 ml. Se transfirieron 50 a 400 µl de cada alícuota a 100 ml de medio en un matraz oscilante de 250 ml, se incubaron a 250 rpm y 37 °C durante 15-24 horas, y se recogieron en el crecimiento exponencial medio-tardío. Se recogieron muestras de 35 a 40 ml y se inocularon en recipientes de fermentación por lotes que contenían diversos medios de hidrolizado. Se recogieron muestras a intervalos de 4 a 8
30 horas durante toda la fermentación y se probaron para DO₆₀₀ usando un espectrofotómetro y para sustratos y niveles de producto usando análisis de HPLC.

La cepa 3415 presentó un aumento del 80 % en el consumo de xilosa y la velocidad de producción de etanol en un 30 % de medio DMDX de hidrolizado de hojas y tallos de maíz (CSH) a pH 5,8 (Figura 13). Estos resultados confirman que la expresión de KHT105 aumenta el consumo de xilosa y el título de etanol en *I. orientalis* cultivado
35 bajo condiciones fermentativas en medio de hidrolizado. Similarmente se observó un aumento del 75 % en la tasa de producción de etanol para 3415 con respecto a 3416 en un 15 % de medio de hidrolizado (15 % de CSH 5 g/l de ácido acético DMDX) a pH 4,9.

Ejemplo 10: Efecto de KHT105 de *K. marxianus* sobre el consumo de arabinosa en cepas de *I. orientalis* que contienen araA, araB y araD de *B. thetaiotaomicron*:

40 Se integró una copia única del gen Kit105 de *K. marxianus* en la cepa 3408 de *I. orientalis* (Ejemplo 1C; contiene dos copias de cada uno de los genes araA, araB y araD de *B. thetaiotaomicron* insertados en el loci XYL2, XYL1 y AXR1, respectivamente) en el locus S141G4546 para evaluar el efecto del transportador sobre el consumo de arabinosa. El locus S141G4546 tiene homología con sorbitol, butanodiol y glicerol deshidrogenasas.

pSK1 es un vector que contiene las regiones en la dirección 5' y en la dirección 3' para el locus S141G4546, separadas por un sitio de restricción NotI. Se unió pSK1 digerido con NotI al fragmento NotI de pJY27 (Ejemplo 8) que contenía el promotor PDC de *I. orientalis*, gen KHT105 de *K. marxianus*, terminador PDC de *I. orientalis* y casete de selección URA3. Se cribó ADN de plásmido de las colonias transformadas con la ligación por digestión con restricción. Los plásmidos con la inserción deseada se llamaron pLUN108 (orientación 1) y pLUN109 (orientación 2).

50 pLUN108 se digirió con Apal y SacI para liberar el fragmento de integración, y se transformó ADN linealizado en células yJY39 (cepa ura⁻ derivada de la cepa 3408/yHJJ47). Se seleccionaron transformantes y se purificaron en placas de ScD-ura. Se cribó por PCR ADN genómico aislado de las colonias para identificar colonias que tenían insertado en el locus S141G4546, y una cepa tal se identificó como yARA19 (3812).

La cepa 3812 se probó para la utilización de arabinosa en un experimento de matraz oscilante. Las células se cultivaron durante la noche en 50 ml de YPD y se inocularon en 50 ml de YP + 40 g/l de arabinosa y 10 g/l de dextrosa. Se inocularon matraces con agitación por duplicado a DO₆₀₀ = 0,4 y se cultivaron a 37 °C y 100 rpm. Se
55 ejecutó la cepa principal 3408 ura⁺ como control.

La dextrosa fue agotada por todas las cepas antes de 25 horas. La adición de una única copia del gen KHT105 produjo un pequeño aumento en la utilización de arabinosa (~5 g superior a la cepa principal) y un rendimiento de etanol ligeramente más alto después de 100 horas (Figura 14).

Ejemplo 11: Integración del gen transportador KHT105 de *K. marxianus* en el locus S141G4546 de una cepa de la vía doble de *I. orientalis*:

Se integró el gen transportador KHT105 de *K. marxianus* en el derivado ura- de la cepa 3937 de *I. orientalis* (Ejemplo 4D; contiene dos copias de cada uno de los genes araA, araB y araD de *B. thetaiotaomicron* insertados en los loci 9091, 1202 y 9091, respectivamente), que habían mostrado la capacidad de fermentar tanto xilosa como arabinosa a etanol (Ejemplo 5).

Se construyó un casete de integración que contenía el gen transportador KHT105 de *K. marxianus* entre las regiones flanqueantes de S141G4546. Para construir el vector de integración, se ligó un fragmento NotI que contiene un promotor PDC, gen KHT105, terminador PDC y casete de selección URA3 en pSK1 desfosforilado cortado con NotI (vector TOPO con S141G4546 en la dirección 5' y en la dirección 3' separado por el sitio NotI). Las colonias transformadas con la ligación se cribaron por PCR para direccionalidad, y vectores con la inserción deseada se llamaron pHJJ86 (orientación 1) y pHJJ87 (orientación 2).

Se digirió pHJJ87 con Apal y SacI para liberar el fragmento de integración, y se transformó ADN linealizado en yLUN011 (versión ura- de la cepa 3937). Se seleccionaron colonias transformantes y se purificaron en placas de ScD-ura y se cribaron en dos reacciones de PCR separadas. Los clones que presentaron productos de PCR que indican la integración correcta de KHT105 en el locus S141G4546 se designaron yLUN015.

Se cultivó yLUN015 durante la noche en YPD y se sembró en placas de ScD-FOA. Las colonias sacadas del bucle se purificaron en placas de YPD y se cribaron en dos reacciones de PCR separadas. Se identificó un clon por tener retenida la integración, pero perdió el gen URA3. Esta cepa (yLUN016) se sembró duplicada en SCD-ura para confirmar la incapacidad de cultivar sin uracilo.

Se digirió pHJJ86 con Apal y SacI para liberar el fragmento de integración, y se transformó ADN linealizado en yLUN016. Se seleccionaron transformantes y se purificaron en placas de ScD-ura y se cribaron por PCR a través de los cuatro empalmes de integración. Los clones identificados por tener ambas copias del gen KHT105 integradas en el locus S141G4546 se designaron la cepa yLUN018 (12038).

Ejemplo 12: Caracterización de la cepa de la vía doble de *I. orientalis* que contiene dos copias del gen transportador KHT105 de *K. marxianus*:

Se evaluó la capacidad de la cepa 12038 de *I. orientalis* (Ejemplo 11) y su cepa principal 3937 para fermentar arabinosa y xilosa a etanol en experimentos en matraz oscilante. Las cepas se cultivaron a 37 °C y 100 rpm en tanto 1) YP + 20 g/l de dextrosa, 80 g/l de xilosa y 10 g/l de arabinosa, pH 5,1 (YP20D/80X/10A) o 2) YP + 10 g/l de dextrosa, 40 g/l de xilosa y 10 g/l de arabinosa, pH 5,1 (YP10D/40X/10A).

En el medio YP20D/80X/10A, la cepa 12038 presentó un aumento significativo en el consumo de xilosa frente a la cepa principal 3937 (Figura 15). Este aumento en el consumo de xilosa se correspondió con un aumento en la producción de etanol (Figura 15). Las velocidades del consumo de xilosa en la cepa 12038 en este medio fueron similares a aquellas observadas en las cepas de levadura sin la vía de arabinosa bacteriana (p. ej., cepa 3922). La cepa 12038 empezó el consumo de arabinosa antes en la fermentación, probablemente debido al agotamiento más temprano de la xilosa, y se usó aproximadamente un 40 % más de arabinosa en comparación con la cepa principal 3937 (Figura 16).

El aumento en las velocidades de consumo de xilosa para la cepa 12038 con respecto a su cepa principal no fue tan grande en el medio YP10D/40X/10A como en el medio de azúcar más alto (Figura 17). La cepa 12038 presentó otra vez consumo de arabinosa más temprano y un aumento en la arabinosa total usada, pero con una ventaja más pequeña que se observó en el medio de azúcar más alto (Figura 18).

Estos resultados establecen que las células de levadura que contienen genes de la vía de arabinosa y xilosa bacterianas y el gen transportador KHT105 son capaces de fermentar tanto arabinosa como xilosa en etanol de una manera eficiente.

Ejemplo 13: Construcción de la cepa inactiva en ALD5680 de *I. orientalis*:

La aldehído deshidrogenasa (ALD) convierte unidireccionalmente el acetaldehído en acetato, y la expresión de ALD pueden desviar el carbono de la producción de etanol. La actividad de ALD es muy importante para el funcionamiento de la derivación de PDH en levadura; el reducir la actividad de ALD puede hacer que la levadura utilice más acetato del medio. En *S. cerevisiae*, las dos ALD principales están codificadas por los genes ALD4 y ALD6. En *I. orientalis*, los homólogos de ALD incluyen S141G5680 ("ALD5680"), S141G9161 ("ALD9161") y S141G6502 ("ALD6502"), presentando ALD9161 la mayor expresión promedio. ALD6502 es más similar a ALD3 de *S. cerevisiae*, que puede funcionar en el metabolismo de la β -alanina, y parece ser citoplásmico basándose en la

secuencia de proteínas. Tanto ALD5680 como ALD9161 codifican proteínas con una secuencia conductora típica, y la expresión de ambas se potenció con crecimiento sobre xilosa y disminuyó por la adición de acetato. La secuencia de nucleótidos de la región codificante de ALD5680 se expone en SEQ ID NO:72, y la secuencia de polipéptidos de ALD5680 se expone en SEQ ID NO:73. Los intentos por inactivar el locus ALD9161 locus en *I. orientalis* fueron poco satisfactorios, sugiriendo que este locus puede ser esencial. Sin embargo, ambas copias del locus ALD5680 estuvieron inactivadas en la cepa 3489 de *I. orientalis* (previamente manipulada para fermentar xilosa a etanol) para evaluar el efecto sobre la utilización de azúcar y acetato y la producción de etanol.

Las regiones en la dirección 5' y en la dirección 3' de ALD5680 se amplificaron a partir de ADN genómico de *I. orientalis*, y los fragmentos resultantes se purificaron en gel. El fragmento en la dirección 3' se digirió con Apal y NotI y el fragmento en la dirección 5' se digirió con NotI y SacI. Los fragmentos digeridos se unieron en el vector TOPO cortado con Apal/SacI en una ligación de tres trozos. Las colonias transformadas con la ligación se cribaron por PCR para la correcta inserción y la secuencia del inserto se confirmó por secuenciación de ADN. Este plásmido se llamó pHJJ75. Un fragmento NotI que contenía el casete de selección URA3 se ligó en pHJJ75 cortado con NotI. Las colonias transformadas con la ligación se cribaron para direccionalidad del inserto. Los vectores que representan las dos orientaciones se llamaron pHJJ78 y pHJJ79.

Se cortó pHJJ79 con Apal y SacI para liberar el fragmento de integración y se transformó ADN linealizado en la cepa 3514, un derivado ura⁻ de la cepa 3489. Los transformantes se extendieron para la purificación y se cribaron por PCR para la integración correcta en el locus ALD5680. yHJJ114 se identificó por tener una copia de la inactivación de ALD5680. Se cultivó yHJJ114 durante la noche en YPD y se sembró en medio ScD-FOA para seleccionar la salida de bucle del gen URA3. Se confirmaron por PCR dos cepas aisladas ura⁻ resultantes que habían retenido la inactivación de ALD5680. Estas cepas aisladas se llamaron yHJJ118 y yHJJ119.

Se cortó pHJJ78 con Apal y SacI para liberar el fragmento de integración y se transformó ADN linealizado en yHJJ118. Los transformantes ura⁺ se extendieron para la purificación y se cribaron colonias individuales por PCR para la integración correcta en el locus ALD5680. Las cepas yHJJ123 y yHJJ124 (3861) se identificaron por tener ambas copias de ALD5680 delecionadas.

Ejemplo 14: Caracterización de la cepa inactivada en ALD5680 de *I. orientalis*:

Se evaluaron las cepas inactivadas en ALD5680 del Ejemplo 13 para determinar el efecto de la inactivación. Se examinaron la utilización de azúcar y la producción o utilización de acetato usando experimentos en matraz oscilante en medio sin acetato a 100 y 135 rpm de aireación y medio con acetato a 135 rpm de aireación. La cepa 3861 (inactivación de ambas copias de ALD5680) y la cepa principal 3489 se cultivaron en medio YP con 20 g/l de dextrosa y 80 g/l de xilosa, pH 4,8 a 37 °C, o medio YP con 20 g/l de dextrosa, 80 g/l de xilosa, 4 g/l de acetato, pH 5,1 a 37 °C.

La cepa de deleción de ALD5680 presentó producción de acetato más baja en medio sin acetato a 100 rpm, aunque la cepa principal solo produjo 0,49 g/l de acetato bajo estas condiciones (Tabla 4). La cepa de deleción de ALD5680 no mostró ningún beneficio significativo con respecto a la utilización de azúcar (Figura 19).

Tabla 4:

	Xilitol (g/l)	Arabitol (g/l)	Glicerol (g/l)	Acetato (g/l)
Cepa 3489	1,42	0,61	2,43	0,49
Cepa 3861	1,29	0,81	2,38	0,06

La cepa de deleción de ALD5680 también presentó producción de acetato más baja en medios sin acetato a 135 rpm (Figura 20). La cepa principal produjo más de 1 g/l de acetato linealmente con el tiempo, mientras que la cepa de deleción solo produjo acetato después de que se hubiera terminado la xilosa. Bajo estas condiciones, la cepa inactivada presentó utilización de xilosa y producción de etanol más alta que la cepa principal, acabando la xilosa aproximadamente 30 horas antes que la cepa principal (Figura 21). Los subproductos producidos por cada cepa se resumen en la Tabla 5.

Tabla 5:

	Xilitol (g/l)	Arabitol (g/l)	Glicerol (g/l)
Cepa 3489	2,02	0,26	1,33
Cepa 3861	1,56	0,47	2,51

En medio sintético con acetato, la cepa de deleción no mostró una ventaja coherente. Usó lentamente el acetato durante aproximadamente las primeras 48 horas, pero entonces empezó a producir acetato (Figura 22, Tabla 6). La utilización de xilosa fue relativamente lineal hasta las 68 horas y entonces se detuvo (Figura 23). A diferencia, la

utilización de acetato por la cepa principal aumentó a aproximadamente 48 horas y la utilización de xilosa permaneció estacionaria hasta el final de la fermentación (Figuras 22-23).

Tabla 6:

Cepa	Xilitol (g/l)	Arabitol (g/l)	Glicerol (g/l)	Acetato (g/l)
yHJJ82/3489	1,06	1,25	0,62	2,18
Cepa 3861	1,24	0,26	0,53	3,72

- 5 La cepa de delección de ALD5680 presentó una ventaja significativa en el medio basado en hidrolizado. Se probaron 19 cepas diferentes que tenían diversas modificaciones o mutaciones de ingeniería genética en medio DM20D80X con 50 % de hojas y tallos de maíz hidrolizado (CSH) a pH 6,2, 37 °C, y 100 rpm. Los dos rasgos que confirieron el mayor beneficio a la producción de etanol en este medio fueron la expresión en exceso de KHT105 y la delección de ALD5680.
- 10 Estos resultados muestran que la delección de ALD5680 puede conferir un aumento de la capacidad para fermentar xilosa a etanol y reducir la producción de acetato por el huésped, pero que estas ventajas son posiblemente dependientes de los medios específicos y las condiciones de aireación. Estas ventajas son particularmente significativas durante el cultivo en medio basado en hidrolizado.

Ejemplo 15: Introducción de copias adicionales del gen transportador KHT105 de *K. marxianus* en cepas inactivadas en S141G9091, S141G4546 y S141G5680 de *I. orientalis*:

- 15 Se manipularon cepas de *I. orientalis* para contener en cualquier parte dos a seis copias del gen KHT105 de *K. marxianus*. La cepa 3849, que es equivalente a la cepa 3489 con la adición de dos copias del gen KHT105 integrado en el locus 9091, se usó como la cepa principal.

- 20 Se linealizó ADN de plásmido del vector pHJJ86 (Ejemplo 11; contiene el promotor PDC, gen KHT105, terminador PDC y casete de selección URA3 entre las regiones flanqueantes de S141G4546) por digestión con restricción, y se transformó ADN linealizado en la cepa yHJJ172 (derivado ura- de la cepa 3849) para producir la cepa 4014, que contuvo tres copias del gen KHT105. La tercera copia del gen se integró en el sitio S141G4546. También se construyó la cepa de control 4141. La cepa 4141 contuvo el marcador URA3 en vez de una tercera copia de KHT105 en el sitio S141G4546. Los vectores usados para construir esta cepa se produjeron uniendo un fragmento NotI que contenía el casete de marcador URA3 en pSK1 cortado con NotI (Ejemplo 10). Colonias de *E. coli* transformadas con la ligación se cribaron por PCR, y se identificaron los vectores pHJJ88 (orientación 1) y pHJJ89 (orientación 2) como que contenían el casete de marcador URA3 en orientaciones opuestas. Se transformó pHJJ88 linealizado en yHJJ172 como se ha descrito previamente para obtener la cepa 4141.

- 30 La cepa 4014 se cultivó durante la noche en YPD y se sembró en placas de SCX-FOA. Se cribaron por PCR colonias sacadas del bucle, y la cepa sacada del bucle correcta se llamó yHJJ182. Se integró una cuarta copia de KHT105 en el locus S141G4546 de yHJJ182 por transformación con pHJJ87 linealizado (Ejemplo 11; contiene promotor PDC, gen KHT105, terminador PDC y casete de selección URA3 entre las regiones flanqueantes de S141G4546). Se confirmó por PCR que los transformantes tenían KHT105 integrado en ambos loci de S141G4546. Estos clones se llamaron la cepa 4085. La cepa de control 4084, que contuvo el marcador URA3 en S141G4546 en vez de una tercera y cuarta copia de KHT105, se produjo transformando un derivado ura- de la cepa 4141 (yHJJ180) con pHJJ89 linealizado y seleccionando en placas de ScD-ura. Los transformantes se cribaron por PCR en los cuatro empalmes de integración para confirmar que dos copias del marcador URA3 se integraron correctamente en el locus S141G4546.

- 40 El plásmido pHJJ93, que contuvo un casete de expresión de KHT105 entre las regiones flanqueantes de ALD5680 (orientación 1), se digirió con Apal y SacI para liberar el fragmento de integración, y se transformó ADN linealizado en células yHJJ172. Las colonias se seleccionaron y se purificaron en placas de ScD-ura y se cribaron por PCR a través de ambos empalmes de integración para confirmar la integración del casete de expresión de KHT105 en el locus ALD5680. Estos clones se llamaron la cepa 4083. El marcador URA3 en la cepa 4083 salió del bucle cultivando durante la noche en YPD y sembrando en medio ScX-FOA. Las colonias resultantes se cribaron para confirmar la retención de la integración, y clones positivos se sembraron por duplicado en ScD-ura para confirmar la ausencia de crecimiento en medio sin uracilo. Estas salidas del bucle se llamaron yLUN005.

- 45 El plásmido pHJJ94, que contuvo un casete de expresión de KHT105 entre las regiones flanqueantes de ALD5680 (orientación 2), se digirió con Apal y SacI para liberar el fragmento de integración y se transformó ADN linealizado en células yLUN005. Se seleccionaron colonias y se purificaron en placas de ScD-ura y se cribaron a través de ambos empalmes de integración para confirmar la integración en el locus ALD5680. Los clones confirmados por PCR que contenían copias de KHT105 en ambos loci ALD5680 se designaron la cepa yLUN007 (4086).

Se realizaron experimentos en matraz oscilante para evaluar la fermentación de xilosa en las diversas cepas. En un experimento, la cepa 4084 (2X KHT105, 2X inactivada en S141G4546) y 4085 (4X KHT105, 2X inactivada en

S141G4546) se cultivaron a 37 °C y 100 rpm en DM20D80X con 50 % de medio CSH, pH 6,2. El consumo de xilosa y las velocidades de producción de etanol aumentaron mediante las copias adicionales de KHT105 (Figura 24). Los subproductos producidos por cada cepa se resumen en la Tabla 7.

Tabla 7:

	Xilitol (g/l)	Arabitol (g/l)	Glicerol (g/l)	Acetato (g/l)
Cepa 4084	5,90	BDL	4,70	4,38
Cepa 4085	3,90	0,76	5,44	4,21

Las cepas 4083 (3X KHT105, 1X inactivada en ALD5680), 4085 (4X KHT105, 2X inactivada en S141G4546), y dos clones de la cepa 4086 (4X KHT105, 2X inactivada en ALD5680) se caracterizaron usando las mismas condiciones en matraz oscilante. Los resultados se resumen en la Tabla 8 y la Figura 25. La adición de una cuarta copia de KHT105 en combinación con la delección de ALD5680 aumentó sustancialmente las velocidades de consumo de xilosa y los títulos de etanol con respecto a cuatro copias del gen KHT105 combinado con la delección de S141G4546 (Figura 25). Los dos clones de la cepa 4086 produjeron aproximadamente 29 g/l de etanol a las 140 horas, mientras que la cepa 4085 y la cepa principal 3849 (datos no mostrados) lograron sus títulos máximos de 17-18 g/l bajo las mismas condiciones en este momento de tiempo. La delección de ALD5680 también condujo a potenciar el consumo de acetato. La cepa 4086 tuvo elevada producción de glicerol y arabitol, junto con producción de xilitol reducida (Tabla 8).

Tabla 8:

	Xilitol (g/l)	Arabitol (g/l)	Glicerol (g/l)	Acetato (g/l)
Cepa 4083	5,26	BDL	3,72	3,50
Cepa 4085	5,68	BDL	4,69	3,75
Cepa 4086 clon N.º 1	2,53	1,28	7,27	2,09
Cepa 4086 clon N.º 3	2,78	1,47	7,32	1,95

Ejemplo 16: Adición de KHT105 y/o delección de ALD5680 en una cepa de *I. orientalis* tolerante a etanol:

La cepa yGP44 de *I. orientalis* (12053) es un mutante tolerante a etanol obtenido por mutagénesis y selección de la cepa 3489, seguido de manipulación de la inactivación de S141G1202. El marcador de selección URA3 salió del bucle cultivando la cepa 12053 durante la noche en YPD y sembrando en medio ScD-FOA. Las colonias se cribaron por PCR y se sembraron en medio ScD-ura para confirmar la pérdida del gen URA3. Las colonias positivas para la salida del bucle se llamaron la cepa yLUN027.

Para insertar URA3 en el primer locus de ALD5680, los fragmentos de integración linealizados a partir del plásmido pHJJ78 (Ejemplo 13) se transformaron en yLUN027. Se seleccionaron transformantes en placas de ScD-ura y se cribaron por PCR a través de ambos empalmes de integración para identificar transformantes positivos para ambos productos de PCR de empalme. Un transformante tal se llamó yLUN030. El marcador URA3 de yLUN030 salió del bucle por crecimiento en YPD durante la noche y sembrando en placas de ScD-FOA. Las colonias se cribaron por PCR y para la ausencia de crecimiento en placas de ScD-ura para identificar aquellas que habían retenido la delección de ALD5680, pero perdieron el gen URA3. Tres de tales colonias se llamaron la cepa yLUN032.

Para inactivar el segundo locus de ALD5680, se digirió ADN de plásmido de pHJJ79 (Ejemplo 13) con Apal y SacI, y los fragmentos de integración linealizados se transformaron en yLUN032. Se seleccionaron transformantes en placas de ScD-ura y se cribaron por PCR a través de ambos empalmes de integración para ambos loci. Se identificaron dos transformantes que fueron positivos para todos los productos de PCR de empalme. Estos transformantes se llamaron la cepa yLUN035 (12124).

Para insertar el transportador KHT105 en el primer locus de ALD5680, se digirió ADN de plásmido de pHJJ93 (Ejemplo 15) con Apal y SacI, y los fragmentos de integración linealizados se transformaron en yLUN027. Se seleccionaron transformantes en placas de ScD-ura y se cribaron por PCR a través de ambos empalmes de integración. Se identificaron cuatro transformantes que fueron positivos para todos los productos de PCR del empalme. Estos transformantes se llamaron la cepa yLUN031.

El marcador URA3 de yLUN031 salió del bucle por crecimiento durante la noche en YPD y sembrando en placas de ScD-FOA. Se cribaron colonias por PCR y para la ausencia de crecimiento en placas de ScD-ura para identificar aquellas que habían retenido el fragmento de integración KHT105 en el sitio de ALD5680, pero perdieron el marcador URA3. Una colonia tal se llamó la cepa yLUN033.

Para añadir la segunda copia de KHT105 a yLUN033, se digirió ADN de plásmido de pHJJ94 (Ejemplo 15) con Apal y SacI, y los fragmentos de integración linealizados se transformaron en yLUN033. Se seleccionaron transformantes en placas de ScD-ura y se cribaron por PCR a través de ambos empalmes de integración para ambos loci. Se identificaron cuatro transformantes que fueron positivos para todos los productos de PCR del empalme. Estos transformantes se llamaron la cepa yLUN036 (12125).

Las cepas 12124 (ambas inactivadas en los loci de ALD5680), 12125 (2X KHT105, ambas inactivadas en los loci de ALD5680) y 12053 (principal) se caracterizaron por matraz oscilante en medio DM + 50 % de hidrolizado de hojas y tallos de maíz con dos concentraciones de azúcar diferentes. Se ejecutó un conjunto de matraces de agitación con 20 g/l de dextrosa y 80 g/l de xilosa, mientras que el segundo contuvo 70 g/l de dextrosa y 40 g/l de xilosa. El pH de todos los medios fue 5,7. Los matraces de agitación se inocularon a una DO_{600} de partida = 0,1 y se cultivaron a 100 rpm y 37 °C.

En los matraces con agitación de dextrosa más baja / xilosa más alta, la expresión en exceso de KHT105 proporcionó una ventaja significativa con respecto a la producción de etanol (Figura 26) y el consumo de xilosa (Figura 27). La delección de ALD5680 proporcionó un aumento del 25 % en el título de etanol, mientras que la combinación de esta delección con la expresión en exceso de KHT105 dio un aumento de aproximadamente el 125 %. Estas ventajas fueron mucho menos pronunciadas en los medios definidos de dextrosa más alta / xilosa más baja.

Ejemplo 17: Integración de araD de *B. animalis* y *L. lactis* en *I. orientalis* y caracterización de cepas resultantes:

Debido a la actividad relativamente baja de los RE previamente probados, se clonaron genes araD alternativos y se ensayaron para la actividad y rendimiento en medio que contenía arabinosa. Estas secuencias de RE se derivaron de *B. animalis* (SEQ ID NO:18) y *L. lactis* (SEQ ID NO:20) y se optimizaron en el codón para la expresión en *I. orientalis* (SEQ ID NOs:17 y 19, respectivamente). Los genes araD de *B. animalis* y *L. lactis* optimizados en el codón se integraron en el sitio cyb2B de la cepa 12038 (Tabla 3; Ejemplos 11 y 12) usando métodos similares a aquellos descritos anteriormente en el Ejemplo 1. Se había mostrado previamente que la inactivación en cyb2b no tenía efecto fenotípico bajo condiciones de prueba relevantes. Se seleccionó la cepa 12038 como cepa principal de manera que fuera menos probable que el transporte y la actividad de araB fueran limitantes.

La cepa yACN170 resultante contuvo una copia de araD de *B. animalis* y dos copias de araD de *B. thetaiotaomicron*, mientras que la cepa yACN172 contuvo una copia de araD de *L. lactis* y dos copias de araD de *B. thetaiotaomicron*. La cepa yACN168 contuvo una copia de araD de *B. thetaiotaomicron* integrada en el sitio cyb2b, de un total de tres copias. Las tres cepas se evaluaron en matraces con agitación para la fermentación de arabinosa con respecto a la cepa principal. El medio de prueba usado fue un medio definido con 20 g/l de dextrosa, 35 g/l de xilosa, 35 g/l de arabinosa, MES 0,2 M y 7 g/l de acetato, pH 5,0. Las células se cultivaron a 37 °C y 125 rpm y se muestrearon con el tiempo para sustratos y productos. Bajo estas condiciones, yACN170 y yACN172 tuvieron consumo de arabinosa significativamente elevado en comparación con la cepa principal y yACN168 (Figura 28).

Ejemplo 18: Integración de araA de *L. sakei* y araA de *B. thetaiotaomicron* alternativo en *I. orientalis*:

Las enzimas celulasa usadas en la hidrólisis de la celulosa tienen óptimos de pH de aproximadamente 4-5,5 y óptimos de temperatura de aproximadamente 40-50 °C. Enzimas de la vía que demuestran alta actividad bajo estas condiciones pueden proporcionar un beneficio al rendimiento fermentativo. Se ha mostrado recientemente que Al/araA de *Lactobacillus sakei* tiene un óptimo de pH de 5-7, que mantiene el 80 % de la máxima actividad a un pH de 3, y estabilidad de la temperatura hasta 40 °C (Rhimi Bioresour Technol 101:9171 (2010)).

B. thetaiotaomicron tiene una segunda posible arabinosa isomerasa (araA2, SEQ ID NOs:7/8) que es solo el 17 % homóloga en una base de aminoácidos al gen araA de esta especie que se usó en los ejemplos previos. Este homólogo es más corto por 35 aminoácidos, localizados en la dirección 3' de una L-arabinofuranosidasa, y similar a las L-arabinosa isomerasas de *Pedobacter* y *Rhizobium*.

Se integraron araA de *L. sakei* y araA2 de *B. thetaiotaomicron* en el sitio cyb2B de la cepa 12038 usando métodos similares a aquellos descritos anteriormente en el Ejemplo 1. La cepa yACN176 resultante contuvo una copia de araA de *L. sakei* y dos copias de araA de *B. thetaiotaomicron*, mientras que la cepa yACN178 contuvo una copia de araA2 de *B. thetaiotaomicron* y dos copias de araA de *B. thetaiotaomicron*. La cepa yACN174 contuvo una copia de araD de *B. thetaiotaomicron* integrada en el sitio cyb2b, de un total de tres copias. Las tres cepas se evaluaron en matraces con agitación para la fermentación de arabinosa con respecto a la cepa principal. El medio de prueba usado fue un medio definido con 20 g/l de dextrosa, 35 g/l de xilosa, 35 g/l de arabinosa, MES 0,2 M y 7 g/l de acetato, pH 5,0. Las células se cultivaron a 37 °C y 125 rpm y se muestrearon con el tiempo para sustratos y productos. Bajo estas condiciones, yACN176 y yACN178 tuvieron consumo de arabinosa significativamente elevado en comparación con la cepa principal y yACN174 (Figura 29).

Ejemplo 19: Integración del gen TAL exógeno de *I. orientalis* en *I. orientalis*:

Se incorporaron genes TAL de la secuencia nativa de *I. orientalis* en la cepa 3099 de *I. orientalis* (Ejemplo 8) para determinar si la expresión en exceso de esta enzima aumentó la fermentación de xilosa a etanol.

Para construir un vector de expresión de TAL, la región codificante del gen TAL de *I. orientalis* (SEQ ID NO:51) más 400 pb adicionales en la dirección 3' del gen se amplificaron a partir del ADN genómico de *I. orientalis* no mutante. El producto de PCR se clonó en pCR-Blunt II-TOPO para formar el plásmido pACN1 y se verificó la secuencia. Se unió un fragmento EcoRI/XbaI de pACN1 que llevaba el gen TAL y terminador en un fragmento de vector similarmente cortado que llevaba el casete de selección URA3 y un promotor PDC de *I. orientalis* para formar el plásmido pACN3. Se unió un fragmento NotI de pACN3 que llevaba el promotor, TAL, terminador y casete URA3 con pHJJ4 cortado con NotI (secuencias que se dirigen a AXR1, Ejemplo 1B) para formar los plásmidos pACN5 (orientación 1) y pACN7 (orientación 2).

Para construir una cepa que expresa en exceso TAL, se transformó la cepa 3099 con ADN linealizado de pACN7 y se sembró en placas de ScD-ura. Se cribaron por PCR de colonias ura+ a través de ambos empalmes de integración; una cepa aislada con la inserción deseada se llamó yACN3. La cepa yACN3 se sembró en placas de ScD-FOA para sacar del bucle el gen URA3. Se usó PCR de colonias para confirmar la retención de la integración deseada; una cepa aislada ura- tal se llamó yACN7. La cepa yACN7 se transformó con ADN linealizado de pACN5 y se sembró en placas de ScD-ura. Se cribaron por PCR de colonias ura+ a través de todos los empalmes de integración; una cepa aislada con la inserción deseada en ambos loci AXR1 se llamó yACN11 (3082). Se generó una cepa de control, 3862, delecionando ambas copias del locus AXR1 en la cepa 3099 sin expresar en exceso el casete de TAL1. La construcción de deleción usada para preparar esta cepa contuvo el casete de selección URA3 entre las secuencias que se dirigen a AXR1.

Las cepas se caracterizaron en matraces con agitación usando medio YP + 20 g/l de glucosa y 80 g/l de xilosa a pH 4,8. Se cultivaron cultivos iniciales (25 ml de medio en matraz de 125 ml) durante el día a 250 rpm. Los cultivos durante la noche (50 ml de medio en 250 ml de matraz) se inocularon a una DO de 0,00002 y se cultivaron a 230 rpm y 35 °C. A la mañana siguiente, todos los cultivos tuvieron DO de 3,8-5,3. Los matraces de producción (50 ml de medio en matraces de 125 ml) se inocularon a una DO de 0,1 y se cultivaron a 37 °C y 100 rpm. Se tomaron muestras con el tiempo, se centrifugaron, y los sobrenadantes se filtraron y se analizaron usando HPLC.

La cepa principal 3118 (versión ura+ de 3099 antes de sacar del bucle el marcador) y la cepa de control del sitio de inserción 3862 produjeron 20-21 g/l de etanol en 55 horas, mientras que la cepa 3082 produjo 28 g/l en el mismo periodo (Figura 33). En las tres cepas, la glucosa se agotó en nueve horas. La formación de etanol adicional en la cepa 3082 se correlacionó con un aumento del consumo de xilosa. La cepa 3082 produjo menos xilitol durante toda la fermentación, y niveles más altos de glicerol y arabitól. La formación de metabolitos se resume en la Tabla 9.

Tabla 9:

	Xilitol (g/l)	Glicerol (g/l)	Arabitól (g/l)
Cepa 3118	2,2	0,9	0,4
Cepa 3082	1,2	2,0	1,3
Cepa 3862	2,0	1,0	0,5

Ejemplo 20: Integración de genes RKI exógenos de *I. orientalis* en *I. orientalis*:

Se incorporaron genes RKI de la secuencia nativa de *I. orientalis* en la cepa yACN23 de *I. orientalis* para determinar si la expresión en exceso de esta enzima aumentó la fermentación de xilosa a etanol.

Para construir un vector de integración que se dirigía al sitio GAL6, las regiones flanqueantes de 5' y 3' del gen GAL6 (SEQ ID NO:87) se amplificaron a partir de ADN genómico de *I. orientalis* no mutante. Los fragmentos de PCR se clonaron en pCR-BluntII-TOPO para formar plásmidos pACN25 (región en la dirección 5') y pACN26 (región en la dirección 3') y se verificó la secuencia. Se unieron un fragmento ApaI/NotI de pACN25, que contenía la región en la dirección 5', y un fragmento SacI/NotI de pACN26, que contenía la región en la dirección 3', en pCRII cortado con ApaI/SacI para formar el plásmido pACN29.

Para construir un vector de expresión de RKI, la región codificante del gen RKI de *I. orientalis* (SEQ ID NO:39) más 400 pb adicionales en la dirección 3' del gen se amplificaron a partir de ADN genómico de *I. orientalis* no mutante. El producto de PCR se clonó en pCR-Blunt II-TOPO para formar los plásmidos pACN27 y pACN28 y se verificó la secuencia. Se unió el trozo de EcoRI/XbaI de pACN27 que llevaba el gen RKI y terminador en un fragmento de vector cortado similarmente que llevaba el casete de selección URA3 y un promotor PDC de *I. orientalis* (fragmento EcoRI/XbaI de pHJJ2, Ejemplo 1A) para formar el plásmido pACN31. Se unió el fragmento NotI de pACN31 con pACN29 cortado con NotI (secuencias que se dirigen a GAL6) para formar plásmidos pACN44 (orientación 1) y pACN45 (orientación 2).

La cepa yACN23 es un derivado de la cepa 3082 (Ejemplo 19) que contiene una deleción para el locus S141G4738 ("AXR4"). Para construir una cepa que expresa en exceso RKI, la cepa yACN23 se transformó con ADN linealizado a partir de pACN44 y se sembró en placas de ScD-ura. Se cribaron por PCR de colonias ura+ a través de ambos empalmes de integración. Una cepa aislada con la inserción deseada se llamó yACN25. La cepa yACN25 se

sembró en medio ScD-FOA para sacar del bucle el gen URA3. Se usó PCR de colonias a través de ambos empalmes de integración para confirmar la retención del inserto; una cepa aislada ura- tal se llamó yACN35. La cepa yACN35 se transformó con ADN linealizado a partir de pACN43. Se cribaron por PCR de colonias colonias ura+ a través de todos los empalmes de integración; una cepa aislada con el inserto deseado en ambos loci AXR4 se llamó yACN45 (3352).

Se hizo la caracterización de la cepa 3352 y la cepa principal ura+ 3083 (versión ura+ de yACN23 con URA3 en el locus AXR4) en matraces con agitación como se describe en el Ejemplo 19. La cepa 3352 demostró una elevada velocidad de crecimiento y DO₆₀₀ máxima en comparación con la cepa principal. A las 24 horas, la cepa 3083 estaba a su DO₆₀₀ máxima de 9,3, mientras que la cepa 3352 había alcanzado una densidad óptica de 13,5. La cepa 3352 usó 20 g/l más de xilosa en 52 horas que la cepa principal (Figura 34). El aumento en la utilización de xilosa produjo niveles de etanol más altos, produciendo la cepa 3352 36,1 g/l de etanol a las 52 horas frente a 26,9 g/l para la principal (34 % de aumento). En comparación con la cepa 3083, la cepa 3352 mostró disminución de los niveles de arabitol (0,3 frente a 1,0 g/l) y glicerol (0,2 frente a 1,5 g/l).

Ejemplo 21: Integración de genes TKL exógenos de *I. orientalis* en *I. orientalis*:

Se incorporaron genes TKL de la secuencia nativa de *I. orientalis* en la cepa yACN55 de *I. orientalis* (Ejemplo 9) para determinar si la expresión en exceso de esta enzima aumentó la fermentación de xilosa a etanol.

Para construir un casete de alteración de AXR4, se amplificaron las regiones flanqueantes de 5' y 3' de S141G4738 a partir del ADN genómico de *I. orientalis* no mutante. Los fragmentos de PCR se clonaron en un esqueleto del vector pCRII con un sitio NotI entre los fragmentos en la dirección 5' y en la dirección 3' y sitios de restricción únicos en los extremos aguas arriba de 5' y aguas abajo de 3'. El plásmido resultante fue de secuencia verificada y se llamó pACN19.

Para construir un vector de expresión de TKL de *I. orientalis*, se amplificó la región codificante del gen TKL de *I. orientalis* (SEQ ID NO:45) a partir de ADN genómico de *I. orientalis* y se clonó en un vector que contenía el promotor TDH3 de *I. orientalis*, terminador TKL, el casete de marcador URA3 y secuencias que se dirigen a AXR4 de forma que el gen TKL estuviera precisamente en la dirección 3' del promotor TDH3. El vector resultante fue de secuencia verificada y se llamó pHJJ113. Se obtuvo un segundo vector que tenía el casete de expresión en orientación opuesta con respecto a las secuencias de direccionamiento uniendo el fragmento NotI de pHJJ113 que llevaba el casete de expresión con un fragmento NotI que llevaba un esqueleto de vector y las secuencias que se dirigen a AXR4. Se confirmaron por PCR la orientación e inserción deseadas en colonias de *E. coli* transformadas con esta ligación. El vector de expresión de TKL resultante se llamó pHJJ114.

Para construir una cepa que expresaba en exceso TKL de *I. orientalis*, se transformó ADN linealizado de pHJJ113 en yACN55 (derivado ura- de la cepa 3356). Se extendieron colonias individuales para la purificación y colonias individuales de cada extensión se aplicaron en parches a ScD-ura. Se cribaron por PCR las colonias para la integridad deseada a través de ambos empalmes de integración. Una cepa que tiene el casete de expresión en exceso de TKL en el sitio AXR4 se llamó yHJJ221.

Los clones de yHJJ221 se cultivaron en YPD y se sembraron en medio ScD-FOA para sacar del bucle el marcador. Se extendieron colonias individuales para la purificación y colonias individuales de cada extensión se aplicaron en parches a YPD. La salida del bucle del marcador y la retención de la integración de TKL se confirmaron por PCR de colonias a través de ambos empalmes de integración. Una cepa de ura- tal se llamó yHJJ226.

Se transformó ADN linealizado de pHJJ114 en yHJJ226 y la transformación se sembró en medio ScD-ura. Se aislaron colonias individuales y se confirmaron por PCR a través de todos los empalmes de integración. La cepa final que contenía copias de TKL de *I. orientalis* en ambos loci AXR4 se llamó la cepa yHJJ242 (12293).

Se realizó una caracterización en matraces oscilantes para comparar la cepa principal 3356 y la cepa 12293 de expresión en exceso de TKL. El medio usado para esta evaluación fue medio YP que contenía 20 g/l de dextrosa, 80 g/l de xilosa y 10 g/l de arabinosa, a pH 5,15. Se usaron células de una placa de ScD-ura nueva para inocular un matraz oscilante primario (50 ml de medio en un matraz de 250 ml). Los matraces se cultivaron a 250 rpm a 37 °C durante aproximadamente 7 horas. Entonces se usaron las células de la semilla primaria para inocular un matraz de semillas secundarias (50 ml de medio en un matraz de 250 ml). La DO₆₀₀ de inoculación objetivo para estos matraces fue 5x10⁶. Estos matraces se cultivaron durante la noche a 37 °C y 250 rpm. Se usaron células de la semilla secundaria para inocular los matraces de producción a una DO₆₀₀ inicial de 0,2. La DO₆₀₀ de las semillas secundarias osciló de 4,4 a 7,2 cuando se tomaron los inóculos. Los matraces de producción se incubaron a 37 °C y 100 rpm de aireación, tomándose muestras una a dos veces por día. Después de usar una porción de la muestra para determinar la DO, el resto de las muestras se centrifugó y los sobrenadantes filtrados se analizaron por HPLC.

La cepa 12293 mostró una gran mejora en el consumo de xilosa y las velocidades de producción de etanol frente a la cepa principal 3356 (Figura 35). Toda la xilosa fue consumida por la cepa 12293 durante las primeras 50 horas de fermentación. Para la cepa 3356, aproximadamente 11 g de xilosa quedaron en el medio después de 120 horas. Además, se redujeron varios subproductos en la cepa 12293 con respecto a 3356: el xilitol fue de 2,8 g/l a 1,5 g/l, acetato de 2,8 a 1,1 g/l, y arabitol de 1,9 a 0 g/l.

Como se ha establecido anteriormente, lo anterior está simplemente previsto para ilustrar diversas realizaciones de la presente invención. Las modificaciones específicas tratadas anteriormente no deben interpretarse como limitaciones del alcance de la invención reivindicada. Será evidente para un experto en la materia que pueden hacerse diversos equivalentes, cambios y modificaciones sin apartarse del alcance de la invención reivindicada, y se entiende que tales realizaciones equivalentes deben incluirse en la presente memoria.

5

LISTA DE SECUENCIAS

<110> Jessen, Holly J
 Yi, Jian
 5 Lundorff, Joshua
 Liao, Hans
 Negrete-Raymond, Ana
 Suominen, Pirkko
 Aristidou, Aristos
 10
 <120> Composiciones y métodos para la elevada producción de etanol a partir de biomasa
 <130> 33449.8055.WO00 (N00095USP1)
 15 <150> US 61/474,035
 <151> 11-04-2011
 <160> 88
 20 <170> PatentIn versión 3.5
 <210> 1
 <211> 1701
 <212> DNA
 25 <213> Kluyveromyces marxianus
 <220>
 <221> CDS
 <222> (1)..(1701)
 30
 <400> 1
 atg tcc gaa gct gct ggt tta caa acg ggc aca gct gct caa agc act 48
 Met Ser Glu Ala Ala Gly Leu Gln Thr Gly Thr Ala Ala Gln Ser Thr
 1 5 10 15
 cct gta gac acc aag tct ttt gag agt tct caa gtt tcg act cca acc 96
 Pro Val Asp Thr Lys Ser Phe Glu Ser Ser Gln Val Ser Thr Pro Thr
 20 25 30
 aac gtt ggc tcg aag gat gag ttg aag gtc gat gag acc aac act gag 144
 Asn Val Gly Ser Lys Asp Glu Leu Lys Val Asp Glu Thr Asn Thr Glu
 35 40 45
 gtt gag ctt cca aag aaa cct gct tcc gct tac atc act gtc tcc att 192
 Val Glu Leu Pro Lys Lys Pro Ala Ser Ala Tyr Ile Thr Val Ser Ile
 50 55 60
 cta tgt tta atg gtt gcc ttt ggt ggt ttc gtt ttc ggt tgg gat acc 240
 Leu Cys Leu Met Val Ala Phe Gly Gly Phe Val Phe Gly Trp Asp Thr
 65 70 75 80
 ggt acc att tct ggt ttt gtt aac caa act gat ttc gtg aga aga ttc 288
 Gly Thr Ile Ser Gly Phe Val Asn Gln Thr Asp Phe Val Arg Arg Phe
 85 90 95
 ggt tct act cat gcc gat ggt act cac tat ttg tct aac gct aga act 336
 Gly Ser Thr His Ala Asp Gly Thr His Tyr Leu Ser Asn Ala Arg Thr
 100 105 110
 ggt atg att gtt tcc att ttc aac att ggt tgt gca ttt ggt ggt atc 384
 Gly Met Ile Val Ser Ile Phe Asn Ile Gly Cys Ala Phe Gly Gly Ile

ES 2 602 808 T3

115	120	125	
ttt ttg tcc aag gtc ggt gac gtt tac ggt cgt cgt att ggt cta atg Phe Leu Ser Lys Val Gly Asp Val Tyr Gly Arg Arg Ile Gly Leu Met 130 135 140			432
gct gtt gtt cta gtt tac gtt gtt ggt att gtt atc caa atc gct tct Ala Val Val Leu Val Tyr Val Val Gly Ile Val Ile Gln Ile Ala Ser 145 150 155 160			480
tct gac aaa tgg tac caa tac ttc atc ggt aga att gtt tcc ggt ttg Ser Asp Lys Trp Tyr Gln Tyr Phe Ile Gly Arg Ile Val Ser Gly Leu 165 170 175			528
ggt gtc ggt ggt atc gct gtc ttg tcc cca atg ttg att tct gaa act Gly Val Gly Gly Ile Ala Val Leu Ser Pro Met Leu Ile Ser Glu Thr 180 185 190			576
gct cca aag caa ttg aga ggt act ttg gtg tct tgt tac caa ttg atg Ala Pro Lys Gln Leu Arg Gly Thr Leu Val Ser Cys Tyr Gln Leu Met 195 200 205			624
att acc ttc ggt atc ttc ttg ggt tac tgt acc aac tac ggt acc aag Ile Thr Phe Gly Ile Phe Leu Gly Tyr Cys Thr Asn Tyr Gly Thr Lys 210 215 220			672
act cac tcc gac tct gtc caa tgg aga gtc cca ttg ggt cta tgt ttc Thr His Ser Asp Ser Val Gln Trp Arg Val Pro Leu Gly Leu Cys Phe 225 230 235 240			720
ttg tgg gcc att ttc atg atc ggt ggt atg ttg ttc gtt cct gaa tcc Leu Trp Ala Ile Phe Met Ile Gly Gly Met Leu Phe Val Pro Glu Ser 245 250 255			768
cca aga tac ttg att gaa aag gac aga att gaa gaa gct aag gct tcc Pro Arg Tyr Leu Ile Glu Lys Asp Arg Ile Glu Glu Ala Lys Ala Ser 260 265 270			816
atc gcc aag tct aac aag gtt tcc atc gaa gac cca gct gtc caa gct Ile Ala Lys Ser Asn Lys Val Ser Ile Glu Asp Pro Ala Val Gln Ala 275 280 285			864
gaa act gat ttg ttg att gcc ggt gtt gaa gct gaa aga cta gct ggt Glu Thr Asp Leu Leu Ile Ala Gly Val Glu Ala Glu Arg Leu Ala Gly 290 295 300			912
tct gct tct ttc aag gag ttg ttc tcc acc aag acc aag gtt ttc caa Ser Ala Ser Phe Lys Glu Leu Phe Ser Thr Lys Thr Lys Val Phe Gln 305 310 315 320			960
cgt ttg gtc atg ggt att atg atc caa tct ttc caa caa ttg acc ggt Arg Leu Val Met Gly Ile Met Ile Gln Ser Phe Gln Gln Leu Thr Gly 325 330 335			1008
aac aac tac ttc ttc tac tac ggt act agt atc ttc aag tcc gtc ggt Asn Asn Tyr Phe Phe Tyr Tyr Gly Thr Ser Ile Phe Lys Ser Val Gly 340 345 350			1056
atg acc gat tct ttc gaa act tct att gtc ttg ggt att gtt aac ttc Met Thr Asp Ser Phe Glu Thr Ser Ile Val Leu Gly Ile Val Asn Phe 355 360 365			1104
gct tcc act ttc ttg ggt atc tac att gtt ggt aga ttt ggc cgt cgt			1152

ES 2 602 808 T3

Ala	Ser	Thr	Phe	Leu	Gly	Ile	Tyr	Ile	Val	Gly	Arg	Phe	Gly	Arg	Arg		
370						375					380						
caa	tgt	ttg	cta	tgg	ggt	gct	gct	cta	atg	acc	tgt	tgt	atg	gtt	gtc	1200	
Gln	Cys	Leu	Leu	Trp	Gly	Ala	Ala	Leu	Met	Thr	Cys	Cys	Met	Val	Val		
385					390					395					400		
ttt	gca	tcc	gtc	ggt	gtt	acc	aag	ttg	tgg	cca	aag	ggt	cca	aac	ggt	1248	
Phe	Ala	Ser	Val	Gly	Val	Thr	Lys	Leu	Trp	Pro	Lys	Gly	Pro	Asn	Gly		
				405					410					415			
ggt	gtt	tct	tct	aag	ggt	gct	ggt	gac	tgt	atg	att	gtc	ttc	acc	tgt	1296	
Gly	Val	Ser	Ser	Lys	Gly	Ala	Gly	Asp	Cys	Met	Ile	Val	Phe	Thr	Cys		
				420				425					430				
ttc	tac	att	cta	tgt	ttc	gct	acc	acc	tgg	gct	cca	att	gct	tac	gtc	1344	
Phe	Tyr	Ile	Leu	Cys	Phe	Ala	Thr	Thr	Trp	Ala	Pro	Ile	Ala	Tyr	Val		
				435				440					445				
gtt	gtt	gct	gaa	tct	tac	cca	ttg	aga	gtc	aag	tcc	aag	tgt	atg	ggt	1392	
Val	Val	Ala	Glu	Ser	Tyr	Pro	Leu	Arg	Val	Lys	Ser	Lys	Cys	Met	Gly		
				450				455					460				
gtc	gct	acc	gct	tct	aac	tgg	gtc	tgg	ggt	ttc	ttg	att	ggt	ttc	ttc	1440	
Val	Ala	Thr	Ala	Ser	Asn	Trp	Val	Trp	Gly	Phe	Leu	Ile	Gly	Phe	Phe		
					470					475					480		
act	cca	ttc	att	act	tct	gac	atc	cac	ttc	tac	tac	ggt	tac	gtc	ttc	1488	
Thr	Pro	Phe	Ile	Thr	Ser	Asp	Ile	His	Phe	Tyr	Tyr	Gly	Tyr	Val	Phe		
				485					490					495			
atg	ggc	tgt	ttg	gtt	gcc	atg	ttc	ttc	tac	gtc	ttc	ttc	ttt	gtc	cca	1536	
Met	Gly	Cys	Leu	Val	Ala	Met	Phe	Phe	Tyr	Val	Phe	Phe	Phe	Val	Pro		
				500				505						510			
gaa	acc	aag	ggt	cta	act	ttg	gaa	gaa	gtc	gat	gaa	atg	tgg	tta	gaa	1584	
Glu	Thr	Lys	Gly	Leu	Thr	Leu	Glu	Glu	Val	Asp	Glu	Met	Trp	Leu	Glu		
				515				520					525				
ggt	gtc	ttg	cca	tgg	aag	tcc	gaa	tca	tgg	gtc	cca	tct	tcc	aga	aga	1632	
Gly	Val	Leu	Pro	Trp	Lys	Ser	Glu	Ser	Trp	Val	Pro	Ser	Ser	Arg	Arg		
				530				535					540				
ggt	gct	gac	tac	aac	gcc	gat	gac	ttg	caa	cac	gat	gac	aag	cca	tgg	1680	
Gly	Ala	Asp	Tyr	Asn	Ala	Asp	Asp	Leu	Gln	His	Asp	Asp	Lys	Pro	Trp		
					550					555					560		
tac	aag	gct	atg	atg	aaa	taa										1701	
Tyr	Lys	Ala	Met	Met	Lys												
				565													

<210> 2
 <211> 566
 <212> PRT
 <213> Kluyveromyces marxianus

<400> 2

Met	Ser	Glu	Ala	Ala	Gly	Leu	Gln	Thr	Gly	Thr	Ala	Ala	Gln	Ser	Thr
1					5				10					15	

ES 2 602 808 T3

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

ES 2 602 808 T3

Ile Ala Lys Ser Asn Lys Val Ser Ile Glu Asp Pro Ala Val Gln Ala
 275 280 285
 Glu Thr Asp Leu Leu Ile Ala Gly Val Glu Ala Glu Arg Leu Ala Gly
 290 295 300
 Ser Ala Ser Phe Lys Glu Leu Phe Ser Thr Lys Thr Lys Val Phe Gln
 305 310 315 320
 Arg Leu Val Met Gly Ile Met Ile Gln Ser Phe Gln Gln Leu Thr Gly
 325 330 335
 Asn Asn Tyr Phe Phe Tyr Tyr Gly Thr Ser Ile Phe Lys Ser Val Gly
 340 345 350
 Met Thr Asp Ser Phe Glu Thr Ser Ile Val Leu Gly Ile Val Asn Phe
 355 360 365
 Ala Ser Thr Phe Leu Gly Ile Tyr Ile Val Gly Arg Phe Gly Arg Arg
 370 375 380
 Gln Cys Leu Leu Trp Gly Ala Ala Leu Met Thr Cys Cys Met Val Val
 385 390 395 400
 Phe Ala Ser Val Gly Val Thr Lys Leu Trp Pro Lys Gly Pro Asn Gly
 405 410 415
 Gly Val Ser Ser Lys Gly Ala Gly Asp Cys Met Ile Val Phe Thr Cys
 420 425 430
 Phe Tyr Ile Leu Cys Phe Ala Thr Thr Trp Ala Pro Ile Ala Tyr Val
 435 440 445
 Val Val Ala Glu Ser Tyr Pro Leu Arg Val Lys Ser Lys Cys Met Gly
 450 455 460
 Val Ala Thr Ala Ser Asn Trp Val Trp Gly Phe Leu Ile Gly Phe Phe
 465 470 475 480
 Thr Pro Phe Ile Thr Ser Asp Ile His Phe Tyr Tyr Gly Tyr Val Phe
 485 490 495
 Met Gly Cys Leu Val Ala Met Phe Phe Tyr Val Phe Phe Phe Val Pro
 500 505 510
 Glu Thr Lys Gly Leu Thr Leu Glu Glu Val Asp Glu Met Trp Leu Glu
 515 520 525
 Gly Val Leu Pro Trp Lys Ser Glu Ser Trp Val Pro Ser Ser Arg Arg
 530 535 540
 Gly Ala Asp Tyr Asn Ala Asp Asp Leu Gln His Asp Asp Lys Pro Trp
 545 550 555 560
 Tyr Lys Ala Met Met Lys
 565

<210> 3
 <211> 2442
 <212> DNA
 5 <213> Kluyveromyces marxianus

<220>
 <221> CDS
 <222> (1)..(2442)

10 <400> 3

atg atc att aac tcc aat cat cga aag aac atc agt aaa gct gcg aga	48
Met Ile Ile Asn Ser Asn His Arg Lys Asn Ile Ser Lys Ala Ala Arg	
1 5 10 15	
ctc gta ttg ctt cga cgg agg gca aat ggc gac aat gag gct ggt gta	96
Leu Val Leu Leu Arg Arg Arg Ala Asn Gly Asp Asn Glu Ala Gly Val	
20 25 30	
tct gga cta cca gct gga aat aaa tca aac aat gga cta tca gga tcg	144
Ser Gly Leu Pro Ala Gly Asn Lys Ser Asn Asn Gly Leu Ser Gly Ser	
35 40 45	
gca tcc ccg caa gag ctt tac caa gaa aat ggc gaa gag ttt gaa cta	192
Ala Ser Pro Gln Glu Leu Tyr Gln Glu Asn Gly Glu Glu Phe Glu Leu	
50 55 60	
agc aat ttt gct tcg ctt acc gac atg gag tcc gag atc ttt tta caa	240
Ser Asn Phe Ala Ser Leu Thr Asp Met Glu Ser Glu Ile Phe Leu Gln	
65 70 75 80	
cct cca gca aaa cag tcg aag aaa ata tca ata ttc gtt ggt tta ttc	288
Pro Pro Ala Lys Gln Ser Lys Lys Ile Ser Ile Phe Val Gly Leu Phe	
85 90 95	
gtt gca gta gga gga ttt cta ttt ggc tac gat acg ggc tta atc aat	336
Val Ala Val Gly Gly Phe Leu Phe Gly Tyr Asp Thr Gly Leu Ile Asn	
100 105 110	
aat gtc agt caa atg cca tat gtt ctt aaa aca atc gct cca aac aaa	384
Asn Val Ser Gln Met Pro Tyr Val Leu Lys Thr Ile Ala Pro Asn Lys	
115 120 125	
cac cag ttc act act tca cag ata tca att cta gta tcc ttt ttg tct	432
His Gln Phe Thr Thr Ser Gln Ile Ser Ile Leu Val Ser Phe Leu Ser	
130 135 140	
ctg ggt act ttc ttc ggg gca tta ttt gca cca ttc ata tct gac cgt	480
Leu Gly Thr Phe Phe Gly Ala Leu Phe Ala Pro Phe Ile Ser Asp Arg	

ES 2 602 808 T3

145	150	155	160	
tat ggg cgg aaa acc acc atg ctc ttt agt act ttt ttt gtc ttt atg				528
Tyr Gly Arg Lys Thr Thr Met Leu Phe Ser Thr Phe Phe Val Phe Met	165	170	175	
ggt ggt aac tcc tta caa gtg gca gcg act tct atg aca tta tta gtc				576
Val Gly Asn Ser Leu Gln Val Ala Ala Thr Ser Met Thr Leu Leu Val	180	185	190	
gta gga agg gta ctg tct gga ctt agt gtt gga cta ata tcc gct gca				624
Val Gly Arg Val Leu Ser Gly Leu Ser Val Gly Leu Ile Ser Ala Ala	195	200	205	
ggt ccc ttg tat cag agt gaa gcc gca cag aaa tct gtt cgt ggt gct				672
Val Pro Leu Tyr Gln Ser Glu Ala Ala Gln Lys Ser Val Arg Gly Ala	210	215	220	
att att tcc acc tat cag tgg gct ata acc tgg ggt tta tta gtt gcc				720
Ile Ile Ser Thr Tyr Gln Trp Ala Ile Thr Trp Gly Leu Leu Val Ala	225	230	235	240
agt gca gtt tca caa gga acc tac aaa aga atg aat gct tca agt tat				768
Ser Ala Val Ser Gln Gly Thr Tyr Lys Arg Met Asn Ala Ser Ser Tyr	245	250	255	
cga att cca ata agt ttg caa tat gtc tgg gct ttt act ctt ggt gtt				816
Arg Ile Pro Ile Ser Leu Gln Tyr Val Trp Ala Phe Thr Leu Gly Val	260	265	270	
ggt gtc tta ttt cta cca gag agt cca cgt tat tat gtt ttc aaa gat				864
Gly Val Leu Phe Leu Pro Glu Ser Pro Arg Tyr Tyr Val Phe Lys Asp	275	280	285	
agg ctt gat cta gct gcc aaa tct tta tca ttc ttg aga gga gtg ccc				912
Arg Leu Asp Leu Ala Ala Lys Ser Leu Ser Phe Leu Arg Gly Val Pro	290	295	300	
gaa gat gat tcc ggt ttg ctc gag gaa tta gta gag ata aag gca acc				960
Glu Asp Asp Ser Gly Leu Leu Glu Glu Leu Val Glu Ile Lys Ala Thr	305	310	315	320
tat gat tat gaa ctg tct ttc ggt aaa aca tca ttt ctt gac tgt ttt				1008
Tyr Asp Tyr Glu Leu Ser Phe Gly Lys Thr Ser Phe Leu Asp Cys Phe	325	330	335	
agg tca acg aaa tca aga tcc aaa caa cgt ctt cgt atg atg acc ggt				1056
Arg Ser Thr Lys Ser Arg Ser Lys Gln Arg Leu Arg Met Met Thr Gly	340	345	350	
att gct ctt caa gca ttt caa caa gta agt ggt att aat ttc att ttt				1104
Ile Ala Leu Gln Ala Phe Gln Gln Val Ser Gly Ile Asn Phe Ile Phe	355	360	365	
tac tat ggt gtt aat ttc ttc aac aaa acg ggt atc aag aat agt tac				1152
Tyr Tyr Gly Val Asn Phe Phe Asn Lys Thr Gly Ile Lys Asn Ser Tyr	370	375	380	
ttg gtg tca ttc att act tac gct gta aat gtg gta ttt aat gtg cca				1200
Leu Val Ser Phe Ile Thr Tyr Ala Val Asn Val Val Phe Asn Val Pro	385	390	395	400
ggt cta ttt tta gtc gaa tat att ggc cgg cgt aaa ctt tta ctt gga				1248

ES 2 602 808 T3

Gly	Leu	Phe	Leu	Val	Glu	Tyr	Ile	Gly	Arg	Arg	Lys	Leu	Leu	Leu	Gly		
				405					410						415		
ggt	ggc	ata	gtg	atg	aca	tta	gca	aat	ttt	acc	atc	gcg	gtt	aca	ggt	1296	
Gly	Gly	Ile	Val	Met	Thr	Leu	Ala	Asn	Phe	Thr	Ile	Ala	Val	Thr	Gly		
			420					425					430				
tta	gtt	gca	gat	tcc	aag	att	gcc	aac	aaa	gtg	atg	ata	gcg	ttt	atc	1344	
Leu	Val	Ala	Asp	Ser	Lys	Ile	Ala	Asn	Lys	Val	Met	Ile	Ala	Phe	Ile		
			435				440					445					
tgt	ttg	ttt	att	gcc	tca	ttc	tca	gca	acc	tgg	ggt	ggt	ggt	gtt	tgg	1392	
Cys	Leu	Phe	Ile	Ala	Ser	Phe	Ser	Ala	Thr	Trp	Gly	Gly	Gly	Val	Trp		
	450					455					460						
gtt	ata	tct	gcc	gaa	ctt	tat	ccc	tta	ggt	gtg	cgt	gct	aag	tgc	act	1440	
Val	Ile	Ser	Ala	Glu	Leu	Tyr	Pro	Leu	Gly	Val	Arg	Ala	Lys	Cys	Thr		
	465				470				475					480			
tct	ata	tgc	gca	gct	tcg	aat	tgg	ctt	ttc	aat	ttc	atc	tgt	gcc	cta	1488	
Ser	Ile	Cys	Ala	Ala	Ser	Asn	Trp	Leu	Phe	Asn	Phe	Ile	Cys	Ala	Leu		
			485					490						495			
ata	acg	ccg	tac	att	gtt	cgt	atc	gac	aat	ggt	caa	cat	tct	tca	acc	1536	
Ile	Thr	Pro	Tyr	Ile	Val	Arg	Ile	Asp	Asn	Gly	Gln	His	Ser	Ser	Thr		
			500					505					510				
atg	ggg	agc	aaa	atc	ttt	ttt	gtg	tgg	ggt	tcg	tta	aac	gct	ata	tcg	1584	
Met	Gly	Ser	Lys	Ile	Phe	Phe	Val	Trp	Gly	Ser	Leu	Asn	Ala	Ile	Ser		
			515				520					525					
gtg	tta	gtc	ggg	tac	ttc	acc	att	tac	gag	act	agt	gga	ctc	tca	tta	1632	
Val	Leu	Val	Gly	Tyr	Phe	Thr	Ile	Tyr	Glu	Thr	Ser	Gly	Leu	Ser	Leu		
	530					535					540						
gaa	gaa	ata	gac	gaa	ttg	tac	aag	aat	tcg	tca	tct	ggt	gtg	gac	tct	1680	
Glu	Glu	Ile	Asp	Glu	Leu	Tyr	Lys	Asn	Ser	Ser	Ser	Gly	Val	Asp	Ser		
	545				550				555					560			
atg	aaa	tgg	aat	aag	aag	ata	agg	tcc	atg	ccg	gaa	ctc	ttc	caa	aga	1728	
Met	Lys	Trp	Asn	Lys	Lys	Ile	Arg	Ser	Met	Pro	Glu	Leu	Phe	Gln	Arg		
			565					570						575			
aac	gca	caa	aat	gat	gat	tca	atc	ggg	gaa	gag	gta	gta	gta	acg	gga	1776	
Asn	Ala	Gln	Asn	Asp	Asp	Ser	Ile	Gly	Glu	Glu	Val	Val	Val	Thr	Gly		
			580					585					590				
aac	aat	gtt	cac	aac	ttt	ggg	gcc	gcg	caa	ggc	tcg	tct	tcg	aac	gag	1824	
Asn	Asn	Val	His	Asn	Phe	Gly	Ala	Ala	Gln	Gly	Ser	Ser	Ser	Asn	Glu		
		595				600						605					
acc	aac	agc	aac	gag	aac	agc	aac	gag	aag	tac	act	tct	cca	ata	gca	1872	
Thr	Asn	Ser	Asn	Glu	Asn	Ser	Asn	Glu	Lys	Tyr	Thr	Ser	Pro	Ile	Ala		
		610				615					620						
atg	cct	caa	ttc	ggt	gca	cgg	agc	atc	gat	cat	cct	tcc	agc	gct	tct	1920	
Met	Pro	Gln	Phe	Gly	Ala	Arg	Ser	Ile	Asp	His	Pro	Ser	Ser	Ala	Ser		
				630					635					640			
gac	atg	ttc	tca	aag	cgt	ctg	ccg	tta	gca	gaa	ctg	aac	ttt	gtg	gac	1968	
Asp	Met	Phe	Ser	Lys	Arg	Leu	Pro	Leu	Ala	Glu	Leu	Asn	Phe	Val	Asp		
				645					650					655			

ttg ggg aac gga ctt gga atc aca acc tac caa cgt ggc cca cct tcc	2016
Leu Gly Asn Gly Leu Gly Ile Thr Thr Tyr Gln Arg Gly Pro Pro Ser	
660 665 670	
gtg ctt aca gat tcg agc gac gag gat gaa gaa gaa caa gac cta gca	2064
Val Leu Thr Asp Ser Ser Asp Glu Asp Glu Glu Glu Asp Leu Ala	
675 680 685	
gat gca tac tct ttg gag cac gcc tcc cag gac aca gag gac ctg cat	2112
Asp Ala Tyr Ser Leu Glu His Ala Ser Gln Asp Thr Glu Asp Leu His	
690 695 700	
cac ctg cat cac ctt acc tcc aac aga cgg aac act aat gga tca gaa	2160
His Leu His His Leu Thr Ser Asn Arg Arg Thr Asn Gly Ser Glu	
705 710 715 720	
cct ttg agc tct aag agc ggc agc agt gct gca gga acg gtg cgc acc	2208
Pro Leu Ser Ser Lys Ser Gly Ser Ser Ala Ala Gly Thr Val Arg Thr	
725 730 735	
tct cca cct aaa cat aac aaa cac aga agg gaa gac ttt aac atg tac	2256
Ser Pro Pro Lys His Asn Lys His Arg Arg Glu Asp Phe Asn Met Tyr	
740 745 750	
atg gct caa tta atc aac cgt ggg tct caa gag gct gta tct tgc tcc	2304
Met Ala Gln Leu Ile Asn Arg Gly Ser Gln Glu Ala Val Ser Cys Ser	
755 760 765	
agc gag cct aaa aac cat cct ata ccc cac gac atc atg agc caa tgg	2352
Ser Glu Pro Lys Asn His Pro Ile Pro His Asp Ile Met Ser Gln Trp	
770 775 780	
aac tcg tct tcc aaa gaa gag tca aac aga cga aat tcc tca aca gac	2400
Asn Ser Ser Ser Lys Glu Glu Ser Asn Arg Arg Asn Ser Ser Thr Asp	
785 790 795 800	
aat agc aac cca tct acg cca aaa aac aca cat cat aaa tag	2442
Asn Ser Asn Pro Ser Thr Pro Lys Asn Thr His His Lys	
805 810	

<210> 4

<211> 813

5 <212> PRT

<213> Kluyveromyces marxianus

<400> 4

Met Ile Ile Asn Ser Asn His Arg Lys Asn Ile Ser Lys Ala Ala Arg
1 5 10 15

Leu Val Leu Leu Arg Arg Arg Ala Asn Gly Asp Asn Glu Ala Gly Val
20 25 30

Ser Gly Leu Pro Ala Gly Asn Lys Ser Asn Asn Gly Leu Ser Gly Ser
35 40 45

Ala Ser Pro Gln Glu Leu Tyr Gln Glu Asn Gly Glu Glu Phe Glu Leu
50 55 60

ES 2 602 808 T3

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

ES 2 602 808 T3

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

ES 2 602 808 T3

565										570					575				
Asn	Ala	Gln	Asn	Asp	Asp	Ser	Ile	Gly	Glu	Glu	Val	Val	Val	Thr	Gly				
			580					585					590						
Asn	Asn	Val	His	Asn	Phe	Gly	Ala	Ala	Gln	Gly	Ser	Ser	Ser	Asn	Glu				
		595					600					605							
Thr	Asn	Ser	Asn	Glu	Asn	Ser	Asn	Glu	Lys	Tyr	Thr	Ser	Pro	Ile	Ala				
	610					615					620								
Met	Pro	Gln	Phe	Gly	Ala	Arg	Ser	Ile	Asp	His	Pro	Ser	Ser	Ala	Ser				
625					630					635						640			
Asp	Met	Phe	Ser	Lys	Arg	Leu	Pro	Leu	Ala	Glu	Leu	Asn	Phe	Val	Asp				
				645					650					655					
Leu	Gly	Asn	Gly	Leu	Gly	Ile	Thr	Thr	Tyr	Gln	Arg	Gly	Pro	Pro	Ser				
			660					665					670						
Val	Leu	Thr	Asp	Ser	Ser	Asp	Glu	Asp	Glu	Glu	Glu	Gln	Asp	Leu	Ala				
		675					680					685							
Asp	Ala	Tyr	Ser	Leu	Glu	His	Ala	Ser	Gln	Asp	Thr	Glu	Asp	Leu	His				
		690				695					700								
His	Leu	His	His	Leu	Thr	Ser	Asn	Arg	Arg	Asn	Thr	Asn	Gly	Ser	Glu				
705					710					715						720			
Pro	Leu	Ser	Ser	Lys	Ser	Gly	Ser	Ser	Ala	Ala	Gly	Thr	Val	Arg	Thr				
				725					730					735					
Ser	Pro	Pro	Lys	His	Asn	Lys	His	Arg	Arg	Glu	Asp	Phe	Asn	Met	Tyr				
			740					745					750						
Met	Ala	Gln	Leu	Ile	Asn	Arg	Gly	Ser	Gln	Glu	Ala	Val	Ser	Cys	Ser				
		755					760					765							
Ser	Glu	Pro	Lys	Asn	His	Pro	Ile	Pro	His	Asp	Ile	Met	Ser	Gln	Trp				
	770					775					780								
Asn	Ser	Ser	Ser	Lys	Glu	Glu	Ser	Asn	Arg	Arg	Asn	Ser	Ser	Thr	Asp				
785					790					795					800				
Asn	Ser	Asn	Pro	Ser	Thr	Pro	Lys	Asn	Thr	His	His	Lys							
				805					810										

<210> 5
<211> 1530
<212> DNA
<213> Secuencia Artificial

<220>
<223> Codón optimizado de B. thetaiotaomicron araA1

<220>
<221> CDS

<222> (1)..(1530)

<400> 5

atg aac aac gtt ttc gat caa tat gaa gtc tgg ttt gtt acc ggt gca	48
Met Asn Asn Val Phe Asp Gln Tyr Glu Val Trp Phe Val Thr Gly Ala	
1 5 10 15	
caa tta tta tat ggt ggt gac gct gtc att gca gtt gac gct cat tct	96
Gln Leu Leu Tyr Gly Gly Asp Ala Val Ile Ala Val Asp Ala His Ser	
20 25 30	
aac gaa atg gtt aac ggt tta aac gaa tct ggt aag ttg cct gtt aag	144
Asn Glu Met Val Asn Gly Leu Asn Glu Ser Gly Lys Leu Pro Val Lys	
35 40 45	
gtt gtc tat aag ggt act gct aac tct tct aag gag gtc gaa gct gtc	192
Val Val Tyr Lys Gly Thr Ala Asn Ser Ser Lys Glu Val Glu Ala Val	
50 55 60	
ttc aag gca gca aac aac gat gac aag tgc gtc ggt gtt att acc tgg	240
Phe Lys Ala Ala Asn Asn Asp Asp Lys Cys Val Gly Val Ile Thr Trp	
65 70 75 80	
atg cat acc ttc tcc cca gct aag atg tgg att cac ggt tta cag caa	288
Met His Thr Phe Ser Pro Ala Lys Met Trp Ile His Gly Leu Gln Gln	
85 90 95	
ttg aag aag cca tta ttg cac ttg cat acc caa ttc aac aag gaa att	336
Leu Lys Lys Pro Leu Leu His Leu His Thr Gln Phe Asn Lys Glu Ile	
100 105 110	
cca tgg gac aca atg gac atg gac ttc atg aat ctt aat caa tct gct	384
Pro Trp Asp Thr Met Asp Met Phe Met Asn Leu Asn Gln Ser Ala	
115 120 125	
cat ggt gat aga gaa ttc ggt cac att tgt acc aga atg aga att aga	432
His Gly Asp Arg Glu Phe Gly His Ile Cys Thr Arg Met Arg Ile Arg	
130 135 140	
aga aag gtc gtc gtc ggt tat tgg aaa gaa gaa gaa aca ttg cat aag	480
Arg Lys Val Val Val Gly Tyr Trp Lys Glu Glu Glu Thr Leu His Lys	
145 150 155 160	
atc gca gtc tgg atg aga gtc tgt gca ggt tgg gct gat tct caa gat	528
Ile Ala Val Trp Met Arg Val Cys Ala Gly Trp Ala Asp Ser Gln Asp	
165 170 175	
atg tta atc att aga ttt ggt gat caa atg aac aac gtt gct gtt act	576
Met Leu Ile Ile Arg Phe Gly Asp Gln Met Asn Asn Val Ala Val Thr	
180 185 190	
gat ggt gat aag gtt gaa gct gaa caa aga atg ggt tac cac gtt gat	624

ES 2 602 808 T3

Asp Gly Asp Lys Val Glu Ala Glu Gln Arg Met Gly Tyr His Val Asp	
195 200 205	
tac tgt cca gct tcc gaa tta atg gaa tat cac aag gat att aag aac	672
Tyr Cys Pro Ala Ser Glu Leu Met Glu Tyr His Lys Asp Ile Lys Asn	
210 215 220	
gct gat gtt gat gca tta gtt gct acc tac ttt aat gat tac gat cat	720
Ala Asp Val Asp Ala Leu Val Ala Thr Tyr Phe Asn Asp Tyr Asp His	
225 230 235 240	
gat gca tct tta gaa gat aag tcc act gag gca tat caa aag gtt tgg	768
Asp Ala Ser Leu Lys Asp Lys Ser Thr Glu Ala Tyr Gln Lys Val Trp	
245 250 255	
aac gca gct aag gct gaa ttg gct tta aga gcc att ttg aag gct aag	816
Asn Ala Ala Lys Glu Ala Glu Leu Ala Asn Arg Ala Ile Leu Lys Ala Lys	
260 265 270	
ggc gca aag ggt ttt act act aac ttc gat gat ttg ggt caa acc gac	864
Gly Ala Lys Gly Phe Thr Thr Asn Phe Asp Asp Leu Gly Gln Thr Asp	
275 280 285	
ggc tct tat ttc gac caa att cca ggt tta gct tct caa aga tta atg	912
Gly Ser Tyr Phe Asp Gln Ile Pro Gly Leu Ala Ser Gln Arg Leu Met	
290 295 300	
gcc gaa ggt tat ggt ttt ggt gct gaa ggt gat tgg aag tct gct gca	960
Ala Glu Gly Tyr Gly Phe Gly Ala Glu Gly Asp Trp Lys Ser Ala Ala	
305 310 315 320	
tta tac aga aca gtt tgg gtt atg aat caa ggt tta cct aag ggt tgt	1008
Leu Tyr Arg Thr Val Trp Val Met Asn Gln Gly Leu Pro Lys Gly Cys	
325 330 335	
tca ttc tta gaa gac tac acc tta aat ttc gat ggt gca aac tct tct	1056
Ser Phe Leu Glu Asp Tyr Thr Leu Asn Phe Asp Gly Ala Asn Ser Ser	
340 345 350	
att tta caa tct cac atg ttg gaa atc tgt cca ttg att gct gca aac	1104
Ile Leu Gln Ser His Met Leu Glu Ile Cys Pro Leu Ile Ala Ala Asn	
355 360 365	
aag cca aga ttg gaa gtt cac ttc ttg ggt atc ggt att aga aag tct	1152
Lys Pro Arg Leu Glu Val His Phe Leu Gly Ile Gly Ile Arg Lys Ser	
370 375 380	
caa aca gct aga ttg gtt ttc act tcc aag acc ggt act ggt tgt act	1200
Gln Thr Ala Arg Leu Val Phe Thr Ser Lys Thr Gly Thr Gly Cys Thr	
385 390 395 400	
gca acc gtc gtt gac atg ggt aac aga ttt aga ttg att gtc aat gac	1248
Ala Thr Val Val Asp Met Gly Asn Arg Phe Arg Leu Ile Val Asn Asp	
405 410 415	
gtc gaa tgt att gaa cca aag cca tta cca aaa ttg cca gtt gct tca	1296
Val Glu Cys Ile Glu Pro Lys Pro Leu Pro Lys Leu Pro Val Ala Ser	
420 425 430	
gct tta tgg att cca atg cca aac tta gag gtt ggt gca ggt gct tgg	1344
Ala Leu Trp Ile Pro Met Pro Asn Leu Glu Val Gly Ala Gly Ala Trp	
435 440 445	

ES 2 602 808 T3

```

atc tta gct ggt ggt aca cac cac tct tgt ttc tct tat gat ttg acc      1392
Ile Leu Ala Gly Gly Thr His His Ser Cys Phe Ser Tyr Asp Leu Thr
    450                      455                      460

gct gaa tac tgg gaa gac tac gct gaa atc gca ggt att gaa atg gtt      1440
Ala Glu Tyr Trp Glu Asp Tyr Ala Glu Ile Ala Gly Ile Glu Met Val
    465                      470                      475                      480

cac att aac aag gat act act att tct tgc ttt aag aag gaa tta aga      1488
His Ile Asn Lys Asp Thr Thr Ile Ser Cys Phe Lys Lys Glu Leu Arg
    485                      490                      495

atg aac gaa gtc tat tat atg ttg aac aaa gct tta tgt taa      1530
Met Asn Glu Val Tyr Tyr Met Leu Asn Lys Ala Leu Cys
    500                      505

```

<210> 6

<211> 509

5 <212> PRT

<213> Secuencia Artificial

<220>

<223> Construcción Sintética

10

<400> 6

```

Met Asn Asn Val Phe Asp Gln Tyr Glu Val Trp Phe Val Thr Gly Ala
1      5      10      15

Gln Leu Leu Tyr Gly Gly Asp Ala Val Ile Ala Val Asp Ala His Ser
    20      25      30

Asn Glu Met Val Asn Gly Leu Asn Glu Ser Gly Lys Leu Pro Val Lys
    35      40      45

Val Val Tyr Lys Gly Thr Ala Asn Ser Ser Lys Glu Val Glu Ala Val
    50      55      60

Phe Lys Ala Ala Asn Asn Asp Asp Lys Cys Val Gly Val Ile Thr Trp
    65      70      75      80

Met His Thr Phe Ser Pro Ala Lys Met Trp Ile His Gly Leu Gln Gln
    85      90      95

Leu Lys Lys Pro Leu Leu His Leu His Thr Gln Phe Asn Lys Glu Ile
    100     105     110

Pro Trp Asp Thr Met Asp Met Asp Phe Met Asn Leu Asn Gln Ser Ala
    115     120     125

His Gly Asp Arg Glu Phe Gly His Ile Cys Thr Arg Met Arg Ile Arg
    130     135     140

```

ES 2 602 808 T3

Arg Lys Val Val Val Gly Tyr Trp Lys Glu Glu Glu Thr Leu His Lys
 145 150 155 160
 Ile Ala Val Trp Met Arg Val Cys Ala Gly Trp Ala Asp Ser Gln Asp
 165 170 175
 Met Leu Ile Ile Arg Phe Gly Asp Gln Met Asn Asn Val Ala Val Thr
 180 185 190
 Asp Gly Asp Lys Val Glu Ala Glu Gln Arg Met Gly Tyr His Val Asp
 195 200 205
 Tyr Cys Pro Ala Ser Glu Leu Met Glu Tyr His Lys Asp Ile Lys Asn
 210 215 220
 Ala Asp Val Asp Ala Leu Val Ala Thr Tyr Phe Asn Asp Tyr Asp His
 225 230 235 240
 Asp Ala Ser Leu Glu Asp Lys Ser Thr Glu Ala Tyr Gln Lys Val Trp
 245 250 255
 Asn Ala Ala Lys Ala Glu Leu Ala Leu Arg Ala Ile Leu Lys Ala Lys
 260 265 270
 Gly Ala Lys Gly Phe Thr Thr Asn Phe Asp Asp Leu Gly Gln Thr Asp
 275 280 285
 Gly Ser Tyr Phe Asp Gln Ile Pro Gly Leu Ala Ser Gln Arg Leu Met
 290 295 300
 Ala Glu Gly Tyr Gly Phe Gly Ala Glu Gly Asp Trp Lys Ser Ala Ala
 305 310 315 320
 Leu Tyr Arg Thr Val Trp Val Met Asn Gln Gly Leu Pro Lys Gly Cys
 325 330 335
 Ser Phe Leu Glu Asp Tyr Thr Leu Asn Phe Asp Gly Ala Asn Ser Ser
 340 345 350
 Ile Leu Gln Ser His Met Leu Glu Ile Cys Pro Leu Ile Ala Ala Asn
 355 360 365
 Lys Pro Arg Leu Glu Val His Phe Leu Gly Ile Gly Ile Arg Lys Ser
 370 375 380
 Gln Thr Ala Arg Leu Val Phe Thr Ser Lys Thr Gly Thr Gly Cys Thr
 385 390 395 400

ES 2 602 808 T3

Ala Thr Val Val Asp Met Gly Asn Arg Phe Arg Leu Ile Val Asn Asp
405 410 415

Val Glu Cys Ile Glu Pro Lys Pro Leu Pro Lys Leu Pro Val Ala Ser
420 425 430

Ala Leu Trp Ile Pro Met Pro Asn Leu Glu Val Gly Ala Gly Ala Trp
435 440 445

Ile Leu Ala Gly Gly Thr His His Ser Cys Phe Ser Tyr Asp Leu Thr
450 455 460

Ala Glu Tyr Trp Glu Asp Tyr Ala Glu Ile Ala Gly Ile Glu Met Val
465 470 475 480

His Ile Asn Lys Asp Thr Thr Ile Ser Cys Phe Lys Lys Glu Leu Arg
485 490 495

Met Asn Glu Val Tyr Tyr Met Leu Asn Lys Ala Leu Cys
500 505

<210> 7

<211> 1425

5 <212> DNA

<213> Secuencia Artificial

<220>

<223> Codón optimizado de B. thetaiotaomicron araA2

10

<220>

<221> CDS

<222> (1)..(1425)

15 <400> 7

atg ata caa caa aaa gta aga gtg gga ctt tta gga gtc gga ttg gat 48
Met Ile Gln Gln Lys Val Arg Val Gly Leu Leu Gly Val Gly Leu Asp
1 5 10 15

act tat tgg gga caa ttt gaa ggg ctt ctt ccc cgt ctg tta act tat 96
Thr Tyr Trp Gly Gln Phe Glu Gly Leu Leu Pro Arg Leu Leu Thr Tyr
20 25 30

cag gat gaa ata gct gcc aaa ata gag gca atg gac gta caa gtt atc 144
Gln Asp Glu Ile Ala Ala Lys Ile Glu Ala Met Asp Val Gln Val Ile
35 40 45

aat acg ggg atg gtg gat tcg cct cta aaa gct aat gaa tgt gtc cta 192
Asn Thr Gly Met Val Asp Ser Pro Leu Lys Ala Asn Glu Cys Val Leu
50 55 60

caa tta aaa caa gct gat gta gaa ttg gtt ttt ctg ttt att tct act 240
Gln Leu Lys Gln Ala Asp Val Glu Leu Val Phe Leu Phe Ile Ser Thr
65 70 75 80

ES 2 602 808 T3

tat gct ctt tct tct acg ata cta cct gta gct caa caa gta ggc aag Tyr Ala Leu Ser Ser Thr Ile Leu Pro Val Ala Gln Gln Val Gly Lys 85 90 95	288
cct att att ata tta aat ata cag cct gca tca gca ata gat tat caa Pro Ile Ile Ile Leu Asn Ile Gln Pro Ala Ser Ala Ile Asp Tyr Gln 100 105 110	336
aaa ctc aat tca atg ggg gat aga gga aga atg aca ggt gaa tgg ttg Lys Leu Asn Ser Met Gly Asp Arg Gly Arg Met Thr Gly Glu Trp Leu 115 120 125	384
gca cat tgt cag gca tgt tct gtt ccg gag ttt gcg agt gtt ttg aat Ala His Cys Gln Ala Cys Ser Val Pro Glu Phe Ala Ser Val Leu Asn 130 135 140	432
aga gcc ggt gtg cga tat gat att atc acc ggc tat tta tca gaa gat Arg Ala Gly Val Arg Tyr Asp Ile Ile Thr Gly Tyr Leu Ser Glu Asp 145 150 155 160	480
tat gtt tgg gag gaa ata gcc tct tgg gtt gac gct gta cgt gta atg Tyr Val Trp Glu Glu Ile Ala Ser Trp Val Asp Ala Val Arg Val Met 165 170 175	528
tat gga atg cgc aca agc cgt ttg gga gtt ttg ggg cat tac tat tgt Tyr Gly Met Arg Thr Ser Arg Leu Gly Val Leu Gly His Tyr Tyr Cys 180 185 190	576
ggt atg cta gat gta tat act gat ctt atg aag cag agc gca gtg ttt Gly Met Leu Asp Val Tyr Thr Asp Leu Met Lys Gln Ser Ala Val Phe 195 200 205	624
ggt act cat ata gag ttg ctg gag atg tgt gag cta aaa gct tat aga Gly Thr His Ile Glu Leu Leu Glu Met Cys Glu Leu Lys Ala Tyr Arg 210 215 220	672
gaa gaa gtt agc gat ggg gaa ctg aag cga aaa ctg gac gaa ttt tat Glu Glu Val Ser Asp Gly Glu Leu Lys Arg Lys Leu Asp Glu Phe Tyr 225 230 235 240	720
gat aag ttt aat gtg gaa gca tca tgt agt tcg gaa gag cta gtg agg Asp Lys Phe Asn Val Glu Ala Ser Cys Ser Ser Glu Glu Leu Val Arg 245 250 255	768
gct gca cgt act tct gtg gct ttg gat aaa tta gtg aat gtg cat caa Ala Ala Arg Thr Ser Val Ala Leu Asp Lys Leu Val Asn Val His Gln 260 265 270	816
ctg gga gcg atg gct tat tat tac gaa gga ttc tgt ggg aat gat tat Leu Gly Ala Met Ala Tyr Tyr Tyr Glu Gly Phe Cys Gly Asn Asp Tyr 275 280 285	864
gag aat att gta act tct gtt att gca ggt aat acg ttg ttg aca gga Glu Asn Ile Val Thr Ser Val Ile Ala Gly Asn Thr Leu Leu Thr Gly 290 295 300	912
tat gga ata ccc gtt gcc gga gaa tgt gaa gtg aaa aat gca cag gcg Tyr Gly Ile Pro Val Ala Gly Glu Cys Glu Val Lys Asn Ala Gln Ala 305 310 315 320	960
atg aaa att atg tca cta tta aaa gct ggt ggt tct ttc tct gaa ttt Met Lys Ile Met Ser Leu Leu Lys Ala Gly Gly Ser Phe Ser Glu Phe 325 330 335	1008

ES 2 602 808 T3

tat gca atg gat ttt aag gat gat att gta tta cta gga cat gat ggc	1056
Tyr Ala Met Asp Phe Lys Asp Asp Ile Val Leu Leu Gly His Asp Gly	
340 345 350	
ccg gca cat ttt gca atc gcc gaa gag aaa gtg aaa cta gtg cct ctt	1104
Pro Ala His Phe Ala Ile Ala Glu Glu Lys Val Lys Leu Val Pro Leu	
355 360 365	
cca tta tat cac ggt aaa ccg ggt aaa ggt ctg tcc atc cag atg agt	1152
Pro Leu Tyr His Gly Lys Pro Gly Lys Gly Leu Ser Ile Gln Met Ser	
370 375 380	
gtt aaa cca ggt gat gtt aca ctt ttg tcg gta tgt gag gga aga gac	1200
Val Lys Pro Gly Asp Val Thr Leu Leu Ser Val Cys Glu Gly Arg Asp	
385 390 395 400	
gga gtc ttt cta ctt gct gct gag gga gag gct gta caa gga gaa act	1248
Gly Val Phe Leu Leu Ala Ala Glu Gly Glu Ala Val Gln Gly Glu Thr	
405 410 415	
cta cat att ggc aat aca aat agt cgt tat cgc ttt ccg tgt ggt gct	1296
Leu His Ile Gly Asn Thr Asn Ser Arg Tyr Arg Phe Pro Cys Gly Ala	
420 425 430	
cgt cgg ttt atg gat caa tgg agt aaa gcg gga cct tca cat cat tgt	1344
Arg Arg Phe Met Asp Gln Trp Ser Lys Ala Gly Pro Ser His His Cys	
435 440 445	
gcg att ggc att gga cat aaa gtc tct gaa ctg aag aaa ctg gca ttt	1392
Ala Ile Gly Ile Gly His Lys Val Ser Glu Leu Lys Lys Leu Ala Phe	
450 455 460	
ctt ctg gat att cca ata ata gta gtt gaa taa	1425
Leu Leu Asp Ile Pro Ile Ile Val Val Glu	
465 470	

<210> 8

<211> 474

5 <212> PRT

<213> Secuencia Artificial

<220>

10 <223> Construcción Sintética

<400> 8

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

ES 2 602 808 T3

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

ES 2 602 808 T3

Met Lys Ile Met Ser Leu Leu Lys Ala Gly Gly Ser Phe Ser Glu Phe
325 330 335

Tyr Ala Met Asp Phe Lys Asp Asp Ile Val Leu Leu Gly His Asp Gly
340 345 350

Pro Ala His Phe Ala Ile Ala Glu Glu Lys Val Lys Leu Val Pro Leu
355 360 365

Pro Leu Tyr His Gly Lys Pro Gly Lys Gly Leu Ser Ile Gln Met Ser
370 375 380

Val Lys Pro Gly Asp Val Thr Leu Leu Ser Val Cys Glu Gly Arg Asp
385 390 395 400

Gly Val Phe Leu Leu Ala Ala Glu Gly Glu Ala Val Gln Gly Glu Thr
405 410 415

Leu His Ile Gly Asn Thr Asn Ser Arg Tyr Arg Phe Pro Cys Gly Ala
420 425 430

Arg Arg Phe Met Asp Gln Trp Ser Lys Ala Gly Pro Ser His His Cys
435 440 445

Ala Ile Gly Ile Gly His Lys Val Ser Glu Leu Lys Lys Leu Ala Phe
450 455 460

Leu Leu Asp Ile Pro Ile Ile Val Val Glu
465 470

<210> 9

<211> 1425

5 <212> DNA

<213> Secuencia Artificial

<220>

<223> Codón optimizado de L. sakeii araA

10

<220>

<221> CDS

<222> (1)..(1425)

15 <400> 9

atg ttg aat acc gag aat tac gag ttt tgg ttc gtt act ggt tcc caa 48
Met Leu Asn Thr Glu Asn Tyr Glu Phe Trp Phe Val Thr Gly Ser Gln
1 5 10 15

tcc ttg tat ggt gag gaa acc tta aga tct gtc gaa aaa gat gca aag 96
Ser Leu Tyr Gly Glu Glu Thr Leu Arg Ser Val Glu Lys Asp Ala Lys
20 25 30

ES 2 602 808 T3

gaa att gtt gag aaa ctt aat gcc tcc cat caa ttg cca tat cca att Glu Ile Val Glu Lys Leu Asn Ala Ser His Gln Leu Pro Tyr Pro Ile 35 40 45	144
gtt ttc aag tta gtt gcc act aca gca gat aac att acc aag gtt atg Val Phe Lys Leu Val Ala Thr Thr Ala Asp Asn Ile Thr Lys Val Met 50 55 60	192
aag gaa gca aac tac aat gat cat gtt gcc ggt gtt atc act tgg atg Lys Glu Ala Asn Tyr Asn Asp His Val Ala Gly Val Ile Thr Trp Met 65 70 75 80	240
cat acc ttt tct cct gcc aaa aac tgg atc aga ggt act aag tta ttg His Thr Phe Ser Pro Ala Lys Asn Trp Ile Arg Gly Thr Lys Leu Leu 85 90 95	288
caa aag cca tta ctt cac tta gca act cag ttc ctt aac aaa atc cca Gln Lys Pro Leu Leu His Leu Ala Thr Gln Phe Leu Asn Lys Ile Pro 100 105 110	336
tac gat act att gat ttc gac tac atg aac ttg aac caa tct gct cat Tyr Asp Thr Ile Asp Phe Asp Tyr Met Asn Leu Asn Gln Ser Ala His 115 120 125	384
ggt gac cgt gaa tac gca ttc atc aat gct aga ttg aga aag aat aac Gly Asp Arg Glu Tyr Ala Phe Ile Asn Ala Arg Leu Arg Lys Asn Asn 130 135 140	432
aag atc att tct ggt tat tgg ggt gat gaa gat gtc caa aag gct atg Lys Ile Ile Ser Gly Tyr Trp Gly Asp Glu Asp Val Gln Lys Ala Met 145 150 155 160	480
gca aaa tgg atg gat gtt gca gtc gct tac aac gaa tct ttc aag att Ala Lys Trp Met Asp Val Ala Val Ala Tyr Asn Glu Ser Phe Lys Ile 165 170 175	528
aag gtt gtc acc ttc gcc gat aag atg aga aat gtt gct gtc acc gac Lys Val Val Thr Phe Ala Asp Lys Met Arg Asn Val Ala Val Thr Asp 180 185 190	576
ggt gat aag gtc gag gca caa atc aag ttc gcc tgg aca gtt gat tac Gly Asp Lys Val Glu Ala Gln Ile Lys Phe Gly Trp Thr Val Asp Tyr 195 200 205	624
tgg ggt gtt ggc gat ctt gtt gct gaa gtt aat gcc gtt tct gaa gct Trp Gly Val Gly Asp Leu Val Ala Glu Val Asn Ala Val Ser Glu Ala 210 215 220	672
gac att gat gca aag tat gct gac ttg caa aag gaa tac gat ttt gtc Asp Ile Asp Ala Lys Tyr Ala Asp Leu Gln Lys Glu Tyr Asp Phe Val 225 230 235 240	720
gaa ggt caa aac act cca gaa aag ttt gaa cac aac gtt aag tat cag Glu Gly Gln Asn Thr Pro Glu Lys Phe Glu His Asn Val Lys Tyr Gln 245 250 255	768
atc aga gaa tac ttt ggt ttg aaa aag ttt atg gat gat aga ggt tat Ile Arg Glu Tyr Phe Gly Leu Lys Lys Phe Met Asp Asp Arg Gly Tyr 260 265 270	816
act gca ttc acc acc aat ttc gag gac tta gtc ggt tta gaa cag tta Thr Ala Phe Thr Thr Asn Phe Glu Asp Leu Val Gly Leu Glu Gln Leu 280 285 290 295	864

ES 2 602 808 T3

275	280	285	
cct ggt tta gct gct caa ttg ttg atg gct gag ggt tat ggt ttt gct			912
Pro Gly Leu Ala Ala Gln Leu Leu Met Ala Glu Gly Tyr Gly Phe Ala			
290	295	300	
ggt gaa ggt gac tgg aaa aca gca gct tta gat aga ctt ttg aag att			960
Gly Glu Gly Asp Trp Lys Thr Ala Ala Leu Asp Arg Leu Leu Lys Ile			
305	310	315	320
atg gct cac aac gaa aag aca gtc ttt atg gaa gat tac acc tta gac			1008
Met Ala His Asn Glu Lys Thr Val Phe Met Glu Asp Tyr Thr Leu Asp			
	325	330	335
ctt aga caa ggt cac gaa gca atc tta ggc tca cat atg ttg gaa gtt			1056
Leu Arg Gln Gly His Glu Ala Ile Leu Gly Ser His Met Leu Glu Val			
	340	345	350
gat cca tca att gct tct gac aaa cct cgt gtc gaa gtt cac cca tta			1104
Asp Pro Ser Ile Ala Ser Asp Lys Pro Arg Val Glu Val His Pro Leu			
	355	360	365
gat att ggt gat aaa gac gat cca gct aga tta gtt ttc acc ggt atg			1152
Asp Ile Gly Asp Lys Asp Asp Pro Ala Arg Leu Val Phe Thr Gly Met			
	370	375	380
caa ggc gac gct gtt gat gtt aca atg gca gat tat ggt gac gag ttc			1200
Gln Gly Asp Ala Val Asp Val Thr Met Ala Asp Tyr Gly Asp Glu Phe			
	385	390	400
aag ttg atg tct tac gat gtc aga ggt aac aaa cca gaa gca gat acc			1248
Lys Leu Met Ser Tyr Asp Val Arg Gly Asn Lys Pro Glu Ala Asp Thr			
	405	410	415
cca cat ttg cca gtt gcc aaa cag ttg tgg act cca aag caa ggt tta			1296
Pro His Leu Pro Val Ala Lys Gln Leu Trp Thr Pro Lys Gln Gly Leu			
	420	425	430
aga gaa ggt gca gtc ggt tgg ctt aca gtt ggc ggt ggt cat cac act			1344
Arg Glu Gly Ala Val Gly Trp Leu Thr Val Gly Gly His His Thr			
	435	440	445
gtt ttg tca ttt gct gtt gat tcc gaa caa ttg caa gac tta tcc cac			1392
Val Leu Ser Phe Ala Val Asp Ser Glu Gln Leu Gln Asp Leu Ser His			
	450	455	460
ttg ttc gac ttg act tat gtc aac att aag taa			1425
Leu Phe Asp Leu Thr Tyr Val Asn Ile Lys			
	465	470	

<210> 10

<211> 474

5 <212> PRT

<213> Secuencia Artificial

<220>

<223> Construcción Sintética

10

<400> 10

Met	Leu	Asn	Thr	Glu	Asn	Tyr	Glu	Phe	Trp	Phe	Val	Thr	Gly	Ser	Gln
1			5					10					15		

ES 2 602 808 T3

Ser Leu Tyr Gly Glu Glu Thr Leu Arg Ser Val Glu Lys Asp Ala Lys
 20 25 30
 Glu Ile Val Glu Lys Leu Asn Ala Ser His Gln Leu Pro Tyr Pro Ile
 35 40 45
 Val Phe Lys Leu Val Ala Thr Thr Ala Asp Asn Ile Thr Lys Val Met
 50 55 60
 Lys Glu Ala Asn Tyr Asn Asp His 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 Lys Leu Leu
 85 90 95
 Gln Lys Pro Leu Leu His Leu Ala Thr Gln Phe Leu Asn Lys Ile Pro
 100 105 110
 Tyr Asp Thr Ile Asp Phe Asp Tyr Met Asn Leu Asn Gln Ser Ala His
 115 120 125
 Gly Asp Arg Glu Tyr Ala Phe Ile Asn Ala Arg Leu Arg Lys Asn Asn
 130 135 140
 Lys Ile Ile Ser Gly Tyr Trp Gly Asp Glu Asp Val Gln Lys Ala Met
 145 150 155 160
 Ala Lys Trp Met Asp Val Ala Val Ala Tyr Asn Glu Ser Phe Lys Ile
 165 170 175
 Lys Val Val Thr Phe Ala Asp Lys Met Arg Asn Val Ala Val Thr Asp
 180 185 190
 Gly Asp Lys Val Glu Ala Gln Ile Lys Phe Gly Trp Thr Val Asp Tyr
 195 200 205
 Trp Gly Val Gly Asp Leu Val Ala Glu Val Asn Ala Val Ser Glu Ala
 210 215 220
 Asp Ile Asp Ala Lys Tyr Ala Asp Leu Gln Lys Glu Tyr Asp Phe Val
 225 230 235 240
 Glu Gly Gln Asn Thr Pro Glu Lys Phe Glu His Asn Val Lys Tyr Gln
 245 250 255
 Ile Arg Glu Tyr Phe Gly Leu Lys Lys Phe Met Asp Asp Arg Gly Tyr

ES 2 602 808 T3

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

<210> 11

<211> 1596

5 <212> DNA

<213> Secuencia Artificial

<220>

<223> Codón optimizado de B. thetaiotaomicron araB

10

<220>

<221> CDS

<222> (1)..(1596)

15

<400> 11

ES 2 602 808 T3

atg aag tta gat gct aag tcc acc att gaa act ggt aag gct att tta Met Lys Leu Asp Ala Lys Ser Thr Ile Glu Thr Gly Lys Ala Ile Leu 1 5 10 15	48
ggt atc gaa ttg ggt tct aca aga att aag gct gtt ttg att gat caa Gly Ile Glu Leu Gly Ser Thr Arg Ile Lys Ala Val Leu Ile Asp Gln 20 25 30	96
gaa aac aaa cca atc gct caa ggt tca cac acc tgg gaa aat caa tta Glu Asn Lys Pro Ile Ala Gln Gly Ser His Thr Trp Glu Asn Gln Leu 35 40 45	144
gtt aac ggt tta tgg act tac tcc atc gac gct att tgg tcc ggc ttg Val Asn Gly Leu Trp Thr Tyr Ser Ile Asp Ala Ile Trp Ser Gly Leu 50 55 60	192
caa gat tgt tat gct gac ttg aga tcc aat gtt aag aag ctt tac gat Gln Asp Cys Tyr Ala Asp Leu Arg Ser Asn Val Lys Lys Leu Tyr Asp 65 70 75 80	240
act gaa att gaa acc tta gca gct att ggt gtc tct gca atg atg cac Thr Glu Ile Glu Thr Leu Ala Ala Ile Gly Val Ser Ala Met Met His 85 90 95	288
ggt tat atg cca ttc aac gaa aag gaa gaa att ttg gtc cca ttc aga Gly Tyr Met Pro Phe Asn Glu Lys Glu Glu Ile Leu Val Pro Phe Arg 100 105 110	336
act tgg aga aat act aac act ggt aga gct gca gct gaa ttg tct gaa Thr Trp Arg Asn Thr Asn Thr Gly Arg Ala Ala Ala Glu Leu Ser Glu 115 120 125	384
ttg ttc gtc tac aat atc cca tta aga tgg tcc att tcc cat ctt tac Leu Phe Val Tyr Asn Ile Pro Leu Arg Trp Ser Ile Ser His Leu Tyr 130 135 140	432
caa gca atc ttg gat aat gaa gct cac gtc aag gac atc aag ttc ctt Gln Ala Ile Leu Asp Asn Glu Ala His Val Lys Asp Ile Lys Phe Leu 145 150 155 160	480
act act tta gct ggt tat gtc cac tgg caa atc act ggt gaa aag gtc Thr Thr Leu Ala Gly Tyr Val His Trp Gln Ile Thr Gly Glu Lys Val 165 170 175	528
tta ggt att ggt gat gct tcc ggt atg ttg cca atc gac cct act acc Leu Gly Ile Gly Asp Ala Ser Gly Met Leu Pro Ile Asp Pro Thr Thr 180 185 190	576
aac aat tac tct gct gaa atg gtt gct aag ttc aac aac ctt att gct Asn Asn Tyr Ser Ala Glu Met Val Ala Lys Phe Asn Asn Leu Ile Ala 195 200 205	624
tcc aaa gaa tac tct tgg aag ctt gaa gat att ttg cca aag gtc tta Ser Lys Glu Tyr Ser Trp Lys Leu Glu Asp Ile Leu Pro Lys Val Leu 210 215 220	672
tct gca ggt gaa aac gct ggt gtt ttg acc cca gaa ggt tgt aag aaa	720

ES 2 602 808 T3

Ser 225	Ala	Gly	Glu	Asn 230	Ala	Gly	Val	Leu	Thr	Pro 235	Glu	Gly	Cys	Lys	Lys 240	
tta gat gct tcc ggt cat tta aag gca ggt atc cca gtt tgt cca cca Leu Asp Ala Ser Gly His Leu Lys Ala Gly Ile Pro Val Cys Pro Pro 245 250 255																768
gaa ggt gat gca ggt act ggc atg gtc gct acc aat gca gtc aag caa Glu Gly Asp Ala Gly Thr Gly Met Val Ala Thr Asn Ala Val Lys Gln 260 265 270																816
aga act ggt aac gtc tct gca ggt act tct tcc ttc tcc atg att gtc Arg Thr Gly Asn Val Ser Ala Gly Thr Ser Ser Phe Ser Met Ile Val 275 280 285																864
tta gag aag gaa tta tct aag cca tac gaa atg att gat atg gtt act Leu Glu Lys Glu Leu Ser Lys Pro Tyr Glu Met Ile Asp Met Val Thr 290 295 300																912
acc cca gat ggt tct tta gtt gcc atg gtt cac tgt aat aac tgt act Thr Pro Asp Gly Ser Leu Val Ala Met Val His Cys Asn Asn Cys Thr 305 310 315 320																960
tct gac tta aac gct tgg gtc aac tta ttc aag gaa tat caa gaa ttg Ser Asp Leu Asn Ala Trp Val Asn Leu Phe Lys Glu Tyr Gln Glu Leu 325 330 335																1008
tta ggt att cca gtc gat atg gac gaa tta tac ggt aag tta tac aat Leu Gly Ile Pro Val Asp Met Asp Glu Leu Tyr Gly Lys Leu Tyr Asn 340 345 350																1056
att gct tta act ggt gat act gat tgt ggc ggt ctt tta tct tat aat Ile Ala Leu Thr Gly Asp Thr Asp Cys Gly Gly Leu Leu Ser Tyr Asn 355 360 365																1104
tat att tcc ggt gaa cca gtc act ggt tta gca gaa ggt aga cca ttg Tyr Ile Ser Gly Glu Pro Val Thr Gly Leu Ala Glu Gly Arg Pro Leu 370 375 380																1152
ttt gtc aga tcc gca aac gac aag ttt aat ttg gct aac ttt atg aga Phe Val Arg Ser Ala Asn Asp Lys Phe Asn Leu Ala Asn Phe Met Arg 385 390 395 400																1200
gca cac ttg tac gca tct gtc ggt gtt ctt aaa atc ggt aac gac atc Ala His Leu Tyr Ala Ser Val Gly Val Leu Lys Ile Gly Asn Asp Ile 405 410 415																1248
ttg ttc aac gaa gaa aag atc aag gtt gac aga att act ggt cat ggt Leu Phe Asn Glu Glu Lys Ile Lys Val Asp Arg Ile Thr Gly His Gly 420 425 430																1296
ggt tta ttc aga act aaa ggt gtt ggt caa aga gtt tta gca gct gct Gly Leu Phe Arg Thr Lys Gly Val Gly Gln Arg Val Leu Ala Ala Ala 435 440 445																1344
att aat tca cca att tcc gtt atg gaa act gca ggt gaa ggt ggt gct Ile Asn Ser Pro Ile Ser Val Met Glu Thr Ala Gly Glu Gly Gly Ala 450 455 460																1392
tgg ggt atc gca tta tta ggc tca tac tta gtc aat aac aag aaa ggt Trp Gly Ile Ala Leu Leu Gly Ser Tyr Leu Val Asn Asn Lys Lys Gly 465 470 475 480																1440

ES 2 602 808 T3

caa tcc tta gct gat ttc tta gac gaa tct gtt ttc gtt tct gac gca	1488
Gln Ser Leu Ala Asp Phe Leu Asp Glu Ser Val Phe Val Ser Asp Ala	
485 490 495	
ggt gtt gaa gtt tct cca acc cca gaa gat gtt gct ggc ttc aat acc	1536
Gly Val Glu Val Ser Pro Thr Pro Glu Asp Val Ala Gly Phe Asn Thr	
500 505 510	
tat atc gaa tcc tat aag gca ggt ttg cca att gag gaa gca gct gtt	1584
Tyr Ile Glu Ser Tyr Lys Ala Gly Leu Pro Ile Glu Glu Ala Ala Val	
515 520 525	
aag ttc aag taa	1596
Lys Phe Lys	
530	

<210> 12

<211> 531

5 <212> PRT

<213> Secuencia Artificial

<220>

10 <223> Construcción Sintética

<400> 12

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

Gly Ile Glu Leu Gly Ser Thr Arg Ile Lys Ala Val Leu Ile Asp Gln
20 25 30

Glu Asn Lys Pro Ile Ala Gln Gly Ser His Thr Trp Glu Asn Gln Leu
35 40 45

Val Asn Gly Leu Trp Thr Tyr Ser Ile Asp Ala Ile Trp Ser Gly Leu
50 55 60

Gln Asp Cys Tyr Ala Asp Leu Arg Ser Asn Val Lys Lys Leu Tyr Asp
65 70 75 80

Thr Glu Ile Glu Thr Leu Ala Ala Ile Gly Val Ser Ala Met Met His
85 90 95

Gly Tyr Met Pro Phe Asn Glu Lys Glu Glu Ile Leu Val Pro Phe Arg
100 105 110

Thr Trp Arg Asn Thr Asn Thr Gly Arg Ala Ala Ala Glu Leu Ser Glu
115 120 125

Leu Phe Val Tyr Asn Ile Pro Leu Arg Trp Ser Ile Ser His Leu Tyr
130 135 140

ES 2 602 808 T3

Gln Ala Ile Leu Asp Asn Glu Ala His Val Lys Asp Ile Lys Phe Leu
 145 150 155 160

Thr Thr Leu Ala Gly Tyr Val His Trp Gln Ile Thr Gly Glu Lys Val
 165 170 175

Leu Gly Ile Gly Asp Ala Ser Gly Met Leu Pro Ile Asp Pro Thr Thr
 180 185 190

Asn Asn Tyr Ser Ala Glu Met Val Ala Lys Phe Asn Asn Leu Ile Ala
 195 200 205

Ser Lys Glu Tyr Ser Trp Lys Leu Glu Asp Ile Leu Pro Lys Val Leu
 210 215 220

Ser Ala Gly Glu Asn Ala Gly Val Leu Thr Pro Glu Gly Cys Lys Lys
 225 230 235 240

Leu Asp Ala Ser Gly His Leu Lys Ala Gly Ile Pro Val Cys Pro Pro
 245 250 255

Glu Gly Asp Ala Gly Thr Gly Met Val Ala Thr Asn Ala Val Lys Gln
 260 265 270

Arg Thr Gly Asn Val Ser Ala Gly Thr Ser Ser Phe Ser Met Ile Val
 275 280 285

Leu Glu Lys Glu Leu Ser Lys Pro Tyr Glu Met Ile Asp Met Val Thr
 290 295 300

Thr Pro Asp Gly Ser Leu Val Ala Met Val His Cys Asn Asn Cys Thr
 305 310 315 320

Ser Asp Leu Asn Ala Trp Val Asn Leu Phe Lys Glu Tyr Gln Glu Leu
 325 330 335

Leu Gly Ile Pro Val Asp Met Asp Glu Leu Tyr Gly Lys Leu Tyr Asn
 340 345 350

Ile Ala Leu Thr Gly Asp Thr Asp Cys Gly Gly Leu Leu Ser Tyr Asn
 355 360 365

Tyr Ile Ser Gly Glu Pro Val Thr Gly Leu Ala Glu Gly Arg Pro Leu
 370 375 380

Phe Val Arg Ser Ala Asn Asp Lys Phe Asn Leu Ala Asn Phe Met Arg
 385 390 395 400

ES 2 602 808 T3

Ala His Leu Tyr Ala Ser Val Gly Val Leu Lys Ile Gly Asn Asp Ile
405 410 415

Leu Phe Asn Glu Glu Lys Ile Lys Val Asp Arg Ile Thr Gly His Gly
420 425 430

Gly Leu Phe Arg Thr Lys Gly Val Gly Gln Arg Val Leu Ala Ala Ala
435 440 445

Ile Asn Ser Pro Ile Ser Val Met Glu Thr Ala Gly Glu Gly Gly Ala
450 455 460

Trp Gly Ile Ala Leu Leu Gly Ser Tyr Leu Val Asn Asn Lys Lys Gly
465 470 475 480

Gln Ser Leu Ala Asp Phe Leu Asp Glu Ser Val Phe Val Ser Asp Ala
485 490 495

Gly Val Glu Val Ser Pro Thr Pro Glu Asp Val Ala Gly Phe Asn Thr
500 505 510

Tyr Ile Glu Ser Tyr Lys Ala Gly Leu Pro Ile Glu Glu Ala Ala Val
515 520 525

Lys Phe Lys
530

<210> 13
<211> 1623
5 <212> DNA
<213> Leuconostoc citreum

<220>
<221> CDS
10 <222> (1)..(1623)

<400> 13
atg aca gca caa caa ata caa gaa aaa gcc att gca acc gga cag gtg 48
Met Thr Ala Gln Gln Ile Gln Glu Lys Ala Ile Ala Thr Gly Gln Val
1 5 10 15

gca tta ggc gtg gaa ttt gga tcg aca aca att aaa gca gta ttg act 96
Ala Leu Gly Val Glu Phe Gly Ser Thr Thr Ile Lys Ala Val Leu Thr
20 25 30

act aac agc gga tta acg att gct tca ggc agt tat gat tgg aca aac 144
Thr Asn Ser Gly Leu Thr Ile Ala Ser Gly Ser Tyr Asp Trp Thr Asn
35 40 45

aat ttt caa aat ggt ctc tgg acg tat agc ctg gat gat gtt tgg cta 192
Asn Phe Gln Asn Gly Leu Trp Thr Tyr Ser Leu Asp Asp Val Trp Leu
50 55 60

ES 2 602 808 T3

ggt tta cag agt gcc tat cga caa ctg aaa gca caa gtt gaa tca aaa Gly Leu Gln Ser Ala Tyr Arg Gln Leu Lys Ala Gln Val Glu Ser Lys 65 70 75 80	240
tat ggc ttg aaa ctt aaa aaa atc aaa aca atg ggt ttt tca gct atg Tyr Gly Leu Lys 85 Lys Lys Ile Lys 90 Thr Met Gly Phe Ser Ala Met 95	288
atg cac ggc tac tta gca ttc gat aat caa gac aca cta cta gtt cca Met His Gly Tyr Leu Ala Phe Asp Asn Gln Asp Thr Leu Leu Val Pro 100 105 110	336
ttt cgt aca tgg cgc aat gcg aca act ggt cgg gca tca cgc gaa ttg Phe Arg Thr Trp Arg Asn Ala Thr Thr Gly Arg Ala Ser Arg Glu Leu 115 120 125	384
acc aag tta ttt ggc ttt aac gtg cca caa cga tgg agc att gca cat Thr Lys Leu Phe Gly Phe Asn Val Pro Gln Arg Trp Ser Ile Ala His 130 135 140	432
ttg tat caa gct atc tta gat cag gag acg cat gtt aaa aat att agt Leu Tyr Gln Ala Ile Leu Asp Gln Glu Thr His Val Lys Asn Ile Ser 145 150 155 160	480
tat ttc aca aca ctt gca gga tat gtt cat tgg caa tta acc ggt gaa Tyr Phe Thr Thr Leu Ala Gly Tyr Val His Trp Gln Leu Thr Gly Glu 165 170 175	528
aaa gtt tta ggt gtt ggc gat gcc tca ggg atg ttt cct atc gat gca Lys Val Leu Gly Val Gly Asp Ala Ser Gly Met Phe Pro Ile Asp Ala 180 185 190	576
gaa act ggt aat tat aat caa aat atg att gat caa ttt tct cat ttg Glu Thr Gly Asn Tyr Asn Gln Asn Met Ile Asp Gln Phe Ser His Leu 195 200 205	624
aaa gcg gtc caa caa tat cag tgg cag ata caa gat ata tta cca gaa Lys Ala Val Gln Gln Tyr Gln Trp Gln Ile Gln Asp Ile Leu Pro Glu 210 215 220	672
ccg cga cat gct ggt gat atg gcg ggt cat tta aca gca gct ggt gct Pro Arg His Ala Gly Asp Met Ala Gly His Leu Thr Ala Ala Gly Ala 225 230 235 240	720
aaa aag tta gac cca aca ggc gat tta atg gca ggt gtc atc gtg gca Lys Lys Leu Asp Pro Thr Gly Asp Leu Met Ala Gly Val Ile Val Ala 245 250 255	768
cca cca gaa ggc gac gct ggg acg ggt atg gta gcg acc aat agt acg Pro Pro Glu Gly Asp Ala Gly Thr Gly Met Val Ala Thr Asn Ser Thr 260 265 270	816
caa gta cga aca ggt aac att tca gtg gga aca tct att ttt tca atg Gln Val Arg Thr Gly Asn Ile Ser Val Gly Thr Ser Ile Phe Ser Met 275 280 285	864
att gtt tta gaa aaa agt tta aag cat gtt tat agt aat atc gat atc Ile Val Leu Glu Lys Ser Leu Lys His Val Tyr Ser Asn Ile Asp Ile 290 295 300	912
gtc act acg cca act ggc tta cct gtt gcc atg gtt cac gcc aat aat Val Thr Thr Pro Thr Gly Leu Pro Val Ala Met Val His Ala Asn Asn 300	960

ES 2 602 808 T3

305	310	315	320	
tct gcc tct gat tta aat gct tgg tcc aaa ctg ttt gcg gaa ttt gca Ser Ala Ser Asp Leu Asn Ala Trp Ser Lys Leu Phe Ala Glu Phe Ala	325	330	335	1008
ggg atg att ggt cag aat tta tca aac gcg gcc ttg tat caa acc ctg Gly Met Ile Gly Gln Asn Leu Ser Asn Ala Ala Leu Tyr Gln Thr Leu	340	345	350	1056
ttc aat gct gct ttg aat gat gct gat gct gat gct ggt ggt tta act Phe Asn Ala Ala Leu Asn Asp Ala Asp Ala Asp Ala Gly Gly Leu Thr	355	360	365	1104
ggg tat ggc tac tat tcc ggt gaa aac att act gca gta ccg gaa gga Gly Tyr Gly Tyr Tyr Ser Gly Glu Asn Ile Thr Ala Val Pro Glu Gly	370	375	380	1152
cga cca tta tta gtc aga caa cca gac tca cat ttt acg att ggc aat Arg Pro Leu Leu Val Arg Gln Pro Asp Ser His Phe Thr Ile Gly Asn	385	390	400	1200
ctt atg cgt tta cat atc ttt agt gca ttc ggt gcg att aaa att ggc Leu Met Arg Leu His Ile Phe Ser Ala Phe Gly Ala Ile Lys Ile Gly	405	410	415	1248
atg cga att tta gca gat gaa aac gta cta acc gat aat att gtg gct Met Arg Ile Leu Ala Asp Glu Asn Val Leu Thr Asp Asn Ile Val Ala	420	425	430	1296
caa ggc ggt gtg ttt aaa aca cca att gtg gct caa aaa ttg tta gca Gln Gly Gly Val Phe Lys Thr Pro Ile Val Ala Gln Lys Leu Leu Ala	435	440	445	1344
gca gca ctc aac aca aac att act gtg atg gcc aat gct ggt gaa ggt Ala Ala Leu Asn Thr Asn Ile Thr Val Met Ala Asn Ala Gly Glu Gly	450	455	460	1392
ggg ccg tgg gga atg gct att ttg gca ctt tat gca gcc aat aaa tta Gly Pro Trp Gly Met Ala Ile Leu Ala Leu Tyr Ala Ala Asn Lys Leu	465	470	475	1440
ggc ggt cag aca cta gat gat tat tta gca aaa aac ata ttc gct gaa Gly Gly Gln Thr Leu Asp Asp Tyr Leu Ala Lys Asn Ile Phe Ala Glu	485	490	495	1488
act aaa gca caa aca ctc gcg cca gaa cca cgt gat gtt gca gga ttt Thr Lys Ala Gln Thr Leu Ala Pro Glu Pro Arg Asp Val Ala Gly Phe	500	505	510	1536
gaa gaa ttt atg aca cgt tat att gat ggc cta caa att gag tta acc Glu Glu Phe Met Thr Arg Tyr Ile Asp Gly Leu Gln Ile Glu Leu Thr	515	520	525	1584
gcc att aaa gca ttg cct agt aat caa ata aag gag taa Ala Ile Lys Ala Leu Pro Ser Asn Gln Ile Lys Glu	530	535	540	1623

<210> 14

<211> 540

5 <212> PRT

<213> Leuconostoc citreum

<400> 14

ES 2 602 808 T3

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

ES 2 602 808 T3

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

Thr Lys Ala Gln Thr Leu Ala Pro Glu Pro Arg Asp Val Ala Gly Phe
500 505 510

Glu Glu Phe Met Thr Arg Tyr Ile Asp Gly Leu Gln Ile Glu Leu Thr
515 520 525

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

<210> 15
<211> 684
5 <212> DNA
<213> Secuencia Artificial

<220>
10 <223> Codón optimizado de B. thetaiotaomicron araD

<220>
<221> CDS
<222> (1)..(684)

15 <400> 15
atg tta gaa gaa tta aag gaa aag gtt ttc cac gct aat ttg gaa tta 48
Met Leu Glu Glu Leu Lys Glu Lys Val Phe His Ala Asn Leu Glu Leu
1 5 10 15
gtt aag cat ggt tta gtt att ttc acc tgg ggt aac gtt tct gct att 96
Val Lys His Gly Leu Val Ile Phe Thr Trp Gly Asn Val Ser Ala Ile
20 25 30
gat aga gaa acc gaa tta gtt gtt att aag cca tct ggt gtt tct tac 144
Asp Arg Glu Thr Glu Leu Val Ile Lys Pro Ser Gly Val Ser Tyr
35 40 45
gac gat atg aag gca gaa gat atg gtt gtt gtt gat tta gat ggt aag 192
Asp Asp Met Lys Ala Glu Asp Met Val Val Val Asp Leu Asp Gly Lys
50 55 60
gtc gtc gaa ggt aga tta aag cca tct tct gac acc cca acc cac gtt 240
Val Val Glu Gly Arg Leu Lys Pro Ser Ser Asp Thr Pro Thr His Val
65 70 75 80
gtt tta tac aag gct ttt cca gaa att ggt ggt gtt gtt cac act cac 288
Val Leu Tyr Lys Ala Phe Pro Glu Ile Gly Gly Val Val His Thr His
85 90 95
tct acc tac gct act gca tgg gct caa gct ggt tgt gat atc cca aat 336
Ser Thr Tyr Ala Thr Ala Trp Ala Gln Ala Gly Cys Asp Ile Pro Asn
100 105 110
att ggt act act cac gca gat tac ttc cac gat gca att cca tgt act 384
Ile Gly Thr Thr His Ala Asp Tyr Phe His Asp Ala Ile Pro Cys Thr
115 120 125
gca gat atg act gaa gct gaa gtt aag ggt gct tat gaa tta gaa acc 432
Ala Asp Met Thr Glu Ala Glu Val Lys Gly Ala Tyr Glu Leu Glu Thr
130 135 140

ES 2 602 808 T3

```

ggt aat gtt atc gtc aag aga ttt gaa ggt ttg aac cca gtc cac acc      480
Gly Asn Val Ile Val Lys Arg Phe Glu Gly Leu Asn Pro Val His Thr
145                      150                      155                      160

cca ggt gtc tta gtt aag aat cat ggt cca ttc tcc tgg ggt aag gat      528
Pro Gly Val Leu Val Lys Asn His Gly Pro Phe Ser Trp Gly Lys Asp
                      165                      170                      175

gct cac gat gca gtt cac aac gca gtt gtc atg gaa caa gtt gca aag      576
Ala His Asp Ala Val His Asn Ala Val Val Met Glu Gln Val Ala Lys
180                      185                      190

atg gct tct att gct tac gct gtt aat cca aac tta act atg aat cca      624
Met Ala Ser Ile Ala Tyr Ala Val Asn Pro Asn Leu Thr Met Asn Pro
195                      200                      205

tta tta gtt gaa aag cac ttc tcc aga aag cac ggt cca aac gct tat      672
Leu Leu Val Glu Lys His Phe Ser Arg Lys His Gly Pro Asn Ala Tyr
210                      215                      220

tac ggt caa taa
Tyr Gly Gln
225

```

<210> 16
 <211> 227
 <212> PRT
 <213> Secuencia Artificial

<220>
 <223> Construcción Sintética

<400> 16

```

Met Leu Glu Glu Leu Lys Glu Lys Val Phe His Ala Asn Leu Glu Leu
1      5      10      15

Val Lys His Gly Leu Val Ile Phe Thr Trp Gly Asn Val Ser Ala Ile
20      25      30

Asp Arg Glu Thr Glu Leu Val Val Ile Lys Pro Ser Gly Val Ser Tyr
35      40      45

Asp Asp Met Lys Ala Glu Asp Met Val Val Val Asp Leu Asp Gly Lys
50      55      60

Val Val Glu Gly Arg Leu Lys Pro Ser Ser Asp Thr Pro Thr His Val
65      70      75      80

Val Leu Tyr Lys Ala Phe Pro Glu Ile Gly Gly Val Val His Thr His
85      90      95

Ser Thr Tyr Ala Thr Ala Trp Ala Gln Ala Gly Cys Asp Ile Pro Asn
100     105     110

```

ES 2 602 808 T3

Ile Gly Thr Thr His Ala Asp Tyr Phe His Asp Ala Ile Pro Cys Thr
115 120 125

Ala Asp Met Thr Glu Ala Glu Val Lys Gly Ala Tyr Glu Leu Glu Thr
130 135 140

Gly Asn Val Ile Val Lys Arg Phe Glu Gly Leu Asn Pro Val His Thr
145 150 155 160

Pro Gly Val Leu Val Lys Asn His Gly Pro Phe Ser Trp Gly Lys Asp
165 170 175

Ala His Asp Ala Val His Asn Ala Val Val Met Glu Gln Val Ala Lys
180 185 190

Met Ala Ser Ile Ala Tyr Ala Val Asn Pro Asn Leu Thr Met Asn Pro
195 200 205

Leu Leu Val Glu Lys His Phe Ser Arg Lys His Gly Pro Asn Ala Tyr
210 215 220

Tyr Gly Gln
225

<210> 17
<211> 693
5 <212> DNA
<213> Secuencia Artificial

<220>
10 <223> Codóon optimizado de B. animalis araD

<220>
<221> CDS
<222> (1)..(693)

15 <400> 17
atg gct aca ttg gca gac tat ggt cca gaa gtt aga gct gaa gtt aag 48
Met Ala Thr Leu Ala Asp Tyr Gly Pro Glu Val Arg Ala Glu Val Lys
1 5 10 15

cag gtc aga gaa gtt gtt gca gct tta cat gaa caa ttg att aag tgg 96
Gln Val Arg Glu Val Val Ala Ala Leu His Glu Gln Leu Ile Lys Trp
20 25 30

aac tta gtt gtt tgg acc gca ggc aat gtc tct caa aga ttg aga act 144
Asn Leu Val Val Trp Thr Ala Gly Asn Val Ser Gln Arg Leu Arg Thr
35 40 45

gca gac ctt atg gtt atc aag cct tca ggc ctt aga tac gaa tac tta 192
Ala Asp Leu Met Val Ile Lys Pro Ser Gly Leu Arg Tyr Glu Tyr Leu
50 55 60

aca cca tcc tca atg gtt gtc tgt gac ttg gat ggt aac gtt gtt gat 240

Thr	Pro	Ser	Ser	Met	Val	Val	Cys	Asp	Leu	Asp	Gly	Asn	Val	Val	Asp	
65					70					75					80	
ggc	gca	gaa	tcc	cct	tcc	tcc	gat	act	gca	tct	cat	gca	tac	atc	tat	288
Gly	Ala	Glu	Ser	Pro	Ser	Ser	Asp	Thr	Ala	Ser	His	Ala	Tyr	Ile	Tyr	
			85						90					95		
aga	cat	atg	cca	gaa	gtt	tac	ggt	gtt	gtc	cac	act	cat	tct	acc	tat	336
Arg	His	Met	Pro	Glu	Val	Tyr	Gly	Val	Val	His	Thr	His	Ser	Thr	Tyr	
			100					105					110			
gct	act	gct	tgg	gca	gcc	acc	ggt	cag	aat	atc	cca	tgt	ggt	tta	acc	384
Ala	Thr	Ala	Trp	Ala	Ala	Thr	Gly	Gln	Asn	Ile	Pro	Cys	Gly	Leu	Thr	
		115					120					125				
atg	atg	ggt	gat	gag	ttc	ggt	ggt	cca	gtt	cca	gtc	ggt	cca	ttt	cgt	432
Met	Met	Gly	Asp	Glu	Phe	Gly	Gly	Pro	Val	Pro	Val	Gly	Pro	Phe	Arg	
	130					135					140					
ttg	att	ggt	tct	gaa	gct	att	ggt	gaa	ggt	gtt	gtc	gag	aca	ttg	aaa	480
Leu	Ile	Gly	Ser	Glu	Ala	Ile	Gly	Glu	Gly	Val	Val	Glu	Thr	Leu	Lys	
145					150					155					160	
gca	tac	cca	aag	tct	cct	gcc	gtc	tta	atg	caa	aat	cac	ggt	cca	ttc	528
Ala	Tyr	Pro	Lys	Ser	Pro	Ala	Val	Leu	Met	Gln	Asn	His	Gly	Pro	Phe	
			165						170					175		
acc	att	ggt	aaa	gat	gcc	gaa	gca	gcc	gtt	aag	gct	gca	gct	atg	act	576
Thr	Ile	Gly	Lys	Asp	Ala	Glu	Ala	Ala	Val	Lys	Ala	Ala	Ala	Met	Thr	
			180					185					190			
gag	gag	gtc	gct	cac	act	atg	tgg	gct	gcc	aaa	caa	ttg	ggt	gat	atc	624
Glu	Glu	Val	Ala	His	Thr	Met	Trp	Ala	Ala	Lys	Gln	Leu	Gly	Asp	Ile	
		195					200					205				
att	cca	att	cca	caa	gag	gat	att	gac	aag	tta	aac	gat	aga	tat	caa	672
Ile	Pro	Ile	Pro	Gln	Glu	Asp	Ile	Asp	Lys	Leu	Asn	Asp	Arg	Tyr	Gln	
	210					215					220					
aac	gtt	tat	ggt	caa	cac	taa										693
Asn	Val	Tyr	Gly	Gln	His											
225					230											

<210> 18

<211> 230

5 <212> PRT

<213> Secuencia Artificial

<220>

10 <223> Construcción Sintética

<400> 18

Met	Ala	Thr	Leu	Ala	Asp	Tyr	Gly	Pro	Glu	Val	Arg	Ala	Glu	Val	Lys
1				5					10					15	

Gln	Val	Arg	Glu	Val	Val	Ala	Ala	Leu	His	Glu	Gln	Leu	Ile	Lys	Trp
		20						25					30		

Asn	Leu	Val	Val	Trp	Thr	Ala	Gly	Asn	Val	Ser	Gln	Arg	Leu	Arg	Thr
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

ES 2 602 808 T3

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

<210> 19

<211> 702

5 <212> DNA

<213> Secuencia Artificial

<220>

<223> Codón optimizado de L. lactis araD

10

<220>

<221> CDS

<222> (1)..(702)

15

<400> 19

ES 2 602 808 T3

atg tta gag gat ttg aaa gag aag gtt ttg aag gca aat ctt gaa ttg	48
Met Leu Glu Asp Leu Lys Glu Lys Val Leu Lys Ala Asn Leu Glu Leu	
1 5 10 15	
cca gaa cat cac tta gtc gag ttc act tgg ggt aat gca tca gca ttt	96
Pro Glu His His Leu Val Glu Phe Thr Trp Gly Asn Ala Ser Ala Phe	
20 25 30	
gat aag gaa acc ggt tac ttt gtc att aag cct tct ggt att gac tat	144
Asp Lys Glu Thr Gly Tyr Phe Val Ile Lys Pro Ser Gly Ile Asp Tyr	
35 40 45	
ggt tcc tta aag gct tcc gat atg gtt gtc gtt gat ttg gaa ggt aaa	192
Gly Ser Leu Lys Ala Ser Asp Met Val Val Val Asp Leu Glu Gly Lys	
50 55 60	
ggt att gaa ggc gag tta aac cca tcc tct gac act cca act cat gct	240
Val Ile Glu Gly Glu Leu Asn Pro Ser Ser Asp Thr Pro Thr His Ala	
65 70 75 80	
ggt ttg tac aaa aag cat cca gaa ttg ggt ggt att gtt cac acc cac	288
Val Leu Tyr Lys Lys His Pro Glu Leu Gly Gly Ile Val His Thr His	
85 90 95	
tcc aat tgg gca acc gcc tgg gca gaa tca ggt gtt gat gtt tct gcc	336
Ser Asn Trp Ala Thr Ala Trp Ala Glu Ser Gly Val Asp Val Ser Ala	
100 105 110	
atg ggc acc act cat gca gac aca ttc tat ggt cca gtt cca tgt act	384
Met Gly Thr Thr His Ala Asp Thr Phe Tyr Gly Pro Val Pro Cys Thr	
115 120 125	
cgt tac tta aca aag gaa gag att gat aag ggt tat gaa tat gaa aca	432
Arg Tyr Leu Thr Lys Glu Glu Ile Asp Lys Gly Tyr Glu Tyr Glu Thr	
130 135 140	
ggc aag ctt atc att gaa acc ttc gag gaa aga ggt att gat atc tta	480
Gly Lys Leu Ile Ile Glu Thr Phe Glu Glu Arg Gly Ile Asp Ile Leu	
145 150 155 160	
gac att cca gct gtc tta ttg aga ggt cac ggt cca ttc act tgg ggt	528
Asp Ile Pro Ala Val Leu Leu Arg Gly His Gly Pro Phe Thr Trp Gly	
165 170 175	
gaa aat gtc gaa gct gcc gtc tac aac gct gtt gtt tta gaa aac gtt	576
Glu Asn Val Glu Ala Val Tyr Asn Ala Val Val Leu Glu Asn Val	
180 185 190	
tgt aag atg aac atc ttt gct aga caa atc aac tct tat gct gct gat	624
Cys Lys Met Asn Ile Phe Ala Arg Gln Ile Asn Ser Tyr Ala Ala Asp	
195 200 205	
ttg cct cag aga atc ttg gat aag cac tat ctt aga aag cat ggt aaa	672
Leu Pro Gln Arg Ile Leu Asp Lys His Tyr Leu Arg Lys His Gly Lys	
210 215 220	
gac gca tac tac ggt cag aaa aac aag taa	702
Asp Ala Tyr Tyr Gly Gln Lys Asn Lys	
225 230	

<210> 20
 <211> 233
 <212> PRT
 <213> Secuencia Artificial

<220>
 <223> Construcción Sintética

<400> 20

ES 2 602 808 T3

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

- 5 <210> 21
 <211> 1317
 <212> DNA
 <213> Secuencia Artificial

- 10 <220>
 <223> Codón optimizado de B. thetaiotaomicron XI

<220>

<221> CDS

<222> (1)..(1317)

5

<400> 21

atg gct act aaa gaa ttt ttc cca ggt att gaa aag att aaa ttt gaa	48
Met Ala Thr Lys Glu Phe Phe Pro Gly Ile Glu Lys Ile Lys Phe Glu	
1 5 10 15	
ggt aaa gac tcc aaa aat cca atg gct ttt aga tat tac gac gca gaa	96
Gly Lys Asp Ser Lys Asn Pro Met Ala Phe Arg Tyr Tyr Asp Ala Glu	
20 25 30	
aag gtt att aac ggt aaa aaa atg aag gac tgg ttg aga ttt gct atg	144
Lys Val Ile Asn Gly Lys Lys Met Lys Asp Trp Leu Arg Phe Ala Met	
35 40 45	
gca tgg tgg cac acc cta tgt gct gaa ggt ggt gat caa ttc ggt ggt	192
Ala Trp Trp His Thr Leu Cys Ala Glu Gly Gly Asp Gln Phe Gly Gly	
50 55 60	
ggt aca aaa caa ttt cct tgg aac ggt aac gct gac gct att caa gct	240
Gly Thr Lys Gln Phe Pro Trp Asn Gly Asn Ala Asp Ala Ile Gln Ala	
65 70 75 80	
gca aag gac aag atg gac gcc ggt ttc gaa ttt atg caa aag atg ggt	288
Ala Lys Asp Lys Met Asp Ala Gly Phe Glu Phe Met Gln Lys Met Gly	
85 90 95	
att gaa tac tac tgt ttt cac gat gtt gat ttg gtc tct gaa ggt gca	336
Ile Glu Tyr Tyr Cys Phe His Asp Val Asp Leu Val Ser Glu Gly Ala	
100 105 110	
tct gtt gaa gaa tat gaa gct aac ttg aag gaa atc gtt gct tac gct	384
Ser Val Glu Glu Tyr Glu Ala Asn Leu Lys Glu Ile Val Ala Tyr Ala	
115 120 125	
aag caa aag caa gct gag aca ggt att aaa ttg ttg tgg ggt aca gca	432
Lys Gln Lys Gln Ala Glu Thr Gly Ile Lys Leu Leu Trp Gly Thr Ala	
130 135 140	
aac gtt ttc ggt cac gcc aga tat atg aat ggt gct gca act aac cca	480
Asn Val Phe Gly His Ala Arg Tyr Met Asn Gly Ala Ala Thr Asn Pro	
145 150 155 160	
gat ttc gat gtt gtt gca aga gcc gca gtt caa atc aaa aac gct att	528
Asp Phe Asp Val Val Ala Arg Ala Ala Val Gln Ile Lys Asn Ala Ile	

ES 2 602 808 T3

165	170	175	
gac gct aca atc gaa cta ggt ggt gaa aac tat gtt ttt tgg ggt ggt Asp Ala Thr Ile Glu Leu Gly Gly Glu Asn Tyr Val Phe Trp Gly Gly 180 185 190			576
aga gaa ggt tac atg tct ttg ttg aac act gac caa aaa aga gaa aag Arg Glu Gly Tyr Met Ser Leu Leu Asn Thr Asp Gln Lys Arg Glu Lys 195 200 205			624
gaa cac ttg gcc caa atg ttg act atc gct aga gat tac gct aga gct Glu His Leu Ala Gln Met Leu Thr Ile Ala Arg Asp Tyr Ala Arg Ala 210 215 220			672
aga ggt ttt aaa ggt act ttt ttg att gaa cca aaa cca atg gaa cca Arg Gly Phe Lys Gly Thr Phe Leu Ile Glu Pro Lys Pro Met Glu Pro 225 230 235 240			720
act aag cac caa tat gac gtt gat act gaa act gtt att ggt ttc ttg Thr Lys His Gln Tyr Asp Val Asp Thr Glu Thr Val Ile Gly Phe Leu 245 250 255			768
aag gca cac ggt ttg gat aag gat ttt aag gtg aat atc gaa gtt aac Lys Ala His Gly Leu Asp Lys Asp Phe Lys Val Asn Ile Glu Val Asn 260 265 270			816
cac gct act ttg gcc ggt cat act ttc gaa cat gaa ttg gct gtt gct His Ala Thr Leu Ala Gly His Thr Phe Glu His Glu Leu Ala Val Ala 275 280 285			864
gtg gat aac ggt atg ttg ggt tct att gac gca aac aga ggt gac tat Val Asp Asn Gly Met Leu Gly Ser Ile Asp Ala Asn Arg Gly Asp Tyr 290 295 300			912
caa aat ggt tgg gac act gat caa ttt cca att gac aac tat gaa ttg Gln Asn Gly Trp Asp Thr Asp Gln Phe Pro Ile Asp Asn Tyr Glu Leu 305 310 315 320			960
aca caa gca atg atg caa att att aga aac ggt ggt ttg ggt act ggt Thr Gln Ala Met Met Gln Ile Ile Arg Asn Gly Gly Leu Gly Thr Gly 325 330 335			1008
ggt act aac ttc gac gct aag act aga aga aat tcc aca gat ttg gaa Gly Thr Asn Phe Asp Ala Lys Thr Arg Arg Asn Ser Thr Asp Leu Glu 340 345 350			1056
gac att ttt atc gct cac atc gct ggt atg gac gct atg gcc aga gct Asp Ile Phe Ile Ala His Ile Ala Gly Met Asp Ala Met Ala Arg Ala 355 360 365			1104
ttg gaa tcc gct gct gct ttg ttg gac gaa tcc cca tac aag aaa atg Leu Glu Ser Ala Ala Ala Leu Leu Asp Glu Ser Pro Tyr Lys Lys Met 370 375 380			1152
ttg gcc gac aga tac gct tct ttc gac ggt ggt aag ggt aag gaa ttt Leu Ala Asp Arg Tyr Ala Ser Phe Asp Gly Gly Lys Gly Lys Glu Phe 385 390 395 400			1200
gag gac ggt aag ttg aca ttg gaa gat gtt gtt gct tac gct aag act Glu Asp Gly Lys Leu Thr Leu Glu Asp Val Val Ala Tyr Ala Lys Thr 405 410 415			1248
aag ggt gaa cca aag caa aca tcc ggt aag caa gaa ttg tac gaa gct			1296
Lys Gly Glu Pro Lys Gln Thr Ser Gly Lys Gln Glu Leu Tyr Glu Ala 420 425 430			
att cta aat atg tac tgt taa Ile Leu Asn Met Tyr Cys 435			1317

5 <210> 22
<211> 438
<212> PRT

<213> Secuencia Artificial

<220>

<223> Construcción Sintética

5

<400> 22

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
 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

ES 2 602 808 T3

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
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
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		
370	375	380
Leu Ala Asp Arg Tyr Ala Ser Phe Asp Gly Gly Lys Gly Lys Glu Phe		
385	390	395
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		

5 <210> 23
 <211> 1782
 <212> DNA
 <213> Issatchenkia orientalis

10 <220>

<221> CDS

<222> (1)..(1782)

<400> 23

atg tct ttg gct cta ggt ttt gac ctc tca aca caa cag ctg aaa atc	48
Met Ser Leu Ala Leu Gly Phe Asp Leu Ser Thr Gln Gln Leu Lys Ile	
1 5 10 15	
gtc tcc tgt tat cag gat ctt agt ctt cat tca aaa tac tct att gat	96
Val Ser Cys Tyr Gln Asp Leu Ser Leu His Ser Lys Tyr Ser Ile Asp	
20 25 30	
ttc gac gaa ttc aag gac att tac ggt atc cat aaa ggc gta ttg tcg	144
Phe Asp Glu Phe Lys Asp Ile Tyr Gly Ile His Lys Gly Val Leu Ser	
35 40 45	
aat aga gat aca ggt gaa gtc gtt act cct gtc aag ttg ttt gta cat	192
Asn Arg Asp Thr Gly Glu Val Val Thr Pro Val Lys Leu Phe Val His	
50 55 60	
gcc ctc cag acc ctc ctg gac cgc atg cac aat gat ggg ttc ccc ttt	240
Ala Leu Gln Thr Leu Leu Asp Arg Met His Asn Asp Gly Phe Pro Phe	
65 70 75 80	
gat tgc gtg aca tca att tca gga tcg tgc caa caa cat gga acg att	288
Asp Cys Val Thr Ser Ile Ser Gly Ser Cys Gln Gln His Gly Thr Ile	
85 90 95	
ttc tgt aca cgt caa ttc gat aca ctg ctc tcg aat ttg aat ccg gct	336
Phe Cys Thr Arg Gln Phe Asp Thr Leu Ser Asn Leu Asn Pro Ala	
100 105 110	
tct gat act tgg cac agc gat ttg tcc aat gcc ttc tcc tac gag aat	384
Ser Asp Thr Trp His Ser Asp Leu Ser Asn Ala Phe Ser Tyr Glu Asn	
115 120 125	
gcc tcc aat tgg caa gac aga tca acg ggc gaa gaa ttg gcg gtg ttt	432
Ala Ser Asn Trp Gln Asp Arg Ser Thr Gly Glu Glu Leu Ala Val Phe	
130 135 140	
gaa aaa gca ttg gga tca gca gag aaa ctc tgt aaa atc act ggt tca	480
Glu Lys Ala Leu Gly Ser Ala Glu Lys Leu Cys Lys Ile Thr Gly Ser	
145 150 155 160	
aag gcg cat ttc agg ttc tct ggt cct caa atg aga agg agg gcc aag	528
Lys Ala His Phe Arg Phe Ser Gly Pro Gln Met Arg Arg Arg Ala Lys	
165 170 175	
gag ggt ggt gtc cat tgg gag gag acg gcc cac ata tcc ctc ata tcc	576
Glu Gly Gly Val His Trp Glu Glu Thr Ala His Ile Ser Leu Ile Ser	
180 185 190	

aat ttt ctc gat tcc atc ttg tcc ggt aag gtt aga ggg gtg gaa att Asn Phe Leu Asp Ser Ile Leu Ser Gly Lys Val Arg Gly Val Glu Ile 195 200 205	624
gga gaa gct tgt ggt aca aac ctc ttt gat att gag cag aac gac tgg Gly Glu Ala Cys Gly Thr Asn Leu Phe Asp Ile Glu Gln Asn Asp Trp 210 215 220	672
aac gat gag ttg ctt tcc ttg atc ttg atg aag aat tcc aat gtt gac Asn Asp Glu Leu Leu Ser Leu Ile Leu Met Lys Asn Ser Asn Val Asp 225 230 235 240	720
gga gtt cct ttg ggt gaa cag caa gag gct tct ttg aaa gcc cgt caa Gly Val Pro Leu Gly Glu Gln Gln Glu Ala Ser Leu Lys Ala Arg Gln 245 250 255	768
ctt cta aaa acc tta gtt gag cct gat gat tat tca aca att gcg cct Leu Leu Lys Thr Leu Val Glu Pro Asp Asp Tyr Ser Thr Ile Ala Pro 260 265 270	816
tac ttg gcc aaa agg tat ggc ttt aaa agg gac tgt aag gtc tgg ccc Tyr Leu Ala Lys Arg Tyr Gly Phe Lys Arg Asp Cys Lys Val Trp Pro 275 280 285	864
att act ggc gat aat ttg gca acc atc atg tcc ttg cca ttg aaa cat Ile Thr Gly Asp Asn Leu Ala Thr Ile Met Ser Leu Pro Leu Lys His 290 295 300	912
gac gat ttg ttg gtg tct atg ggg acc agt aca acg gtg ttg ttg ttg Asp Asp Leu Leu Val Ser Met Gly Thr Ser Thr Thr Val Leu Leu Leu 305 310 315 320	960
acg aaa aac tac ctt cca agt gtg aac tat cac ctc ttt aag cat cct Thr Lys Asn Tyr Leu Pro Ser Val Asn Tyr His Leu Phe Lys His Pro 325 330 335	1008
gtt gta agg gat atc tat atg ggt atg ttg tgc tat tca aat ggt gct Val Val Arg Asp Ile Tyr Met Gly Met Leu Cys Tyr Ser Asn Gly Ala 340 345 350	1056
ctg gca cgt gag gaa att agg gat gaa att aac gac aag tat aaa acg Leu Ala Arg Glu Glu Ile Arg Asp Glu Ile Asn Asp Lys Tyr Lys Thr 355 360 365	1104
gta aag tgg gat aaa ttc aac gag att tta gac act aga aag tct ccc Val Lys Trp Asp Lys Phe Asn Glu Ile Leu Asp Thr Arg Lys Ser Pro 370 375 380	1152
gac aga gag gtt gga atc tat ttc ccc cta ggc gaa atc att ccc aac Asp Arg Glu Val Gly Ile Tyr Phe Pro Leu Gly Glu Ile Ile Pro Asn 385 390 395 400	1200
gtc aag ccc tgt aag cgt atc ttc aag tat tcg gca gcg aag ggg ctt Val Lys Pro Cys Lys Arg Ile Phe Lys Tyr Ser Ala Ala Lys Gly Leu 405 410 415	1248
gtg gaa gtg gac aga gaa gtc gag ctg gac gac caa gtg aag ctt atc Val Glu Val Asp Arg Glu Val Glu Leu Asp Asp Gln Val Lys Leu Ile 420 425 430	1296
att gag tcg cag gcg tta tcc aat cga ctc cgt gta gca cca ctt cta Ile Glu Ser Gln Ala Leu Ser Asn Arg Leu Arg Val Ala Pro Leu Leu 435 440 445	1344

ES 2 602 808 T3

```

acc gat gtt gaa acc gtg aag gag aag tcg gtg acc aga gac att gag      1392
Thr Asp Val Glu Thr Val Lys Glu Lys Ser Val Thr Arg Asp Ile Glu
    450                      455                      460

agt gca agg aag att gtt ggt gac tcg gtt aca att gac cat gtc gct      1440
Ser Ala Arg Lys Ile Val Gly Asp Ser Val Thr Ile Asp His Val Ala
465                      470                      475                      480

tac acg ttt gcc gat att atc aag cgt ccc aat agt gta tac tat gct      1488
Tyr Thr Phe Ala Asp Ile Ile Lys Arg Pro Asn Ser Val Tyr Tyr Ala
                      485                      490                      495

gga ggt tct tca cag aat gca tcg att ctc aag att tac aat gac att      1536
Gly Gly Ser Ser Gln Asn Ala Ser Ile Leu Lys Ile Tyr Asn Asp Ile
                      500                      505                      510

cta gga cct aaa cat ggt ggc tac aag gtt gaa gtc ggt gat gcc tgt      1584
Leu Gly Pro Lys His Gly Gly Tyr Lys Val Glu Val Gly Asp Ala Cys
                      515                      520                      525

gcg cta ggc ggt tgt ttc cga gca atc tat gga tac aac gac agc ata      1632
Ala Leu Gly Gly Cys Phe Arg Ala Ile Tyr Gly Tyr Asn Asp Ser Ile
                      530                      535                      540

tca ttt cag gat tgg ttg gag agc aag ttt gat ttc cac aga cat acc      1680
Ser Phe Gln Asp Trp Leu Glu Ser Lys Phe Asp Phe His Arg His Thr
545                      550                      555                      560

tct ccc att gag agg gac gaa acc cat gcc att tcc acg tgg gca agt      1728
Ser Pro Ile Glu Arg Asp Glu Thr His Ala Ile Ser Thr Trp Ala Ser
                      565                      570                      575

tat ctc gac aag gtt gcc ata ttg acc ttg gca gag cag caa tta gat      1776
Tyr Leu Asp Lys Val Ala Ile Leu Thr Leu Ala Glu Gln Gln Leu Asp
                      580                      585                      590

tgt tga
Cys
                                                    1782

```

<210> 24
 <211> 593
 <212> PRT
 <213> Issatchenkia orientalis

<400> 24

```

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

Val Ser Cys Tyr Gln Asp Leu Ser Leu His Ser Lys Tyr Ser Ile Asp
          20          25          30

Phe Asp Glu Phe Lys Asp Ile Tyr Gly Ile His Lys Gly Val Leu Ser
          35          40          45

Asn Arg Asp Thr Gly Glu Val Val Thr Pro Val Lys Leu Phe Val His
          50          55          60

```

ES 2 602 808 T3

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

ES 2 602 808 T3

305		310		315		320
Thr Lys Asn Tyr	Leu Pro Ser Val	Asn Tyr His Leu Phe	Lys His Pro			
	325		330		335	
Val Val Arg Asp	Ile Tyr Met Gly	Met Leu Cys Tyr	Ser Asn Gly Ala			
	340		345		350	
Leu Ala Arg Glu	Glu Ile Arg Asp	Glu Ile Asn Asp	Lys Tyr Lys Thr			
	355		360		365	
Val Lys Trp Asp	Lys Phe Asn Glu	Ile Leu Asp Thr	Arg Lys Ser Pro			
	370		375		380	
Asp Arg Glu Val	Gly Ile Tyr Phe	Pro Leu Gly Glu	Ile Ile Pro Asn			
	385		390		395	
Val Lys Pro Cys	Lys Arg Ile Phe	Lys Tyr Ser Ala	Ala Lys Gly Leu			
	405		410		415	
Val Glu Val Asp	Arg Glu Val Glu	Leu Asp Asp Gln	Val Lys Leu Ile			
	420		425		430	
Ile Glu Ser Gln	Ala Leu Ser Asn	Arg Leu Arg Val	Ala Pro Leu Leu			
	435		440		445	
Thr Asp Val Glu	Thr Val Lys Glu	Lys Ser Val Thr	Arg Asp Ile Glu			
	450		455		460	
Ser Ala Arg Lys	Ile Val Gly Asp	Ser Val Thr Ile	Asp His Val Ala			
	465		470		475	
Tyr Thr Phe Ala	Asp Ile Ile Lys	Arg Pro Asn Ser	Val Tyr Tyr Ala			
	485		490		495	
Gly Gly Ser Ser	Gln Asn Ala Ser	Ile Leu Lys Ile	Tyr Asn Asp Ile			
	500		505		510	
Leu Gly Pro Lys	His Gly Gly Tyr	Lys Val Glu Val	Gly Asp Ala Cys			
	515		520		525	
Ala Leu Gly Gly	Cys Phe Arg Ala	Ile Tyr Gly Tyr	Asn Asp Ser Ile			
	530		535		540	
Ser Phe Gln Asp	Trp Leu Glu Ser	Lys Phe Asp Phe	His Arg His Thr			
	545		550		555	
Ser Pro Ile Glu	Arg Asp Glu Thr	His Ala Ile Ser	Thr Trp Ala Ser			
	565		570		575	
Tyr Leu Asp Lys	Val Ala Ile Leu	Thr Leu Ala Glu	Gln Gln Leu Asp			
	580		585		590	

Cys

ES 2 602 808 T3

<210> 25
 <211> 1803
 <212> DNA
 5 <213> *Saccharomyces cerevisiae*

<220>
 <221> CDS
 <222> (1)..(1803)

10 <400> 25
 atg ttg tgt tca gta att cag aga cag aca aga gag gtt tcc aac aca 48
 Met Leu Cys Ser Val Ile Gln Arg Gln Thr Arg Glu Val Ser Asn Thr
 1 5 10 15
 atg tct tta gac tca tac tat ctt ggg ttt gat ctt tcg acc caa caa 96
 Met Ser Leu Asp Ser Tyr Tyr Leu Gly Phe Asp Leu Ser Thr Gln Gln
 20 25 30
 ctg aaa tgt ctc gcc att aac cag gac cta aaa att gtc cat tca gaa 144
 Leu Lys Cys Leu Ala Ile Asn Gln Asp Leu Lys Ile Val His Ser Glu
 35 40 45
 aca gtg gaa ttt gaa aag gat ctt ccg cat tat cac aca aag aag ggt 192
 Thr Val Glu Phe Glu Lys Asp Leu Pro His Tyr His Thr Lys Lys Gly
 50 55 60
 gtc tat ata cac ggc gac act atc gaa tgt ccc gta gcc atg tgg tta 240
 Val Tyr Ile His Gly Asp Thr Ile Glu Cys Pro Val Ala Met Trp Leu
 65 70 75 80
 gag gct cta gat ctg gtt ctc tcg aaa tat cgc gag gct aaa ttt cca 288
 Glu Ala Leu Asp Leu Val Leu Ser Lys Tyr Arg Glu Ala Lys Phe Pro
 85 90 95
 ttg aac aaa gtt atg gcc gtc tca ggg tcc tgc cag cag cac ggg tct 336
 Leu Asn Lys Val Met Ala Val Ser Gly Ser Cys Gln Gln His Gly Ser
 100 105 110
 gtc tac tgg tcc tcc caa gcc gaa tct ctg tta gag caa ttg aat aag 384
 Val Tyr Trp Ser Ser Gln Ala Glu Ser Leu Leu Glu Gln Leu Asn Lys
 115 120 125
 aaa ccg gaa aaa gat tta ttg cac tac gtg agc tct gta gca ttt gca 432
 Lys Pro Glu Lys Asp Leu Leu His Tyr Val Ser Ser Val Ala Phe Ala
 130 135 140
 agg caa acc gcc ccc aat tgg caa gac cac agt act gca aag caa tgt 480
 Arg Gln Thr Ala Pro Asn Trp Gln Asp His Ser Thr Ala Lys Gln Cys
 145 150 155 160

ES 2 602 808 T3

caa gag ttt gaa gag tgc ata ggt ggg cct gaa aaa atg gct caa tta Gln Glu Phe Glu Glu Cys Ile Gly Gly Pro Glu Lys Met Ala Gln Leu 165 170 175	528
aca ggg tcc aga gcc cat ttt aga ttt act ggt cct caa att ctg aaa Thr Gly Ser Arg Ala His Phe Arg Phe Thr Gly Pro Gln Ile Leu Lys 180 185 190	576
att gca caa tta gaa cca gaa gct tac gaa aaa aca aag acc att tct Ile Ala Gln Leu Glu Pro Glu Ala Tyr Glu Lys Thr Lys Thr Ile Ser 195 200 205	624
tta gtg tct aat ttt ttg act tct atc tta gtg ggc cat ctt gtt gaa Leu Val Ser Asn Phe Leu Thr Ser Ile Leu Val Gly His Leu Val Glu 210 215 220	672
ttg gag gag gca gat gcc tgt ggt atg aac ctt tat gat ata cgt gaa Leu Glu Glu Ala Asp Ala Cys Gly Met Asn Leu Tyr Asp Ile Arg Glu 225 230 235 240	720
aga aaa ttc agt gat gag cta cta cat cta att gat agt tct tct aag Arg Lys Phe Ser Asp Glu Leu Leu His Leu Ile Asp Ser Ser Ser Lys 245 250 255	768
gat aaa act atc aga caa aaa tta atg aga gca ccc atg aaa aat ttg Asp Lys Thr Ile Arg Gln Lys Leu Met Arg Ala Pro Met Lys Asn Leu 260 265 270	816
ata gcg ggt acc atc tgt aaa tat ttt att gag aag tac ggt ttc aat Ile Ala Gly Thr Ile Cys Lys Tyr Phe Ile Glu Lys Tyr Gly Phe Asn 275 280 285	864
aca aac tgc aag gtc tct ccc atg act ggg gat aat tta gcc act ata Thr Asn Cys Lys Val Ser Pro Met Thr Gly Asp Asn Leu Ala Thr Ile 290 295 300	912
tgt tct tta ccc ctg cgg aag aat gac gtt ctc gtt tcc cta gga aca Cys Ser Leu Pro Leu Arg Lys Asn Asp Val Leu Val Ser Leu Gly Thr 305 310 315 320	960
agt act aca gtt ctt ctg gtc acc gat aag tat cac ccc tct ccg aac Ser Thr Thr Val Leu Leu Val Thr Asp Lys Tyr His Pro Ser Pro Asn 325 330 335	1008
tat cat ctt ttc att cat cca act ctg cca aac cat tat atg ggt atg Tyr His Leu Phe Ile His Pro Thr Leu Pro Asn His Tyr Met Gly Met 340 345 350	1056
att tgt tat tgt aat ggt tct ttg gca agg gag agg ata aga gac gag Ile Cys Tyr Cys Asn Gly Ser Leu Ala Arg Glu Arg Ile Arg Asp Glu 355 360 365	1104
tta aac aaa gaa cgg gaa aat aat tat gag aag act aac gat tgg act Leu Asn Lys Glu Arg Glu Asn Asn Tyr Glu Lys Thr Asn Asp Trp Thr 370 375 380	1152
ctt ttt aat caa gct gtg cta gat gac tca gaa agt agt gaa aat gaa Leu Phe Asn Gln Ala Val Leu Asp Asp Ser Glu Ser Ser Glu Asn Glu 385 390 395 400	1200
tta ggt gta tat ttt cct ctg ggg gag atc gtt cct agc gta aaa gcc Leu Gly Val Tyr Phe Pro Leu Gly Glu Ile Val Pro Ser Val Lys Ala 405 410 415	1248

ata aac aaa agg gtt atc ttc aat cca aaa acg ggt atg att gaa aga Ile Asn Lys Arg Val Ile Phe Asn Pro Lys Thr Gly Met Ile Glu Arg 420 425 430	1296
gag gtg gcc aag ttc aaa gac aag agg cac gat gcc aaa aat att gta Glu Val Ala Lys Phe Lys Asp Lys Arg His Asp Ala Lys Asn Ile Val 435 440 445	1344
gaa tca cag gct tta agt tgc agg gta aga ata tct ccc ctg ctt tcg Glu Ser Gln Ala Leu Ser Cys Arg Val Arg Ile Ser Pro Leu Leu Ser 450 455 460	1392
gat tca aac gca agc tca caa cag aga ctg aac gaa gat aca atc gtg Asp Ser Asn Ala Ser Ser Gln Gln Arg Leu Asn Glu Asp Thr Ile Val 465 470 475 480	1440
aag ttt gat tac gat gaa tct ccg ctg cgg gac tac cta aat aaa agg Lys Phe Asp Tyr Asp Glu Ser Pro Leu Arg Asp Tyr Leu Asn Lys Arg 485 490 495	1488
cca gaa agg act ttt ttt gta ggt ggg gct tct aaa aac gat gct att Pro Glu Arg Thr Phe Phe Val Gly Gly Ala Ser Lys Asn Asp Ala Ile 500 505 510	1536
gtg aag aag ttt gct caa gtc att ggt gct aca aag ggt aat ttt agg Val Lys Lys Phe Ala Gln Val Ile Gly Ala Thr Lys Gly Asn Phe Arg 515 520 525	1584
cta gaa aca cca aac tca tgt gcc ctt ggt ggt tgt tat aag gcc atg Leu Glu Thr Pro Asn Ser Cys Ala Leu Gly Gly Cys Tyr Lys Ala Met 530 535 540	1632
tgg tca ttg tta tat gac tct aat aaa att gca gtt cct ttt gat aaa Trp Ser Leu Leu Tyr Asp Ser Asn Lys Ile Ala Val Pro Phe Asp Lys 545 550 555 560	1680
ttt ctg aat gac aat ttt cca tgg cat gta atg gaa agc ata tcc gat Phe Leu Asn Asp Asn Phe Pro Trp His Val Met Glu Ser Ile Ser Asp 565 570 575	1728
gtg gat aat gaa aat tgg gat cgc tat aat tcc aag att gtc ccc tta Val Asp Asn Glu Asn Trp Asp Arg Tyr Asn Ser Lys Ile Val Pro Leu 580 585 590	1776
agc gaa ctg gaa aag act ctc atc taa Ser Glu Leu Glu Lys Thr Leu Ile 595 600	1803

<210> 26
 <211> 600
 <212> PRT
 <213> Saccharomyces cerevisiae

<400> 26

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

ES 2 602 808 T3

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

ES 2 602 808 T3

275	280	285
Thr Asn Cys Lys Val Ser Pro Met Thr Gly Asp Asn Leu Ala Thr Ile 290 295 300		
Cys Ser Leu Pro Leu Arg Lys Asn Asp Val Leu Val Ser Leu Gly Thr 305 310 315 320		
Ser Thr Thr Val Leu Leu Val Thr Asp Lys Tyr His Pro Ser Pro Asn 325 330 335		
Tyr His Leu Phe Ile His Pro Thr Leu Pro Asn His Tyr Met Gly Met 340 345 350		
Ile Cys Tyr Cys Asn Gly Ser Leu Ala Arg Glu Arg Ile Arg Asp Glu 355 360 365		
Leu Asn Lys Glu Arg Glu Asn Asn Tyr Glu Lys Thr Asn Asp Trp Thr 370 375 380		
Leu Phe Asn Gln Ala Val Leu Asp Asp Ser Glu Ser Ser Glu Asn Glu 385 390 395 400		
Leu Gly Val Tyr Phe Pro Leu Gly Glu Ile Val Pro Ser Val Lys Ala 405 410 415		
Ile Asn Lys Arg Val Ile Phe Asn Pro Lys Thr Gly Met Ile Glu Arg 420 425 430		
Glu Val Ala Lys Phe Lys Asp Lys Arg His Asp Ala Lys Asn Ile Val 435 440 445		
Glu Ser Gln Ala Leu Ser Cys Arg Val Arg Ile Ser Pro Leu Leu Ser 450 455 460		
Asp Ser Asn Ala Ser Ser Gln Gln Arg Leu Asn Glu Asp Thr Ile Val 465 470 475 480		
Lys Phe Asp Tyr Asp Glu Ser Pro Leu Arg Asp Tyr Leu Asn Lys Arg 485 490 495		
Pro Glu Arg Thr Phe Phe Val Gly Gly Ala Ser Lys Asn Asp Ala Ile 500 505 510		
Val Lys Lys Phe Ala Gln Val Ile Gly Ala Thr Lys Gly Asn Phe Arg 515 520 525		

ES 2 602 808 T3

Leu Glu Thr Pro Asn Ser Cys Ala Leu Gly Gly Cys Tyr Lys Ala Met
530 535 540

Trp Ser Leu Leu Tyr Asp Ser Asn Lys Ile Ala Val Pro Phe Asp Lys
545 550 555 560

Phe Leu Asn Asp Asn Phe Pro Trp His Val Met Glu Ser Ile Ser Asp
565 570 575

Val Asp Asn Glu Asn Trp Asp Arg Tyr Asn Ser Lys Ile Val Pro Leu
580 585 590

Ser Glu Leu Glu Lys Thr Leu Ile
595 600

<210> 27

<211> 1809

5 <212> DNA

<213> Kluyveromyces marxianus

<220>

<221> CDS

10 <222> (1)..(1809)

<400> 27

atg tct acc cca tac tat tta ggt ttc gat ctt tct acg cag caa ttg 48
Met Ser Thr Pro Tyr Tyr Leu Gly Phe Asp Leu Ser Thr Gln Gln Leu
1 5 10 15

aaa tgt ctc gct atc gat gac caa ttg aac atc gtc acc agt gtt tcc 96
Lys Cys Leu Ala Ile Asp Asp Gln Leu Asn Ile Val Thr Ser Val Ser
20 25 30

atc gag ttc gac agg gac ttc ccg gct tac aac acc aag aag ggt gtg 144
Ile Glu Phe Asp Arg Asp Phe Pro Ala Tyr Asn Thr Lys Lys Gly Val
35 40 45

tac atc aag aat ggc ggt gtt att gat gct cca gtg gcc atg tgg cta 192
Tyr Ile Lys Asn Gly Gly Val Ile Asp Ala Pro Val Ala Met Trp Leu
50 55 60

gaa gct gtg gac ttg tgt ttc tca cag cta gcg gaa cga atc gat ttg 240
Glu Ala Val Asp Leu Cys Phe Ser Gln Leu Ala Glu Arg Ile Asp Leu
65 70 75 80

aaa cga gtc cag tcg atg tct ggc tcg tgc cag caa cac gga acc gtg 288
Lys Arg Val Gln Ser Met Ser Gly Ser Cys Gln Gln His Gly Thr Val
85 90 95

tac tgg aat tgc gaa cac ttg ccc tca aac ctg gac ccg gca tcc acg 336
Tyr Trp Asn Cys Glu His Leu Pro Ser Asn Leu Asp Pro Ala Ser Thr
100 105 110

ctt cgc gaa caa ttg caa ggg tcg ctc agc aga ccg gtg gcc ccc aat 384
Leu Arg Glu Gln Leu Gln Gly Ser Leu Ser Arg Pro Val Ala Pro Asn
115 120 125

tgg caa gac cac agc aca aag aaa cag tgc gat gag ttg gca gaa agc Trp Gln Asp His Ser Thr Lys Lys Gln Cys Asp Glu Leu Ala Glu Ser 130 135 140	432
gtt gga ggt cca gaa gaa ttg gca cga atc act ggt tct ggt gcc cat Val Gly Gly Pro Glu Glu Leu Ala Arg Ile Thr Gly Ser Gly Ala His 145 150 155 160	480
tac agg ttt tcc ggc tgc caa ata gcc aag atc cac gag acc gag ccc Tyr Arg Phe Ser Gly Ser Gln Ile Ala Lys Ile His Glu Thr Glu Pro 165 170 175	528
gag gtt tac gaa gct acc aag agg atc tgc ctt gtg tgc tct ttc cta Glu Val Tyr Glu Ala Thr Lys Arg Ile Ser Leu Val Ser Ser Phe Leu 180 185 190	576
gcc tct gtg ctt gtt ggg gac att gtc cca ttg gaa gaa gcg gat gcg Ala Ser Val Leu Val Gly Asp Ile Val Pro Leu Glu Glu Ala Asp Ala 195 200 205	624
tgc ggc atg aac ttg tac gac ttg agc aag cac gat ttc gac gag act Cys Gly Met Asn Leu Tyr Asp Leu Ser Lys His Asp Phe Asp Glu Thr 210 215 220	672
tta ctg gca gtg gta gac cac gac acg gct cgt ctc agg aga aag ttg Leu Leu Ala Val Val Asp His Asp Thr Ala Arg Leu Arg Arg Lys Leu 225 230 235 240	720
agc gat cca ccg gtg gga gct ccg act cga gag tcc cct ttg acc agt Ser Asp Pro Pro Val Gly Ala Pro Thr Arg Glu Ser Pro Leu Thr Ser 245 250 255	768
ttg ggt aaa gtg tcc aag tac ttc cag gac aag tac ggc gtg aac tgc Leu Gly Lys Val Ser Lys Tyr Phe Gln Asp Lys Tyr Gly Val Asn Cys 260 265 270	816
gag tgc gag atc ttc ccc ttc act gga gac aac cta gcc acc ata tgc Glu Cys Glu Ile Phe Pro Phe Thr Gly Asp Asn Leu Ala Thr Ile Cys 275 280 285	864
tgc ctt ccc ttg cag aag aac gac gtg ttg atc tgc ttg ggc acc tgc Ser Leu Pro Leu Gln Lys Asn Asp Val Leu Ile Ser Leu Gly Thr Ser 290 295 300	912
acc acg att ctc ttg gtc acg gac cag tac cac tgc tgc ccc aat tac Thr Thr Ile Leu Leu Val Thr Asp Gln Tyr His Ser Ser Pro Asn Tyr 305 310 315 320	960
cac ttg ttc atc cat cca acg gtg ccc gga tac tac atg ggc atg att His Leu Phe Ile His Pro Thr Val Pro Gly Tyr Tyr Met Gly Met Ile 325 330 335	1008
tgc tac tgc aac ggg tct ttg gcg cgc gag cgt gtt cgc gac gac ttg Cys Tyr Cys Asn Gly Ser Leu Ala Arg Glu Arg Val Arg Asp Asp Leu 340 345 350	1056
gcg gga cca cag gcc tgc cag gcc ccc gga gaa cag gtg ccc tgg acc Ala Gly Pro Gln Ala Ser Gln Ala Pro Gly Glu Gln Val Pro Trp Thr 355 360 365	1104
caa ttc aac gac gct ttg ctc gac gac agt tta tcc aac gac aac gaa Gln Phe Asn Asp Ala Leu Leu Asp Asp Ser Leu Ser Asn Asp Asn Glu 370 375 380	1152

atc ggg ttg tac ttc cca ttg ggc gag atc gtg ccc aac gtc gac gcc Ile Gly Leu Tyr Phe Pro Leu Gly Glu Ile Val Pro Asn Val Asp Ala 385 390 395 400	1200
gtg acc aag cgc tgg acg ttc gag cgc aag gaa aac cat ccc aac aaa Val Thr Lys Arg Trp Thr Phe Glu Arg Lys Glu Asn His Pro Asn Lys 405 410 415	1248
acc att gtg cta cac gag ctc gac caa ttc acc cca aaa cgc aag gac Thr Ile Val Leu His Glu Leu Asp Gln Phe Thr Pro Lys Arg Lys Asp 420 425 430	1296
gcc aag aat atc gtc gag tcg caa gcc tta agc tgc agg gtc cgc att Ala Lys Asn Ile Val Glu Ser Gln Ala Leu Ser Cys Arg Val Arg Ile 435 440 445	1344
tct cca cta ttg tcc gac gaa acg gac gcc ctg agc gag acc cag gtg Ser Pro Leu Leu Ser Asp Glu Thr Asp Ala Leu Ser Glu Thr Gln Val 450 455 460	1392
cta tcc aag aaa gaa aac acc caa gtg aca ttc gac tac gac gca ttc Leu Ser Lys Lys Glu Asn Thr Gln Val Thr Phe Asp Tyr Asp Ala Phe 465 470 475 480	1440
cca ctc tgg act tac gca aag aga ccc aac cgt gcc ttc ttc gtt ggt Pro Leu Trp Thr Tyr Ala Lys Arg Pro Asn Arg Ala Phe Phe Val Gly 485 490 495	1488
ggt gcc tcc aag aac gat gcc att gtc cgg aca atg gcc aac gtc att Gly Ala Ser Lys Asn Asp Ala Ile Val Arg Thr Met Ala Asn Val Ile 500 505 510	1536
ggc gcc aga aac ggc aac tac aga cta gaa acg ccc aat tcg tgc gca Gly Ala Arg Asn Gly Asn Tyr Arg Leu Glu Thr Pro Asn Ser Cys Ala 515 520 525	1584
ttg ggc ggc tgc tac aag gcg atg tgg tcc tgg ttg aag gtc cac gag Leu Gly Gly Cys Tyr Lys Ala Met Trp Ser Trp Leu Lys Val His Glu 530 535 540	1632
ccc acg acc acc cca tcc ttc gac gtt tgg ctg aat gcc agt ttc aac Pro Thr Thr Thr Pro Ser Phe Asp Val Trp Leu Asn Ala Ser Phe Asn 545 550 555 560	1680
tgg caa aga gac tgc gaa ttt gtg tgt caa tct gac gct gcc aaa tgg Trp Gln Arg Asp Cys Glu Phe Val Cys Gln Ser Asp Ala Ala Lys Trp 565 570 575	1728
gag caa tgc aac ggc aag ata cag gca ttg agc gaa gca gag gcg tac Glu Gln Cys Asn Gly Lys Ile Gln Ala Leu Ser Glu Ala Glu Ala Tyr 580 585 590	1776
gtc aag gcc ctg gcc cac gac caa ggc cag tga Val Lys Ala Leu Ala His Asp Gln Gly Gln 595 600	1809

<210> 28

<211> 602

<212> PRT

<213> Kluyveromyces marxianus

<400> 28

ES 2 602 808 T3

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

ES 2 602 808 T3

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

Gly Ala Ser Lys Asn Asp Ala Ile Val Arg Thr Met Ala Asn Val Ile
500 505 510

Gly Ala Arg Asn Gly Asn Tyr Arg Leu Glu Thr Pro Asn Ser Cys Ala
515 520 525

Leu Gly Gly Cys Tyr Lys Ala Met Trp Ser Trp Leu Lys Val His Glu
530 535 540

Pro Thr Thr Thr Pro Ser Phe Asp Val Trp Leu Asn Ala Ser Phe Asn
545 550 555 560

Trp Gln Arg Asp Cys Glu Phe Val Cys Gln Ser Asp Ala Ala Lys Trp
565 570 575

Glu Gln Cys Asn Gly Lys Ile Gln Ala Leu Ser Glu Ala Glu Ala Tyr
580 585 590

Val Lys Ala Leu Ala His Asp Gln Gly Gln
595 600

<210> 29
<211> 957
5 <212> NA
<213> Pichia stipitis

<220>
<221> CDS
10 <222> (1)..(957)

<400> 29
atg cct tct att aag ttg aac tct ggt tac gac atg cca gcc gtc ggt 48
Met Pro Ser Ile Lys Leu Asn Ser Gly Tyr Asp Met Pro Ala Val Gly
1 5 10 15
ttc ggc tgt tgg aaa gtc gac gtc gac acc tgt tct gaa cag atc tac 96
Phe Gly Cys Trp Lys Val Asp Val Asp Thr Cys Ser Glu Gln Ile Tyr
20 25 30
cgt gct atc aag acc ggt tac aga ttg ttc gac ggt gcc gaa gat tac 144
Arg Ala Ile Lys Thr Gly Tyr Arg Leu Phe Asp Gly Ala Glu Asp Tyr
35 40 45
gcc aac gaa aag tta gtt ggt gcc ggt gtc aag aag gcc att gac gaa 192
Ala Asn Glu Lys Leu Val Gly Ala Gly Val Lys Lys Ala Ile Asp Glu
50 55 60
ggt atc gtc aag cgt gaa gac ttg ttc ctt acc tcc aag ttg tgg aac 240
Gly Ile Val Lys Arg Glu Asp Leu Phe Leu Thr Ser Lys Leu Trp Asn
65 70 75 80
aac tac cac cac cca gac aac gtc gaa aag gcc ttg aac aga acc ctt 288
Asn Tyr His His Pro Asp Asn Val Glu Lys Ala Leu Asn Arg Thr Leu
85 90 95

ES 2 602 808 T3

tct gac ttg caa gtt gac tac gtt gac ttg ttc ttg atc cac ttc cca Ser Asp Leu Gln Val Asp Tyr Val Asp Leu Phe Leu Ile His Phe Pro 100 105 110	336
gtc acc ttc aag ttc gtt cca tta gaa gaa aag tac cca cca gga ttc Val Thr Phe Lys Phe Val Pro Leu Glu Glu Lys Tyr Pro Pro Gly Phe 115 120 125	384
tac tgt ggt aag ggt gac aac ttc gac tac gaa gat gtt cca att tta Tyr Cys Gly Lys Gly Asp Asn Phe Asp Tyr Glu Asp Val Pro Ile Leu 130 135 140	432
gag acc tgg aag gct ctt gaa aag ttg gtc aag gcc ggt aag atc aga Glu Thr Trp Lys Ala Leu Glu Lys Leu Val Lys Ala Gly Lys Ile Arg 145 150 155 160	480
tct atc ggt gtt tct aac ttc cca ggt gct ttg ctc ttg gac ttg ttg Ser Ile Gly Val Ser Asn Phe Pro Gly Ala Leu Leu Leu Asp Leu Leu 165 170 175	528
aga ggt gct acc atc aag cca tct gtc ttg caa gtt gaa cac cac cca Arg Gly Ala Thr Ile Lys Pro Ser Val Leu Gln Val Glu His His Pro 180 185 190	576
tac ttg caa caa cca aga ttg atc gaa ttc gct caa tcc cgt ggt att Tyr Leu Gln Gln Pro Arg Leu Ile Glu Phe Ala Gln Ser Arg Gly Ile 195 200 205	624
gct gtc acc gct tac tct tcg ttc ggt cct caa tct ttc gtt gaa ttg Ala Val Thr Ala Tyr Ser Ser Phe Gly Pro Gln Ser Phe Val Glu Leu 210 215 220	672
aac caa ggt aga gct ttg aac act tct cca ttg ttc gag aac gaa act Asn Gln Gly Arg Ala Leu Asn Thr Ser Pro Leu Phe Glu Asn Glu Thr 225 230 235 240	720
atc aag gct atc gct gct aag cac ggt aag tct cca gct caa gtc ttg Ile Lys Ala Ile Ala Ala Lys His Gly Lys Ser Pro Ala Gln Val Leu 245 250 255	768
ttg aga tgg tct tcc caa aga ggc att gcc atc att cca aag tcc aac Leu Arg Trp Ser Ser Gln Arg Gly Ile Ala Ile Ile Pro Lys Ser Asn 260 265 270	816
act gtc cca aga ttg ttg gaa aac aag gac gtc aac agc ttc gac ttg Thr Val Pro Arg Leu Leu Glu Asn Lys Asp Val Asn Ser Phe Asp Leu 275 280 285	864
gac gaa caa gat ttc gct gac att gcc aag ttg gac atc aac ttg aga Asp Glu Gln Asp Phe Ala Asp Ile Ala Lys Leu Asp Ile Asn Leu Arg 290 295 300	912
ttc aac gac cca tgg gac tgg gac aag att cct atc ttc gtc taa Phe Asn Asp Pro Trp Asp Trp Asp Lys Ile Pro Ile Phe Val 305 310 315	957

<210> 30
 <211> 318
 <212> PRT
 <213> *Pichia stipitis*
 <400> 30

ES 2 602 808 T3

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

Leu Arg Trp Ser Ser Gln Arg Gly Ile Ala Ile Ile Pro Lys Ser Asn
260 265 270

Thr Val Pro Arg Leu Leu Glu Asn Lys Asp Val Asn Ser Phe Asp Leu
275 280 285

Asp Glu Gln Asp Phe Ala Asp Ile Ala Lys Leu Asp Ile Asn Leu Arg
290 295 300

Phe Asn Asp Pro Trp Asp Trp Asp Lys Ile Pro Ile Phe Val
305 310 315

<210> 31

<211> 1092

5 <212> DNA

<213> Pichia stipitis

<220>

<221> CDS

10 <222> (1)..(1092)

<400> 31

atg act gct aac cct tcc ttg gtg ttg aac aag atc gac gac att tcg 48
Met Thr Ala Asn Pro Ser Leu Val Leu Asn Lys Ile Asp Asp Ile Ser
1 5 10 15

ttc gaa act tac gat gcc cca gaa atc tct gaa cct acc gat gtc ctc 96
Phe Glu Thr Tyr Asp Ala Pro Glu Ile Ser Glu Pro Thr Asp Val Leu
20 25 30

gtc cag gtc aag aaa acc ggt atc tgt ggt tcc gac atc cac ttc tac 144
Val Gln Val Lys Lys Thr Gly Ile Cys Gly Ser Asp Ile His Phe Tyr
35 40 45

gcc cat ggt aga atc ggt aac ttc gtt ttg acc aag cca atg gtc ttg 192
Ala His Gly Arg Ile Gly Asn Phe Val Leu Thr Lys Pro Met Val Leu
50 55 60

ggc cac gaa tcc gcc ggt act gtt gtc cag gtt ggt aag ggt gtc acc 240
Gly His Glu Ser Ala Gly Thr Val Val Gln Val Gly Lys Gly Val Thr
65 70 75 80

tct ctt aag gtt ggt gac aac gtc gct atc gaa cca ggt att cca tcc 288
Ser Leu Lys Val Gly Asp Asn Val Ala Ile Glu Pro Gly Ile Pro Ser
85 90 95

aga ttc tcc gac gaa tac aag agc ggt cac tac aac ttg tgt cct cac 336
Arg Phe Ser Asp Glu Tyr Lys Ser Gly His Tyr Asn Leu Cys Pro His
100 105 110

atg gcc ttc gcc gct act cct aac tcc aag gaa ggc gaa cca aac cca 384
Met Ala Phe Ala Ala Thr Pro Asn Ser Lys Glu Gly Glu Pro Asn Pro
115 120 125

cca ggt acc tta tgt aag tac ttc aag tcg cca gaa gac ttc ttg gtc 432
Pro Gly Thr Leu Cys Lys Tyr Phe Lys Ser Pro Glu Asp Phe Leu Val

ES 2 602 808 T3

130	135	140	
aag ttg cca gac cac gtc agc ttg gaa ctc ggt gct ctt gtt gag cca Lys Leu Pro Asp His Val Ser Leu Glu Leu Gly Ala Leu Val Glu Pro 145 150 155 160			480
ttg tct gtt ggt gtc cac gcc tct aag ttg ggt tcc gtt gct ttc ggc Leu Ser Val Gly Val His Ala Ser Lys Leu Gly Ser Val Ala Phe Gly 165 170 175			528
gac tac gtt gcc gtc ttt ggt gct ggt cct gtt ggt ctt ttg gct gct Asp Tyr Val Ala Val Phe Gly Ala Gly Pro Val Gly Leu Leu Ala Ala 180 185 190			576
gct gtc gcc aag acc ttc ggt gct aag ggt gtc atc gtc gtt gac att Ala Val Ala Lys Thr Phe Gly Ala Lys Gly Val Ile Val Val Asp Ile 195 200 205			624
ttc gac aac aag ttg aag atg gcc aag gac att ggt gct gct act cac Phe Asp Asn Lys Leu Lys Met Ala Lys Asp Ile Gly Ala Ala Thr His 210 215 220			672
acc ttc aac tcc aag acc ggt ggt tct gaa gaa ttg atc aag gct ttc Thr Phe Asn Ser Lys Thr Gly Gly Ser Glu Glu Leu Ile Lys Ala Phe 225 230 235 240			720
ggt ggt aac gtg cca aac gtc gtt ttg gaa tgt act ggt gct gaa cct Gly Gly Asn Val Pro Asn Val Val Leu Glu Cys Thr Gly Ala Glu Pro 245 250 255			768
tgt atc aag ttg ggt gtt gac gcc att gcc cca ggt ggt cgt ttc gtt Cys Ile Lys Leu Gly Val Asp Ala Ile Ala Pro Gly Gly Arg Phe Val 260 265 270			816
caa gtc ggt aac gct gct ggt cca gtc agc ttc cca atc acc gtt ttc Gln Val Gly Asn Ala Ala Gly Pro Val Ser Phe Pro Ile Thr Val Phe 275 280 285			864
gcc atg aag gaa ttg act ttg ttc ggt tct ttc aga tac gga ttc aac Ala Met Lys Glu Leu Thr Leu Phe Gly Ser Phe Arg Tyr Gly Phe Asn 290 295 300			912
gac tac aag act gct gtt gga atc ttt gac act aac tac caa aac ggt Asp Tyr Lys Thr Ala Val Gly Ile Phe Asp Thr Asn Tyr Gln Asn Gly 305 310 315 320			960
aga gaa aat gct cca att gac ttt gaa caa ttg atc acc cac aga tac Arg Glu Asn Ala Pro Ile Asp Phe Glu Gln Leu Ile Thr His Arg Tyr 325 330 335			1008
aag ttc aag gac gct att gaa gcc tac gac ttg gtc aga gcc ggt aag Lys Phe Lys Asp Ala Ile Glu Ala Tyr Asp Leu Val Arg Ala Gly Lys 340 345 350			1056
ggt gct gtc aag tgt ctc att gac gcc cct gag taa Gly Ala Val Lys Cys Leu Ile Asp Gly Pro Glu 355 360			1092

<210> 32
 <211> 363
 <212> PRT
 <213> Pichia stipitis

<400> 32

ES 2 602 808 T3

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

ES 2 602 808 T3

Gly Gly Asn Val Pro Asn Val Val Leu Glu Cys Thr Gly Ala Glu Pro
245 250 255

Cys Ile Lys Leu Gly Val Asp Ala Ile Ala Pro Gly Gly Arg Phe Val
260 265 270

Gln Val Gly Asn Ala Ala Gly Pro Val Ser Phe Pro Ile Thr Val Phe
275 280 285

Ala Met Lys Glu Leu Thr Leu Phe Gly Ser Phe Arg Tyr Gly Phe Asn
290 295 300

Asp Tyr Lys Thr Ala Val Gly Ile Phe Asp Thr Asn Tyr Gln Asn Gly
305 310 315 320

Arg Glu Asn Ala Pro Ile Asp Phe Glu Gln Leu Ile Thr His Arg Tyr
325 330 335

Lys Phe Lys Asp Ala Ile Glu Ala Tyr Asp Leu Val Arg Ala Gly Lys
340 345 350

Gly Ala Val Lys Cys Leu Ile Asp Gly Pro Glu
355 360

<210> 33
<211> 690
5 <212> DNA
<213> Issatchenkia orientalis

<220>
<221> CDS
10 <222> (1)..(690)

<400> 33
atg gtg aaa tca att att gcg cca tct ttg cta gcc ggt gac ttt gca 48
Met Val Lys Ser Ile Ile Ala Pro Ser Leu Leu Ala Gly Asp Phe Ala
1 5 10 15

cac ctt gga tgc gac tgt caa cgt atg ttt gat aat ggc tct gac tgg 96
His Leu Gly Cys Asp Cys Gln Arg Met Phe Asp Asn Gly Ser Asp Trp
20 25 30

gtc cat tta gac gtc atg gat gga cat ttt gtg cca aac att aca atg 144
Val His Leu Asp Val Met Asp Gly His Phe Val Pro Asn Ile Thr Met
35 40 45

ggg cca ccg gtt ata tca tgt ttg aga aaa gca gtg cca aga aag gag 192
Gly Pro Pro Val Ile Ser Cys Leu Arg Lys Ala Val Pro Arg Lys Glu
50 55 60

gat cag cct ggt aag aac ttc ttt gat tgc cat atg atg gtc tcc aat 240
Asp Gln Pro Gly Lys Asn Phe Phe Asp Cys His Met Met Val Ser Asn
65 70 75 80

cct gaa cag tgg gta ccg gag att gcc aaa gct gga ggt gat caa tac 288
Pro Glu Gln Trp Val Pro Glu Ile Ala Lys Ala Gly Gly Asp Gln Tyr
85 90 95

acc ttc cat tac gaa tcc acc aag gat cct gtt gga ttg gtc aag tcc 336
Thr Phe His Tyr Glu Ser Thr Lys Asp Pro Val Gly Leu Val Lys Ser
100 105 110

ata aag gcg cac ggt atg aag gcg gca tgc gcc gtt aaa ccg ggc act 384
Ile Lys Ala His Gly Met Lys Ala Cys Ala Val Lys Pro Gly Thr
115 120 125

gat gtc tct gtc ttg tat gag ctg gcg cca atg cta gat atg gct ttg 432
Asp Val Ser Val Leu Tyr Glu Leu Ala Pro Met Leu Asp Met Ala Leu
130 135 140

gtt atg act gtc gag cca ggt ttt ggt gga caa agt ttc atg cca gat 480
Val Met Thr Val Glu Pro Gly Phe Gly Gly Gln Ser Phe Met Pro Asp
145 150 155 160

atg atg cag aag gtt agg gac cta agg gcc aag ttt cca gat ctc aac 528
Met Met Gln Lys Val Arg Asp Leu Arg Ala Lys Phe Pro Asp Leu Asn
165 170 175

atc caa gtc gac gga ggg ttg ggt aaa ggt act gtt gaa gtt gca gcc 576
Ile Gln Val Asp Gly Gly Leu Gly Lys Gly Thr Val Glu Val Ala Ala
180 185 190

gag gca ggg gcc aat gtt att gtt gct gga aca tcg gtg ttc aag gct 624
Glu Ala Gly Ala Asn Val Ile Val Ala Gly Thr Ser Val Phe Lys Ala
195 200 205

gaa gat cct agt gca atg att tct tat ttg aga gaa gaa gtt gaa aag 672
Glu Asp Pro Ser Ala Met Ile Ser Tyr Leu Arg Glu Glu Val Glu Lys
210 215 220

aat ctg aag aaa gat tag 690
Asn Leu Lys Lys Asp
225

<210> 34

<211> 229

5 <212> PRT

<213> Issatchenkia orientalis

<400> 34

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

His Leu Gly Cys Asp Cys Gln Arg Met Phe Asp Asn Gly Ser Asp Trp
20 25 30

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

Gly Pro Pro Val Ile Ser Cys Leu Arg Lys Ala Val Pro Arg Lys Glu
50 55 60

ES 2 602 808 T3

Asp Gln Pro Gly Lys Asn Phe Phe Asp Cys His Met Met Val Ser Asn
 65 70 75 80
 Pro Glu Gln Trp Val Pro Glu Ile Ala Lys Ala Gly Gly Asp Gln Tyr
 85 90 95
 Thr Phe His Tyr Glu Ser Thr Lys Asp Pro Val Gly Leu Val Lys Ser
 100 105 110
 Ile Lys Ala His Gly Met Lys Ala Ala Cys Ala Val Lys Pro Gly Thr
 115 120 125
 Asp Val Ser Val Leu Tyr Glu Leu Ala Pro Met Leu Asp Met Ala Leu
 130 135 140
 Val Met Thr Val Glu Pro Gly Phe Gly Gly Gln Ser Phe Met Pro Asp
 145 150 155 160
 Met Met Gln Lys Val Arg Asp Leu Arg Ala Lys Phe Pro Asp Leu Asn
 165 170 175
 Ile Gln Val Asp Gly Gly Leu Gly Lys Gly Thr Val Glu Val Ala Ala
 180 185 190
 Glu Ala Gly Ala Asn Val Ile Val Ala Gly Thr Ser Val Phe Lys Ala
 195 200 205
 Glu Asp Pro Ser Ala Met Ile Ser Tyr Leu Arg Glu Glu Val Glu Lys
 210 215 220
 Asn Leu Lys Lys Asp
 225

<210> 35
 <211> 717
 5 <212> DNA
 <213> Saccharomyces cerevisiae

<220>
 <221> CDS
 10 <222> (1)..(717)

<400> 35
 atg gtc aaa cca att ata gct ccc agt atc ctt gct tct gac ttc gcc 48
 Met Val Lys Pro Ile Ile Ala Pro Ser Ile Leu Ala Ser Asp Phe Ala
 1 5 10 15
 aac ttg ggt tgc gaa tgt cat aag gtc atc aac gcc ggc gca gat tgg 96
 Asn Leu Gly Cys Glu Cys His Lys Val Ile Asn Ala Gly Ala Asp Trp
 20 25 30

tta cat atc gat gtc atg gac ggc cat ttt gtt cca aac att act ctg	144
Leu His Ile Asp Val Met Asp Gly His Phe Val Pro Asn Ile Thr Leu	
35 40 45	
ggc caa cca att gtt acc tcc cta cgt cgt tct gtg cca cgc cct ggc	192
Gly Gln Pro Ile Val Thr Ser Leu Arg Arg Ser Val Pro Arg Pro Gly	
50 55 60	
gat gct agc aac aca gaa aag aag ccc act gcg ttc ttc gat tgt cac	240
Asp Ala Ser Asn Thr Glu Lys Lys Pro Thr Ala Phe Phe Asp Cys His	
65 70 75 80	
atg atg gtt gaa aat cct gaa aaa tgg gtc gac gat ttt gct aaa tgt	288
Met Met Val Glu Asn Pro Glu Lys Trp Val Asp Asp Phe Ala Lys Cys	
85 90 95	
ggg gct gac caa ttt acg ttc cac tac gag gcc aca caa gac cct ttg	336
Gly Ala Asp Gln Phe Thr Phe His Tyr Glu Ala Thr Gln Asp Pro Leu	
100 105 110	
cat tta gtt aag ttg att aag tct aag ggc atc aaa gct gca tgc gcc	384
His Leu Val Lys Leu Ile Lys Ser Lys Gly Ile Lys Ala Ala Cys Ala	
115 120 125	
atc aaa cct ggt act tct gtt gac gtt tta ttt gaa cta gct cct cat	432
Ile Lys Pro Gly Thr Ser Val Asp Val Leu Phe Glu Leu Ala Pro His	
130 135 140	
ttg gat atg gct ctt gtt atg act gtg gaa cct ggg ttt gga ggc caa	480
Leu Asp Met Ala Leu Val Met Thr Val Glu Pro Gly Phe Gly Gly Gln	
145 150 155 160	
aaa ttc atg gaa gac atg atg cca aaa gtg gaa act ttg aga gcc aag	528
Lys Phe Met Glu Asp Met Met Pro Lys Val Glu Thr Leu Arg Ala Lys	
165 170 175	
ttc ccc cat ttg aat atc caa gtc gat ggt ggt ttg ggc aag gag acc	576
Phe Pro His Leu Asn Ile Gln Val Asp Gly Gly Leu Gly Lys Glu Thr	
180 185 190	
atc ccg aaa gcc gcc aaa gcc ggt gcc aac gtt att gtc gct ggt acc	624
Ile Pro Lys Ala Ala Lys Ala Gly Ala Asn Val Ile Val Ala Gly Thr	
195 200 205	
agt gtt ttc act gca gct gac ccg cac gat gtt atc tcc ttc atg aaa	672
Ser Val Phe Thr Ala Ala Asp Pro His Asp Val Ile Ser Phe Met Lys	
210 215 220	
gaa gaa gtc tcg aag gaa ttg cgt tct aga gat ttg cta gat tag	717
Glu Glu Val Ser Lys Glu Leu Arg Ser Arg Asp Leu Leu Asp	
225 230 235	

<210> 36

<211> 238

5 <212> PRT

<213> *Saccharomyces cerevisiae*

<400> 36

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

ES 2 602 808 T3

```

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
      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> 37
 <211> 711
 5 <212> DNA
 <213> Kluyveromyces marxianus

<220>
 <221> CDS
 10 <222> (1)..(711)

<400> 37

ES 2 602 808 T3

atg gtc caa cct atc att gct cct tct atc tta gcc agt gat ttt tgt	48
Met Val Gln Pro Ile Ile Ala Pro Ser Ile Leu Ala Ser Asp Phe Cys	
1 5 10 15	
aat ctt ggg tgc gag tgc cac aag gtg atc aac tcc ggt gcc gaa tgg	96
Asn Leu Gly Cys Glu Cys His Lys Val Ile Asn Ser Gly Ala Glu Trp	
20 25 30	
ttg cac att gac att atg gac ggc cat ttc gtt cca aat atg tct ttg	144
Leu His Ile Asp Ile Met Asp Gly His Phe Val Pro Asn Met Ser Leu	
35 40 45	
ggg cag cca gtt gtc gag tcg ttg cgt aag gtg att ggg aag tac aac	192
Gly Gln Pro Val Val Glu Leu Arg Lys Val Ile Gly Lys Tyr Asn	
50 55 60	
gat cca gat acc aag ttg ccc aag gcg ttt ttc gac tgc cac atg atg	240
Asp Pro Asp Thr Lys Leu Pro Lys Ala Phe Phe Asp Cys His Met Met	
65 70 75 80	
gtg agc gag cct gag aaa tgg gta gag gac ttt gcg agg att ggg tgc	288
Val Ser Glu Pro Glu Lys Trp Val Glu Asp Phe Ala Arg Ile Gly Cys	
85 90 95	
gac caa ttc acg ttc cat tac gag gcc acg aag gat cca aag ggg ttg	336
Asp Gln Phe Thr Phe His Tyr Glu Ala Thr Lys Asp Pro Lys Gly Leu	
100 105 110	
gtg gag ttg atc aag aag aac ggg atg aag gcc gcg tgt gct gtg aaa	384
Val Glu Leu Ile Lys Lys Asn Gly Met Lys Ala Ala Cys Ala Val Lys	
115 120 125	
ccg ggc acc ccg gtt gat gtg cta tac gag ttg gca cct gag ttg gat	432
Pro Gly Thr Pro Val Asp Val Leu Tyr Glu Leu Ala Pro Glu Leu Asp	
130 135 140	
atg gcg ctt gtg atg acg gtg gag ccc ggt ttt ggt ggc cag aag ttc	480
Met Ala Leu Val Met Thr Val Glu Pro Gly Phe Gly Gly Gln Lys Phe	
145 150 155 160	
atg agc gat atg atg agc aag gtg aag gat ttg aga gag agg ttc cca	528
Met Ser Asp Met Met Ser Lys Val Lys Asp Leu Arg Glu Arg Phe Pro	
165 170 175	
acg ttg aac atc cag gtc gac ggt ggg ttg ggt aag caa aac gtc gaa	576
Thr Leu Asn Ile Gln Val Asp Gly Gly Leu Gly Lys Gln Asn Val Glu	
180 185 190	
cag gcc tcg gaa gcg ggt gcg aat gtg att gtc gcg ggt acg tcc gtg	624
Gln Ala Ser Glu Ala Gly Ala Asn Val Ile Val Ala Gly Thr Ser Val	
195 200 205	
ttc cga tcg gac gac ccg gcg gac gtt atc ggc ttc atg aaa agt aag	672
Phe Arg Ser Asp Asp Pro Ala Asp Val Ile Gly Phe Met Lys Ser Lys	
210 215 220	
gtc aag gat gcg ttg gtg gcc aaa gac ctc ttg aca tag	711
Val Lys Asp Ala Leu Val Ala Lys Asp Leu Leu Thr	
225 230 235	

5 <210> 38
 <211> 236
 <212> PRT
 <213> Kluyveromyces marxianus

10 <400> 38

ES 2 602 808 T3

```

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

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

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

Gly Gln Pro Val Val Glu Ser Leu Arg Lys Val Ile Gly Lys Tyr Asn
          50           55           60

Asp Pro Asp Thr Lys Leu Pro Lys Ala Phe Phe Asp Cys His Met Met
65           70           75           80

Val Ser Glu Pro Glu Lys Trp Val Glu Asp Phe Ala Arg Ile Gly Cys
          85           90           95

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

Val Glu Leu Ile Lys Lys Asn Gly Met Lys Ala Ala Cys Ala Val Lys
          115          120          125

Pro Gly Thr Pro Val Asp Val Leu Tyr Glu Leu Ala Pro Glu Leu Asp
          130          135          140

Met Ala Leu Val Met Thr Val Glu Pro Gly Phe Gly Gly Gln Lys Phe
145           150           155           160

Met Ser Asp Met Met Ser Lys Val Lys Asp Leu Arg Glu Arg Phe Pro
          165          170          175

Thr Leu Asn Ile Gln Val Asp Gly Gly Leu Gly Lys Gln Asn Val Glu
          180          185          190

Gln Ala Ser Glu Ala Gly Ala Asn Val Ile Val Ala Gly Thr Ser Val
          195          200          205

Phe Arg Ser Asp Asp Pro Ala Asp Val Ile Gly Phe Met Lys Ser Lys
          210          215          220

Val Lys Asp Ala Leu Val Ala Lys Asp Leu Leu Thr
225           230           235

```

5 <210> 39
 <211> 780
 <212> DNA
 <213> Issatchenkia orientalis

10 <220>
 <221> CDS
 <222> (1)..(780)

<400> 39

atg ttt gga cga tat tcc cgt cat att tct cgt cta aaa cca acc ttc	48
Met Phe Gly Arg Tyr Ser Arg His Ile Ser Arg Leu Lys Pro Thr Phe	
1 5 10 15	
tct cac cta cta tcg aaa ggg atg tct gac ttg gtc gaa aaa tct aag	96
Ser His Leu Leu Ser Lys Gly Met Ser Asp Leu Val Glu Lys Ser Lys	
20 25 30	
aaa ctt tgc gct tat act gcg gtt gac aag aat tta aag cca agc cat	144
Lys Leu Cys Ala Tyr Thr Ala Val Asp Lys Asn Leu Lys Pro Ser His	
35 40 45	
aaa gtt att ggt atc ggt tct ggc tct aca gta gtt tat gtt gct gag	192
Lys Val Ile Gly Ile Gly Ser Gly Ser Thr Val Val Tyr Val Ala Glu	
50 55 60	
cgt atc ggt cag cta gaa aac aaa aac cac tat gtt tgc att ccc act	240
Arg Ile Gly Gln Leu Glu Asn Lys Asn His Tyr Val Cys Ile Pro Thr	
65 70 75 80	
ggt ttc caa tct aaa cag cta atc att gaa aat ggt ttg agg ttg ggt	288
Gly Phe Gln Ser Lys Gln Leu Ile Ile Glu Asn Gly Leu Arg Leu Gly	
85 90 95	
ggt ttg gaa gaa tat cca act atc gat att gca ttt gat ggt gct gat	336
Gly Leu Glu Glu Tyr Pro Thr Ile Asp Ile Ala Phe Asp Gly Ala Asp	
100 105 110	
gag atc gat acc ggc ttg aac tgt atc aaa ggt ggt ggt gct tgt caa	384
Glu Ile Asp Thr Gly Leu Asn Cys Ile Lys Gly Gly Gly Ala Cys Gln	
115 120 125	
tta cag gaa aag ttg gtt gct gac tca gca act gag ttt att att gtt	432
Leu Gln Glu Lys Leu Val Ala Asp Ser Ala Thr Glu Phe Ile Ile Val	
130 135 140	
gca gat gat cgt aaa aat acc ggt gtg ttg ggt aaa ggt tgg aag aaa	480
Ala Asp Asp Arg Lys Asn Thr Gly Val Leu Gly Lys Gly Trp Lys Lys	
145 150 155 160	
ggg att cca att gaa gtc atc cca aat gca tat gct aag atc agt aaa	528
Gly Ile Pro Ile Glu Val Ile Pro Asn Ala Tyr Ala Lys Ile Ser Lys	
165 170 175	
gag cta aga gaa cta ggt ggc aat cca gta gta agg tct ggt gct ccg	576
Glu Leu Arg Glu Leu Gly Gly Asn Pro Val Val Arg Ser Gly Ala Pro	
180 185 190	
tca aaa gca ggt cct gca atc act gat aat ggc aac ttt gtc atc gat	624
Ser Lys Ala Gly Pro Ala Ile Thr Asp Asn Gly Asn Phe Val Ile Asp	
195 200 205	
tgt gat ttt gga gaa atc gaa gtc tct aaa gta gaa gcg ctc aac agt	672
Cys Asp Phe Gly Glu Ile Glu Val Ser Lys Val Glu Ala Leu Asn Ser	
210 215 220	
aaa atc aaa gct atg att ggt gtt gtg gaa aca ggt tta ttt gtt aat	720
Lys Ile Lys Ala Met Ile Gly Val Val Glu Thr Gly Leu Phe Val Asn	
225 230 235 240	
atg gcc aag aag gct tat att ggt aat gca gac ggt agt gtt act aca	768
Met Ala Lys Lys Ala Tyr Ile Gly Asn Ala Asp Gly Ser Val Thr Thr	
245 250 255	
tta tct gtt taa	780
Leu Ser Val	

5

<210> 40

ES 2 602 808 T3

<211> 259
 <212> PRT
 <213> Issatchenkia orientalis

5 <400> 40

```

Met Phe Gly Arg Tyr Ser Arg His Ile Ser Arg Leu Lys Pro Thr Phe
 1           5           10           15

Ser His Leu Leu Ser Lys Gly Met Ser Asp Leu Val Glu Lys Ser Lys
      20           25           30

Lys Leu Cys Ala Tyr Thr Ala Val Asp Lys Asn Leu Lys Pro Ser His
 35           40           45

Lys Val Ile Gly Ile Gly Ser Gly Ser Thr Val Val Tyr Val Ala Glu
 50           55           60

Arg Ile Gly Gln Leu Glu Asn Lys Asn His Tyr Val Cys Ile Pro Thr
 65           70           75           80

Gly Phe Gln Ser Lys Gln Leu Ile Ile Glu Asn Gly Leu Arg Leu Gly
      85           90           95

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

Glu Ile Asp Thr Gly Leu Asn Cys Ile Lys Gly Gly Gly Ala Cys Gln
      115           120           125

Leu Gln Glu Lys Leu Val Ala Asp Ser Ala Thr Glu Phe Ile Ile Val
      130           135           140

Ala Asp Asp Arg Lys Asn Thr Gly Val Leu Gly Lys Gly Trp Lys Lys
      145           150           155           160

Gly Ile Pro Ile Glu Val Ile Pro Asn Ala Tyr Ala Lys Ile Ser Lys
      165           170           175

Glu Leu Arg Glu Leu Gly Gly Asn Pro Val Val Arg Ser Gly Ala Pro
      180           185           190

Ser Lys Ala Gly Pro Ala Ile Thr Asp Asn Gly Asn Phe Val Ile Asp
      195           200           205

Cys Asp Phe Gly Glu Ile Glu Val Ser Lys Val Glu Ala Leu Asn Ser
      210           215           220

Lys Ile Lys Ala Met Ile Gly Val Val Glu Thr Gly Leu Phe Val Asn
      225           230           235           240

Met Ala Lys Lys Ala Tyr Ile Gly Asn Ala Asp Gly Ser Val Thr Thr
      245           250           255

Leu Ser Val
  
```

```

<210> 41
<211> 777
<212> DNA
5 <213> Saccharomyces cerevisiae

<220>
<221> CDS
10 <222> (1)..(777)

<400> 41
  atg gct gcc ggt gtc cca aaa att gat gcg tta gaa tct ttg ggc aat      48
  Met Ala Ala Gly Val Pro Lys Ile Asp Ala Leu Glu Ser Leu Gly Asn
   1             5             10             15

  cct ttg gag gat gcc aag aga gct gca gca tac aga gca gtt gat gaa      96
  Pro Leu Glu Asp Ala Lys Arg Ala Ala Ala Tyr Arg Ala Val Asp Glu
           20           25           30

  aat tta aaa ttt gat gat cac aaa att att gga att ggt agt ggt agc     144
  Asn Leu Lys Phe Asp Asp His Lys Ile Ile Gly Ile Gly Ser Gly Ser
           35           40           45

  aca gtg gtt tat gtt gcc gaa aga att gga caa tat ttg cat gac cct     192
  Thr Val Val Tyr Val Ala Glu Arg Ile Gly Gln Tyr Leu His Asp Pro
   50             55             60

```

ES 2 602 808 T3

aaa ttt tat gaa gta gcg tct aaa ttc att tgc att cca aca gga ttc	240
Lys Phe Tyr Glu Val Ala Ser Lys Phe Ile Cys Ile Pro Thr Gly Phe	
65 70 75 80	
caa tca aga aac ttg att ttg gat aac aag ttg caa tta ggc tcc att	288
Gln Ser Arg Asn Leu Ile Leu Asp Asn Lys Leu Gln Leu Gly Ser Ile	
85 90 95	
gaa cag tat cct cgc att gat ata gcg ttt gac ggt gct gat gaa gtg	336
Glu Gln Tyr Pro Arg Ile Asp Ile Ala Phe Asp Gly Ala Asp Glu Val	
100 105 110	
gat gag aat tta caa tta att aaa ggt ggt ggt gct tgt cta ttt caa	384
Asp Glu Asn Leu Gln Leu Ile Lys Gly Gly Gly Ala Cys Leu Phe Gln	
115 120 125	
gaa aaa ttg gtt agt act agt gct aaa acc ttc att gtc gtt gct gat	432
Glu Lys Leu Val Ser Thr Ser Ala Lys Thr Phe Ile Val Val Ala Asp	
130 135 140	
tca aga aaa aag tca cca aaa cat tta ggt aag aac tgg agg caa ggt	480
Ser Arg Lys Lys Ser Pro Lys His Leu Gly Lys Asn Trp Arg Gln Gly	
145 150 155 160	
gtt ccc att gaa att gta cct tcc tca tac gtg agg gtc aag aat gat	528
Val Pro Ile Glu Ile Val Pro Ser Ser Tyr Val Arg Val Lys Asn Asp	
165 170 175	
cta tta gaa caa ttg cat gct gaa aaa gtt gac atc aga caa gga ggt	576
Leu Leu Glu Gln Leu His Ala Glu Lys Val Asp Ile Arg Gln Gly Gly	
180 185 190	
tct gct aaa gca ggt cct gtt gta act gac aat aat aac ttc att atc	624
Ser Ala Lys Ala Gly Pro Val Val Thr Asp Asn Asn Phe Ile Ile	
195 200 205	
gat gcg gat ttc ggt gaa att tcc gat cca aga aaa ttg cat aga gaa	672
Asp Ala Asp Phe Gly Glu Ile Ser Asp Pro Arg Lys Leu His Arg Glu	
210 215 220	
atc aaa ctg tta gtg ggc gtg gtg gaa aca ggt tta ttc atc gac aac	720
Ile Lys Leu Leu Val Gly Val Val Glu Thr Gly Leu Phe Ile Asp Asn	
225 230 235 240	
gct tca aaa gcc tac ttc ggt aat tct gac ggt agt gtt gaa gtt acc	768
Ala Ser Lys Ala Tyr Phe Gly Asn Ser Asp Gly Ser Val Glu Val Thr	
245 250 255	
gaa aag tga	777
Glu Lys	

<210> 42

<211> 258

5 <212> PRT

<213> Saccharomyces cerevisiae

<400> 42

Met	Ala	Ala	Gly	Val	Pro	Lys	Ile	Asp	Ala	Leu	Glu	Ser	Leu	Gly	Asn
1				5					10					15	

ES 2 602 808 T3

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

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

Glu Lys

<210> 43

<211> 819

5 <212> DNA

<213> Kluyveromyces marxianus

<220>

<221> CDS

10 <222> (1)..(819)

<400> 43

ES 2 602 808 T3

atg tac tgt gct gta agc agg cgt gtt cat aga ctc ata tta cca agc Met Tyr Cys Ala Val Ser Arg Arg Val His Arg Leu Ile Leu Pro Ser 1 5 10 15	48
ttt cca cac gtt ttc aag atg cca tta agc gat ctc tcg aaa ctg cca Phe Pro His Val Phe Lys Met Pro Leu Ser Asp Leu Ser Lys Leu Pro 20 25 30	96
cct ctt tcg gac caa ttg gag cag gct aaa cgt act gct gca tac cgt Pro Leu Ser Asp Gln Leu Glu Ala Lys Arg Thr Ala Ala Tyr Arg 35 40 45	144
gcg gtt gat gag aat ttt gat gcc aag att cat aaa gta gtg ggt gtt Ala Val Asp Glu Asn Phe Asp Ala Lys Ile His Lys Val Val Gly Val 50 55 60	192
ggt agt ggt act act gta gtt tat gtt gcc gaa agg ctt gga gag tat Gly Ser Gly Thr Thr Val Val Tyr Val Ala Glu Arg Leu Gly Glu Tyr 65 70 75 80	240
gtg aac gat agc agc aac ttt gtg tgt ata cct act gga ttc cag tcc Val Asn Asp Ser Ser Asn Phe Val Cys Ile Pro Thr Gly Phe Gln Ser 85 90 95	288
aaa cag ttg att cta tcg aac agg ttg cag tta ggg agc att gag cag Lys Gln Leu Ile Leu Ser Asn Arg Leu Gln Leu Gly Ser Ile Glu Gln 100 105 110	336
tat cct gag atc gat att gcg ttc gat ggt gca gat gag gtc gat gag Tyr Pro Glu Ile Asp Ile Ala Phe Asp Gly Ala Asp Glu Val Asp Glu 115 120 125	384
aat tta cag ttg att aaa ggt ggt gga gcg tgc ttg ttt cag gaa aag Asn Leu Gln Leu Ile Lys Gly Gly Gly Ala Cys Leu Phe Gln Glu Lys 130 135 140	432
ttg gtg agc act agt gcg aag aag ttt att gtt gtt gca gac tcg aga Leu Val Ser Thr Ser Ala Lys Lys Phe Ile Val Val Ala Asp Ser Arg 145 150 155 160	480
aaa cgg tcg cca aag cac ttg ggg acc aac tgg aaa cgc ggt gtg ccg Lys Arg Ser Pro Lys His Leu Gly Thr Asn Trp Lys Arg Gly Val Pro 165 170 175	528
att gaa gtg gtg ccc agt tcc tat gtg cat gtg ctt act gcg ttg aaa Ile Glu Val Val Pro Ser Ser Tyr Val His Val Leu Thr Ala Leu Lys 180 185 190	576
gat aga cta cat tgc aag agc gct att gtg aga cag ggc ggt agc gcg	624

ES 2 602 808 T3

Asp	Arg	Leu	His	Cys	Lys	Ser	Ala	Ile	Val	Arg	Gln	Gly	Gly	Ser	Ala		
	195						200					205					
aaa	gcg	ggg	ccc	gtg	gtt	acg	gat	aac	tgc	aac	ttc	atc	att	gac	gct	672	
Lys	Ala	Gly	Pro	Val	Val	Thr	Asp	Asn	Cys	Asn	Phe	Ile	Ile	Asp	Ala		
	210						215				220						
gac	ttt	ggc	gag	atc	gcg	gac	cct	aga	aag	cta	cat	caa	gac	atc	aag	720	
Asp	Phe	Gly	Glu	Ile	Ala	Asp	Pro	Arg	Lys	Leu	His	Gln	Asp	Ile	Lys		
	225					230					235				240		
atg	ttg	gtt	ggg	gtt	gtt	gag	acc	ggg	tta	ttc	att	gac	aat	gcg	gaa	768	
Met	Leu	Val	Gly	Val	Val	Glu	Thr	Gly	Leu	Phe	Ile	Asp	Asn	Ala	Glu		
				245					250					255			
aag	gcg	tac	ttc	ggt	tgc	cca	gat	ggg	tcc	gtc	gag	ctg	cag	gtg	ttg	816	
Lys	Ala	Tyr	Phe	Gly	Ser	Pro	Asp	Gly	Ser	Val	Glu	Leu	Gln	Val	Leu		
			260					265					270				
tag																819	

<210> 44
 <211> 272
 <212> PRT
 <213> Kluyveromyces marxianus

<400> 44

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

ES 2 602 808 T3

Leu Val Ser Thr Ser Ala Lys Lys Phe Ile Val Val Ala Asp Ser Arg
145 150 155 160

Lys Arg Ser Pro Lys His Leu Gly Thr Asn Trp Lys Arg Gly Val Pro
165 170 175

Ile Glu Val Val Pro Ser Ser Tyr Val His Val Leu Thr Ala Leu Lys
180 185 190

Asp Arg Leu His Cys Lys Ser Ala Ile Val Arg Gln Gly Gly Ser Ala
195 200 205

Lys Ala Gly Pro Val Val Thr Asp Asn Cys Asn Phe Ile Ile Asp Ala
210 215 220

Asp Phe Gly Glu Ile Ala Asp Pro Arg Lys Leu His Gln Asp Ile Lys
225 230 235 240

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

Lys Ala Tyr Phe Gly Ser Pro Asp Gly Ser Val Glu Leu Gln Val Leu
260 265 270

<210> 45

<211> 2031

5 <212> DNA

<213> Issatchenkia orientalis

<220>

<221> CDS

10 <222> (1)..(2031)

<400> 45

atg tcg gac atc acc acc aag gca gtc aac acg atc cgt gtg tta gcg 48
Met Ser Asp Ile Thr Thr Lys Ala Val Asn Thr Ile Arg Val Leu Ala
1 5 10 15

gca gac gtt gtt gca aag gcg aat tcc ggg cat cca ggt gcc cca atg 96
Ala Asp Val Val Ala Lys Ala Asn Ser Gly His Pro Gly Ala Pro Met
20 25 30

gga atg gct cca gca gcc cat gtt tta ttt tcc cag tta aag act aat 144
Gly Met Ala Pro Ala Ala His Val Leu Phe Ser Gln Leu Lys Thr Asn
35 40 45

cca aag aat cca gaa tgg atc aac aga gac cgg ttt gtg ctc tcc aat 192
Pro Lys Asn Pro Glu Trp Ile Asn Arg Asp Arg Phe Val Leu Ser Asn
50 55 60

ggc cat gcg gtt gca tta ttg tac gtc atg ctg cat ctc tcc ggg tac 240
Gly His Ala Val Ala Leu Leu Tyr Val Met Leu His Leu Ser Gly Tyr

ES 2 602 808 T3

65	70	75	80	
cct att tct atg gag gat ttg aaa cag ttt aga cag ttg gat tca aag				288
Pro Ile Ser Met	Glu Asp Leu Lys	Gln Phe Arg Gln Leu Asp	Ser Lys	
	85	90	95	
act cct ggc cat cca gag agt gag act gtt ggt gtt gac gtc act aca				336
Thr Pro Gly His	Pro Glu Ser Glu Thr	Val Gly Val Asp	Val Thr Thr	
	100	105	110	
ggg cca tta ggt caa ggt atc tcc aat gcc gtt ggt ctg gcc att gcg				384
Gly Pro	Gly Gln Gly Ile	Ser Asn Ala Val Gly	Leu Ala Ile Ala	
	115	120	125	
cag gcg aat ttc ggc gca acc tac aac aaa ccg ggc tac acc atc tca				432
Gln Ala Asn Phe	Gly Ala Thr Tyr	Asn Lys Pro Gly Tyr	Thr Ile Ser	
	130	135	140	
aat aac tac aca tac acg ttt ttc ggt gac ggt tgt atg atg gaa ggt				480
Asn Asn Tyr Thr Tyr Thr Phe Phe Gly Asp	Gly Cys Met Met	Glu Gly		
	145	150	155	160
gtt gcc tct gag gca gca tct ctt gct ggc cat tta cag ttg ggc aac				528
Val Ala Ser Glu Ala Ala Ser Leu Ala Gly His	Leu Gln Leu Gly Asn			
	165	170	175	
ttg att gcc ttt tat gat gat aac aag atc tca att gat ggt tcc acc				576
Leu Ile Ala Phe Tyr Asp Asp Asn Lys Ile Ser Ile Asp	Gly Ser Thr			
	180	185	190	
aat atg gcc ttc acc gag gat gtt tcc aag agg tta gaa tct tac ggt				624
Asn Met Ala Phe Thr Glu Asp Val Ser Lys Arg	Leu Glu Ser Tyr Gly			
	195	200	205	
tgg gaa gtt att gaa gtt aag gat gca gac act gat ttt gat gcc ctg				672
Trp Glu Val Ile Glu Val Lys Asp Ala Asp Thr Asp Phe Asp	Ala Leu			
	210	215	220	
gct ctt gcc att gaa aag gca aag tcc aac aag aac caa cca tct tgt				720
Ala Leu Ala Ile Glu Lys Ala Lys Ser Asn Lys Asn Gln Pro Ser Cys				
	225	230	235	240
att aga atg tcc act acc att ggt tat ggt tcc ttg aaa cag gga act				768
Ile Arg Met Ser Thr Thr Ile Gly Tyr Gly Ser Leu Lys Gln Gly Thr				
	245	250	255	
gcc ggt gtc cat ggt tct cca tta aag gca gac gac att gcc caa ttg				816
Ala Gly Val His Gly Ser Pro Leu Lys Ala Asp Asp Ile Ala Gln Leu				
	260	265	270	
aag gag aaa tgg ggg ttt gat cct gca aag tca ttt aat gtc gaa gat				864
Lys Glu Lys Trp Gly Phe Asp Pro Ala Lys Ser Phe Asn Val Glu Asp				
	275	280	285	
gat gtc tat gat tat tgg aag tct gtt gct gct agg ggg gaa gag gaa				912
Asp Val Tyr Asp Tyr Trp Lys Ser Val Ala Ala Arg Gly Glu Glu Glu				
	290	295	300	
aat aga aaa tgg gat tct ctc ttt gaa gct tat tca aag gag tac cct				960
Asn Arg Lys Trp Asp Ser Leu Phe Glu Ala Tyr Ser Lys Glu Tyr Pro				
	305	310	315	320
aag gaa gcc gaa gaa atc aag agg aga gtc tct tat aag tta cca caa				1008

ES 2 602 808 T3

Lys	Glu	Ala	Glu	Glu	Ile	Lys	Arg	Arg	Val	Ser	Tyr	Lys	Leu	Pro	Gln	
				325					330					335		
ggg	tgg	gag	aag	gtc	tta	cca	act	tac	acc	aag	gat	gat	aaa	cct	ctt	1056
Gly	Trp	Glu	Lys	Val	Leu	Pro	Thr	Tyr	Thr	Lys	Asp	Asp	Lys	Pro	Leu	
			340					345					350			
gca	tca	aga	aaa	tta	tca	gaa	att	gtc	ctg	ggg	aaa	atc	gaa	gaa	tct	1104
Ala	Ser	Arg	Lys	Leu	Ser	Glu	Ile	Val	Leu	Gly	Lys	Ile	Glu	Glu	Ser	
		355					360					365				
tta	cct	gaa	cta	att	ggg	ggg	tca	gcg	gat	ttg	act	cct	tcc	aac	ttg	1152
Leu	Pro	Glu	Leu	Ile	Gly	Gly	Ser	Ala	Asp	Leu	Thr	Pro	Ser	Asn	Leu	
	370					375					380					
acc	aga	tgg	ggg	ggg	gca	gtt	gat	ttc	caa	cct	cct	caa	acc	gga	ttg	1200
Thr	Arg	Trp	Gly	Gly	Ala	Val	Asp	Phe	Gln	Pro	Pro	Gln	Thr	Gly	Leu	
	385				390					395					400	
gga	gat	tat	gcc	ggc	aga	tac	atc	aga	ttt	ggg	gtt	aga	gaa	cac	ggg	1248
Gly	Asp	Tyr	Ala	Gly	Arg	Tyr	Ile	Arg	Phe	Gly	Val	Arg	Glu	His	Gly	
			405					410						415		
atg	ggg	gca	atc	atg	aat	ggg	att	gca	gca	tac	ggg	gct	aac	tac	aag	1296
Met	Gly	Ala	Ile	Met	Asn	Gly	Ile	Ala	Ala	Tyr	Gly	Ala	Asn	Tyr	Lys	
			420				425						430			
cct	tat	ggg	ggg	acc	ttt	ttg	aac	ttt	gtc	tct	tat	gct	gcc	ggg	gca	1344
Pro	Tyr	Gly	Gly	Thr	Phe	Leu	Asn	Phe	Val	Ser	Tyr	Ala	Ala	Gly	Ala	
		435					440					445				
gtt	aga	tta	tct	gca	ttg	tcg	ggc	cac	cct	gtc	att	tgg	gtt	gct	act	1392
Val	Arg	Leu	Ser	Ala	Leu	Ser	Gly	His	Pro	Val	Ile	Trp	Val	Ala	Thr	
	450					455					460					
cat	gac	tct	att	ggg	ttg	ggg	gaa	gac	ggg	cca	acc	cat	caa	cca	att	1440
His	Asp	Ser	Ile	Gly	Leu	Gly	Glu	Asp	Gly	Pro	Thr	His	Gln	Pro	Ile	
	465				470				475						480	
gaa	acc	ttg	gct	cac	ttt	aga	gct	act	cca	aac	ttg	atg	gtc	tgg	aga	1488
Glu	Thr	Leu	Ala	His	Phe	Arg	Ala	Thr	Pro	Asn	Leu	Met	Val	Trp	Arg	
			485					490						495		
cca	gca	gat	ggg	aat	gaa	gtt	tct	gca	gct	tat	aag	gtt	gcc	ctg	gaa	1536
Pro	Ala	Asp	Gly	Asn	Glu	Val	Ser	Ala	Ala	Tyr	Lys	Val	Ala	Leu	Glu	
			500					505					510			
tct	ctt	gca	act	cct	tcc	atc	att	gca	ttg	acc	aga	caa	aac	ttg	cca	1584
Ser	Leu	Ala	Thr	Pro	Ser	Ile	Ile	Ala	Leu	Thr	Arg	Gln	Asn	Leu	Pro	
		515					520					525				
caa	ttg	gaa	aac	tcc	tct	att	gaa	aag	gca	act	aaa	ggg	ggg	tac	atc	1632
Gln	Leu	Glu	Asn	Ser	Ser	Ile	Glu	Lys	Ala	Thr	Lys	Gly	Gly	Tyr	Ile	
	530					535					540					
tta	aac	gac	att	gat	aat	gcc	aag	ttg	att	att	gcc	gca	aca	ggg	tcc	1680
Leu	Asn	Asp	Ile	Asp	Asn	Ala	Lys	Leu	Ile	Ile	Ala	Ala	Thr	Gly	Ser	
	545				550				555						560	
gaa	gtt	tcc	ctt	gca	gtt	gaa	gct	gct	gac	aaa	tta	acc	agt	gag	ggc	1728
Glu	Val	Ser	Leu	Ala	Val	Glu	Ala	Ala	Asp	Lys	Leu	Thr	Ser	Glu	Gly	
			565					570						575		

ES 2 602 808 T3

att cca acc aag gtt gtc tct gtt cca gac ttc tac acc ttt gac aag Ile Pro Thr Lys Val Val Ser Val Pro Asp Phe Tyr Thr Phe Asp Lys 580 585 590	1776
caa tcc cac gag tac aag atg tct gtc ttc ccg cac ggc att cca gtt Gln Ser His Glu Tyr Lys Met Ser Val Phe Pro His Gly Ile Pro Val 595 600 605	1824
gtt tct ctc gaa gtt atg tca tcc ttt ggt tgg tcc aaa tat gct cat Val Ser Leu Glu Val Met Ser Ser Phe Gly Trp Ser Lys Tyr Ala His 610 615 620	1872
gca cat att tca tta gac aga ttt ggt gca tct gct cca gct gat aaa Ala His Ile Ser Leu Asp Arg Phe Gly Ala Ser Ala Pro Ala Asp Lys 625 630 635 640	1920
ttg ttt gaa aag ttt ggt ttc acc aag gac gtt gtt gct gac aag gca Leu Phe Glu Lys Phe Gly Phe Thr Lys Asp Val Val Ala Asp Lys Ala 645 650 655	1968
aag aag gtg gtc tcg tat tac gcg ggc aag gag gtc ttg tcc caa ttg Lys Lys Val Val Ser Tyr Tyr Ala Gly Lys Glu Val Leu Ser Gln Leu 660 665 670	2016
gaa act cct ttt taa Glu Thr Pro Phe 675	2031

<210> 46
<211> 676
<212> PRT
<213> Issatchenkia orientalis

<400> 46

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

ES 2 602 808 T3

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

ES 2 602 808 T3

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

Thr Arg Trp Gly Gly Ala Val Asp Phe Gln Pro Pro Gln Thr Gly Leu
385 390 395 400

Gly Asp Tyr Ala Gly Arg Tyr Ile Arg Phe Gly Val Arg Glu His Gly
405 410 415

Met Gly Ala Ile Met Asn Gly Ile Ala Ala Tyr Gly Ala Asn Tyr Lys
420 425 430

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

Val Arg Leu Ser Ala Leu Ser Gly His Pro Val Ile Trp Val Ala Thr
450 455 460

His Asp Ser Ile Gly Leu Gly Glu Asp Gly Pro Thr His Gln Pro Ile
465 470 475 480

Glu Thr Leu Ala His Phe Arg Ala Thr Pro Asn Leu Met Val Trp Arg
485 490 495

Pro Ala Asp Gly Asn Glu Val Ser Ala Ala Tyr Lys Val Ala Leu Glu
500 505 510

Ser Leu Ala Thr Pro Ser Ile Ile Ala Leu Thr Arg Gln Asn Leu Pro
515 520 525

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

Leu Asn Asp Ile Asp Asn Ala Lys Leu Ile Ile Ala Ala Thr Gly Ser
545 550 555 560

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

Ile Pro Thr Lys Val Val Ser Val Pro Asp Phe Tyr Thr Phe Asp Lys
580 585 590

Gln Ser His Glu Tyr Lys Met Ser Val Phe Pro His Gly Ile Pro Val
595 600 605

Val Ser Leu Glu Val Met Ser Ser Phe Gly Trp Ser Lys Tyr Ala His

ES 2 602 808 T3

610

615

620

Ala His Ile Ser Leu Asp Arg Phe Gly Ala Ser Ala Pro Ala Asp Lys
625 630 635 640

Leu Phe Glu Lys Phe Gly Phe Thr Lys Asp Val Val Ala Asp Lys Ala
645 650 655

Lys Lys Val Val Ser Tyr Tyr Ala Gly Lys Glu Val Leu Ser Gln Leu
660 665 670

Glu Thr Pro Phe
675

<210> 47

<211> 2043

5 <212> DNA

<213> Saccharomyces cerevisiae

<220>

<221> CDS

10 <222> (1)..(2043)

<400> 47

atg act caa ttc act gac att gat aag cta gcc gtc tcc acc ata aga 48
Met Thr Gln Phe Thr Asp Ile Asp Lys Leu Ala Val Ser Thr Ile Arg
1 5 10 15

att ttg gct gtg gac acc gta tcc aag gcc aac tca ggt cac cca ggt 96
Ile Leu Ala Val Asp Thr Val Ser Lys Ala Asn Ser Gly His Pro Gly
20 25 30

gct cca ttg ggt atg gca cca gct gca cac gtt cta tgg agt caa atg 144
Ala Pro Leu Gly Met Ala Pro Ala Ala His Val Leu Trp Ser Gln Met
35 40 45

cgc atg aac cca acc aac cca gac tgg atc aac aga gat aga ttt gtc 192
Arg Met Asn Pro Thr Asn Pro Asp Trp Ile Asn Arg Asp Arg Phe Val
50 55 60

ttg tct aac ggt cac gcg gtc gct ttg ttg tat tct atg cta cat ttg 240
Leu Ser Asn Gly His Ala Val Ala Leu Leu Tyr Ser Met Leu His Leu
65 70 75 80

act ggt tac gat ctg tct att gaa gac ttg aaa cag ttc aga cag ttg 288
Thr Gly Tyr Asp Leu Ser Ile Glu Asp Leu Lys Gln Phe Arg Gln Leu
85 90 95

ggt tcc aga aca cca ggt cat cct gaa ttt gag ttg cca ggt gtt gaa 336
Gly Ser Arg Thr Pro Gly His Pro Glu Phe Glu Leu Pro Gly Val Glu
100 105 110

gtt act acc ggt cca tta ggt caa ggt atc tcc aac gct gtt ggt atg 384
Val Thr Thr Gly Pro Leu Gly Gln Gly Ile Ser Asn Ala Val Gly Met
115 120 125

gcc atg gct caa gct aac ctg gct gcc act tac aac aag ccg ggc ttt 432

ES 2 602 808 T3

Ala Met Ala Gln Ala Asn Leu Ala Ala Thr Tyr Asn Lys Pro Gly Phe 130 135 140	
acc ttg tct gac aac tac acc tat gtt ttc ttg ggt gac ggt tgt ttg Thr Leu Ser Asp Asn Tyr Thr Tyr Val Phe Leu Gly Asp Gly Cys Leu 145 150 155 160	480
caa gaa ggt att tct tca gaa gct tcc tcc ttg gct ggt cat ttg aaa Gln Glu Gly Ile Ser Ser Glu Ala Ser Ser Leu Ala Gly His Leu Lys 165 170 175	528
ttg ggt aac ttg att gcc atc tac gat gac aac aag atc act atc gat Leu Gly Asn Leu Ile Ala Ile Tyr Asp Asp Asn Lys Ile Thr Ile Asp 180 185 190	576
ggg gct acc agt atc tca ttc gat gaa gat gtt gct aag aga tac gaa Gly Ala Thr Ser Ile Ser Phe Asp Glu Asp Val Ala Lys Arg Tyr Glu 195 200 205	624
gcc tac ggt tgg gaa gtt ttg tac gta gaa aat ggt aac gaa gat cta Ala Tyr Gly Trp Glu Val Leu Tyr Val Glu Asn Gly Asn Glu Asp Leu 210 215 220	672
gcc ggt att gcc aag gct att gct caa gct aag tta tcc aag gac aaa Ala Gly Ile Ala Lys Ala Ile Ala Gln Ala Lys Leu Ser Lys Asp Lys 225 230 235 240	720
cca act ttg atc aaa atg acc aca acc att ggt tac ggt tcc ttg cat Pro Thr Leu Ile Lys Met Thr Thr Thr Ile Gly Tyr Gly Ser Leu His 245 250 255	768
gcc ggc tct cac tct gtg cac ggt gcc cca ttg aaa gca gat gat gtt Ala Gly Ser His Ser Val His Gly Ala Pro Leu Lys Ala Asp Asp Val 260 265 270	816
aaa caa cta aag agc aaa ttc ggt ttc aac cca gac aag tcc ttt gtt Lys Gln Leu Lys Ser Lys Phe Gly Phe Asn Pro Asp Lys Ser Phe Val 275 280 285	864
gtt cca caa gaa gtt tac gac cac tac caa aag aca att tta aag cca Val Pro Gln Glu Val Tyr Asp His Tyr Gln Lys Thr Ile Leu Lys Pro 290 295 300	912
ggg gtc gaa gcc aac aac aag tgg aac aag ttg ttc agc gaa tac caa Gly Val Glu Ala Asn Asn Lys Trp Asn Lys Leu Phe Ser Glu Tyr Gln 305 310 315 320	960
aag aaa ttc cca gaa tta ggt gct gaa ttg gct aga aga ttg agc ggc Lys Lys Phe Pro Glu Leu Gly Ala Glu Leu Ala Arg Arg Leu Ser Gly 325 330 335	1008
caa cta ccc gca aat tgg gaa tct aag ttg cca act tac acc gcc aag Gln Leu Pro Ala Asn Trp Glu Ser Lys Leu Pro Thr Tyr Thr Ala Lys 340 345 350	1056
gac tct gcc gtg gcc act aga aaa tta tca gaa act gtt ctt gag gat Asp Ser Ala Val Ala Thr Arg Lys Leu Ser Glu Thr Val Leu Glu Asp 355 360 365	1104
gtt tac aat caa ttg cca gag ttg att ggt ggt tct gcc gat tta aca Val Tyr Asn Gln Leu Pro Glu Leu Ile Gly Gly Ser Ala Asp Leu Thr 370 375 380	1152

ES 2 602 808 T3

cct tct aac ttg acc aga tgg aag gaa gcc ctt gac ttc caa cct cct	1200
Pro Ser Asn Leu Thr Arg Trp Lys Glu Ala Leu Asp Phe Gln Pro Pro	
385 390 395 400	
tct tcc ggt tca ggt aac tac tct ggt aga tac att agg tac ggt att	1248
Ser Ser Gly Ser Gly Asn Tyr Ser Gly Arg Tyr Ile Arg Tyr Gly Ile	
405 410 415	
aga gaa cac gct atg ggt gcc ata atg aac ggt att tca gct ttc ggt	1296
Arg Glu His Ala Met Gly Ala Ile Met Asn Gly Ile Ser Ala Phe Gly	
420 425 430	
gcc aac tac aaa cca tac ggt ggt act ttc ttg aac ttc gtt tct tat	1344
Ala Asn Tyr Lys Pro Tyr Gly Gly Thr Phe Leu Asn Phe Val Ser Tyr	
435 440 445	
gct gct ggt gcc gtt aga ttg tcc gct ttg tct ggc cac cca gtt att	1392
Ala Ala Gly Ala Val Arg Leu Ser Ala Leu Ser Gly His Pro Val Ile	
450 455 460	
tgg gtt gct aca cat gac tct atc ggt gtc ggt gaa gat ggt cca aca	1440
Trp Val Ala Thr His Asp Ser Ile Gly Val Gly Glu Asp Gly Pro Thr	
465 470 475 480	
cat caa cct att gaa act tta gca cac ttc aga tcc cta cca aac att	1488
His Gln Pro Ile Glu Thr Leu Ala His Phe Arg Ser Leu Pro Asn Ile	
485 490 495	
caa gtt tgg aga cca gct gat ggt aac gaa gtt tct gcc gcc tac aag	1536
Gln Val Trp Arg Pro Ala Asp Gly Asn Glu Val Ser Ala Ala Tyr Lys	
500 505 510	
aac tct tta gaa tcc aag cat act cca agt atc att gct ttg tcc aga	1584
Asn Ser Leu Glu Ser Lys His Thr Pro Ser Ile Ile Ala Leu Ser Arg	
515 520 525	
caa aac ttg cca caa ttg gaa ggt agc tct att gaa agc gct tct aag	1632
Gln Asn Leu Pro Gln Leu Gly Ser Ser Ile Glu Ser Ala Ser Lys	
530 535 540	
ggt ggt tac gta cta caa gat gtt gct aac cca gat att att tta gtg	1680
Gly Gly Tyr Val Leu Gln Asp Val Ala Asn Pro Asp Ile Ile Leu Val	
545 550 555 560	
gct act ggt tcc gaa gtg tct ttg agt gtt gaa gct gct aag act ttg	1728
Ala Thr Gly Ser Glu Val Ser Leu Ser Val Glu Ala Ala Lys Thr Leu	
565 570 575	
gcc gca aag aac atc aag gct cgt gtt gtt tct cta cca gat ttc ttc	1776
Ala Ala Lys Asn Ile Lys Ala Arg Val Val Ser Leu Pro Asp Phe Phe	
580 585 590	
act ttt gac aaa caa ccc cta gaa tac aga cta tca gtc tta cca gac	1824
Thr Phe Asp Lys Gln Pro Leu Glu Tyr Arg Leu Ser Val Leu Pro Asp	
595 600 605	
aac gtt cca atc atg tct gtt gaa gtt ttg gct acc aca tgt tgg ggc	1872
Asn Val Pro Ile Met Ser Val Glu Val Leu Ala Thr Thr Cys Trp Gly	
610 615 620	
aaa tac gct cat caa tcc ttc ggt att gac aga ttt ggt gcc tcc ggt	1920
Lys Tyr Ala His Gln Ser Phe Gly Ile Asp Arg Phe Gly Ala Ser Gly	
625 630 635 640	

ES 2 602 808 T3

aag gca cca gaa gtc ttc aag ttc ttc ggt ttc acc cca gaa ggt gtt 1968
 Lys Ala Pro Glu Val Phe Lys Phe Phe Gly Phe Thr Pro Glu Gly Val
 645 650 655

gct gaa aga gct caa aag acc att gca ttc tat aag ggt gac aag cta 2016
 Ala Glu Arg Ala Gln Lys Thr Ile Ala Phe Tyr Lys Gly Asp Lys Leu
 660 665 670

att tct cct ttg aaa aaa gct ttc taa 2043
 Ile Ser Pro Leu Lys Lys Ala Phe
 675 680

<210> 48
 <211> 680
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 48

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

ES 2 602 808 T3

Leu Gly Asn Leu Ile Ala Ile Tyr Asp Asp Asn Lys Ile Thr Ile Asp
 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

ES 2 602 808 T3

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		
450	455	460
Trp Val Ala Thr His Asp Ser Ile Gly Val Gly Glu Asp Gly Pro Thr		
465	470	475
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
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
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	

5 <210> 49
 <211> 2040
 <212> DNA
 <213> Kluyveromyces marxianus

10 <220>

<221> CDS

<222> (1)..(2040)

<400> 49

atg tct caa tat tcc gat atc gat cgt cta gct atc aac act atc aga	48
Met Ser Gln Tyr Ser Asp Ile Asp Arg Leu Ala Ile Asn Thr Ile Arg	
1 5 10 15	
ttg ttg tct gtt gac caa gtt gca tca gcc aac tct ggt cac cca ggt	96
Leu Leu Ser Val Asp Gln Val Ala Ser Ala Asn Ser Gly His Pro Gly	
20 25 30	
gct cca ttg ggt cta gct cca gca gcc cat gtt gtc tgg aag caa atg	144
Ala Pro Leu Gly Leu Ala Pro Ala Ala His Val Val Trp Lys Gln Met	
35 40 45	
aga ttg aac cca aag aac cca gaa tgg atc aac cgt gac aga ttt gtt	192
Arg Leu Asn Pro Lys Asn Pro Glu Trp Ile Asn Arg Asp Arg Phe Val	
50 55 60	
ttg tct aac ggt cac gca tgt gct ttg ttg tac tcc ttg ttg cat ctg	240
Leu Ser Asn Gly His Ala Cys Ala Leu Leu Tyr Ser Leu Leu His Leu	
65 70 75 80	
ttt gga tac gac ttc tcc att gag gat ttg aag caa ttc aga cac ttg	288
Phe Gly Tyr Asp Phe Ser Ile Glu Asp Leu Lys Gln Phe Arg His Leu	
85 90 95	
ggc tcc aag act cca ggt cac cca gaa ttc gaa ttg cca ggt gtc gaa	336
Gly Ser Lys Thr Pro Gly His Pro Glu Phe Glu Leu Pro Gly Val Glu	
100 105 110	
gtg act act ggt cca ttg ggt caa ggt atc tct aac gct gtt ggt ttg	384
Val Thr Thr Gly Pro Leu Gly Gln Gly Ile Ser Asn Ala Val Gly Leu	
115 120 125	
gct att gct cag gcc aac ttt gct gcc act tac aac aag cca gac ttt	432
Ala Ile Ala Gln Ala Asn Phe Ala Ala Thr Tyr Asn Lys Pro Asp Phe	
130 135 140	
gaa ttg tct gac tca ttc aca tac gtg ttt ttg ggt gac ggt tgt ttg	480
Glu Leu Ser Asp Ser Phe Thr Tyr Val Phe Leu Gly Asp Gly Cys Leu	
145 150 155 160	
caa gaa ggt gtt tca tct gaa gct tgt tct ttg gct ggc cat ttg aaa	528
Gln Glu Gly Val Ser Ser Glu Ala Cys Ser Leu Ala Gly His Leu Lys	
165 170 175	
ttg aag aac ttg att gcc ttt tac gac gac aac aag atc acc atc gat	576
Leu Lys Asn Leu Ile Ala Phe Tyr Asp Asp Asn Lys Ile Thr Ile Asp	
180 185 190	

ggt aac acc aac gtg tct ttc gac gaa gat gtt ggc aag aga tac gag Gly Asn Thr Asn Val Ser Phe Asp Glu Asp Val Gly Lys Arg Tyr Glu 195 200 205	624
gcc tac ggc tgg gaa gtg ttg agc gtc gaa aac ggt aac gac gat ctg Ala Tyr Gly Trp Glu Val Leu Ser Val Glu Asn Gly Asn Asp Asp Leu 210 215 220	672
gac tcc atc agc aag gcc ttg gag caa gcc aag cgc tcc gac aag cca Asp Ser Ile Ser Lys Ala Leu Glu Gln Ala Lys Arg Ser Asp Lys Pro 225 230 235 240	720
act ttg atc aag ttg gtc acc acc atc ggt ttc ggt tcc cta caa gcc Thr Leu Ile Lys Leu Val Thr Thr Ile Gly Phe Gly Ser Leu Gln Ala 245 250 255	768
ggt acc cac gcc gtc cac ggt gcc cca ttg aag gcc gac gat atc aag Gly Thr His Ala Val His Gly Ala Pro Leu Lys Ala Asp Asp Ile Lys 260 265 270	816
caa ttg aag acc aag ttc ggc ttc aac cca gaa gaa tcc ttc gtc gtc Gln Leu Lys Thr Lys Phe Gly Phe Asn Pro Glu Glu Ser Phe Val Val 275 280 285	864
cca caa gaa gtc tac gac cta tac aac aag tcc acc atc gaa cct ggt Pro Gln Glu Val Tyr Asp Leu Tyr Asn Lys Ser Thr Ile Glu Pro Gly 290 295 300	912
atc gaa gcc aac aac aag tgg aac gct cta ttc gag gcc tac tcc gcc Ile Glu Ala Asn Asn Lys Trp Asn Ala Leu Phe Glu Ala Tyr Ser Ala 305 310 315 320	960
aag ttc cca gaa ttg ggt gcc gaa atc aag aga aga ctt gca ggt gaa Lys Phe Pro Glu Leu Gly Ala Glu Ile Lys Arg Arg Leu Ala Gly Glu 325 330 335	1008
ttg cca act ggt tgg gaa aac aag ttg cca act tac act cca aag gac Leu Pro Thr Gly Trp Glu Asn Lys Leu Pro Thr Tyr Thr Pro Lys Asp 340 345 350	1056
tcc gcc gtc gct tcc aga aag ttg tcc gag atc cta ttg caa aac atc Ser Ala Val Ala Ser Arg Lys Leu Ser Glu Ile Leu Leu Gln Asn Ile 355 360 365	1104
ttc gaa gac att cca gaa ttg att ggt ggt tcc gcc gat ttg acc cca Phe Glu Asp Ile Pro Glu Leu Ile Gly Gly Ser Ala Asp Leu Thr Pro 370 375 380	1152
tct aac ttg act aga acc aag gaa gtt gtc gac ttc caa gct cct tcc Ser Asn Leu Thr Arg Thr Lys Glu Val Val Asp Phe Gln Ala Pro Ser 385 390 395 400	1200
tcc ggt cta ggt gac tac acc ggt aga tac atc aga tac ggt gtc cgt Ser Gly Leu Gly Asp Tyr Thr Gly Arg Tyr Ile Arg Tyr Gly Val Arg 405 410 415	1248
gaa cac ggt atg ggt gcc atc atg aac ggt atc tcc gct ttc ggt gct Glu His Gly Met Gly Ala Ile Met Asn Gly Ile Ser Ala Phe Gly Ala 420 425 430	1296
aac tac aag cca tac ggt ggt act ttc ttg aac ttc gtc tct tac gcc Asn Tyr Lys Pro Tyr Gly Gly Thr Phe Leu Asn Phe Val Ser Tyr Ala 435 440 445	1344

ES 2 602 808 T3

gct ggt gcc ttg aga ttg tcc gct ttg tcc ggt cac cca gtc atc tgg Ala Gly Ala Leu Arg Leu Ser Ala Leu Ser Gly His Pro Val Ile Trp 450 455 460	1392
gtc gcc act cac gac tcc atc ggt cta ggt gaa gat ggt cca acc cat Val Ala Thr His Asp Ser Ile Gly Leu Gly Asp Gly Pro Thr His 465 470 475 480	1440
caa cct atc gaa act ttg gcc cac ttc aga gct ctt cca aac ttg caa Gln Pro Ile Glu Thr Leu Ala His Phe Arg Ala Leu Pro Asn Leu Gln 485 490 495	1488
gtt tgg aga cca gct gac ggt aac gaa gtt tcc gct gct tac aag gtc Val Trp Arg Pro Ala Asp Gly Asn Glu Val Ser Ala Ala Tyr Lys Val 500 505 510	1536
gcc ttg aag agc aag cac act cca gcc gtc att gct cta tct aga caa Ala Leu Lys Ser Lys His Thr Pro Ala Val Ile Ala Leu Ser Arg Gln 515 520 525	1584
aac ttg cct caa ttg gaa ggc tct tcc atc gaa aag gcc gcc aag ggt Asn Leu Pro Gln Leu Glu Gly Ser Ser Ile Glu Lys Ala Ala Lys Gly 530 535 540	1632
ggt tac gtc ttg caa gac gtc gaa caa cca gac atc gcc atc gtc tcc Gly Tyr Val Leu Gln Asp Val Glu Gln Pro Asp Ile Ala Ile Val Ser 545 550 555 560	1680
act ggt tcc gaa gtt ggt att gcc gtc gaa gct gct aag gtt ttg gct Thr Gly Ser Glu Val Gly Ile Ala Val Glu Ala Ala Lys Val Leu Ala 565 570 575	1728
gaa aag aac atc aag gcc cgt atc gtc tct cta cca gac ttc cac acc Glu Lys Asn Ile Lys Ala Arg Ile Val Ser Leu Pro Asp Phe His Thr 580 585 590	1776
ttt gac caa caa cca aag gaa tac caa ttg tcc gtt cta cca gat ggt Phe Asp Gln Gln Pro Lys Glu Tyr Gln Leu Ser Val Leu Pro Asp Gly 595 600 605	1824
gtc cca atc ttg tcc gtc gaa gtc ttg tcc act tcc ggt tgg gct aag Val Pro Ile Leu Ser Val Glu Val Leu Ser Thr Ser Gly Trp Ala Lys 610 615 620	1872
tac tct cac caa caa ttc ggt ttg aac aga ttc ggt gcc tcc ggt aag Tyr Ser His Gln Gln Phe Gly Leu Asn Arg Phe Gly Ala Ser Gly Lys 625 630 635 640	1920
ggt cca gcc gtc tac gag aag ttc gat ttc act cca caa ggt att gct Gly Pro Ala Val Tyr Glu Lys Phe Asp Phe Thr Pro Gln Gly Ile Ala 645 650 655	1968
tcc aga gct gaa aag act gtt gaa ttc tac aag ggt aag caa gtc tat Ser Arg Ala Glu Lys Thr Val Glu Phe Tyr Lys Gly Lys Gln Val Tyr 660 665 670	2016
tct cct ttg aac act gct ttc taa Ser Pro Leu Asn Thr Ala Phe 675	2040

<210> 50

<211> 679

<212> PRT

<213> Kluyveromyces marxianus

<400> 50

ES 2 602 808 T3

Met Ser Gln Tyr Ser Asp Ile Asp Arg Leu Ala Ile Asn Thr Ile Arg
1 5 10 15

Leu Leu Ser Val Asp Gln Val Ala Ser Ala Asn Ser Gly His Pro Gly
20 25 30

Ala Pro Leu Gly Leu Ala Pro Ala Ala His Val Val Trp Lys Gln Met
35 40 45

Arg Leu Asn Pro Lys Asn Pro Glu Trp Ile Asn Arg Asp Arg Phe Val
50 55 60

Leu Ser Asn Gly His Ala Cys Ala Leu Leu Tyr Ser Leu Leu His Leu
65 70 75 80

Phe Gly Tyr Asp Phe Ser Ile Glu Asp Leu Lys Gln Phe Arg His Leu
85 90 95

Gly Ser Lys 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 Leu
115 120 125

Ala Ile Ala Gln Ala Asn Phe Ala Ala Thr Tyr Asn Lys Pro Asp Phe
130 135 140

Glu Leu Ser Asp Ser Phe Thr Tyr Val Phe Leu Gly Asp Gly Cys Leu
145 150 155 160

Gln Glu Gly Val Ser Ser Glu Ala Cys Ser Leu Ala Gly His Leu Lys
165 170 175

Leu Lys Asn Leu Ile Ala Phe Tyr Asp Asp Asn Lys Ile Thr Ile Asp
180 185 190

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

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

Asp Ser Ile Ser Lys Ala Leu Glu Gln Ala Lys Arg Ser Asp Lys Pro

ES 2 602 808 T3

225		230		235		240
Thr Leu Ile Lys	Leu Val Thr Thr Ile Gly Phe Gly Ser Leu Gln Ala					
	245			250		255
Gly Thr His Ala Val His Gly Ala Pro Leu Lys Ala Asp Asp Ile Lys						
	260			265		270
Gln Leu Lys Thr Lys Phe Gly Phe Asn Pro Glu Glu Ser Phe Val Val						
	275			280		285
Pro Gln Glu Val Tyr Asp Leu Tyr Asn Lys Ser Thr Ile Glu Pro Gly						
	290			295		300
Ile Glu Ala Asn Asn Lys Trp Asn Ala Leu Phe Glu Ala Tyr Ser Ala						
305		310		315		320
Lys Phe Pro Glu Leu Gly Ala Glu Ile Lys Arg Arg Leu Ala Gly Glu						
	325			330		335
Leu Pro Thr Gly Trp Glu Asn Lys Leu Pro Thr Tyr Thr Pro Lys Asp						
	340			345		350
Ser Ala Val Ala Ser Arg Lys Leu Ser Glu Ile Leu Leu Gln Asn Ile						
	355			360		365
Phe Glu Asp Ile Pro Glu Leu Ile Gly Gly Ser Ala Asp Leu Thr Pro						
	370			375		380
Ser Asn Leu Thr Arg Thr Lys Glu Val Val Asp Phe Gln Ala Pro Ser						
385		390		395		400
Ser Gly Leu Gly Asp Tyr Thr Gly Arg Tyr Ile Arg Tyr Gly Val Arg						
	405			410		415
Glu His Gly Met Gly Ala Ile Met Asn Gly Ile Ser Ala Phe Gly Ala						
	420			425		430
Asn Tyr Lys Pro Tyr Gly Gly Thr Phe Leu Asn Phe Val Ser Tyr Ala						
	435			440		445
Ala Gly Ala Leu Arg Leu Ser Ala Leu Ser Gly His Pro Val Ile Trp						
	450			455		460
Val Ala Thr His Asp Ser Ile Gly Leu Gly Glu Asp Gly Pro Thr His						
465		470		475		480

ES 2 602 808 T3

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

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

Ala Leu Lys Ser Lys His Thr Pro Ala Val Ile Ala Leu Ser Arg Gln
515 520 525

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

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

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

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

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

Val Pro Ile Leu Ser Val Glu Val Leu Ser Thr Ser Gly Trp Ala Lys
610 615 620

Tyr Ser His Gln Gln Phe Gly Leu Asn Arg Phe Gly Ala Ser Gly Lys
625 630 635 640

Gly Pro Ala Val Tyr Glu Lys Phe Asp Phe Thr Pro Gln Gly Ile Ala
645 650 655

Ser Arg Ala Glu Lys Thr Val Glu Phe Tyr Lys Gly Lys Gln Val Tyr
660 665 670

Ser Pro Leu Asn Thr Ala Phe
675

<210> 51
<211> 972
5 <212> DNA
<213> Issatchenkia orientalis

<220>
<221> CDS
10 <222> (1)..(972)

<400> 51

ES 2 602 808 T3

atg tct caa tta gaa caa ctt aag aag tca ggt acc act gtt gtt gca Met Ser Gln Leu Glu Gln Leu Lys Lys Ser Gly Thr Thr Val Val Ala 1 5 10 15	48
gat act ggt gat ttt gaa tct att gca aag ttc act cca caa gat gca Asp Thr Gly Asp Phe Glu Ser Ile Ala Lys Phe Thr Pro Gln Asp Ala 20 25 30	96
act acg aat cca tcg ttg att ttg gca gct act aag caa cca caa tac Thr Thr Asn Pro Ser Leu Ile Leu Ala Ala Thr Lys Gln Pro Gln Tyr 35 40 45	144
gcc aag tta att gac att gct att gac tat gcc aag agc aag ggt tcc Ala Lys Leu Ile Asp Ile Ala Ile Asp Tyr Ala Lys Ser Lys Gly Ser 50 55 60	192
acc aag gag gaa cag gct agt cta gca ctt gac aga ttg ttg gtt gaa Thr Lys Glu Glu Gln Ala Ser Leu Ala Leu Asp Arg Leu Leu Val Glu 65 70 75 80	240
ttt ggt gct gag att ctc aag att gtg cca ggt cgg gtg tcc acc gag Phe Gly Ala Glu Ile Leu Lys Ile Val Pro Gly Arg Val Ser Thr Glu 85 90 95	288
gtt gat gca agg tta tcc ttc aac agg gac gcc aca att aag aag gcc Val Asp Ala Arg Leu Ser Phe Asn Arg Asp Ala Thr Ile Lys Lys Ala 100 105 110	336
att gag att att gaa tta tac aag gca cag gga atc gac aag gag aga Ile Glu Ile Ile Glu Leu Tyr Lys Ala Gln Gly Ile Asp Lys Glu Arg 115 120 125	384
gta ttg atc aag att gcc tcc act tgg gaa ggt atc caa gct gca agg Val Leu Ile Lys Ile Ala Ser Thr Trp Glu Gly Ile Gln Ala Ala Arg 130 135 140	432
gag ttg gaa tcc aag cat ggt atc cac tgt aat ttg acg ctt ttg ttc Glu Leu Glu Ser Lys His Gly Ile His Cys Asn Leu Thr Leu Leu Phe 145 150 155 160	480
tcc ttt gtt cag gct gtt gca tgt gca gaa gct aat gtc aca ctt att Ser Phe Val Gln Ala Val Ala Cys Ala Glu Ala Asn Val Thr Leu Ile 165 170 175	528
tct cca ttt gtc ggt aga att ttg gac tgg tac aag gct tcc act ggt Ser Pro Phe Val Gly Arg Ile Leu Asp Trp Tyr Lys Ala Ser Thr Gly 180 185 190	576
aag gaa tac acc tct gaa acc gat cct ggt gtt ttg tcc gtt aga aac Lys Glu Tyr Thr Ser Glu Thr Asp Pro Gly Val Leu Ser Val Arg Asn 195 200 205	624
atc ttc aac tac tac aag aag ttt ggc tac aag aca att gtc atg ggt Ile Phe Asn Tyr Tyr Lys Lys Phe Gly Tyr Lys Thr Ile Val Met Gly 210 215 220	672
gcc tcc ttc aga aac act ggt gaa att gca gct ttg gca ggt tgt gac Ala Ser Phe Arg Asn Thr Gly Glu Ile Ala Ala Leu Ala Gly Cys Asp 225 230 235 240	720
tac ttg acc att tct cca agt ttg ttg gat aaa ctg gct aat tcc aat Tyr Leu Thr Ile Ser Pro Ser Leu Leu Asp Lys Leu Ala Asn Ser Asn 245 250 255	768

ES 2 602 808 T3

gac cca ctt cca aag gtt tta gac gct tct aag gca aag gaa ttg gat	816
Asp Pro Leu Pro Lys Val Leu Asp Ala Ser Lys Ala Lys Glu Leu Asp	
260 265 270	
ctt gaa aaa gtc tcc tat gtt gac gat gag cca gat ttt aga ttc ctc	864
Leu Glu Lys Val Ser Tyr Val Asp Asp Glu Pro Asp Phe Arg Phe Leu	
275 280 285	
ttg aat gag gat gca atg gca act gaa aaa ttg tct gaa ggt atc aga	912
Leu Asn Glu Asp Ala Met Ala Thr Glu Lys Leu Ser Glu Gly Ile Arg	
290 295 300	
aag ttc tct gct gat tgt gag gct ctt tac aat gaa tta tta aag aga	960
Lys Phe Ser Ala Asp Cys Glu Ala Leu Tyr Asn Glu Leu Leu Lys Arg	
305 310 315 320	
gtt tct gct tga	972
Val Ser Ala	

<210> 52
 <211> 323
 <212> PRT
 <213> Issatchenkia orientalis

<400> 52

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

ES 2 602 808 T3

Glu Leu Glu Ser Lys His Gly Ile His Cys Asn Leu Thr Leu Leu Phe
 145 150 155 160
 Ser Phe Val Gln Ala Val Ala Cys Ala Glu Ala Asn Val Thr Leu Ile
 165 170 175
 Ser Pro Phe Val Gly Arg Ile Leu Asp Trp Tyr Lys Ala Ser Thr Gly
 180 185 190
 Lys Glu Tyr Thr Ser Glu Thr Asp Pro Gly Val Leu Ser Val Arg Asn
 195 200 205
 Ile Phe Asn Tyr Tyr Lys Lys Phe Gly Tyr Lys Thr Ile Val Met Gly
 210 215 220
 Ala Ser Phe Arg Asn Thr Gly Glu Ile Ala Ala Leu Ala Gly Cys Asp
 225 230 235 240
 Tyr Leu Thr Ile Ser Pro Ser Leu Leu Asp Lys Leu Ala Asn Ser Asn
 245 250 255
 Asp Pro Leu Pro Lys Val Leu Asp Ala Ser Lys Ala Lys Glu Leu Asp
 260 265 270
 Leu Glu Lys Val Ser Tyr Val Asp Asp Glu Pro Asp Phe Arg Phe Leu
 275 280 285
 Leu Asn Glu Asp Ala Met Ala Thr Glu Lys Leu Ser Glu Gly Ile Arg
 290 295 300
 Lys Phe Ser Ala Asp Cys Glu Ala Leu Tyr Asn Glu Leu Leu Lys Arg
 305 310 315 320
 Val Ser Ala

<210> 53
 <211> 1008
 5 <212> DNA
 <213> *Saccharomyces cerevisiae*

<220>
 <221> CDS
 10 <222> (1)..(1008)

<400> 53
 atg tct gaa cca gct caa aag aaa caa aag gtt gct aac aac tct cta
 Met Ser Glu Pro Ala Gln Lys Lys Gln Lys Val Ala Asn Asn Ser Leu

48

ES 2 602 808 T3

1	5	10	15	
gaa caa ttg aaa gcc tcc ggc act gtc gtt gtt gcc gac act ggt gat				96
Glu Gln Leu Lys Ala Ser Gly Thr Val Val Val Ala Asp Thr Gly Asp	20	25	30	
ttc ggc tct att gcc aag ttt caa cct caa gac tcc aca act aac cca				144
Phe Gly Ser Ile Ala Lys Phe Gln Pro Gln Asp Ser Thr Asn Pro	35	40	45	
tca ttg atc ttg gct gct gcc aag caa cca act tac gcc aag ttg atc				192
Ser Leu Ile Leu Ala Ala Tyr Gln Lys Gln Pro Thr Tyr Ala Lys Leu Ile	50	55	60	
gat gtt gcc gtg gaa tac ggt aag aag cat ggt aag acc acc gaa gaa				240
Asp Val Ala Val Glu Tyr Gly Lys Lys His Gly Lys Thr Thr Glu Glu	65	70	75	80
caa gtc gaa aat gct gtg gac aga ttg tta gtc gaa ttc ggt aag gag				288
Gln Val Glu Asn Val Asp Arg Leu Val Glu Phe Gly Lys Glu	85	90	95	
atc tta aag att gtt cca ggc aga gtc tcc acc gaa gtt gat gct aga				336
Ile Leu Lys Ile Val Pro Gly Arg Val Ser Thr Glu Val Asp Ala Arg	100	105	110	
ttg tct ttt gac act caa gct acc att gaa aag gct aga cat atc att				384
Leu Ser Phe Asp Thr Gln Ala Thr Ile Glu Lys Ala Arg His Ile Ile	115	120	125	
aaa ttg ttt gaa caa gaa ggt gtc tcc aag gaa aga gtc ctt att aaa				432
Lys Leu Phe Glu Gln Glu Gly Val Ser Lys Glu Arg Val Leu Ile Lys	130	135	140	
att gct tcc act tgg gaa ggt att caa gct gcc aaa gaa ttg gaa gaa				480
Ile Ala Ser Thr Trp Glu Gly Ile Gln Ala Ala Lys Glu Leu Glu Glu	145	150	155	160
aag gac ggt atc cac tgt aat ttg act cta tta ttc tcc ttc gtt caa				528
Lys Asp Gly Ile His Cys Asn Leu Thr Leu Leu Phe Ser Phe Val Gln	165	170	175	
gca gtt gcc tgt gcc gag gcc caa gtt act ttg att tcc cca ttt gtt				576
Ala Val Ala Cys Ala Glu Ala Gln Val Thr Leu Ile Ser Pro Phe Val	180	185	190	
ggt aga att cta gac tgg tac aaa tcc agc act ggt aaa gat tac aag				624
Gly Arg Ile Leu Asp Trp Tyr Lys Ser Ser Thr Gly Lys Asp Tyr Lys	195	200	205	
ggt gaa gcc gac cca ggt gtt att tcc gtc aag aaa atc tac aac tac				672
Gly Glu Ala Asp Pro Gly Val Ile Ser Val Lys Lys Ile Tyr Asn Tyr	210	215	220	
tac aag aag tac ggt tac aag act att gtt atg ggt gct tct ttc aga				720
Tyr Lys Lys Tyr Gly Tyr Lys Thr Ile Val Met Gly Ala Ser Phe Arg	225	230	235	240
agc act gac gaa atc aaa aac ttg gct ggt gtt gac tat cta aca att				768
Ser Thr Asp Glu Ile Lys Asn Leu Ala Gly Val Asp Tyr Leu Thr Ile	245	250	255	
tct cca gct tta ttg gac aag ttg atg aac agt act gaa cct ttc cca				816

ES 2 602 808 T3

```

Ser Pro Ala Leu Leu Asp Lys Leu Met Asn Ser Thr Glu Pro Phe Pro
      260                      265                      270

aga gtt ttg gac cct gtc tcc gct aag aag gaa gcc gcc gac aag att      864
Arg Val Leu Asp Pro Val Ser Ala Lys Lys Glu Ala Gly Asp Lys Ile
      275                      280                      285

tct tac atc agc gac gaa tct aaa ttc aga ttc gac ttg aat gaa gac      912
Ser Tyr Ile Ser Asp Glu Ser Lys Phe Arg Phe Asp Leu Asn Glu Asp
      290                      295                      300

gct atg gcc act gaa aaa ttg tcc gaa ggt atc aga aaa ttc tct gcc      960
Ala Met Ala Thr Glu Lys Leu Ser Glu Gly Ile Arg Lys Phe Ser Ala
      305                      310                      315                      320

gat att gtt act cta ttc gac ttg att gaa aag aaa gtt acc gct taa      1008
Asp Ile Val Thr Leu Phe Asp Leu Ile Glu Lys Lys Val Thr Ala
      325                      330                      335

```

<210> 54
 <211> 335
 5 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 54

```

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

```

ES 2 602 808 T3

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
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> 55

<211> 1005

5 <212> DNA

<213> Kluyveromyces marxianus

<220>

<221> CDS

10 <222> (1)..(1005)

<400> 55

atg tct gaa cca gct gct aag aaa caa aag ttc gcc aac tca ttg gaa
Met Ser Glu Pro Ala Ala Lys Lys Gln Lys Phe Ala Asn Ser Leu Glu
1 5 10 15

48

ES 2 602 808 T3

gcc ttg aag gct acc ggc acc gtt gtt gtt gcc gac act ggt gat ttc Ala Leu Lys Ala Thr Gly Thr Val Val Val Ala Asp Thr Gly Asp Phe 20 25 30	96
gaa tca att gcc aag ttc aca cca caa gac gcc acc acc aac cca tct Glu Ser Ile Ala Lys Phe Thr Pro Gln Asp Ala Thr Thr Asn Pro Ser 35 40 45	144
ttg att ttg gct gct gca aag caa gat gca tac gcc aag ttg atc gat Leu Ile Leu Ala Ala Ala Lys Gln Asp Ala Tyr Ala Lys Leu Ile Asp 50 55 60	192
gct gct gtc gaa tac ggt aag aag cac gga tca aac att gac gag caa Ala Ala Val Glu Tyr Gly Lys Lys His Gly Ser Asn Ile Asp Glu Gln 65 70 75 80	240
gtt gag att tcc gtc gac aag ctt ttg gtc gag ttt ggt act gcc atc Val Glu Ile Ser Val Asp Lys Leu Leu Val Glu Phe Gly Thr Ala Ile 85 90 95	288
ttg aag gtt gtt cca ggc aga gtc tcc acc gaa gtt gat gct aga ttg Leu Lys Val Val Pro Gly Arg Val Ser Thr Glu Val Asp Ala Arg Leu 100 105 110	336
tcc ttc gac aag gaa gcc act gtc aag aag gct ttg gaa atc atc aag Ser Phe Asp Lys Glu Ala Thr Val Lys Lys Ala Leu Glu Ile Ile Lys 115 120 125	384
cta tac gaa gcc gaa ggt atc tcc aag gaa aga gtg cta atc aag att Leu Tyr Glu Ala Glu Gly Ile Ser Lys Glu Arg Val Leu Ile Lys Ile 130 135 140	432
gcc tcc acc tgg gaa ggt atc caa gcc gca caa gag cta gag aag gat Ala Ser Thr Trp Glu Gly Ile Gln Ala Ala Gln Glu Leu Glu Lys Asp 145 150 155 160	480
cac ggt att cac gtc aac ttg acc ttg ttg ttc tct ttc tcc caa gcc His Gly Ile His Val Asn Leu Thr Leu Leu Phe Ser Phe Ser Gln Ala 165 170 175	528
gtt gct gct gct gag gcc aat gtc acc ttg atc tct cca ttt gtt ggt Val Ala Ala Ala Glu Ala Asn Val Thr Leu Ile Ser Pro Phe Val Gly 180 185 190	576
aga atc ttg gac tgg tac aag gcc aag acc ggt gaa act tac act gcc Arg Ile Leu Asp Trp Tyr Lys Ala Lys Thr Gly Glu Thr Tyr Thr Ala 195 200 205	624
gaa acc gac cca ggt gtc gaa tcc gtc aag aac atc tac aac tac tac Glu Thr Asp Pro Gly Val Glu Ser Val Lys Asn Ile Tyr Asn Tyr Tyr 210 215 220	672
aag aag cac ggc tac aag acc atc gtt atg ggt gct tcc ttc aga aac Lys Lys His Gly Tyr Lys Thr Ile Val Met Gly Ala Ser Phe Arg Asn 225 230 235 240	720
gtt ggt gaa atc aag gct ctt gcc ggt gtc gac tac ttg act att tct Val Gly Glu Ile Lys Ala Leu Ala Gly Val Asp Tyr Leu Thr Ile Ser 245 250 255	768
cca aag ttg ttg gac gaa ttg atg gcc tcc cag gac cct gtc cca caa Pro Lys Leu Leu Asp Glu Leu Met Ala Ser Gln Asp Pro Val Pro Gln 260 265 270	816

ES 2 602 808 T3

gtc ttg gac cct gaa tct gcc aag gat caa ggt tcc gga aga gtc tcc 864
Val Leu Asp Pro Glu Ser Ala Lys Asp Gln Gly Ser Gly Arg Val Ser
275 280 285

ttc atc aac gac gaa tcc aag ttc aga ttc gag ttg aac gaa gac gcc 912
Phe Ile Asn Asp Glu Ser Lys Phe Arg Phe Glu Leu Asn Glu Asp Ala
290 295 300

atg gcc act gaa aag ttg tct gaa ggt atc aga aag ttc tct gcc gac 960
Met Ala Thr Glu Lys Leu Ser Glu Gly Ile Arg Lys Phe Ser Ala Asp
305 310 315 320

atc gtc act ctg ttc gac ttg atc aag gct aag atc caa gct taa 1005
Ile Val Thr Leu Phe Asp Leu Ile Lys Ala Lys Ile Gln Ala
325 330

<210> 56

<211> 334

5 <212> PRT

<213> Kluyveromyces marxianus

<400> 56

Met Ser Glu Pro Ala Ala Lys Lys Gln Lys Phe Ala Asn Ser Leu Glu
1 5 10 15

Ala Leu Lys Ala Thr Gly Thr Val Val Val Ala Asp Thr Gly Asp Phe
20 25 30

Glu Ser Ile Ala Lys Phe Thr Pro Gln Asp Ala Thr Thr Asn Pro Ser
35 40 45

Leu Ile Leu Ala Ala Ala Lys Gln Asp Ala Tyr Ala Lys Leu Ile Asp
50 55 60

Ala Ala Val Glu Tyr Gly Lys Lys His Gly Ser Asn Ile Asp Glu Gln
65 70 75 80

Val Glu Ile Ser Val Asp Lys Leu Leu Val Glu Phe Gly Thr Ala Ile
85 90 95

Leu Lys Val Val Pro Gly Arg Val Ser Thr Glu Val Asp Ala Arg Leu
100 105 110

Ser Phe Asp Lys Glu Ala Thr Val Lys Lys Ala Leu Glu Ile Ile Lys
115 120 125

Leu Tyr Glu Ala Glu Gly Ile Ser Lys Glu Arg Val Leu Ile Lys Ile
130 135 140

Ala Ser Thr Trp Glu Gly Ile Gln Ala Ala Gln Glu Leu Glu Lys Asp
145 150 155 160

ES 2 602 808 T3

His Gly Ile His Val Asn Leu Thr Leu Leu Phe Ser Phe Ser Gln Ala
 165 170 175
 Val Ala Ala Ala Glu Ala Asn Val Thr Leu Ile Ser Pro Phe Val Gly
 180 185 190
 Arg Ile Leu Asp Trp Tyr Lys Ala Lys Thr Gly Glu Thr Tyr Thr Ala
 195 200 205
 Glu Thr Asp Pro Gly Val Glu Ser Val Lys Asn Ile Tyr Asn Tyr Tyr
 210 215 220
 Lys Lys His Gly Tyr Lys Thr Ile Val Met Gly Ala Ser Phe Arg Asn
 225 230 235 240
 Val Gly Glu Ile Lys Ala Leu Ala Gly Val Asp Tyr Leu Thr Ile Ser
 245 250 255
 Pro Lys Leu Leu Asp Glu Leu Met Ala Ser Gln Asp Pro Val Pro Gln
 260 265 270
 Val Leu Asp Pro Glu Ser Ala Lys Asp Gln Gly Ser Gly Arg Val Ser
 275 280 285
 Phe Ile Asn Asp Glu Ser Lys Phe Arg Phe Glu Leu Asn Glu Asp Ala
 290 295 300
 Met Ala Thr Glu Lys Leu Ser Glu Gly Ile Arg Lys Phe Ser Ala Asp
 305 310 315 320
 Ile Val Thr Leu Phe Asp Leu Ile Lys Ala Lys Ile Gln Ala
 325 330

<210> 57

<211> 2110

5 <212> DNA

<213> Issatchenkia orientalis

<220>

<221> CDS

10 <222> (536)..(1360)

<400> 57

ccccctttt gccgagcctc tgcgcacgtt tctttttcct gcgtgtgcgc cccccctct 60

cggagaggga agcacaaact ccgtcggaga caccgcgcaa agaaaagata agcgagaaag 120

ccctctgttt gaggacgcta tagaggacgc taccgagaag cgagaggatg ctcgagaatg 180

ES 2 602 808 T3

tctgctgcag gggtcggttc ttttggcgca gcgctggttc ttttggggcc atgcgggtgt	240
tgctattctc aatggggagg gtgagccaag atagatatga cgtaaggat agagtgtgt	300
tgatcaatag catatgtttg gcaacaaaca gtgactccgc taagaaatgc gtgtggattg	360
caggggtata aataggaagg tgtgtcttac aattacagta ttctccatat agtatttatt	420
ctagactata ctatttattc tagaataaat tacactacac tagactccac tagactcact	480
acactacact ccattacact acactacact acactocatt acactacatt atacc atg Met 1	538
tca gtt cca act ttt cga cta aca aac gag ctc acc gtc gtg aca ggc Ser Val Pro Thr Phe Arg Leu Thr Asn Glu Leu Thr Val Val Thr Gly 5 10 15	586
gcc agc ggc ggc att gcc cac gcc ctc gtg gag acg ctt ctg gtc tac Ala Ser Gly Gly Ile Ala His Ala Leu Val Glu Thr Leu Leu Val Tyr 20 25 30	634
ggg gcg ccg ctg gcg ctc gtg gac cgt aac atg gag gcg ttg cac cgc Gly Ala Pro Leu Ala Leu Val Asp Arg Asn Met Glu Ala Leu His Arg 35 40 45	682
aca cgg gac gcc atg gtg cgt ttc tgt gtt gag gag gca aac att aag Thr Arg Asp Ala Met Val Arg Phe Cys Val Glu Glu Ala Asn Ile Lys 50 55 60 65	730
gag gag gat gtt cca aag atg gag tgc ttt aca tgt aat ata gga gac Glu Glu Asp Val Pro Lys Met Glu Cys Phe Thr Cys Asn Ile Gly Asp 70 75 80	778
gcg ggg gaa gtt gaa act cta ttt ggc gag ata tac aat gtc ttc caa Ala Gly Glu Val Glu Thr Leu Phe Gly Glu Ile Tyr Asn Val Phe Gln 85 90 95	826
cgg tac cca cta cac atg gtg aat tgt gca ggt tat tgc gag aac ttt Arg Tyr Pro Leu His Met Val Asn Cys Ala Gly Tyr Cys Glu Asn Phe 100 105 110	874
gcc gct gtc gac tat cca gca caa aat gcc cat gat ttg atg gga gtt Ala Ala Val Asp Tyr Pro Ala Gln Asn Ala His Asp Leu Met Gly Val 115 120 125	922
aac cta ttg ggt gcc ttt tat ctc tca caa tgt ttt gca aag cct ttg Asn Leu Leu Gly Ala Phe Tyr Leu Ser Gln Cys Phe Ala Lys Pro Leu 130 135 140 145	970
att gaa cat aac atc tca gga ggt tgc att gtc ctt att gca tca atg Ile Glu His Asn Ile Ser Gly Gly Ser Ile Val Leu Ile Ala Ser Met 150 155 160	1018
agt ggg aaa att gta aat acc ccg cag aat cag tgc atc tat aac gct Ser Gly Lys Ile Val Asn Thr Pro Gln Asn Gln Cys Ile Tyr Asn Ala 165 170 175	1066
agc aaa gcc ggc gtt att cat ttg gca aaa tct ctc gca gcg gaa tgg Ser Lys Ala Gly Val Ile His Leu Ala Lys Ser Leu Ala Ala Glu Trp 180 185 190	1114
ggc gcc ctc atg cac ccc att cga gtc aat acg ctc tcc ccc ggg tac	1162

ES 2 602 808 T3

Gly	Ala	Leu	Met	His	Pro	Ile	Arg	Val	Asn	Thr	Leu	Ser	Pro	Gly	Tyr		
195						200					205						
acc	gcc	acc	cca	tta	acc	aga	aac	gtg	gtg	agc	ggc	gac	gcg	tcg	ctc	1210	
Thr	Ala	Thr	Pro	Leu	Thr	Arg	Asn	Val	Val	Ser	Gly	Asp	Ala	Ser	Leu		
210					215					220					225		
gcc	gcg	gaa	tgg	aca	aga	cgt	gtc	ccc	ctg	ggg	aga	atg	gcg	cac	ccg	1258	
Ala	Ala	Glu	Trp	Thr	Arg	Arg	Val	Pro	Leu	Gly	Arg	Met	Ala	His	Pro		
				230					235					240			
cgt	gaa	atg	gcg	ggc	gcc	gtt	ctc	ttt	ctc	ctt	gca	aac	gac	gca	agt	1306	
Arg	Glu	Met	Ala	Gly	Ala	Val	Leu	Phe	Leu	Leu	Ala	Asn	Asp	Ala	Ser		
			245					250					255				
tct	tac	acc	acg	ggg	gag	gat	gtt	ctc	gtt	gat	gga	ggg	tac	tct	gtg	1354	
Ser	Tyr	Thr	Thr	Gly	Glu	Asp	Val	Leu	Val	Asp	Gly	Gly	Tyr	Ser	Val		
		260					265					270					
tgg	tga	atggaggggg	acgaggggaca	ggggcgctcgc	gacggccgcg	ccgtgcaggg										1410	
Trp																	
oataataaac	tatccactac	acactggcac	totatagata	ttccaaccac	tcaacactag											1470	
acgacactgt	aacctataag	taatactgta	tatgctatct	actacaatac	accacacaca											1530	
cactaggtta	cacataacac	catatcggat	acagtgtata	ccctccatct	acatatggac											1590	
accacttttg	gaagaataca	tgactgtcta	tgcgacggca	agcgggttca	aacggccgca											1650	
cccacaacgc	cgccaccgct	acaaggggtgc	ccgtctcatt	atcaacgacg	aaatgcgcag											1710	
ggtaaattcc	ctccccgcag	cgacgcgtcc	gtccacgacg	tacatgtcca	ttgccgcccc											1770	
tccctctctg	cgtcctcctc	gtaaatactg	cgatatcact	ggcctcccag	cacactacac											1830	
cgctccacat	aatcaaattcc	gatactttga	cagtgaatgc	tatcagctcg	tgaaaaacat											1890	
gccaccaggt	gtcgatcagc	agtacttatc	ccttaggggt	gctaacgtta	tacttaaata											1950	
gtaatactcg	ttcgctagtc	tattctacat	atgtctttaa	aaccatttta	taccatgcc											2010	
aaacgatact	atatactata	ctatactata	ctatactata	ctatattaca	ctaacactct											2070	
cgtataatgg	ctctgagctt	ctcgctaagc	acagcgccct													2110	

<210> 58

<211> 274

5 <212> PRT

<213> Issatchenkia orientalis

<400> 58

Met	Ser	Val	Pro	Thr	Phe	Arg	Leu	Thr	Asn	Glu	Leu	Thr	Val	Val	Thr
1				5					10					15	

Gly	Ala	Ser	Gly	Gly	Ile	Ala	His	Ala	Leu	Val	Glu	Thr	Leu	Leu	Val
			20					25					30		

10

ES 2 602 808 T3

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

Val Trp

<210> 59
 <211> 3098
 5 <212> DNA
 <213> Issatchenkia orientalis

<220>
 <221> CDS
 10 <222> (1001)..(2098)

<400> 59

ES 2 602 808 T3

tcgtttgttg ttgtccctt tgagttggtg aagatccggt tacaggataa gtccaaggcg	60
tccatgtata gcggtcccat tgatgttcta atgaagacag tcaaaaatga aggtatcttg	120
gccttgata acggattaga ggcaacactg tggagacata ttgtatggaa tgccggctat	180
tttggtgtga ttttccaagt tcgtgacatg ttaccaagg caaaggacaa gaccagaaa	240
accatcaacg atttggtggg tggatgatt ggtggtatcg ttggtactgc cttgaacatt	300
ccattcgatg ttgtcaaata cagaatccaa agtgcaacca ttatggaagg tcaggtccgt	360
aaatacaatt ggacttgcc ctctcttggc attgtcttga gggaagaagg tgtctctgca	420
ttgtataaag gattccttcc aaaagtcctg agattgggtc caggtggtgg tattctactt	480
gttgtcttca caaactgtat ggacttcttt agaggccaat actatggaga caagaaatag	540
tagctgggtt cttcaccatc aggtttatac accctccacg acgtccattc tatactatac	600
tatgctattc gatgctattc gatgctatcc tgcctatct tatctaactc tatcttatct	660
tacaaaatta tatactttcc ttgtttcttt cacctcctcc tttatagatc aattgatttg	720
ataatacact tatacacatt gacgtctgtt gacatcttca tacaaaaaac cttaaaacat	780
agtgc aaagt cacgtgcacg ccttaaaaat gcagctgagc ccctttccca cttctctccc	840
ttcggatgcc ccacgtgact attttcactt cccaattcga gcatcctcct ccagtcctccg	900
cagactcaag agataagaaa ccttacagag actgttccca tcctcctatg ctcaacgtgt	960
cctgtatcta cattacacac cctctcgatt cgctacagct atg acc att gac cct	1015
	Met Thr Ile Asp Pro
	1 5
aca ctt gat tta aac aac ttg aaa gag gac aat ccc agt gtt gta ctt	1063
Thr Leu Asp Leu Asn Asn Leu Lys Glu Asp Asn Pro Ser Val Val Leu	
	10 15 20
gag aaa atc ggg gaa atc cgc ttt gag gag aga cct gtt cct gaa atc	1111
Glu Lys Ile Gly Glu Ile Arg Phe Glu Glu Arg Pro Val Pro Glu Ile	
	25 30 35
tcc gag cca aac tat gtc aag att gca atc aca cat act gga cta tgt	1159
Ser Glu Pro Asn Tyr Val Lys Ile Ala Ile Thr His Thr Gly Leu Cys	
	40 45 50
gga tcc gat gtc cac tat tac gag cac ggt tct tgt gga tcc ttc aag	1207
Gly Ser Asp Val His Tyr Tyr Glu His Gly Ser Cys Gly Ser Phe Lys	
	55 60 65

ES 2 602 808 T3

gtc gaa tct cca atg gtg tta ggc cat gaa tcg gca ggg ata att gtc Val Glu Ser Pro Met Val Leu Gly His Glu Ser Ala Gly Ile Ile Val 70 75 80 85	1255
caa gtt ggc gac agc gtt aca cgg ttg aag ccg gga gac cga gtt gca Gln Val Gly Asp Ser Val Thr Arg Leu Lys Pro Gly Asp Arg Val Ala 90 95 100	1303
tgc gaa cca ggt gtt ccc tca aga ctc tcc aag gaa tac aag gcc ggc Cys Glu Pro Gly Val Pro Ser Arg Leu Ser Lys Glu Tyr Lys Ala Gly 105 110 115	1351
aac tac aac ttg tgc ccc cac atg gcg ttt gca gca aca ccg ccc tac Asn Tyr Asn Leu Cys Pro His Met Ala Phe Ala Ala Thr Pro Pro Tyr 120 125 130	1399
gac ggc aca ctc tgc aga tac tat gta ttg ccg gag gat ttt gtc gtc Asp Gly Thr Leu Cys Arg Tyr Tyr Val Leu Pro Glu Asp Phe Val Val 135 140 145	1447
aaa tta cca gat cat gtc tca cta gaa gaa ggt gcg ttg gtg gaa cct Lys Leu Pro Asp His Val Ser Leu Glu Glu Gly Ala Leu Val Glu Pro 150 155 160 165	1495
cta tct gtt ggt gta cat gca aac aga cta att gat gtg aaa ttc gga Leu Ser Val Gly Val His Ala Asn Arg Leu Ile Asp Val Lys Phe Gly 170 175 180	1543
gat tcg atg gtg gtg ttt gga gca gga cca gtt gga ctt ctt gcc gct Asp Ser Met Val Val Phe Gly Ala Gly Pro Val Gly Leu Leu Ala Ala 185 190 195	1591
ggt gtc gcc aag gcg ttt ggc tgc gat aag gtt ctc att gtc gat att Gly Val Ala Lys Ala Phe Gly Cys Asp Lys Val Leu Ile Val Asp Ile 200 205 210	1639
gtg aat gag aaa tta gac ttt gca gtg caa cac aag ttg gca aca cat Val Asn Glu Lys Leu Asp Phe Ala Val Gln His Lys Leu Ala Thr His 215 220 225	1687
tgt ttc aac tca aaa ggg aaa act ttt gaa gat ttg ctt gca tgt att Cys Phe Asn Ser Lys Gly Lys Thr Phe Glu Asp Leu Leu Ala Cys Ile 230 235 240 245	1735
aaa gac atc tgg gac gag gac gaa ttg cca act tgt ggc atc gac gcc Lys Asp Ile Trp Asp Glu Asp Glu Leu Pro Thr Cys Gly Ile Asp Ala 250 255 260	1783
acg ggg aac cag tat tgc atc aac atg tgc att cga tcc cta gcc aag Thr Gly Asn Gln Tyr Cys Ile Asn Met Cys Ile Arg Ser Leu Ala Lys 265 270 275	1831
aag gga aga ttt gtg cag gtc ggg atg ggc ggc gac act ctg gac aag Lys Gly Arg Phe Val Gln Val Gly Met Gly Gly Asp Thr Leu Asp Lys 280 285 290	1879
ttc ccc ata gcg gcc gtt ttg gag aag gag ttg acg gtg aag ggc tcg Phe Pro Ile Ala Ala Val Leu Glu Lys Glu Leu Thr Val Lys Gly Ser 295 300 305	1927
ttt agg tac tct gtt gat gac tac aag tat tct gtc cag ttg ttg aag Phe Arg Tyr Ser Val Asp Asp Tyr Lys Tyr Ser Val Gln Leu Leu Lys	1975

310	315	320	325	
gat ggg aaa atc aat gtt cgg cct ttg atc acc cat cgc ttc aag ttt				2023
Asp Gly Lys Ile Asn Val Arg Pro Leu Ile Thr His Arg Phe Lys Phe	330	335	340	
gaa cag gcc gtg gag gcg tat gag ttt tct aaa caa ggc aag agt atc				2071
Glu Gln Ala Val Glu Ala Tyr Glu Phe Ser Lys Gln Gly Lys Ser Ile	345	350	355	
aag atc atg atc gaa gga cct cca taa agagggtctg atggatcagt				2118
Lys Ile Met Ile Glu Gly Pro Pro	360	365		
tggtctcttat atatatataa tatatatata tatatatata tatagatgta tacaatatgt				2178
ctctttcata tacctatata tctacgtgtt tagcatttcc tatatacatg actctggttt				2238
tcatttcggtt tggttctcat tcctcttggc agcttcaact aacaactggg cgtgttggtc				2298
gtcgtgtttt gccttgaaga atgtatagt caacacaacg tcttcgatgt ttctcattgc				2358
cggatctctg gaaaactctg gatcgataaa gaaaaacaag ggcatatcaa cctcctcacc				2418
cttggccaac cgctgctctt caaagcagaa acactggatc ttgttgaagt aaggcgctac				2478
atgatcgga gtcactgagt atgtggccat gccagtaatg tccttggtcac ttatattctt				2538
ggctttgtag aaggccaagg cagtctctcc ggggacaaca taaacttctc tttgttgagg				2598
tacaaacttc catggtaacg caccacttgt ctccgccgta aaggataccc gcagtcttct				2658
ctctgtagct actggagtta gcttgccct cgtgaacctg ctcttggtcg tgattggtgt				2718
accaccccat ccagtacgtt gacaaattgc acgatacaag gggacactcg catacgataa				2778
tgcaaggaaa atcatcatca tggataacga ataataaatg gtggtttgcc tctcatacct				2838
cttcttttct ctatggtact tatctctcaa tgcttggaaac tctgccaag acatctttgg				2898
aagctccttc cggtttgctc gtggtgatac ctgatgttct gacgaccac caccaggaac				2958
ttcgtatattt gcaatacaac tggcatgtac atatctccta tgaagggcaa gtccgggaat				3018
cagcccaaca tcccgaagg ggcgttgat actagtcttg aaaatccgcc ttaacatcac				3078
cgtacagaga caccttcacc				3098

<210> 60
 <211> 365
 <212> PRT
 <213> Issatchenkia orientalis

<400> 60

Met Thr Ile Asp Pro Thr Leu Asp Leu Asn Asn Leu Lys Glu Asp Asn
1 5 10 15
Pro Ser Val Val Leu Glu Lys Ile Gly Glu Ile Arg Phe Glu Glu Arg
20 25 30

ES 2 602 808 T3

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

ES 2 602 808 T3

Asp Thr Leu Asp Lys Phe Pro Ile Ala Ala Val Leu Glu Lys Glu Leu
290 295 300

Thr Val Lys Gly Ser Phe Arg Tyr Ser Val Asp Asp Tyr Lys Tyr Ser
305 310 315 320

Val Gln Leu Leu Lys Asp Gly Lys Ile Asn Val Arg Pro Leu Ile Thr
325 330 335

His Arg Phe Lys Phe Glu Gln Ala Val Glu Ala Tyr Glu Phe Ser Lys
340 345 350

Gln Gly Lys Ser Ile Lys Ile Met Ile Glu Gly Pro Pro
355 360 365

<210> 61

<211> 3143

5 <212> DNA

<213> Issatchenkia orientalis

<220>

<221> CDS

10 <222> (1001)..(2143)

<400> 61

agtagatttc ccaaactcctg catgactcta gcaagaagac attagaaaag cagtgcgggg	60
aatgaaggaa cctttcattg tatttctttt tttttttttt tttttttttt accgtttacg	120
actaaaataa ccccatggag agactatttt ggaatgacta ctaaagtggg aggaaaaaaa	180
aataacgagc tgcatatagt acgactcaac tggaactaga tggaactaga tggaactaaa	240
tggaacaaca aaagaatggg ctgtctatac tgtgtcatcc tttaacattt tttttccctt	300
tgtccattta tcctttttaca tagtgcacag agagactgcc ttaatgacac tatttcagta	360
gactaaacgc attcgaccta gaatgttccg gcccggaagg aacggcacgg ccgagaacaa	420
acaaggagag agggagtcga aaaagggaaa aaaaaaagga aaaaaaaaaa aggaaaaaaa	480
aaaaaagaaa ggcccttagt tcctcttccc tttcttcccc cctccctgtg ttcaattccc	540
gcaaggacat gaacaccaca tcgtaaatac ctttcatttt cttatttccc ttactccacc	600
ccagccccc cccattttcg ctgcattttac ctccatttct gtccctctct ataagaggcg	660
ctgatctcaa tgacagtcaa tttcaatggt tctttgtttt tccccatatt ggtgcataca	720
cattgttaat ctcttaattt aaaagtgcgc ttctctaat ggtgttcttt tcatttaaatt	780
aggtctaggt atcctttgtc tctccattct ttacttggtt ttattaaact tgctctctct	840
tgttttttac ttgtttttac ttgtctttac taatctttgc ttctttcctt tcttacttcc	900

ataactaata aacaatagac acgctggtcc atctactgat cattaccttc tgtttctatc	960
cattcgtaaa cataaaaaaca agcacgactc caacattacc atg aaa ggc cta tta Met Lys Gly Leu Leu 1 5	1015
tat tat gga aga gaa gag atc cgc tac tca gaa gac att ccc gaa cca Tyr Tyr Gly Arg Glu Glu Ile Arg Tyr Ser Glu Asp Ile Pro Glu Pro 10 15 20	1063
caa att aaa aac cca aac gac gtc aag gtc aag att gcc tat tgt ggc Gln Ile Lys Asn Pro Asn Asp Val Lys Val Lys Ile Ala Tyr Cys Gly 25 30 35	1111
atc tgt ggt act gat ttg cac gag ttt tta gat ggt cct att ttt ttc Ile Cys Gly Thr Asp Leu His Glu Phe Leu Asp Gly Pro Ile Phe Phe 40 45 50	1159
cca cag cca aat ggc cgt tcc gaa ata tcc ggt aaa aaa ttg cct ctt Pro Gln Pro Asn Gly Arg Ser Glu Ile Ser Gly Lys Lys Leu Pro Leu 55 60 65	1207
tgt cca ggt cat gaa ttt tcc ggt gtt att gaa gaa gtt ggc act ggt Cys Pro Gly His Glu Phe Ser Gly Val Ile Glu Glu Val Gly Thr Gly 70 75 80 85	1255
gtc acc aag ttt caa agg gga gac cgt gtt gtt gtt gaa gca acg tcc Val Thr Lys Phe Gln Arg Gly Asp Arg Val Val Val Glu Ala Thr Ser 90 95 100	1303
cat tgc tcc gac aga gaa cgg tat aag gac gaa att gag gat aag gac His Cys Ser Asp Arg Glu Arg Tyr Lys Asp Glu Ile Glu Asp Lys Asp 105 110 115	1351
ctc tcc ttt tgt gca gcg tgt aag gca gaa aag cca aat tgc tgt aaa Leu Ser Phe Cys Ala Ala Cys Lys Ala Glu Lys Pro Asn Cys Cys Lys 120 125 130	1399
cgt ttg tcc ttt gtt gga tta ggt act gac cat ggt gcc ttt ggt caa Arg Leu Ser Phe Val Gly Leu Gly Thr Asp His Gly Ala Phe Gly Gln 135 140 145	1447
tat gtc gtt tat ggt gaa gac cat atc ttg aag att cca gat gat ttg Tyr Val Val Tyr Gly Glu Asp His Ile Leu Lys Ile Pro Asp Asp Leu 150 155 160 165	1495
cct ctt gac ttg gct gcc ttg gtg gaa cct cta tcg gtt gca tgg cac Pro Leu Asp Leu Ala Ala Leu Val Glu Pro Leu Ser Val Ala Trp His 170 175 180	1543
gct gtc agc ttg gcc aac ttt aaa ccg gga caa acg gca gtt gtt tta Ala Val Ser Leu Ala Asn Phe Lys Pro Gly Gln Thr Ala Val Val Leu 185 190 195	1591
gga ggc ggt cca att gga tta tgt act att ctt gca ttg aag ggc cat Gly Gly Gly Pro Ile Gly Leu Cys Thr Ile Leu Ala Leu Lys Gly His 200 205 210	1639
cag gct ggt aag att gtc tgt tcg gaa cct gca gct att aga aga gaa Gln Ala Gly Lys Ile Val Cys Ser Glu Pro Ala Ala Ile Arg Arg Glu 215 220 225	1687
ttg gct gaa aaa ttg gga gct gaa act ttc aat cca atg gat cat gaa	1735

ES 2 602 808 T3

Leu Ala Glu Lys Leu Gly Ala Glu Thr Phe Asn Pro Met Asp His Glu 230 235 240 245	
gac cct att gca gaa ttg aaa aac tta tta cct gaa act gaa ggc ttt Asp Pro Ile Ala Glu Leu Lys Asn Leu Leu Pro Glu Thr Glu Gly Phe 250 255 260	1783
acc gct tca ttt gat tgt tct ggt att cag aaa act ttt gat act tct Thr Ala Ser Phe Asp Cys Ser Gly Ile Gln Lys Thr Phe Asp Thr Ser 265 270 275	1831
atc gac gtt ttg ggt cca gga ggt tct gca gtt aat gtt gca att tgg Ile Asp Val Leu Gly Pro Gly Ser Ala Val Asn Val Ala Ile Trp 280 285 290	1879
cct aat gtc cct atc caa tat gtc cca atg tgt ttg acc tat caa gag Pro Asn Val Pro Ile Gln Tyr Val Pro Met Cys Leu Thr Tyr Gln Glu 295 300 305	1927
aaa acc gct aca ggt tcc atg tgt tat gtc act aag gat ttc agg gaa Lys Thr Ala Thr Gly Ser Met Cys Tyr Val Thr Lys Asp Phe Arg Glu 310 315 320 325	1975
gtt ctt gat gcc att gca gca ggt ctc att gac cag aag tca atg agg Val Leu Asp Ala Ile Ala Ala Gly Leu Ile Asp Gln Lys Ser Met Arg 330 335 340	2023
ttg ttg gtg act ggt aaa gtt gaa gct aag gac gga atc gaa ggc ggg Leu Leu Val Thr Gly Lys Val Glu Ala Lys Asp Gly Ile Glu Gly Gly 345 350 355	2071
ttc atg caa ttg atc aac cac aag gaa aca aac gtc aag att ttg att Phe Met Gln Leu Ile Asn His Lys Glu Thr Asn Val Lys Ile Leu Ile 360 365 370	2119
gcg cca aac gga ctc gat atg tga tggagaatga ccatgttgct catcattatg Ala Pro Asn Gly Leu Asp Met 375 380	2173
tttgtcgttt gctctcattt gtctaagttg caacatgtaa tctcgtgcaa cactttcaac	2233
tttccttctt ctctttaact ttttcaattc ctgtattaat ctaattaatt taatgaactc	2293
tttataaagg taaagctaaa ccaagcgcaa aagtatcaat aataacctca aataaagtaa	2353
aaaattgaaa gaaaataacc atggggcggtg atattaacgg aacaaaatta ttatctgtgt	2413
agtcacatgt accgcgtcaa gtacttgcaa ataatacaag tgggtgtttaa acatcaccac	2473
catgcacct tcttctctta tgttgaagaa taccctattg gttagacgtt gttgatggcc	2533
aggatgcaag cttcgacccc tgtattaccg ccaattggat cagtataag tttgccta	2593
tcaagcgata ttgcaatatt aaagtatgtc ggatcagtcg acaatcatga tgggtgtgtc	2653
tgtggactgg agctatgtgg agccctggca tcaagtggtg aaaacgacgg catcatcaat	2713
ggcattgagt atttccaagt cagtgtgccg aaaagtggct tatttgtacc attgagaaaa	2773
atacttggtt gggtatcaca tactcatact caaccacaac cacaaccaca accacaatca	2833
ttaccgcaat cacaaccact attacaaccg caattactac cccaaccaa ctcgggtatca	2893
atagagtcaa cttcttcagg gtctgttgct gctactaagg aaatagagga actaaagcga	2953
catatcattt cgcttgagaa acagctccta ctacgtgaaa atgacctgaa agagttagat	3013
atccagttgg acgagttgga tgcaactctg agatctaata atgcacgatt ggccaggaag	3073
gaggagcgtt tcaatagata caagggtgaa aaggaggagg aaatctcaat gttgctaaca	3133
acaattgaat	3143

ES 2 602 808 T3

<210> 62

<211> 380

<212> PRT

5 <213> Issatchenkia orientalis

<400> 62

Met Lys Gly Leu Leu Tyr Tyr Gly Arg Glu Glu Ile Arg Tyr Ser Glu
1 5 10 15

Asp Ile Pro Glu Pro Gln Ile Lys Asn Pro Asn Asp Val Lys Val Lys
20 25 30

Ile Ala Tyr Cys Gly Ile Cys Gly Thr Asp Leu His Glu Phe Leu Asp
35 40 45

Gly Pro Ile Phe Phe Pro Gln Pro Asn Gly Arg Ser Glu Ile Ser Gly
50 55 60

Lys Lys Leu Pro Leu Cys Pro Gly His Glu Phe Ser Gly Val Ile Glu
65 70 75 80

Glu Val Gly Thr Gly Val Thr Lys Phe Gln Arg Gly Asp Arg Val Val
85 90 95

Val Glu Ala Thr Ser His Cys Ser Asp Arg Glu Arg Tyr Lys Asp Glu
100 105 110

Ile Glu Asp Lys Asp Leu Ser Phe Cys Ala Ala Cys Lys Ala Glu Lys
115 120 125

Pro Asn Cys Cys Lys Arg Leu Ser Phe Val Gly Leu Gly Thr Asp His
130 135 140

Gly Ala Phe Gly Gln Tyr Val Val Tyr Gly Glu Asp His Ile Leu Lys
145 150 155 160

Ile Pro Asp Asp Leu Pro Leu Asp Leu Ala Ala Leu Val Glu Pro Leu
165 170 175

Ser Val Ala Trp His Ala Val Ser Leu Ala Asn Phe Lys Pro Gly Gln

ES 2 602 808 T3

[illegible]

<210> 63
<211> 2879
<212> DNA
<213> *Issatchenkia orientalis*

<220>
<221> CDS
<222> (1025)..(1882)

<400> 63
atctagaagg ttgggcaaaa tatgaacctc aaaatggtgt cattaacgat tccttcccat

60

ES 2 602 808 T3

tgtcagcatt agatagagag gtttattcaa ggggcgtctc aagaattatt ttgacccaag	120
taggtgaggg cccaactcaa acaaacgatg ttttaattga cttagagagc ggattgcca	180
agtccaatta gacaaaccac cagttacatc aatattctat agactgtttc agttgtttgg	240
cactgcttta tatagtcaga aacttttcgc caatagataa tgttgtagat tttgtttttt	300
gtcgcattac aacagcatgg tttgtttgtg tcacaaaagt aatgttttct ttgctttact	360
atcaaattct acgcttaatc tccgtatttt accaagtaga tattaatgca atcgtgtttt	420
cgggtgtatct ggttttagaa gctttattac atcaagaaaa gagcagccag ataacatttc	480
tgaacaaaaa aaaactgttt tttaaattac gaaaaacaga aagtttttaa atagttgtag	540
tgtatgacat tgtcgttttg gcatccaatt tcactttttc aatcacaca cgaattgcaa	600
cttcattcat acagaagaaa aaaaaaagtt ggacaacatg cacacaacgt tgatgtotat	660
tacaaaattg aatggatgtg aatagtgatg acaagtgat tctgactttg taaggaaaga	720
aaaaacatta tcatgaggaa aaactctacg tcatgttaaa tcttttgtac gtaaaggtta	780
tttctagatc tcgaacttga taggttactg acaaggtctt aatgcgggga ggtgtagttt	840
gtggcagata aggtgtcttg gctgtgagtg cgcttcggg aggtgttttc tcggagagac	900
tgaggtacat taatggttta tttcgcacga tattttacgg caaaagataa gtccgtcccc	960
gatgggtacg ccagtaagaa acacactcac tgctgcttag ttctttgccg caaaccaatt	1020
gtgt atg ata aaa aga gcc atg act atc gga aag cca att act tta aac	1069
Met Ile Lys Arg Ala Met Thr Ile Gly Lys Pro Ile Thr Leu Asn	
1 5 10 15	
aac gga aca aag atc cca ttc atg ggg tta ggt acc tgg gaa atc agt	1117
Asn Gly Thr Lys Ile Pro Phe Met Gly Leu Gly Thr Trp Glu Ile Ser	
20 25 30	
aat gca gat gtg gtt gtt cgg gaa gca tta aat gtc ggc tat aga tgc	1165
Asn Ala Asp Val Val Val Arg Glu Ala Leu Asn Val Gly Tyr Arg Cys	
35 40 45	
atc gat acc gct gtt ttg tat gga aat gag aag ttg tgc gga gat ggt	1213
Ile Asp Thr Ala Val Leu Tyr Gly Asn Glu Lys Leu Cys Gly Asp Gly	
50 55 60	
att atc aaa tgg ttg gag tcc gat cct aac aat aag agg gaa gat gtg	1261
Ile Ile Lys Trp Leu Glu Ser Asp Pro Asn Asn Lys Arg Glu Asp Val	
65 70 75	
tat tac att acc aag tta tgg aac cat cag aat gga tac gag aaa gcg	1309
Tyr Tyr Ile Thr Lys Leu Trp Asn His Gln Asn Gly Tyr Glu Lys Ala	
80 85 90 95	
aaa cgt gcc atc agg gag tgc ttt gaa aaa gtc aag ggg tta gga tac	1357
Lys Arg Ala Ile Arg Glu Cys Phe Glu Lys Val Lys Gly Leu Gly Tyr	
100 105 110	
att gat tta ctt ttg atc cat tcg ccg acc gaa ggg cca aga atg agg	1405
Ile Asp Leu Leu Leu Ile His Ser Pro Thr Glu Gly Pro Arg Met Arg	

ES 2 602 808 T3

115	120	125	
ttg gag act tgg aaa gcc atg caa gag gca gtt gac gag ggt atc gtc Leu Glu Thr Trp Lys Ala Met Gln Glu Ala Val Asp Glu Gly Ile Val 130 135 140			1453
aag tct att ggc gtt tct aac tat ggt att aag cat ttg caa gaa ttg Lys Ser Ile Gly Val Ser Asn Tyr Gly Ile Lys His Leu Gln Glu Leu 145 150 155			1501
tta tca tgg gaa ggc act tat atc aag ccc gtt gcc aac gag atc gag Leu Ser Trp Glu Gly Thr Tyr Ile Lys Pro Val Ala Asn Glu Ile Glu 160 165 170 175			1549
gtt tct cct tgg tgt atg aga caa gaa cta tgt gac ttt act aaa aaa Val Ser Pro Trp Cys Met Arg Gln Glu Leu Cys Asp Phe Thr Lys Lys 180 185 190			1597
cac gat att gtg gtt att gcc tat gct ccg cta tca cac agt tat cgt His Asp Ile Val Val Ile Ala Tyr Ala Pro Leu Ser His Ser Tyr Arg 195 200 205			1645
ctc caa gat aag gac gct gtg gaa atc gca aag aaa aag aac gtt act Leu Gln Asp Lys Asp Ala Val Glu Ile Ala Lys Lys Lys Asn Val Thr 210 215 220			1693
gtt gcc caa gtc cta atc aga tgg tct cta cag aag ggg tac att cca Val Ala Gln Val Leu Ile Arg Trp Ser Leu Gln Lys Gly Tyr Ile Pro 225 230 235			1741
atc ccg aaa aca aag acc cta gct aga ctt cct gta aat ttg gat gtt Ile Pro Lys Thr Lys Thr Leu Ala Arg Leu Pro Val Asn Leu Asp Val 240 245 250 255			1789
ttg agt ttc gag ttg agc act gac gaa atg aaa cag ctt gac cac cca Leu Ser Phe Glu Leu Ser Thr Asp Glu Met Lys Gln Leu Asp His Pro 260 265 270			1837
ttg gaa cac gat cca tcc gat tgg gag gtt aca atg tgt cca taa Leu Glu His Asp Pro Ser Asp Trp Glu Val Thr Met Cys Pro 275 280 285			1882
ccctcaccat atatagggtg gccatacatt tgtaaacat gtatctctct tgtgaaaccc			1942
ttctgaaacc cccttgtttt cttacagggt tgggtccattt cctcctcctc ttcttcttct			2002
tctttttccc ctttaaatca gataaatagt tacattgtgc atgcagtgtc gattttgttt			2062
tgaaatatgg tatctaacgg gatagtaaatt cttctcactc aaggaaattt tccaatat			2122
ttcttgccgg catttccctc aactgcaaag aagtttggaa ttgtattata tgtcgacacg			2182
gtacttacag aaccgaacaa ctacacaagg atcacctgaa ggatgtcaga taccgagtct			2242
ctctttgacg gtgggttgtt tgaagaacct gaggagtttc aaagaaagga aattccttca			2302
cattttgcc aatacgaaag aaaataccaa acaccagagt tccatcattc tattgatgaa			2362
atccagctga gggttggttg caaatcccc ctatgggggc acttggtatg gaatgcaggt			2422
acgtatactg ccaattacat tgaaaagcat cccaaagaag ttaggggaag aaggtggttg			2482
agtttggtgc agcgtctgca ctaccatcgt tactatgtgc cttgaacgga gcagagaggg			2542

tgatatgtac agactatcca gatccagatt tgttgtcgaa catcaagtac aacgttgaac 2602
atctggagta ttccgccgca caagagatca tagacgttga aggtttcatt tggggtaatc 2662
cggttgagga catttcccgcc aaactaggcg gcaacggcaa ggccgacttc ctcattatga 2722
gtgacttggt gttcaaccac agcgagcacc acaactcct caaatcatgc aaggaactca 2782
tccaaccact ggaggagggc aaaccagaa gtggtggtag atgcttggtt gtctggtcac 2842
cgcacccgacc agtgccaaag atggtggaaa acgattt 2879

<210> 64

<211> 285

5 <212> PRT

<213> Issatchenkia orientalis

<400> 64

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

ES 2 602 808 T3

Ser Pro Trp Cys Met Arg Gln Glu Leu Cys Asp Phe Thr Lys Lys His
180 185 190

Asp Ile Val Val Ile Ala Tyr Ala Pro Leu Ser His Ser Tyr Arg Leu
195 200 205

Gln Asp Lys Asp Ala Val Glu Ile Ala Lys Lys Lys Asn Val Thr Val
210 215 220

Ala Gln Val Leu Ile Arg Trp Ser Leu Gln Lys Gly Tyr Ile Pro Ile
225 230 235 240

Pro Lys Thr Lys Thr Leu Ala Arg Leu Pro Val Asn Leu Asp Val Leu
245 250 255

Ser Phe Glu Leu Ser Thr Asp Glu Met Lys Gln Leu Asp His Pro Leu
260 265 270

Glu His Asp Pro Ser Asp Trp Glu Val Thr Met Cys Pro
275 280 285

<210> 65
<211> 2873
5 <212> DNA
<213> Issatchenkia orientalis

<220>
<221> CDS
10 <222> (1001)..(1873)

<400> 65
accacaatgc ggtatgcggt gaatggactg ctgctctcac tgattgtgct aaatggctca 60
atggaatggc gagaggcgca aggtatgatg caacatatca gaacgacact ccaattgggtt 120
cttgttctaa cttgtacttg gctgattatt catacttcca acaggaagat gttagacaaa 180
cctacagaag gtatgttgaa gctcagatgg acgcatacct tcatgaaaag atgaatggat 240
gggttttttg gtgttgaaa actgaaaact tgattgagtg ggacttccaa aggttggttg 300
agttgaatat cattcctcag ccgttgaact ctagagaatt ctacaatcaa tgtgggttatt 360
gattcatctt actttctttt gaaacaaagg aaataggtag aaatcataaa accaacctaa 420
aaacatcaat caaaacaata atactttata ccttacatgc cggtttacct cattttatcc 480
agcatacttt tgctttcggg ggggttggtc taatcagata tttttgtctg tctttactat 540
taatatgcgt tctaaaaggg aacgattgac tgtatacctt aataacttta tatttcataa 600
actagaaaaa aagactatct actatactcg gatatgcagt tgaatgggtg acgcagttgc 660

tggagcctta gtagaccaga ctaacagtac ggtacgaaga cctgtaaaat agaactagcc	720
gtttctgtgt tgactaatct aggggaagaa aaaaactaga cagacaattg cattttaagc	780
tgcagggttc cctttctgtg gtgaatctcc ggggtgagaa caatagaaat gggttttagc	840
gattggcctg aaattgtctc gggcaggaga tatccttttg gcattcatgc ttgcattaca	900
agtatataat tgaaagcttg caaccacaac ctattttttg caattagagg tcaggtagaa	960
acttttccac aatgtataac taaacatttc aatcctctca atg acg atc cct tca	1015
Met Thr Ile Pro Ser	1
tat cgt aca ctt aat tca ggt cac aaa ata cca agt att gca ttg ggt	1063
Tyr Arg Thr Leu Asn Ser Gly His Lys Ile Pro Ser Ile Ala Leu Gly	10
gtc tat cag aca ccg cca cac gag act gca gca gtg gtg ttt gcc gcc	1111
Val Tyr Gln Thr Pro Pro His Glu Thr Ala Ala Val Val Phe Ala Ala	25
tta gaa agt gga tat cgg cac att gat tgt gca cag ttt tat gaa aat	1159
Leu Glu Ser Gly Tyr Arg His Ile Asp Cys Ala Gln Phe Tyr Glu Asn	40
gag gag gaa gca tgc cga ggt ata gca aaa tgg att gcc aaa gac ccc	1207
Glu Glu Glu Ala Cys Arg Gly Ile Ala Lys Trp Ile Ala Lys Asp Pro	55
agc aga aac aag aga gag cat gtc ttc tac acg act aag atc ttt gac	1255
Ser Arg Asn Lys Arg Glu His Val Phe Tyr Thr Thr Lys Ile Phe Asp	70
cca gat cac ggt tac gca agg aca aac aaa gcc att gaa ctt tca ctt	1303
Pro Asp His Gly Tyr Ala Arg Thr Asn Lys Ala Ile Glu Leu Ser Leu	90
gag aga gca aaa gaa att ggt tac att gat cta ctc tta ctt cat tct	1351
Glu Arg Ala Lys Glu Ile Gly Tyr Ile Asp Leu Leu Leu Leu His Ser	105
cca cag tct gat tac gag aga aga cat ggt tcg tgg atg gcg ttt cag	1399
Pro Gln Ser Asp Tyr Glu Arg Arg His Gly Ser Trp Met Ala Phe Gln	120
gag ttt gta gaa tca ggt aaa gtc aag agt atc ggc gtt tcc aat tat	1447
Glu Phe Val Glu Ser Gly Lys Val Lys Ser Ile Gly Val Ser Asn Tyr	135
ggt att aaa cac ctc aag gaa tta ctt gaa tac cct gac ttg aaa acc	1495
Gly Ile Lys His Leu Lys Glu Leu Leu Glu Tyr Pro Asp Leu Lys Thr	150
aag cca gca gtg aac caa ctc gaa ctt cat cca tgg ttg aca agg aat	1543
Lys Pro Ala Val Asn Gln Leu Glu Leu His Pro Trp Leu Thr Arg Asn	170
gat ttg acc gcg tac act gcg aac cag ggt cta cta gtg gaa gct tat	1591
Asp Leu Thr Ala Tyr Thr Ala Asn Gln Gly Leu Leu Val Glu Ala Tyr	185
acc ccg ttg gta agg gct agg aag atg gat gat ccc acc ctg ctc aaa	1639

ES 2 602 808 T3

Thr	Pro	Leu	Val	Arg	Ala	Arg	Lys	Met	Asp	Asp	Pro	Thr	Leu	Leu	Lys	
		200					205					210				
gta	gca	gag	gac	cat	aac	aga	aca	cct	gca	cag	atc	ctc	atc	aat	tgg	1687
Val	Ala	Glu	Asp	His	Asn	Arg	Thr	Pro	Ala	Gln	Ile	Leu	Ile	Asn	Trp	
	215					220					225					
tcg	ctt	tcc	aag	ggg	ttt	att	cct	ctt	cca	aag	acc	gcc	aca	gtt	tca	1735
Ser	Leu	Ser	Lys	Gly	Phe	Ile	Pro	Leu	Pro	Lys	Thr	Ala	Thr	Val	Ser	
	230				235					240					245	
aga	ttg	gca	tct	aac	ttt	gag	gcc	atg	caa	ttc	caa	cta	tct	aaa	aaa	1783
Arg	Leu	Ala	Ser	Asn	Phe	Glu	Ala	Met	Gln	Phe	Gln	Leu	Ser	Lys	Lys	
				250					255					260		
cag	gtt	gac	acc	cta	gat	gcc	ctc	aat	gag	ggc	atg	cac	att	tgc	tgg	1831
Gln	Val	Asp	Thr	Leu	Asp	Ala	Leu	Asn	Glu	Gly	Met	His	Ile	Cys	Trp	
			265					270					275			
aat	cct	agc	act	tat	cct	cta	gat	aat	gag	aga	caa	gct	tag			1873
Asn	Pro	Ser	Thr	Tyr	Pro	Leu	Asp	Asn	Glu	Arg	Gln	Ala				
	280						285				290					
tcgatgtaag	tacgtctata	aagcccgtag	gcatacacgta	acagctatat	gtgtcactct											1933
tctgtaacaa	tcatgtatat	gcattcatag	gaatggaaac	gtcaaaagga	attggttggt											1993
cgcctaagta	atgtaaaacg	tctcgggtat	taccogataa	aagcgaggtc	tttttttttc											2053
ttttcagatt	atttgagttt	tctataagaa	gcagacattg	ctttatcttc	ctgtagcaat											2113
agattcattc	atacccttgg	atgtactctg	tacaagctat	acatttctct	ctaaataaga											2173
taactctaac	gtgtttactt	ggacttgaag	aagacagaca	ttaggaagga	aaaaaaagga											2233
tttttaattt	catctcaaca	aaaaatggcc	aaaaaaagag	aatgccccat	ttgtttagaa											2293
gatatcacct	ctaatagatcc	ttcatataca	ttgacaatac	cctgcaagca	tttttatcac											2353
aagtcgtgca	ttctttcatg	gacctcaaaa	tcagcatcta	cctgtcctca	gtgccgaaat											2413
gagctaacgt	cattattcac	gccagctgat	cagaagacta	taaagatcaa	ccataaagta											2473
caggataaac	tgggttgactt	gatcaataat	cacccatctg	aaccgtcgtc	gtctatcatt											2533
tctacaaatg	gactatcaca	tatagaaatt	aatacagaat	cggctttatc	aagaccaaac											2593
ggcccacttt	tttcaaatac	acaccaacag	gtacaacaac	agttaaacac	aaatatcagg											2653
catttgtcta	atcagcaatg	cagtatttgc	gataatacag	tactcattac	acaactgata											2713
atatgtccac	agtgttcagg	tttgtaccat	cgtcatgttt	gtgatggtct	taattgcccc											2773
ttttgcgaag	aatggatoga	tgatttagct	tgttcaaccg	tgaccacaaa	gaagaggaaa											2833
actttggatc	gatcagctga	tgataactcaa	tactatacaa													2873

<210> 66

<211> 290

5 <212> PRT

<213> Issatchenkia orientalis

<400> 66

ES 2 602 808 T3

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

ES 2 602 808 T3

245

250

255

Gln Leu Ser Lys Lys Gln Val Asp Thr Leu Asp Ala Leu Asn Glu Gly
260 265 270

Met His Ile Cys Trp Asn Pro Ser Thr Tyr Pro Leu Asp Asn Glu Arg
275 280 285

Gln Ala
290

<210> 67
<211> 5331
5 <212> DNA
<213> Issatchenkia orientalis

<220>
<221> CDS
10 <222> (1001)..(1882)

<220>
<221> CDS
15 <222> (3469)..(4332)

<400> 67
taatatggga aagtccttgc gtggttgtgc acacttctct atttgacatc tagatttccc 60
caattgtgga caaattatac attgaagaca caaaaggggt atcactgaaa ttgatcactt 120
ttgcattgtt tgggaactta ttctattcaa tgtctctact attatctgag aattcattga 180
gaggcgggga agaatcaaag gagttttgga aggccgaatt gagttacttt ttaggggcaa 240
tcggaacagt attgtttgat ttatttgcaa ttttacaatg gattcattat gacagccaca 300
gtaatcgtac caatcatatc caatctgtga ggttgaaagc ttacaccctt aatcattaa 360
aaagccagac aattcccaaa tcggtgccat tgatacatc acgtacatcg tccatgagag 420
atggtacaaa gatagatccc atcgaaatgg cggctagcgt caagtcaaca ttgtcacccc 480
agaatgtacg caaactcaat gagttcacac cattgtctcc tatggattta ttgctagatg 540
aacatatttc acgcagttat gtttcctcta ctgatacaaa aactatacct cagaagaaga 600
gacctgatag tatcaagtct gtacacaggc acaacgagga cctgctaattg acattcgaag 660
aatagaagca gtcccaattt aaaccgtggc cgttgtaaca gccataactg tagccacaat 720
tggaaattat ggatgtattg tctgatttgg acctccgggg cagggacaat ggacttgccc 780
aaagagtcga aaaaaatggt caacagacga gataattggt ctttaattgt ctcggacatg 840
tgatttcctt aaaagtttaa tttcacaccc gcaggtttat ttatataaaa gtgtggccac 900
aagtcttggg aagatgaaca tcttgatatt catgtcccct ctcatcttct gagactggca 960
taagataagt agaaagcttt tgtaatcgaa taacatcata atg gca att gca acc 1015

ES 2 602 808 T3

											Met 1	Ala 2	Ile 3	Ala 4	Thr 5	
cac His	ctt Leu	aca Thr	ttg Leu	aat Asn 10	tca Ser	ggt Gly	tac Tyr	cag Gln	att Ile 15	cca Pro	acc Thr	att Ile	gcg Ala	ttg Leu 20	ggt Gly	1063
gtc Val	tac Tyr	caa Gln	aca Thr 25	cca Pro	cca Pro	gag Glu	gag Glu	act Thr 30	gag Glu	aag Lys	atc Ile	gtt Val	ctt Leu 35	gct Ala	gct Ala	1111
ttg Leu	gag Glu	gcc Ala 40	ggt Gly	tat Tyr	cgc Arg	cat His	att Ile 45	gac Asp	agt Ser	gcc Ala	caa Gln 50	tac Tyr	tac Tyr	cac His	aac Asn	1159
gaa Glu	gaa Glu 55	gat Asp	gtt Val	gca Ala	aag Lys 60	gca Ala	att Ile	gca Ala	aag Lys	tgg Trp 65	att Ile	gcc Ala	gaa Glu	gat Asp	cct Pro	1207
gct Ala 70	cac His	aac Asn	aaa Lys	agg Arg	gag Glu 75	gac Asp	att Ile	ttt Phe	tac Tyr	acc Thr 80	acc Thr	aag Lys	atc Ile	tac Tyr	gac Asp 85	1255
caa Gln	gac Asp	cac His	ggg Gly	tat Tyr 90	gag Glu	aag Lys	acc Thr	aaa Lys 95	aaa Lys	gct Ala	gtt Val	gaa Glu	gtt Val	tct Ser 100	ctt Leu	1303
gaa Glu	agg Arg	gca Ala 105	aag Lys	gac Asp	att Ile	gac Asp	tac Tyr	atc Ile 110	gac Asp	ttg Leu	ctc Leu	cta Leu	ctc Leu 115	cat His	tcc Ser	1351
ccc Pro	caa Gln	tcc Ser 120	gac Asp	tac Tyr	gaa Glu	agg Arg 125	aga Arg	cat His	ggc Gly	tcc Ser	tgg Trp 130	ttg Leu	gca Ala	ttc Phe	cag Gln	1399
gaa Glu 135	ttt Phe	gtg Val	gag Glu	tcc Ser	ggc Gly 140	aag Lys	gtc Val	aga Arg	agc Ser	atc Ile 145	ggt Gly	gtc Val	tcc Ser	aac Asn	tac Tyr	1447
ggt Gly 150	gtc Val	aag Lys	cac His	atc Ile	aag Lys 155	gag Glu	ttg Leu	ctt Leu	gaa Glu	tat Tyr 160	ccc Pro	gac Asp	ttg Leu	aag Lys	att Ile 165	1495
aaa Lys	ccc Pro	gcc Ala	gtc Val	aac Asn 170	cag Gln	gtc Val	gag Glu	ctt Leu	cat His 175	cca Pro	tgg Trp	ttg Leu	acc Thr	aga Arg 180	gag Glu	1543
gac Asp	att Ile	gtg Val	gac Asp 185	tat Tyr	gct Ala	gcc Ala	aaa Lys	cac His 190	gga Gly	atc Ile	atc Ile	atc Ile	gag Glu 195	gcg Ala	tac Tyr	1591
tct Ser	ccc Pro	ttg Leu 200	gta Val	aga Arg	ggc Gly	cag Gln	aag Lys 205	atg Met	gac Asp	gac Asp	cct Pro	acg Thr 210	ctt Leu	gtg Val	aag Lys	1639
att Ile 215	gca Ala	gag Glu	aag Lys	tac Tyr	aac Asn 220	aag Lys	act Thr	ccc Pro	gca Ala	cag Gln 225	atc Ile	ctc Leu	atc Ile	aac Asn	tgg Trp	1687
tcc Ser 230	tta Leu	tcc Ser	aag Lys	ggt Gly 235	tat Tyr 235	att Ile	ccg Pro	ctt Leu	cca Pro	aaa Lys 240	aca Thr	agc Ser	aag Lys	gtg Val	tct Ser 245	1735

aga cta gct tct aac ctt gaa gca acg caa ttt gaa tta tcc aag gaa	1783
Arg Leu Ala Ser Asn Leu Glu Ala Thr Gln Phe Glu Leu Ser Lys Glu	
250 255 260	
gac gtt gac atc ttg agt gca cta aac gaa aat ctc cac acc tgt tgg	1831
Asp Val Asp Ile Leu Ser Ala Leu Asn Glu Asn Leu His Thr Cys Trp	
265 270 275	
gat cct act gtc tac cca cta gat aac gaa aag gaa gcc gaa aag aag	1879
Asp Pro Thr Val Tyr Pro Leu Asp Asn Glu Lys Glu Ala Glu Lys Lys	
280 285 290	
taa atacgtacct tctgtactta atgttttagtt ttctttcaac tgaatcccta	1932
tgttattgga ttagtatttt caggaatag ctctgacaga tccttacatt agtttctttg	1992
tagaagtgtg gccttttttt tagcggaggt tgcgttttgt cttctcctgt ggaaacctcc	2052
tgaaggttct aagtgacact tgtatctgta tcgaactaag atgttaatcc attggctatg	2112
aatcaaaaca aacagaaagc aaagacgtga tgtaagcat tagtttatgt taaccctaac	2172
tgaaactcac gtggttaactg cagacagaat tatagcttcg ataggtttga atgacagcaa	2232
agggaaaaaa agttaaaacta tagttgcagt gaagtcactc catagcggga ataaaatggg	2292
cgtaacaatg aaaaaggtta actagaccac tgcaaacctg tgatagtgc acttttttgt	2352
tttcccatgc ttacgttgat acaagtttac aagcagttca atggcagcat gggtaacaga	2412
attctgggta aaccagtaat tcattagcaa aacaaaacct gcagtttact aaaagtttc	2472
aaaacaaaaa ttttacccaa gattggaact tatactactt gaaccaaag gacaattttt	2532
ttcttgtttc atagtcgtga gtgtcaagaa tcctttttaa atataaacag cccataggta	2592
gcttctaagt taaaatcccg gtagtgaaaa ttatgaggta acatgaagtt aacaggatca	2652
agcccaaaaa acaagtacta ctgtaaacac catctttttg tgttttattt tagatcaaag	2712
cagtaacctt tacgggacat ccaggatta tttcttgcca atacaaagtt gtgttcagga	2772
aaaagtaaaa ttgtcaaata tatttcatat tgatagaaac aaaactctac tcgttctagc	2832
tttgctgttc ttcttattag ttctacctc tgacctaaaa caacttgatt agataagctt	2892
taagttcctt tacattcatg acactcttag ataaagtatc caatttgcta ggcatttgaa	2952
caaatcacac catcacagtc acttgatacc acaaaacctt catacataa atgtttcatc	3012
gtttcataag caacaaaaa gaataaaacg gtcacagaag ggctctagtc taatcaaaag	3072
tcaataactc taaggtccat ttacttagc acaataacaa cgataagcga ctaaaatata	3132
acgtactctt ttttatactg tcggcatgaa tccggcttat actgaatttt gtgtagctag	3192
tccattacca tttccacctt ctgggcaaat atagttgaaa gccgtttgaa ttatgataga	3252
tttactctta atttcaccaa aaaaaggttg ctctgaaaga gcaatagtta tagggttcag	3312
aaaatataag gtgagagtag tgggttgata acaaagagtg attaccttaa tagtgtacct	3372
acgaagactt taaaagggcc caagaatgcc cttatcatto gtttagtatt gatttaatgt	3432

ctgcaaattc agctacagct aaaaaacgta gtccta atg gca att gca acc cac	3486
Met Ala Ile Ala Thr His	
295	
ctt aca ttg aat tca ggt tac cag att cca acc att gcg ttg ggt gtc	3534
Leu Thr Leu Asn Ser Gly Tyr Gln Ile Pro Thr Ile Ala Leu Gly Val	
300 305 310 315	
tat caa aca cca cca gag gag act gag aag atc gtt ctt gct gct ttg	3582
Tyr Gln Thr Pro Pro Glu Glu Thr Glu Lys Ile Val Leu Ala Ala Leu	
320 325 330	
gag gcc ggt tat cgc cat att gac agt gcc caa tac tac cac aat gaa	3630
Glu Ala Gly Tyr Arg His Ile Asp Ser Ala Gln Tyr Tyr His Asn Glu	
335 340 345	
gaa gat gtt gcc agg gca att gca aag tgg atc gcc aaa aat cca gct	3678
Glu Asp Val Ala Arg Ala Ile Ala Lys Trp Ile Ala Lys Asn Pro Ala	
350 355 360	
cac aac aaa agg aag aat atc ttc tat act acg aag atc tat gac cag	3726
His Asn Lys Arg Lys Asn Ile Phe Tyr Thr Thr Lys Ile Tyr Asp Gln	
365 370 375	
cat gta tat gag aaa acc aaa aaa gct gtt gaa gtt tct ctt gaa agg	3774
His Val Tyr Glu Lys Thr Lys Lys Ala Val Glu Val Ser Leu Glu Arg	
380 385 390 395	
gca aag gac att gac tac atc gac ttg ctc cta ctc cat tcc ccc caa	3822
Ala Lys Asp Ile Asp Tyr Ile Asp Leu Leu Leu Leu His Ser Pro Gln	
400 405 410	
tcc gac tac gaa agg aga cat ggc tcc tgg ttg gca ttc cag gaa ttt	3870
Ser Asp Tyr Glu Arg Arg His Gly Ser Trp Leu Ala Phe Gln Glu Phe	
415 420 425	
gtg gag tcc ggc aag gtc aga agc atc ggt gtc tcc aac tat ggt gtc	3918
Val Glu Ser Gly Lys Val Arg Ser Ile Gly Val Ser Asn Tyr Gly Val	
430 435 440	
aag cac atc aag gag ttg ctt gaa tat ccc gac ttg aag att aaa ccc	3966
Lys His Ile Lys Glu Leu Leu Glu Tyr Pro Asp Leu Lys Ile Lys Pro	
445 450 455	
gcc gtc aac cag gtc gag ctt cat cca tgg ttg acc aga gag gac att	4014
Ala Val Asn Gln Val Glu Leu His Pro Trp Leu Thr Arg Glu Asp Ile	
460 465 470 475	
gtg gac tat gct gcc aaa cac gga atc atc atc gag gcg tac tct ccc	4062
Val Asp Tyr Ala Ala Lys His Gly Ile Ile Ile Glu Ala Tyr Ser Pro	
480 485 490	
ttg gtg aga ggc cag aag atg gac gac cct acg ctt gtg aag att gca	4110
Leu Val Arg Gly Gln Lys Met Asp Asp Pro Thr Leu Val Lys Ile Ala	
495 500 505	
gag aag tac aac aag act ccc gca cag atc ctc atc aac tgg tcc tta	4158
Glu Lys Tyr Asn Lys Thr Pro Ala Gln Ile Leu Ile Asn Trp Ser Leu	
510 515 520	
tcc aag ggt tat att cca ctt cca aat acg agc caa atc tct aga tta	4206
Ser Lys Gly Tyr Ile Pro Leu Pro Asn Thr Ser Gln Ile Ser Arg Leu	

525					530					535						
gcg tct aac cac gat gcc gca caa ttc gaa tta tcc aaa gaa gac att															4254	
Ala Ser Asn His Asp Ala Ala Gln Phe Glu Leu Ser Lys Glu Asp Ile																
540	545					550					555					
ggt act tta agt tct ctc aat gag cgc ctt cac acc tgt tgg gat cct															4302	
Gly Thr Leu Ser Ser Leu Asn Glu Arg Leu His Thr Cys Trp Asp Pro																
	560					565					570					
acc ctc tac ccc ctt ggt gac gaa aag taa aacaacaaga attgatggta															4352	
Thr Leu Tyr Pro Leu Gly Asp Glu Lys																
	575					580										
tattggtaag gcggtgtaac ataccagtca gtaaattctat ccctactagc tttttttttc															4412	
tatatattta cacaaccaa cagctacatg ttccaataca taaacatgga gaaccgctcc															4472	
ccttttatatt tttttttttc acacacacct tttatcttat cgctttacat ttccgggtggc															4532	
aaattgatta aaaaaagtac agaaatgctc agctocaaat agccttgaat tgggggttgct															4592	
tcctttctct gataaccatt tttcctttct caattgctag ctaacagtag caaaacaact															4652	
agccctatac caaatgaaca ttcactcgtc agtattgaca tcctgtagtcc tcttgctcgc															4712	
ttcaattacg ggctccgatg ctaagggttca ttctgccagc atccacaaga atccgttcca															4772	
agacaattat aaagatattt cctatctaga atatgttgac tccatcaaga acaagtatgt															4832	
taacaatttt gtcaagaact tcaatgcacc ttttgtcca tttgttgaag atgcgggtcat															4892	
tgaggacact catgaactac cottaacca ctatatgaat gcccaatact tcactgagat															4952	
tcaacttggt acccctggcc agccattcaa ggtgattcta gacactgggt cttctaattt															5012	
gtgggttctt tccacaaaat gtacatcttt ggcatgttat ttgactcta aatatgatca															5072	
cgatgcaagt tccacataca aacaaaatgg taccgattct ctatcagata tggttctggt															5132	
tccttggaag gttttatttc acaagattta ctaacttttg gtgacttggt cattccagag															5192	
caggatttcg ctgaggcaac aagtgaaccg ggcttggcgt ttgctttcgg aaaattcgac															5252	
ggtattctag gtttagctta tgataccatc tcggtggaca aggttgttcc tccaatttac															5312	
aatgccattg acaagggtc															5331	

<210> 68
<211> 293
<212> PRT
<213> *Issatchenkia orientalis*

<400> 68

Met Ala Ile Ala Thr His Leu Thr Leu Asn Ser Gly Tyr Gln Ile Pro
1 5 10 15

Thr Ile Ala Leu Gly Val Tyr Gln Thr Pro Pro Glu Glu Thr Glu Lys
20 25 30

ES 2 602 808 T3

```

Ile Val Leu Ala Ala Leu Glu Ala Gly Tyr Arg His Ile Asp Ser Ala
  35                                40                                45

Gln Tyr Tyr His Asn Glu Glu Asp Val Ala Lys Ala Ile Ala Lys Trp
  50                                55                                60

Ile Ala Glu Asp Pro Ala His Asn Lys Arg Glu Asp Ile Phe Tyr Thr
  65                                70                                75                                80

Thr Lys Ile Tyr Asp Gln Asp His Gly Tyr Glu Lys Thr Lys Lys Ala
                                85                                90                                95

Val Glu Val Ser Leu Glu Arg Ala Lys Asp Ile Asp Tyr Ile Asp Leu
                                100                                105                                110

Leu Leu Leu His Ser Pro Gln Ser Asp Tyr Glu Arg Arg His Gly Ser
  115                                120                                125

Trp Leu Ala Phe Gln Glu Phe Val Glu Ser Gly Lys Val Arg Ser Ile
  130                                135                                140

Gly Val Ser Asn Tyr Gly Val Lys His Ile Lys Glu Leu Leu Glu Tyr
  145                                150                                155                                160

Pro Asp Leu Lys Ile Lys Pro Ala Val Asn Gln Val Glu Leu His Pro
                                165                                170                                175

Trp Leu Thr Arg Glu Asp Ile Val Asp Tyr Ala Ala Lys His Gly Ile
                                180                                185                                190

Ile Ile Glu Ala Tyr Ser Pro Leu Val Arg Gly Gln Lys Met Asp Asp
  195                                200                                205

Pro Thr Leu Val Lys Ile Ala Glu Lys Tyr Asn Lys Thr Pro Ala Gln
  210                                215                                220

Ile Leu Ile Asn Trp Ser Leu Ser Lys Gly Tyr Ile Pro Leu Pro Lys
  225                                230                                235                                240

Thr Ser Lys Val Ser Arg Leu Ala Ser Asn Leu Glu Ala Thr Gln Phe
                                245                                250                                255

Glu Leu Ser Lys Glu Asp Val Asp Ile Leu Ser Ala Leu Asn Glu Asn
                                260                                265                                270

Leu His Thr Cys Trp Asp Pro Thr Val Tyr Pro Leu Asp Asn Glu Lys
  275                                280                                285

Glu Ala Glu Lys Lys
  290

```

5 <210> 69
 <211> 287
 <212> PRT
 <213> Issatchenkia orientalis

10 <400> 69

ES 2 602 808 T3

Met Ala Ile Ala Thr His Leu Thr Leu Asn Ser Gly Tyr Gln Ile Pro
1 5 10 15

Thr Ile Ala Leu Gly Val Tyr Gln Thr Pro Pro Glu Glu Thr Glu Lys
20 25 30

Ile Val Leu Ala Ala Leu Glu Ala Gly Tyr Arg His Ile Asp Ser Ala
35 40 45

Gln Tyr Tyr His Asn Glu Glu Asp Val Ala Arg Ala Ile Ala Lys Trp
50 55 60

Ile Ala Lys Asn Pro Ala His Asn Lys Arg Lys Asn Ile Phe Tyr Thr
65 70 75 80

Thr Lys Ile Tyr Asp Gln His Val Tyr Glu Lys Thr Lys Lys Ala Val
85 90 95

Glu Val Ser Leu Glu Arg Ala Lys Asp Ile Asp Tyr Ile Asp Leu Leu
100 105 110

Leu Leu His Ser Pro Gln Ser Asp Tyr Glu Arg Arg His Gly Ser Trp
115 120 125

Leu Ala Phe Gln Glu Phe Val Glu Ser Gly Lys Val Arg Ser Ile Gly
130 135 140

Val Ser Asn Tyr Gly Val Lys His Ile Lys Glu Leu Leu Glu Tyr Pro
145 150 155 160

Asp Leu Lys Ile Lys Pro Ala Val Asn Gln Val Glu Leu His Pro Trp
165 170 175

Leu Thr Arg Glu Asp Ile Val Asp Tyr Ala Ala Lys His Gly Ile Ile
180 185 190

Ile Glu Ala Tyr Ser Pro Leu Val Arg Gly Gln Lys Met Asp Asp Pro
195 200 205

ES 2 602 808 T3

Thr Leu Val Lys Ile Ala Glu Lys Tyr Asn Lys Thr Pro Ala Gln Ile
210 215 220

Leu Ile Asn Trp Ser Leu Ser Lys Gly Tyr Ile Pro Leu Pro Asn Thr
225 230 235 240

Ser Gln Ile Ser Arg Leu Ala Ser Asn His Asp Ala Ala Gln Phe Glu
245 250 255

Leu Ser Lys Glu Asp Ile Gly Thr Leu Ser Ser Leu Asn Glu Arg Leu
260 265 270

His Thr Cys Trp Asp Pro Thr Leu Tyr Pro Leu Gly Asp Glu Lys
275 280 285

<210> 70

<211> 2963

5 <212> DNA

<213> Issatchenkia orientalis

<220>

<221> CDS

10 <222> (1001)..(1963)

<400> 70

ttttctctct gtctctttgt tttttttttc caatctgatt tgacgtgcaa ggcaaagaca	60
tcacatgttt gagaatggca agagaagggg cgtggttagtg tataccaagc cgggttagag	120
agtgtgattt tagagtgaat ccatccatga acacgagtag aggagatgta tgagcaaadc	180
cagggtgttt gtaatggtcc aagccgcaag gcggcgtaat ggaatgcaag aaacaagggg	240
cactaatgaa ggggtaagag gtgtctagtt gagaagtaca tartaaaaga tgaatagttg	300
agawgtacat rgtaaaagat gaatagttga gacaaatgaa ggtgtcaatg ttcctgataa	360
tgacactgca agraacaaat accgtgcagt tggaaggggg aaagagatgr ccgagataag	420
tggtgttgag gccaaaggat gttggaacct gctacaatag gagatggagc ggcctataac	480
tccggcgtgt ttgtgttgac agccctatac atcagccaat acgagagttt ggcatgtcct	540
ttaaagggtt tgctaccccc actcccgtaa tcatcgtaa aatcatcatc attgaaatca	600
ttataattaa cctcatcacc attcccacta ttatcacctt atattctcca ctccaggag	660
atgcatcggt gtaaagggca tggctgtttg tttattttac ccgacaagcc aataccaaga	720
gcggacaaac cgcacagaa tgcaacagaa ggttgagaa acgtgatgtc attttttccg	780
caaacggaga tctcgcacag cggtagagata taaaaggcgg agatgtggac accttcttta	840
tacaattccc ctctactga ttgttcata ttcctaacat ctagttacaa ctctgaacat	900

cataattatt ttaaaattct caaccctaact gcaattggat tgaactaaat caaattatat	960
caagttaaac caaactaaac tatttttaaag cttaaacaca atg tcc caa gtc tac Met Ser Gln Val Tyr 1 5	1015
gtt act ttg aac aac ggt atc aag atc cct caa gtt ggc ttt gga tgc Val Thr Leu Asn Asn Gly Ile Lys Ile Pro Gln Val Gly Phe Gly Cys 10 15 20	1063
tggtt aaa ctt gtc aat gag gtt gca gct gac caa atc tac gag gcc atc Trp Lys Leu Val Asn Glu Val Ala Ala Asp Gln Ile Tyr Glu Ala Ile 25 30 35	1111
aaa att gga tac aga ctg ttt gat ggt gcc caa gac tat ggt aac gag Lys Ile Gly Tyr Arg Leu Phe Asp Gly Ala Gln Asp Tyr Gly Asn Glu 40 45 50	1159
aag gag att ggc cag ggc atc aag aga gca att aag gag gga att gtc Lys Glu Ile Gly Gln Gly Val Lys Arg Ala Ile Lys Glu Gly Ile Val 55 60 65	1207
aag agg gaa gac cta gtt gta gtt tcc aag cta tgg aac agt ttc cac Lys Arg Glu Asp Leu Val Val Ser Lys Leu Trp Asn Ser Phe His 70 75 80 85	1255
gat ccc aag aat gtg gag gtt gcc atc aac aag gtt ttg tgc gac ttg Asp Pro Lys Asn Val Glu Val Ala Ile Asn Lys Val Leu Ser Asp Leu 90 95 100	1303
gac ttg gat tac ctt gac atc ttt tac att cat ttc cca att gcg caa Asp Leu Asp Tyr Leu Asp Ile Phe Tyr Ile His Phe Pro Ile Ala Gln 105 110 115	1351
aag ttt gtt cca att gag aag aag tac cca cct gga ttt tac tgt ggt Lys Phe Val Pro Ile Glu Lys Lys Tyr Pro Pro Gly Phe Tyr Cys Gly 120 125 130	1399
gaa aat gga tgg gaa ttt gaa gat gtt cct ctc tct gtt act tgg aag Glu Asn Gly Trp Glu Phe Glu Asp Val Pro Leu Ser Val Thr Trp Lys 135 140 145	1447
gca atg gag aat ttg gtt gac caa ggt aaa gtt aaa tgc att ggt atc Ala Met Glu Asn Leu Val Asp Gln Gly Lys Val Lys Ser Ile Gly Ile 150 155 160 165	1495
tca aac tgc aat ggt gcc cta gtt cag gat ttg cta agg agc gcc aga Ser Asn Cys Asn Gly Ala Leu Val Gln Asp Leu Leu Arg Ser Ala Arg 170 175 180	1543
atc aag cca cag ttg tta cag att gaa cac cat cca tac ctc gtt caa Ile Lys Pro Gln Leu Leu Gln Ile Glu His His Pro Tyr Leu Val Gln 185 190 195	1591
cca agg ttg gtt aaa tac gca cag gat aac ggc atc cat gtt gta gca Pro Arg Leu Val Lys Tyr Ala Gln Asp Asn Gly Ile His Val Val Ala 200 205 210	1639
tat tcc tcc ttt ggc ccg caa tca ttc ctt gaa ttg gac cat cca aag Tyr Ser Ser Phe Gly Pro Gln Ser Phe Leu Glu Leu Asp His Pro Lys 215 220 225	1687
gct aag gat act gtg tca ttg ttt gaa cac gac acc atc aag gag att	1735

ES 2 602 808 T3

Ala Lys Asp Thr Val Ser Leu Phe Glu His Asp Thr Ile Lys Glu Ile 230 235 240 245	
gca gcc aaa cat aat gtt tcc aca tcc aag gtg ttg ttg aga tgg gcc Ala Ala Lys His Asn Val Ser Thr Ser Lys Val Leu Leu Arg Trp Ala 250 255 260	1783
acc caa aat ggc gtt ttg gtt att cca aag tcc aac agg aag gag aga Thr Gln Asn Gly Val Leu Val Ile Pro Lys Ser Asn Arg Lys Glu Arg 265 270 275	1831
ctc ttg gag aac ttt tcg gtg aat gac ttc cag ttg gat gag gaa gat Leu Leu Glu Asn Phe Ser Val Asn Asp Phe Gln Leu Asp Glu Glu Asp 280 285 290	1879
atg aac aag atc acc ggt cta gac atg aat ctg aga ttc aac gac cca Met Asn Lys Ile Thr Gly Leu Asp Met Asn Leu Arg Phe Asn Asp Pro 295 300 305	1927
tgg aca tgg ggg gct gag att cca act ttt gtg taa atggtgttag Trp Thr Trp Gly Ala Glu Ile Pro Thr Phe Val 310 315 320	1973
tctgatctaa tgacaactaa ttacgcactt acgactgtaa tgcctttatt tttctttata	2033
tttcccagcg tgttgttctt tcaaataatc gatgagtata aattaatttt acaaagcaga	2093
aacaacagga tctttagaaa cgtcactgta aacatcgaat cttctttgaa cactgaaggg	2153
aatattttctt ctcgtttctt caacaacgtc cttcttcagt tctgcataaa cgatggtttc	2213
ctcatggccg gcctcaacga ggatctcacc atctggatcg accaccatgc tatggccata	2273
agcctgatag ccgcccctgtg gggttacgagc gggggaacac atcaacacgt agttttggtt	2333
gtcaatagct ctggcaacgg caaactttga ccagaattta ggacctgtca cggatttgaa	2393
tgcaccggga taagccataa taccagcgcc acgtctggct gcaatcatgg ccaattccgg	2453
gaacctgata tcatagcaaa tacctaagcc gaatctgggtg tcgattttctg gaatgtcgaa	2513
aactgtaacc ttgttgcccg gttttaaaga atcagactcc ttgaacgtga ttccgcccgg	2573
aatagaaatg tcaaagaggt gcaccttacg atgcttggtc acgatttccc cttggggatt	2633
gaaaacaaga gaggtgttgt agataaccgcc gtcattgtcg tcgatttccg gaatcgaacc	2693
tccaatgata gagacattgt actttttcgc ctgttcactt aaaaacgtgc tagtttcccc	2753
ctctgggata cgttctgcat aatttgcaaa ttggtctacg gcatatggag attggaaaca	2813
ttcaggtaga acaagaagtt gtggtttttg atcgtgttg atcgccctct cgatgaattg	2873
ggtcactttg gcgagattgg cttcttgtc tccaccacag tggaattgca gcagtgccac	2933
ttggagagtc ttggagagag taacggcaga	2963

<210> 71

<211> 320

5 <212> PRT

<213> Issatchenkia orientalis

<400> 71

ES 2 602 808 T3

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

ES 2 602 808 T3

245 250 255

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

Asn Arg Lys Glu Arg Leu Leu Glu Asn Phe Ser Val Asn Asp Phe Gln
275 280 285

Leu Asp Glu Glu Asp Met Asn Lys Ile Thr Gly Leu Asp Met Asn Leu
290 295 300

Arg Phe Asn Asp Pro Trp Thr Trp Gly Ala Glu Ile Pro Thr Phe Val
305 310 315 320

<210> 72

<211> 3966

5 <212> DNA

<213> Issatchenkia orientalis

<220>

<221> CDS

10 <222> (1201)..(2766)

<400> 72

aaagaacgag gttttactca aatccggtgt tgtgatggcc aagggagaat tggagacaga	60
ggataaaciaa cgggtgcaca gggccaagaa gagaaagcag cacatgcaac acaaggatga	120
ttctctgaag aagaagaagg tggaggtgaa gcagttatga agtgtaaaac acatatagcc	180
accctatagc tcgaattgtc taatctttgt acattattat atttctctct atctgtatag	240
gagagtagcc cacaacacag tccaacccaa cataggctcc aacatctcgt agattctgtt	300
tggtttcggc ctcagagttg tccaggaaag agcagcatat gacgccaacg tggttcatat	360
gagtcocctag tgtaacttaa tagcgcaaac tccgacaacg aggagacact gacgtgccct	420
gttcaaggaa ttacatacag ccacttgtcc gagatgtaga accattgggc aacacgactc	480
tctgtagcaa ggggcaaaga cactagtgtt tctcaaggat acacttgtag atagacaaac	540
tagttgataa acaaatactc aaacgcatat acgcttactg gccttaatac tacagtaaac	600
gtgctccccg aatccccgaa cacaaggtcg aaccggctac tctatccctg cctttgcccc	660
tcctccctag ccatttcacc ccgtcaatgc agacaccctc tttggacttc ccctgttggc	720
cctcctaagg ccccccccc cccaccccc cctttcttcc ccgcaatacc caaacttata	780
gccgcgtacc cgttttccca agtttggtgc tatttcgctt acacaccctc cacgtgctct	840
tacctctagc tcctctttct tgacacatca ccccccccc gtctccgaat accgaaaata	900
gaaaagcca ccccggggtg caaagttccg gagatctcgg ccctcagtg accggctcat	960
gccacgcact ggaacccac actgtaccgc ttaccggtga tgactctttg taatttttca	1020

gtggatgcta ctcaacaaaa accaaggggg gaagtgcaca tgcttactaa accccgggag	1080
aaggggaagtg tctggaattt acacttctag ggggggacat ttcttcacgt ggatataaaa	1140
gcaaacaccc attcctacca gagaggttac aattagactt ccaatatccc actttttata	1200
atg ttg tcc ctc tct aaa cag tca aga aac ttt ttc aaa ttg aac tat	1248
Met Leu Ser Leu Ser Lys Gln Ser Arg Asn Phe Phe Lys Leu Asn Tyr	
1 5 10 15	
ttt tca gtc acc caa ata gca aaa atg tct gca act tcc gtc act ttc	1296
Phe Ser Val Thr Gln Ile Ala Lys Met Ser Ala Thr Ser Val Thr Phe	
20 25 30	
cca att atc aac gaa act tac caa cag cca acc ggg ctt ttc atc aac	1344
Pro Ile Ile Asn Glu Thr Tyr Gln Gln Pro Thr Gly Leu Phe Ile Asn	
35 40 45	
aat gaa ttt gtt agt gca aag tca ggt aag act ttt gat gtt aac acc	1392
Asn Glu Phe Val Ser Ala Lys Ser Gly Lys Thr Phe Asp Val Asn Thr	
50 55 60	
cca att gat gag tct ctc att tgt aaa gtc caa cag gcc gat gct gaa	1440
Pro Ile Asp Glu Ser Leu Ile Cys Lys Val Gln Gln Ala Asp Ala Glu	
65 70 75 80	
gat gtt gaa att gcc gtt caa gca gca tct aaa gct tac aag act tgg	1488
Asp Val Glu Ile Ala Val Gln Ala Ala Ser Lys Ala Tyr Lys Thr Trp	
85 90 95	
aga ttt aca ccg cca aat gaa aga ggc aga tac ttg aac aaa ttg gcc	1536
Arg Phe Thr Pro Pro Asn Glu Arg Gly Arg Tyr Leu Asn Lys Leu Ala	
100 105 110	
gat ttg atg gac gaa aag aga gac tta ctt gcc aaa att gaa tcc ctt	1584
Asp Leu Met Asp Glu Lys Arg Asp Leu Leu Ala Lys Ile Glu Ser Leu	
115 120 125	
gat aat ggt aag gcc tta cat tgt gca aaa ttc gat gtc aat ctt gtc	1632
Asp Asn Gly Lys Ala Leu His Cys Ala Lys Phe Asp Val Asn Leu Val	
130 135 140	
att gaa tat ttc aga tac tgt gca ggt tac tgt gat aaa atc gat ggt	1680
Ile Glu Tyr Phe Arg Tyr Cys Ala Gly Tyr Cys Asp Lys Ile Asp Gly	
145 150 155 160	
aga aca att aca acc gat gta gaa cat ttt acc tac act aga aag gaa	1728
Arg Thr Ile Thr Thr Asp Val Glu His Phe Thr Tyr Thr Arg Lys Glu	
165 170 175	
cct tta ggt gtc tgt ggt gca att aca cct tgg aac ttc cca ttg ctg	1776
Pro Leu Gly Val Cys Gly Ala Ile Thr Pro Trp Asn Phe Pro Leu Leu	
180 185 190	
atg ttt gct tgg aaa atc ggc ccg gct tta gca acc ggt aat acc att	1824
Met Phe Ala Trp Lys Ile Gly Pro Ala Leu Ala Thr Gly Asn Thr Ile	
195 200 205	
atc ttg aag cct gcc agt gca aca cct cta tca aac ctc ttt act tgt	1872
Ile Leu Lys Pro Ala Ser Ala Thr Pro Leu Ser Asn Leu Phe Thr Cys	
210 215 220	

ES 2 602 808 T3

acc ttg atc aag gag gcg ggc att cca gcc ggt gtt gtt aat gtt gtt	1920
Thr Leu Ile Lys Glu Ala Gly Ile Pro Ala Gly Val Val Asn Val Val	
225 230 235 240	
cca ggt tcc ggt aga ggc tgt ggt aac tcc att tta caa cat cct aaa	1968
Pro Gly Ser Gly Arg Gly Cys Gly Asn Ser Ile Leu Gln His Pro Lys	
245 250 255	
att aag aag gtt gcg ttt acc gga tct aca gaa gtt ggt aaa act gtt	2016
Ile Lys Lys Val Ala Phe Thr Gly Ser Thr Glu Val Gly Lys Thr Val	
260 265 270	
atg aag gaa tgt gct aat tcc atc aaa aag gtt act ctg gaa ttg ggt	2064
Met Lys Glu Cys Ala Asn Ser Ile Lys Lys Val Thr Leu Glu Leu Gly	
275 280 285	
ggt aag tct cca aac att gtt ttc aaa gac tgt aac gtt gaa caa acc	2112
Gly Lys Ser Pro Asn Ile Val Phe Lys Asp Cys Asn Val Glu Gln Thr	
290 295 300	
att caa aat ttg att act ggt att ttc ttc aat ggt ggt gaa gtc tgt	2160
Ile Gln Asn Leu Ile Thr Gly Ile Phe Phe Asn Gly Gly Glu Val Cys	
305 310 315 320	
tgt gct ggt tct aga att tac att gaa gca acc gat gag aaa tgg tat	2208
Cys Ala Gly Ser Arg Ile Tyr Ile Glu Ala Thr Asp Glu Lys Trp Tyr	
325 330 335	
act gaa ttc ttg acc aaa ttc aag gag act gtt gaa aaa tta aag att	2256
Thr Glu Phe Leu Thr Lys Phe Lys Glu Thr Val Glu Lys Leu Lys Ile	
340 345 350	
ggt aac cca ttt gaa gag ggt gtt ttc caa ggt gca caa acc act cca	2304
Gly Asn Pro Phe Glu Glu Gly Val Phe Gln Gly Ala Gln Thr Thr Pro	
355 360 365	
gat caa ttc caa act gtc ttg gac tac atc acc gct gct aac gaa tcc	2352
Asp Gln Phe Gln Thr Val Leu Asp Tyr Ile Thr Ala Ala Asn Glu Ser	
370 375 380	
agc ttg aaa cta tta act ggt ggt aaa aga att ggc aat aag gga tac	2400
Ser Leu Lys Leu Leu Thr Gly Gly Lys Arg Ile Gly Asn Lys Gly Tyr	
385 390 395 400	
ttt gtt gag cca act atc ttc tac gat gtt cct caa aat tcc aag tta	2448
Phe Val Glu Pro Thr Ile Phe Tyr Asp Val Pro Gln Asn Ser Lys Leu	
405 410 415	
act caa gaa gaa atc ttt ggt cca gtt gct gtt gtt tta cct ttc aag	2496
Thr Gln Glu Glu Ile Phe Gly Pro Val Ala Val Val Leu Pro Phe Lys	
420 425 430	
tcc act gaa gaa ttg att gaa aag gca aat gat tcc gat ttt ggc tta	2544
Ser Thr Glu Glu Leu Ile Glu Lys Ala Asn Asp Ser Asp Phe Gly Leu	
435 440 445	
ggt tcc ggt att cac act gaa gat ttc aac aag gca att tgg gtt tcc	2592
Gly Ser Gly Ile His Thr Glu Asp Phe Asn Lys Ala Ile Trp Val Ser	
450 455 460	
gaa agg ctt gaa gca ggt tct gtt tgg atc aac act tac aat gat ttc	2640
Glu Arg Leu Glu Ala Gly Ser Val Trp Ile Asn Thr Tyr Asn Asp Phe	
465 470 475 480	

cac cca gct gct cca ttc ggt ggt tac aag gaa tcc ggt att ggc aga	2688
His Pro Ala Ala Pro Phe Gly Gly Tyr Lys Glu Ser Gly Ile Gly Arg	
485 490 495	
gaa atg ggt att gaa gct ttc gac aac tat act caa acc aag tta gtt	2736
Glu Met Gly Ile Glu Ala Phe Asp Asn Tyr Thr Gln Thr Lys Leu Val	
500 505 510	
aga gct aga gtt aac aag cca gct ttt tag atgccccggg tctaaaataa	2786
Arg Ala Arg Val Asn Lys Pro Ala Phe	
515 520	
aatcaatac caatcccctt gtttgcatta tgtattttga aaatatgtat ataggtcctc	2846
tcactttcct ttgacttttc taatctctac tattcagata attgatttat cagaatggtg	2906
tatttattta tattgtttggt cttaaaattt gaaataagta caataacaaa ataaaaata	2966
aagtttttca aagcttaaat ttcggagttg ccattttcgg ccaaataatga ttttctotta	3026
aaaagtttat acctagttat cttattttgtc atcagctaaa acataggaaa atctgccttt	3086
gagaaggcta cattctccat cttatcttaa tattctttgt agcagtcaaa tcagaagcaa	3146
tacagtgttt cattgagtta ctaagggatg ggtttttgtt gaagttaactc tagtatttta	3206
agacatctct aaaagatata atctctggtc ttaagaccaa gctgtataat tccccaaac	3266
tttgaataca aagtagtatg tgatgaaata attgtctcgg atgctgcaa acatggcctg	3326
aaaatcgggt gaaaattgca cctaaaagat cagcacatat atatatatga tactttaaac	3386
aatgtagatg ttaactataa cattgatatt ctcaaattgg ttctcaaaag tggaacacat	3446
ttgtcgttgc ttttgattgc ttgattgcaa ataactgtat atacctacgt caagcactaa	3506
tacaccgtct ttgtttcaag tgtgctaatt aaaattaatt atgcctgcat ataccaaaca	3566
ggttgtgact attgactggt gttcttcctt tccgggcaaa gtcttgattc tattaggaaa	3626
aattaaacaa ttattctgta cttgatgttt gtggaaatga taaacctaat agttatacta	3686
actgctttga gtataatggc aagaatcgga gtcaatgaat attatttttc tatcattagc	3746
cgtagtgata taatttgaaa aaggacaaat aactagaagt cacttttctg ggaaccatcc	3806
catcatagca tagatttgga tattacttgc ctcaatgcgt gaccgttact aaaatcccaa	3866
gactacttat actgtagtta tacggatcat ttaaattggc aactaaatct gcgagcaaaa	3926
attgttaagc tttgtaactg tgtagtttga agagtctgaa	3966

<210> 73

<211> 521

5 <212> PRT

<213> Issatchenkia orientalis

<400> 73

Met	Leu	Ser	Leu	Ser	Lys	Gln	Ser	Arg	Asn	Phe	Phe	Lys	Leu	Asn	Tyr
1				5				10					15		

ES 2 602 808 T3

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

ES 2 602 808 T3

260	265	270
Met Lys Glu Cys Ala Asn Ser Ile Lys Lys Val Thr Leu Glu Leu Gly		
275	280	285
Gly Lys Ser Pro Asn Ile Val Phe Lys Asp Cys Asn Val Glu Gln Thr		
290	295	300
Ile Gln Asn Leu Ile Thr Gly Ile Phe Phe Asn Gly Gly Glu Val Cys		
305	310	315
Cys Ala Gly Ser Arg Ile Tyr Ile Glu Ala Thr Asp Glu Lys Trp Tyr		
325	330	335
Thr Glu Phe Leu Thr Lys Phe Lys Glu Thr Val Glu Lys Leu Lys Ile		
340	345	350
Gly Asn Pro Phe Glu Glu Gly Val Phe Gln Gly Ala Gln Thr Thr Pro		
355	360	365
Asp Gln Phe Gln Thr Val Leu Asp Tyr Ile Thr Ala Ala Asn Glu Ser		
370	375	380
Ser Leu Lys Leu Leu Thr Gly Gly Lys Arg Ile Gly Asn Lys Gly Tyr		
385	390	395
Phe Val Glu Pro Thr Ile Phe Tyr Asp Val Pro Gln Asn Ser Lys Leu		
405	410	415
Thr Gln Glu Glu Ile Phe Gly Pro Val Ala Val Val Leu Pro Phe Lys		
420	425	430
Ser Thr Glu Glu Leu Ile Glu Lys Ala Asn Asp Ser Asp Phe Gly Leu		
435	440	445
Gly Ser Gly Ile His Thr Glu Asp Phe Asn Lys Ala Ile Trp Val Ser		
450	455	460
Glu Arg Leu Glu Ala Gly Ser Val Trp Ile Asn Thr Tyr Asn Asp Phe		
465	470	475
His Pro Ala Ala Pro Phe Gly Gly Tyr Lys Glu Ser Gly Ile Gly Arg		
485	490	495
Glu Met Gly Ile Glu Ala Phe Asp Asn Tyr Thr Gln Thr Lys Leu Val		
500	505	510
Arg Ala Arg Val Asn Lys Pro Ala Phe		
515	520	

- 5 <210> 74
 <211> 3179
 <212> DNA
 <213> Issatchenkia orientalis

<220>
<221> CDS
<222> (1052)..(2182)

5 <400> 74
gatttggacc tacaaggtgc tgtaaagagt atgaacactt ctggggagga ggaatggaac 60
agtgatgacg atgatgatga agaaagtgac gaaagtaacg aaagtgatta ctattcttac 120
gatgaaggcg aagaaacaga tgatagttag ggagcccaag agggagagga agacgaaaat 180
gaacgaatca ttgaagctct aagtagtggt gttggtgaac tcaagatgga ctcttttaggt 240
aattatattc ttgaatagtt gtgtaaagcg aatatgcaaa tagatttggt ttataattat 300
gcatctcttt gaaagaggtt tagaggcaaa gttcttgcac acaatattgt gattgtttta 360
atgtcattct tgattttcat aaagagatta aaaaaaaaaa aaaaaaactt ataaaattga 420
gtagaaccat ttatatataa gacaaagatt gtctgtatta gtcctcaaca cactaaacct 480
tacatactta gggtaaattt gctaatagag tgatatgttc atgagaactc caacgacaac 540
acaaccacct atttgcacaa caaacaccat tgcgcacgc tgcgcgccct agaagtagaa 600
agaaagggaa atgacattaa gagaatcata ccccgtagcc gtaacgccga aaaaatcaca 660
cccgtagccc cacaccttaa aacctcaacc gcttaacacc gccacacctt ttctctttat 720
aaacgcggtt tgcattactc attctcttta taaacgcgac cccccaaaac ggggaatagc 780
ttcaaccccc caatcagata tgagtttccc gggaaacccg cttttccgga cagccccaca 840
aggggttggt ctataaaaga ggacgttttc cccgtcatcg agattgaaga ttcttacagg 900
cccatttatt caaattggag ttgattcttc ttgtctttac tttctttctc tctttttctt 960
ccttttttaa tattatcttt tgtcaagcct ggttccttaa gttgaactct cttttcttgt 1020
gatcctocta tatagatacg ccttgccaaa t atg ttt gca tca acc ttc aga 1072
Met Phe Ala Ser Thr Phe Arg
1 5
agt caa gct gta aga gct gca aga ttt act aga ttc caa tcc act ttt 1120
Ser Gln Ala Val Arg Ala Ala Arg Phe Thr Arg Phe Gln Ser Thr Phe
10 15 20
gcc att cct gag aag caa atg ggt gtt atc ttt gaa act cat ggt ggt 1168
Ala Ile Pro Glu Lys Gln Met Gly Val Ile Phe Glu Thr His Gly Gly
25 30 35
cct tta caa tac aag gaa att cca gtt cca aaa cca aaa cca act gaa 1216
Pro Leu Gln Tyr Lys Glu Ile Pro Val Pro Lys Pro Lys Pro Thr Glu

ES 2 602 808 T3

40	45	50	55	
att tta atc aat gtt aaa tac tct ggt gtc tgc cat acc gat tta cac Ile Leu Ile Asn Val Lys Tyr Ser Gly Val Cys His Thr Asp Leu His 60 65 70				1264
gca tgg aaa ggt gac tgg cca tta cca gca aag tta ccc cta gtt ggt Ala Trp Lys Gly Asp Trp Pro Leu Pro Ala Lys Leu Pro Leu Val Gly 75 80 85				1312
ggt cac gaa ggt gcg ggc att gtt gtt gcg aaa ggt tct gca gtt acc Gly His Glu Gly Ala Gly Ile Val Ala Lys Gly Ser Ala Val Thr 90 95 100				1360
aac ttt gag att ggc gat tat gct ggt att aag tgg tta aac ggt tca Asn Phe Glu Ile Gly Asp Tyr Ala Gly Ile Lys Trp Leu Asn Gly Ser 105 110 115				1408
tgt atg tca tgt gaa ttc tgt gaa caa ggt gat gaa tct aac tgt gaa Cys Met Ser Cys Glu Phe Cys Glu Gln Gly Asp Glu Ser Asn Cys Glu 120 125 130 135				1456
cat gcc gat ttg agt ggt tat act cat gat ggt tct ttc caa caa tat His Ala Asp Leu Ser Gly Tyr Thr His Asp Gly Ser Phe Gln Gln Tyr 140 145 150				1504
gcc act gct gac gct att caa gct gca aag atc cca aag ggt acc gac Ala Thr Ala Asp Ala Ile Gln Ala Ala Lys Ile Pro Lys Gly Thr Asp 155 160 165				1552
tta tct gaa gtt gcg cca att tta tgt gct ggt gtt act gtc tat aaa Leu Ser Glu Val Ala Pro Ile Leu Cys Ala Gly Val Thr Val Tyr Lys 170 175 180				1600
gct ttg aaa act gct gat tta aga gca ggt caa tgg gtt gcg att tct Ala Leu Lys Thr Ala Asp Leu Arg Ala Gly Gln Trp Val Ala Ile Ser 185 190 195				1648
ggt gcc gct ggt ggt cta ggt tct ctt gct gtc caa tat gca aag gca Gly Ala Ala Gly Gly Leu Gly Ser Leu Ala Val Gln Tyr Ala Lys Ala 200 205 210 215				1696
atg ggt cta aga gtt tta ggt atc gat ggt ggt gaa ggt aaa aag gaa Met Gly Leu Arg Val Leu Gly Ile Asp Gly Gly Glu Gly Lys Lys Glu 220 225 230				1744
ctt ttt gaa caa tgt ggt ggt gat gtg ttt atc gat ttc acc aga tac Leu Phe Glu Gln Cys Gly Gly Asp Val Phe Ile Asp Phe Thr Arg Tyr 235 240 245				1792
cca aga gat gca cct gaa aag atg gtt gct gat att aag gct gca act Pro Arg Asp Ala Pro Glu Lys Met Val Ala Asp Ile Lys Ala Ala Thr 250 255 260				1840
aac ggt ttg ggt cca cac ggt gtt atc aat gtc tct gtc tcc cca gct Asn Gly Leu Gly Pro His Gly Val Ile Asn Val Ser Val Ser Pro Ala 265 270 275				1888
gct atc tct caa tca tgt gac tat gtt aga gca act ggt aag gtt gtc Ala Ile Ser Gln Ser Cys Asp Tyr Val Arg Ala Thr Gly Lys Val Val 280 285 290 295				1936
ctt gtc ggt atg cca tct ggt gct gtc tgt aag tct gat gtc ttc act				1984

ES 2 602 808 T3

Leu Val Gly Met Pro Ser Gly Ala Val Cys Lys Ser Asp Val Phe Thr	
300 305 310	
cat gtt gtt aaa tcc tta caa att aaa ggt tct tat gtt ggt aac aga	2032
His Val Val Lys Ser Leu Gln Ile Lys Gly Ser Tyr Val Gly Asn Arg	
315 320 325	
gca gat acc aga gaa gct ttg gaa ttc ttt aat gaa ggt aag gtc aga	2080
Ala Asp Thr Arg Glu Ala Leu Glu Phe Phe Asn Glu Gly Lys Val Arg	
330 335 340	
tct cca atc aag gtt gtc cca tta tct act tta cct gaa att tac gaa	2128
Ser Pro Ile Lys Val Val Pro Leu Ser Thr Leu Pro Glu Ile Tyr Glu	
345 350 355	
ttg atg gag caa ggt aag att tta ggt aga tac gtt gtt gat act tct	2176
Leu Met Glu Gln Gly Lys Ile Leu Gly Arg Tyr Val Val Asp Thr Ser	
360 365 370 375	
aaa taa tgaagatgaa gaaaacagca aactttttat gactaccccc aaccatctaa	2232
Lys	
cgatttatga tctatatata gctttctaga acatccattt atttattcac ttactcatgt	2292
atztatatta tataatacaa aataactaat tacaatgtgt acattttttt ttttcattac	2352
cataatgtat gcgttgagcc tcttgcacct tctttattag gaaatcagtt gaaaaatttc	2412
cggattgtct ttattattgg cccatttttt tttggtcaca cctttatttt tgtacacttc	2472
tcgggcaaag caaaaactat agtacgggat aggcctttat aaaactccag tgtgtatgat	2532
tttagttggt gtgccatcta cacgttctct tagtttcttt atcatgtcac agaaagcaag	2592
catgcaaac cttacaaaaa ataacaacat acaaatgcct aaacaactgg actataatga	2652
tggtgagtca gttacgaaaa gagcaagtgg gttaatacga tttcgtaagg gacagtctga	2712
ggaagactac aattttcaaa aggagcagtt ctgggtccacg ggtccttttag tacagaatca	2772
cacatttgtg actgaatttg ttgaaaagtt tattgaaaac acaattagtg aagattattc	2832
aatcacagat agatcgaaaa tagaacgtga aacaatcata cacggattgg agaagctgta	2892
ttttcaaagg gaatatgagc gatgtctaaa agatgttcaa ctattgaagg acaatatcga	2952
taagttcaat cctaatttgg atcttaatga aaagaattta taatgagctg aattatattt	3012
cttggtatgt catcaaaaag atccatgaga gtaacgaaaa gaaactgggg gaaatcta	3072
aatttacaat ttcaatatac acttctatat cttttaatgt aatggcttta taaataaaca	3132
cgaacttcta cagcaccgac gtttcttttt cttaccagct cctcttc	3179

<210> 75

<211> 376

<212> PRT

<213> Issatchenkia orientalis

<400> 75

ES 2 602 808 T3

Met Phe Ala Ser Thr Phe Arg Ser Gln Ala Val Arg Ala Ala Arg Phe
1 5 10 15

Thr Arg Phe Gln Ser Thr Phe Ala Ile Pro Glu Lys Gln Met Gly Val
20 25 30

Ile Phe Glu Thr His Gly Gly Pro Leu Gln Tyr Lys Glu Ile Pro Val
35 40 45

Pro Lys Pro Lys Pro Thr Glu Ile Leu Ile Asn Val Lys Tyr Ser Gly
50 55 60

Val Cys His Thr Asp Leu His Ala Trp Lys Gly Asp Trp Pro Leu Pro
65 70 75 80

Ala Lys Leu Pro Leu Val Gly Gly His Glu Gly Ala Gly Ile Val Val
85 90 95

Ala Lys Gly Ser Ala Val Thr Asn Phe Glu Ile Gly Asp Tyr Ala Gly
100 105 110

Ile Lys Trp Leu Asn Gly Ser Cys Met Ser Cys Glu Phe Cys Glu Gln
115 120 125

Gly Asp Glu Ser Asn Cys Glu His Ala Asp Leu Ser Gly Tyr Thr His
130 135 140

Asp Gly Ser Phe Gln Gln Tyr Ala Thr Ala Asp Ala Ile Gln Ala Ala
145 150 155 160

Lys Ile Pro Lys Gly Thr Asp Leu Ser Glu Val Ala Pro Ile Leu Cys
165 170 175

Ala Gly Val Thr Val Tyr Lys Ala Leu Lys Thr Ala Asp Leu Arg Ala
180 185 190

Gly Gln Trp Val Ala Ile Ser Gly Ala Ala Gly Gly Leu Gly Ser Leu
195 200 205

Ala Val Gln Tyr Ala Lys Ala Met Gly Leu Arg Val Leu Gly Ile Asp
210 215 220

Gly Gly Glu Gly Lys Lys Glu Leu Phe Glu Gln Cys Gly Gly Asp Val
225 230 235 240

Phe Ile Asp Phe Thr Arg Tyr Pro Arg Asp Ala Pro Glu Lys Met Val
245 250 255

Ala Asp Ile Lys Ala Ala Thr Asn Gly Leu Gly Pro His Gly Val Ile
260 265 270

Asn Val Ser Val Ser Pro Ala Ala Ile Ser Gln Ser Cys Asp Tyr Val
275 280 285

Arg Ala Thr Gly Lys Val Val Leu Val Gly Met Pro Ser Gly Ala Val
290 295 300

Cys Lys Ser Asp Val Phe Thr His Val Val Lys Ser Leu Gln Ile Lys
305 310 315 320

Gly Ser Tyr Val Gly Asn Arg Ala Asp Thr Arg Glu Ala Leu Glu Phe
325 330 335

Phe Asn Glu Gly Lys Val Arg Ser Pro Ile Lys Val Val Pro Leu Ser
340 345 350

Thr Leu Pro Glu Ile Tyr Glu Leu Met Glu Gln Gly Lys Ile Leu Gly
355 360 365

Arg Tyr Val Val Asp Thr Ser Lys
370 375

<210> 76
<211> 696
5 <212> DNA
<213> Escherichia coli

<220>
<221> CDS
10 <222> (1)..(696)

<400> 76
atg tta gaa gat ctc aaa cgc cag gta tta gaa gcc aac ctg gcg ctg 48
Met Leu Glu Asp Leu Lys Arg Gln Val Leu Glu Ala Asn Leu Ala Leu
1 5 10 15

cca aaa cac aac ctg gtc acg ctc aca tgg ggc aac gtc agc gcc gtt 96
Pro Lys His Asn Leu Val Thr Leu Thr Trp Gly Asn Val Ser Ala Val
20 25 30

gat cgc gag cgc ggc gtc ttt gtg atc aaa cct tcc ggc gtc gat tac 144
Asp Arg Glu Arg Gly Val Phe Val Ile Lys Pro Ser Gly Val Asp Tyr
35 40 45

agc gtc atg acc gct gac gat atg gtc gtg gtt agc atc gaa acc ggt 192
Ser Val Met Thr Ala Asp Asp Met Val Val Val Ser Ile Glu Thr Gly
50 55 60

gaa gtg gtt gaa ggt acg aaa aag ccc tcc tcc gac acg cca act cac 240
Glu Val Val Glu Gly Thr Lys Lys Pro Ser Ser Asp Thr Pro Thr His
65 70 75 80

ES 2 602 808 T3

cgg ctg ctc tat cag gca ttc ccc tcc att ggc ggc att gtg cat acg 288
 Arg Leu Leu Tyr Gln Ala Phe Pro Ser Ile Gly Gly Ile Val His Thr
 85 90 95
 cac tcg cgc cac gcc acc atc tgg gcg cag gcg ggt cag tcg att cca 336
 His Ser Arg His Ala Thr Ile Trp Ala Gln Ala Gly Gln Ser Ile Pro
 100 105 110
 gca acc ggc acc acc cac gcc gac tat ttc tac ggc acc att ccc tgc 384
 Ala Thr Gly Thr Thr His Ala Asp Tyr Phe Tyr Gly Thr Ile Pro Cys
 115 120 125
 acc cgc aaa atg acc gac gca gaa atc aac ggc gaa tat gag tgg gaa 432
 Thr Arg Lys Met Thr Asp Ala Glu Ile Asn Gly Glu Tyr Glu Trp Glu
 130 135 140
 acc ggt aac gtc atc gta gaa acc ttt gaa aaa cag ggt atc gat gca 480
 Thr Gly Asn Val Ile Val Glu Thr Phe Glu Lys Gln Gly Ile Asp Ala
 145 150 155 160
 gcg caa atg ccc ggc gtt ctg gtc cat tcc cac ggc ccg ttt gca tgg 528
 Ala Gln Met Pro Gly Val Leu Val His Ser His Gly Pro Phe Ala Trp
 165 170 175
 ggc aaa aat gcc gaa gat gcg gtg cat aac gcc atc gtg ctg gaa gag 576
 Gly Lys Asn Ala Glu Asp Ala Val His Asn Ala Ile Val Leu Glu Glu
 180 185 190
 gtc gct tat atg ggg ata ttc tgc cgt cag tta gcg ccg cag tta ccg 624
 Val Ala Tyr Met Gly Ile Phe Cys Arg Gln Leu Ala Pro Gln Leu Pro
 195 200 205
 gat atg cag caa acg ctg ctg gat aaa cac tat ctg cgt aag cat ggc 672
 Asp Met Gln Gln Thr Leu Leu Asp Lys His Tyr Leu Arg Lys His Gly
 210 215 220
 gcg aag gca tat tac ggg cag taa 696
 Ala Lys Ala Tyr Tyr Gly Gln
 225 230

<210> 77
 <211> 231
 <212> PRT
 <213> Escherichia coli

<400> 77

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 Val Met Thr Ala Asp Asp Met Val Val Val Ser Ile Glu Thr Gly
 50 55 60

ES 2 602 808 T3

Glu Val Val Glu Gly Thr 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 Asp Lys His Tyr Leu Arg Lys His Gly
210 215 220

Ala Lys Ala Tyr Tyr Gly Gln
225 230

<210> 78

<211> 729

5 <212> DNA

<213> Secuencia Artificial

<220>

10 <223> Codón optimizado L. plantarum araD

<220>

<221> CDS

<222> (1)..(729)

15 <400> 78

atg ttg gaa gca tta aag caa gaa gtt tac gaa gct aat atg cag ttg
Met Leu Glu Ala Leu Lys Gln Glu Val Tyr Glu Ala Asn Met Gln Leu
1 5 10 15

48

ES 2 602 808 T3

cca aaa tta ggt tta gtt act ttc acc tgg ggt aat gtt tct ggt att	96
Pro Lys Leu Gly Leu Val Thr Phe Thr Trp Gly Asn Val Ser Gly Ile	
20 25 30	
gat aga gag aaa ggt ttg ttc gtt atc aag cca tca ggt gtc gac tat	144
Asp Arg Glu Lys Gly Leu Phe Val Ile Lys Pro Ser Gly Val Asp Tyr	
35 40 45	
ggt gaa ctt aag cca tcc gat tta gtt gtt gtc aat tta caa ggt gaa	192
Gly Glu Leu Lys Pro Ser Asp Leu Val Val Val Asn Leu Gln Gly Glu	
50 55 60	
gtt gtt gaa ggt aaa tta aac cca tct tct gat aca cct acc cac act	240
Val Val Glu Gly Lys Leu Asn Pro Ser Ser Asp Thr Pro Thr His Thr	
65 70 75 80	
gtt tta tat aac gca ttt cca aac atc ggt ggt atc gtt cat acc cat	288
Val Leu Tyr Asn Ala Phe Pro Asn Ile Gly Gly Ile Val His Thr His	
85 90 95	
tcc cca tgg gcc gtc gca tac gca gct gca caa atg gac gtc cct gca	336
Ser Pro Trp Ala Val Ala Tyr Ala Ala Ala Gln Met Asp Val Pro Ala	
100 105 110	
atg aac acc acc cac gct gat acc ttc tac ggt gat gtc cct gca gct	384
Met Asn Thr Thr His Ala Asp Thr Phe Tyr Gly Asp Val Pro Ala Ala	
115 120 125	
gac gca ttg act aag gaa gaa att gaa gca gac tac gag ggt aat acc	432
Asp Ala Leu Thr Lys Glu Glu Ile Glu Ala Asp Tyr Glu Gly Asn Thr	
130 135 140	
ggt aaa act att gtc aag acc ttt caa gaa aga ggt tta gat tac gaa	480
Gly Lys Thr Ile Val Lys Thr Phe Gln Glu Arg Gly Leu Asp Tyr Glu	
145 150 155 160	
gcc gtt cca gct tct ttg gtc tct caa cat ggt cct ttc gca tgg ggt	528
Ala Val Pro Ala Ser Leu Val Ser Gln His Gly Pro Phe Ala Trp Gly	
165 170 175	
cca aca cca gct aaa gca gtt tat aat gct aag gtt tta gag gtt gtc	576
Pro Thr Pro Ala Lys Ala Val Tyr Asn Ala Lys Val Leu Glu Val Val	
180 185 190	
gca gag gaa gac tat cac acc gca caa ttg acc aga gct tcc tcc gaa	624
Ala Glu Glu Asp Tyr His Thr Ala Gln Leu Thr Arg Ala Ser Ser Glu	
195 200 205	
ctt cca caa tac ttg tta gat aag cac tat tta aga aag cac ggt gct	672
Leu Pro Gln Tyr Leu Leu Asp Lys His Tyr Leu Arg Lys His Gly Ala	
210 215 220	
tct gct tat tac ggt caa aac aat gca cat tcc aag gat cac gct gtt	720
Ser Ala Tyr Tyr Gly Gln Asn Asn Ala His Ser Lys Asp His Ala Val	
225 230 235 240	
aga aag taa	729
Arg Lys	

<210> 79

<211> 242

5 <212> PRT

<213> Secuencia Artificial

<220>

<223> Construcción Sintética

10

<400> 79

ES 2 602 808 T3

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> 80
- <211> 1425
- <212> DNA
- <213> Secuencia Artificial

10 <220>

<223> Codón optimizado de L. plantarum araA

<220>

<221> CDS

5 <222> (1)..(1425)

<400> 80

atg tta tcc gtt cca gat tac gaa ttt tgg ttc gtc act ggt tct caa	48
Met Leu Ser Val Pro Asp Tyr Glu Phe Trp Phe Val Thr Gly Ser Gln	
1 5 10 15	
cac ttg tac ggt gaa gaa caa ctt aag tct gtc gct aag gat gca caa	96
His Leu Tyr Gly Glu Glu Gln Leu Lys Ser Val Ala Lys Asp Ala Gln	
20 25 30	
gac atc gca gac aag tta aac gct tct ggt aag ttg cca tac aag gtt	144
Asp Ile Ala Asp Lys Leu Asn Ala Ser Gly Lys Leu Pro Tyr Lys Val	
35 40 45	
gtc ttc aag gat gtt atg act acc gca gaa tct atc act aac ttt atg	192
Val Phe Lys Asp Val Met Thr Ala Glu Ser Ile Thr Asn Phe Met	
50 55 60	
aag gaa gtt aat tac aac gat aag gtt gct ggt gtt att acc tgg atg	240
Lys Glu Val Asn Tyr Asn Asp Lys Val Ala Gly Val Ile Thr Trp Met	
65 70 75 80	
cat acc ttc tct cca gct aaa aac tgg att aga ggt aca gaa ttg ttg	288
His Thr Phe Ser Pro Ala Lys Asn Trp Ile Arg Gly Thr Glu Leu Leu	
85 90 95	
caa aag cca ttg tta cac ctt gct act caa tat ttg aac aac att cca	336
Gln Lys Pro Leu Leu His Leu Ala Thr Gln Tyr Leu Asn Asn Ile Pro	
100 105 110	
tac gca gat att gat ttt gat tat atg aac ttg aac caa tct gct cac	384
Tyr Ala Asp Ile Asp Phe Asp Tyr Met Asn Leu Asn Gln Ser Ala His	
115 120 125	
ggt gac aga gaa tat gca tat atc aat gca aga tta caa aag cat aac	432
Gly Asp Arg Glu Tyr Ala Tyr Ile Asn Ala Arg Leu Gln Lys His Asn	
130 135 140	
aag att gtt tac ggt tat tgg ggt gat gaa gat gtt caa gaa caa att	480
Lys Ile Val Tyr Gly Tyr Trp Gly Asp Glu Asp Val Gln Glu Gln Ile	
145 150 155 160	

ES 2 602 808 T3

gct aga tgg gaa gat gtt gca gtt gca tac aat gaa tct ttt aag gtt Ala Arg Trp Glu Asp Val Ala Val Ala Tyr Asn Glu Ser Phe Lys Val 165 170 175	528
aag gtc gca aga ttc ggt gat act atg aga aac gtc gca gtt act gaa Lys Val Ala Arg Phe Gly Asp Thr Met Arg Asn Val Ala Val Thr Glu 180 185 190	576
ggt gat aag gtt gaa gct caa att aag atg ggt tgg act gtt gac tac Gly Asp Lys Val Glu Ala Gln Ile Lys Met Gly Trp Thr Val Asp Tyr 195 200 205	624
tac ggt att ggt gat ttg gtt gag gaa att aat aag gtt tcc gat gct Tyr Gly Ile Gly Asp Leu Val Glu Glu Ile Asn Lys Val Ser Asp Ala 210 215 220	672
gat gtc gat aag gaa tac gca gat tta gaa tcc aga tac gaa atg gtc Asp Val Asp Lys Glu Tyr Ala Asp Leu Glu Ser Arg Tyr Glu Met Val 225 230 235 240	720
caa ggt gat aat gac gcc gat act tac aag cat tcc gtc aga gtt caa Gln Gly Asp Asn Asp Ala Asp Thr Tyr Lys His Ser Val Arg Val Gln 245 250 255	768
ctt gct caa tac tta ggt atc aag aga ttt ttg gaa aga ggt ggt tac Leu Ala Gln Tyr Leu Gly Ile Lys Arg Phe Leu Glu Arg Gly Gly Tyr 260 265 270	816
act gct ttc act act aat ttc gaa gat tta tgg ggt atg gaa caa tta Thr Ala Phe Thr Thr Asn Phe Glu Asp Leu Trp Gly Met Glu Gln Leu 275 280 285	864
cca ggt ctt gca tcc caa tta tta att aga gat ggt tat ggt ttc ggt Pro Gly Leu Ala Ser Gln Leu Leu Ile Arg Asp Gly Tyr Gly Phe Gly 290 295 300	912
gca gaa ggt gac tgg aag act gca gca tta ggt aga gtc atg aag att Ala Glu Gly Asp Trp Lys Thr Ala Ala Leu Gly Arg Val Met Lys Ile 305 310 315 320	960
atg tct cac aac aaa caa acc gct ttt atg gag gat tat act tta gat Met Ser His Asn Lys Gln Thr Ala Phe Met Glu Asp Tyr Thr Leu Asp 325 330 335	1008
tta aga cac ggt cac gaa gca att tta ggt tcc cac atg ctt gaa gtt Leu Arg His Gly His Glu Ala Ile Leu Gly Ser His Met Leu Glu Val 340 345 350	1056
gat cca tcc att gct tct gat aaa cca aga gtt gaa gtt cac cca tta Asp Pro Ser Ile Ala Ser Asp Lys Pro Arg Val Glu Val His Pro Leu 355 360 365	1104
gat atc ggt ggt aaa gac gac cca gct aga tta gtt ttt act ggt tct Asp Ile Gly Gly Lys Asp Asp Pro Ala Arg Leu Val Phe Thr Gly Ser 370 375 380	1152
gag ggt gaa gct att gat gtt act gtc gca gac ttt aga gat ggt ttt Glu Gly Glu Ala Ile Asp Val Thr Val Ala Asp Phe Arg Asp Gly Phe 385 390 395 400	1200
aag atg att tcc tac gct gtt gac gca aac aag cca gag gct gaa aca Lys Met Ile Ser Tyr Ala Val Asp Ala Asn Lys Pro Glu Ala Glu Thr	1248

ES 2 602 808 T3

	405	410	415	
	cca aat ttg cca gtc gcc aag caa ttg tgg acc cca aag atg ggt tta			1296
	Pro Asn Leu Pro Val Ala Lys Gln Leu Trp Thr Pro Lys Met Gly Leu			
	420	425	430	
	aag aag ggt gct tta gaa tgg atg caa gct ggt ggt ggt cac cat act			1344
	Lys Lys Gly Ala Leu Glu Trp Met Gln Ala Gly Gly Gly His His Thr			
	435	440	445	
	atg tta tct ttc tct tta act gaa gaa caa atg gaa gat tat gct act			1392
	Met Leu Ser Phe Ser Leu Thr Glu Glu Gln Met Glu Asp Tyr Ala Thr			
	450	455	460	
	atg gtt ggt atg acc aaa gcc ttc tta aag tag			1425
	Met Val Gly Met Thr Lys Ala Phe Leu Lys			
	465	470		

<210> 81
 <211> 474
 <212> PRT
 <213> Secuencia Artificial

<220>
 <223> Construcción Sintética

<400> 81

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				

ES 2 602 808 T3

Lys Ile Val Tyr Gly Tyr Trp Gly Asp Glu Asp Val Gln Glu Gln Ile
 145 150 155 160
 Ala Arg Trp Glu Asp Val Ala Val Ala Tyr Asn Glu Ser Phe Lys Val
 165 170 175
 Lys Val Ala Arg Phe Gly Asp Thr Met Arg Asn Val Ala Val Thr Glu
 180 185 190
 Gly Asp Lys Val Glu Ala Gln Ile Lys Met Gly Trp Thr Val Asp Tyr
 195 200 205
 Tyr Gly Ile Gly Asp Leu Val Glu Glu Ile Asn Lys Val Ser Asp Ala
 210 215 220
 Asp Val Asp Lys Glu Tyr Ala Asp Leu Glu Ser Arg Tyr Glu Met Val
 225 230 235 240
 Gln Gly Asp Asn Asp Ala Asp Thr Tyr Lys His Ser Val Arg Val Gln
 245 250 255
 Leu Ala Gln Tyr Leu Gly Ile Lys Arg Phe Leu Glu Arg Gly Gly Tyr
 260 265 270
 Thr Ala Phe Thr Thr Asn Phe Glu Asp Leu Trp Gly Met Glu Gln Leu
 275 280 285
 Pro Gly Leu Ala Ser Gln Leu Leu Ile Arg Asp Gly Tyr Gly Phe Gly
 290 295 300
 Ala Glu Gly Asp Trp Lys Thr Ala Ala Leu Gly Arg Val Met Lys Ile
 305 310 315 320
 Met Ser His Asn Lys Gln Thr Ala Phe Met Glu Asp Tyr Thr Leu Asp
 325 330 335
 Leu Arg His Gly His Glu Ala Ile Leu Gly Ser His Met Leu Glu Val
 340 345 350
 Asp Pro Ser Ile Ala Ser Asp Lys Pro Arg Val Glu Val His Pro Leu
 355 360 365
 Asp Ile Gly Gly Lys Asp Asp Pro Ala Arg Leu Val Phe Thr Gly Ser
 370 375 380
 Glu Gly Glu Ala Ile Asp Val Thr Val Ala Asp Phe Arg Asp Gly Phe

ES 2 602 808 T3

[illegible]

<210> 82
<211> 1482
<212> DNA
<213> Secuencia Artificial

<220>
<223> Codón optimizado de *B. licheniformis* araA

<220>
<221> CDS
<222> (1)..(1482)

<400> 82																	
atg	atc	caa	gca	aag	acc	cac	gtc	ttc	tgg	ttc	gtc	aca	ggt	tcc	caa		48
Met	Ile	Gln	Ala	Lys	Thr	His	Val	Phe	Trp	Phe	Val	Thr	Gly	Ser	Gln		
1				5				10						15			
cac	tta	tac	ggt	gaa	gaa	gca	gtt	caa	gaa	gtc	gaa	gaa	cac	tcc	aag		96
His	Leu	Tyr	Gly	Glu	Glu	Ala	Val	Gln	Glu	Val	Glu	Glu	His	Ser	Lys		
			20					25					30				
atg	atc	tgt	aac	ggt	tta	aac	gat	ggt	gat	tta	aga	ttt	caa	ggt	gaa		144
Met	Ile	Cys	Asn	Gly	Leu	Asn	Asp	Gly	Asp	Leu	Arg	Phe	Gln	Val	Glu		
		35					40					45					
tac	aag	gct	gtc	gct	act	tcc	tta	gac	ggt	gtc	aga	aag	ttg	ttc	gaa		192
Tyr	Lys	Ala	Val	Ala	Thr	Ser	Leu	Asp	Gly	Val	Arg	Lys	Leu	Phe	Glu		
	50					55					60						
gaa	gct	aac	aga	gat	gaa	gaa	tgt	gcc	ggt	att	att	acc	tgg	atg	cat		240
Glu	Ala	Asn	Arg	Asp	Glu	Glu	Cys	Ala	Gly	Ile	Ile	Thr	Trp	Met	His		
65					70					75					80		
acc	ttc	tct	cct	gca	aag	atg	tgg	att	cca	ggt	tta	tct	gaa	ttg	aac		288
Thr	Phe	Ser	Pro	Ala	Lys	Met	Trp	Ile	Pro	Gly	Leu	Ser	Glu	Leu	Asn		
				85					90					95			
aag	cca	tta	ttg	cac	ttt	cac	act	caa	ttt	aac	aga	gac	atc	cct	tgg		336
Lys	Pro	Leu	Leu	His	Phe	His	Thr	Gln	Phe	Asn	Arg	Asp	Ile	Pro	Trp		

ES 2 602 808 T3

100	105	110	
gac aag att gat atg gac ttc atg aat atc aac caa tct gct cac ggt Asp Lys Ile Asp Met Asp Phe Met Asn Ile Asn Gln Ser Ala His Gly 115 120 125			384
gac aga gaa tac ggt ttt atc ggt gct aga ttg ggt atc cca aga aag Asp Arg Glu Tyr Gly Phe Ile Gly Ala Arg Leu Gly Ile Pro Arg Lys 130 135 140			432
gtc atc gcc ggt tac tgg gaa gat aga gaa gtt aaa aga tcc att gat Val Ile Ala Gly Tyr Trp Glu Asp Arg Glu Val Lys Arg Ser Ile Asp 145 150 155 160			480
aaa tgg atg tcc gcc gct gtt gct tac atc gaa tca aga cac att aag Lys Trp Met Ser Ala Ala Val Ala Tyr Ile Glu Ser Arg His Ile Lys 165 170 175			528
gtt gct aga ttt ggt gat aat atg aga aat gtt gct gtt act gaa ggt Val Ala Arg Phe Gly Asp Asn Met Arg Asn Val Ala Val Thr Glu Gly 180 185 190			576
gat aag att gaa gct caa att caa ctt ggt tgg tca gtt gat ggc tac Asp Lys Ile Glu Ala Gln Ile Gln Leu Gly Trp Ser Val Asp Gly Tyr 195 200 205			624
ggt att ggt gat tta gtt acc gaa atc aat gca gtc tct gaa caa tcc Gly Ile Gly Asp Leu Val Thr Glu Ile Asn Ala Val Ser Glu Gln Ser 210 215 220			672
tta tcc gaa tta att tcc gaa tac gaa gaa tta tac gag tgg cca gaa Leu Ser Glu Leu Ile Ser Glu Tyr Glu Leu Tyr Glu Trp Pro Glu 225 230 235 240			720
ggt gaa gct gct aga gaa tct gtt aag gaa caa gca aga att gaa ctt Gly Glu Ala Ala Arg Glu Ser Val Lys Glu Gln Ala Arg Ile Glu Leu 245 250 255			768
ggt tta aag aga ttt ctt tct tct ggt ggt tat acc gca ttt acc act Gly Leu Lys Arg Phe Leu Ser Ser Gly Gly Tyr Thr Ala Phe Thr Thr 260 265 270			816
acc ttc gaa gac tta cac ggt atg aag caa tta cct ggt tta gca gtt Thr Phe Glu Asp Leu His Gly Met Lys Gln Leu Pro Gly Leu Ala Val 275 280 285			864
caa aga ctt atg gct gaa ggt tac ggt ttc ggt ggt gaa ggt gac tgg Gln Arg Leu Met Ala Glu Gly Tyr Gly Phe Gly Gly Glu Gly Asp Trp 290 295 300			912
aag acc gct gct ttg gtc aga atg atg aag atg atg gct ggt ggt aag Lys Thr Ala Ala Leu Val Arg Met Met Lys Met Met Ala Gly Gly Lys 305 310 315 320			960
gaa act tct ttc atg gaa gat tac acc tac cac ttc gaa cca ggt aat Glu Thr Ser Phe Met Glu Asp Tyr Thr Tyr His Phe Glu Pro Gly Asn 325 330 335			1008
gaa atg att tta ggt tcc cac atg tta gaa gtt tgc cct tct atc gct Glu Met Ile Leu Gly Ser His Met Leu Glu Val Cys Pro Ser Ile Ala 340 345 350			1056
gag cac aag cca aga atc gaa gtt cat cct tta tct atg ggt gca aag			1104

ES 2 602 808 T3

Glu His Lys Pro Arg Ile Glu Val His Pro Leu Ser Met Gly Ala Lys	
355 360 365	
gat gac cca gct aga tta gtc ttc gat ggt att gca ggt cca gca gtt	1152
Asp Asp Pro Ala Arg Leu Val Phe Asp Gly Ile Ala Gly Pro Ala Val	
370 375 380	
aac gtc tcc tta atc gac ctt ggt ggt aga ttc aga tta gtt att aat	1200
Asn Val Ser Leu Ile Asp Leu Gly Gly Arg Phe Arg Leu Val Ile Asn	
385 390 395 400	
aag gtc gaa gct gtt aag gtt cca cat gat atg cct aat cta cca gtt	1248
Lys Val Glu Ala Val Lys Val Pro His Asp Met Pro Asn Leu Pro Val	
405 410 415	
gct aga gtt tta tgg aag cca caa cct tcc ttg aga act tcc gct gaa	1296
Ala Arg Val Leu Trp Lys Pro Gln Pro Ser Leu Arg Thr Ser Ala Glu	
420 425 430	
gcc tgg att ttg gca ggt ggt gcc cat cac acc tgt ttg tct tat caa	1344
Ala Trp Ile Leu Ala Gly Gly Ala His His Thr Cys Leu Ser Tyr Gln	
435 440 445	
ttg act gca gaa caa atg tta gac tgg gct gaa atg tcc ggt att gaa	1392
Leu Thr Ala Glu Gln Met Leu Asp Trp Ala Glu Met Ser Gly Ile Glu	
450 455 460	
gct gtc tta atc aac aga gat act act att ctt aat ctt aga aac gaa	1440
Ala Val Leu Ile Asn Arg Asp Thr Thr Ile Leu Asn Leu Arg Asn Glu	
465 470 475 480	
ctt aag tgg tcc gaa gca gct tac aga ttg aga aag ttt tag	1482
Leu Lys Trp Ser Glu Ala Ala Tyr Arg Leu Arg Lys Phe	
485 490	

<210> 83

<211> 493

5 <212> PRT

<213> Secuencia Artificial

<220>

<223> Construcción Sintética

10

<400> 83

Met Ile Gln Ala Lys Thr His Val Phe Trp Phe Val Thr Gly Ser Gln
1 5 10 15

His Leu Tyr Gly Glu Glu Ala Val Gln Glu Val Glu Glu His Ser Lys
20 25 30

Met Ile Cys Asn Gly Leu Asn Asp Gly Asp Leu Arg Phe Gln Val Glu
35 40 45

Tyr Lys Ala Val Ala Thr Ser Leu Asp Gly Val Arg Lys Leu Phe Glu
50 55 60

Glu Ala Asn Arg Asp Glu Glu Cys Ala Gly Ile Ile Thr Trp Met His

ES 2 602 808 T3

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

ES 2 602 808 T3

Glu Thr Ser Phe Met Glu Asp Tyr Thr Tyr His Phe Glu Pro Gly Asn
325 330 335

Glu Met Ile Leu Gly Ser His Met Leu Glu Val Cys Pro Ser Ile Ala
340 345 350

Glu His Lys Pro Arg Ile Glu Val His Pro Leu Ser Met Gly Ala Lys
355 360 365

Asp Asp Pro Ala Arg Leu Val Phe Asp Gly Ile Ala Gly Pro Ala Val
370 375 380

Asn Val Ser Leu Ile Asp Leu Gly Gly Arg Phe Arg Leu Val Ile Asn
385 390 395 400

Lys Val Glu Ala Val Lys Val Pro His Asp Met Pro Asn Leu Pro Val
405 410 415

Ala Arg Val Leu Trp Lys Pro Gln Pro Ser Leu Arg Thr Ser Ala Glu
420 425 430

Ala Trp Ile Leu Ala Gly Gly Ala His His Thr Cys Leu Ser Tyr Gln
435 440 445

Leu Thr Ala Glu Gln Met Leu Asp Trp Ala Glu Met Ser Gly Ile Glu
450 455 460

Ala Val Leu Ile Asn Arg Asp Thr Thr Ile Leu Asn Leu Arg Asn Glu
465 470 475 480

Leu Lys Trp Ser Glu Ala Ala Tyr Arg Leu Arg Lys Phe
485 490

<210> 84

<211> 3134

5 <212> DNA

<213> Issatchenkia orientalis

<220>

<221> CDS

10 <222> (1001)..(2134)

<400> 84

atgtatttgg agatttcgaa aagagtttgt atagagtctg taattgggtg tgtatttcaa	60
gaccacttt aaactgcgcc attaggagag ggagaggggg gggggggggg ggaagacggt	120
gaagtgtata caggatcgaa gaatagaagt tgtgtgtgtg ttttattacc cgtttcgatg	180
ggattcccag aagtggatac tatactgtct gcaatgcact acactctaaa aaagtattat	240

acattaccat acattagcaa atcaccaata ctctgcaactg tttcagtgtg tgcacattgc	300
taccaaatg ggaattgca gggaaatga gacccccct ccattccgta ttacgtaaga	360
caatatcagg gctgccgaat tcggcagaaa agccgagccg gccgagtcct cttgcacgga	420
gtgtgtccga aaagggcagc tctgcagtgg gggagaggag gtcgcacgtc tatgcggtgt	480
tggcatggcc tgtgcgtgta cctgtcccct ccctgggcat ccccaactgc gcgccttctc	540
cattggggcgc tgcgggcact ccgcgccgtt aatacaggag gggggggggg aaagcttaag	600
attagagcgg gtacagtcag tgggtgtatt gacccattt ctgtcagtat aaacccccg	660
ttgagccgcc ggtttggttg tttatggata aaattttttt tccccgcatg gagaagattg	720
agggggagaa ggaatgggaa aaaggccaga gccatctcca cagcggaaac cgaccgttaa	780
tggggtgaaa cacccccacc aggtagagca ggaagaatgg ggaacaagg tggagagatg	840
gtcattgttg ggaatagtgg gaaaatgagg gggaagagaa tgactataaa atgggaaggg	900
ggtccaagtt atccaagcag tccatttaga gaagggaaaa taaagctata gatagaaacc	960
aaccaaacaa ccaacaatt aaacaaacaa ttaaacgaac atg tta tcc aag acc	1015
	Met Leu Ser Lys Thr
	1 5
atc act gct gca ttg agg ggc aat aca act cgt act gca ttc aga atc	1063
Ile Thr Ala Ala Leu Arg Gly Asn Thr Thr Arg Thr Ala Phe Arg Ile	
	10 15 20
aat gcc att aga agt tta gcg atc cca gct att cca gag aca caa aag	1111
Asn Ala Ile Arg Ser Leu Ala Ile Pro Ala Ile Pro Glu Thr Gln Lys	
	25 30 35
ggc gtt atc ttt tat gag aac gga ggt gaa cta ttt tac aag gac att	1159
Gly Val Ile Phe Tyr Glu Asn Gly Gly Glu Leu Phe Tyr Lys Asp Ile	
	40 45 50
cca gtt cca aag cca aag cca aat gag att ttg gtg aat gtc aag tat	1207
Pro Val Pro Lys Pro Lys Pro Asn Glu Ile Leu Val Asn Val Lys Tyr	
	55 60 65
tct ggt gtt tgt cat acc gat tta cac gca tgg aaa ggt gac tgg cct	1255
Ser Gly Val Cys His Thr Asp Leu His Ala Trp Lys Gly Asp Trp Pro	
	70 75 80 85
ttg gcg acc aag ttg cca ttg gtt ggt gga cat gaa ggt gcc gga gtt	1303
Leu Ala Thr Lys Leu Pro Leu Val Gly Gly His Glu Gly Ala Gly Val	
	90 95 100
gtt gtt gct aag ggg gac aat gtc acc aac ttt gaa att ggc gat tat	1351
Val Val Ala Lys Gly Asp Asn Val Thr Asn Phe Glu Ile Gly Asp Tyr	
	105 110 115
gcc ggt atc aag tgg ttg aat ggt tca tgt atg ggg tgt gaa ttt tgc	1399
Ala Gly Ile Lys Trp Leu Asn Gly Ser Cys Met Gly Cys Glu Phe Cys	
	120 125 130
caa caa ggt gca gag cca aac tgt cca cag gcc gac ttg agt ggt tac	1447
Gln Gln Gly Ala Glu Pro Asn Cys Pro Gln Ala Asp Leu Ser Gly Tyr	
	135 140 145

ES 2 602 808 T3

acc cat gac ggg tcc ttt caa caa tat gcc act gcc gat gct gtt cag Thr His Asp Gly Ser Phe Gln Gln Tyr Ala Thr Ala Asp Ala Val Gln 150 155 160 165	1495
gca gcc aag att cct cag ggc act gat ttg gct caa gtt gcg cca att Ala Ala Lys Ile Pro Gln Gly Thr Asp Leu Ala Gln Val Ala Pro Ile 170 175 180	1543
tta tgt gca ggt att act gtc tat aag gct tta aag act gca gaa tta Leu Cys Ala Gly Ile Thr Val Tyr Lys Ala Leu Lys Thr Ala Glu Leu 185 190 195	1591
aga cca ggt caa tgg gtt gcc att tct ggt gct gct gga ggt tta ggt Arg Pro Gly Gln Trp Val Ala Ile Ser Gly Ala Ala Gly Gly Leu Gly 200 205 210	1639
tct ctt gct gtt caa tat gcc aag gcc atg ggt ttg aga gtt ttg ggt Ser Leu Ala Val Gln Tyr Ala Lys Ala Met Gly Leu Arg Val Leu Gly 215 220 225	1687
att gat ggt ggt gag gag aag ggc aag ttt gca aag tct ctt gga gct Ile Asp Gly Gly Glu Glu Lys Gly Lys Phe Ala Lys Ser Leu Gly Ala 230 235 240 245	1735
gaa gtt ttc att gat ttc acc aaa tcc aag gac att gtc aag gat atc Glu Val Phe Ile Asp Phe Thr Lys Ser Lys Asp Ile Val Lys Asp Ile 250 255 260	1783
caa gag gcc acc aat ggt ggt cca cat ggt gtc att aat gtt tct gtt Gln Glu Ala Asn Gly Gly Pro His Gly Val Ile Asn Val Ser Val 265 270 275	1831
tct cca gct gct att tct caa agt acc cag tat gtc aga acc ttg ggt Ser Pro Ala Ala Ile Ser Gln Ser Thr Gln Tyr Val Arg Thr Leu Gly 280 285 290	1879
aag gtt gtc ctt gtt gga tta cca gcg cat gct gta tgc gag tct tcg Lys Val Val Leu Val Gly Leu Pro Ala His Ala Val Cys Glu Ser Ser 295 300 305	1927
gtt ttc gac cat gtt gtc aag tcg att caa att aga ggc tct tat gtt Val Phe Asp His Val Val Lys Ser Ile Gln Ile Arg Gly Ser Tyr Val 310 315 320 325	1975
ggt aac agg gaa gat act agt gag gct att gat ttt ttc acc agg ggt Gly Asn Arg Glu Asp Thr Ser Glu Ala Ile Asp Phe Phe Thr Arg Gly 330 335 340	2023
tta gtg aag tca cca att aag att gtt ggt ttg agt gag ttg cca aag Leu Val Lys Ser Pro Ile Lys Ile Val Gly Leu Ser Glu Leu Pro Lys 345 350 355	2071
atc tat gaa ttg atg gag caa ggt aag att tta ggc aga tat gtt gtt Ile Tyr Glu Leu Met Glu Gln Gly Lys Ile Leu Gly Arg Tyr Val Val 360 365 370	2119
gac act tcg aaa tga tgggctgact tgggtgtact ggtgtgacgt ttttatgtgt Asp Thr Ser Lys	2174
atattgatat gcatggggga tgtatagtga tgaggagtag agtatataac gaaatgaaat	2234

gaaataatat gatatgataa gataagatga gatcaaatac gataatataa gatgacacat 2294
gaggagttca atgtagcata ctacacgatg ctgcagtaca actctgatac gctagactat 2354
actatacaaa actgtagtac actatacgtt agtgtagtat ccagaaacaa cactgcttta 2414
tagtacaata caactctata atactatagt atactatgcc aaaccacgta ataccataat 2474
atgctocacg acatggtaca atgtgctata cttcatacta ttataccata tatactccga 2534
tatattattg atatactatt ttatactata ataccatacc acacaacact acattacaac 2594
gagcaacctt accataaatg tcagttatgt ggcccgaga ctctctcgag gagcgtgttc 2654
acctcgttgt agacgttctg cacatcctct ccgagcaggg cacgtgctcc catagtggga 2714
ggggcctctt ccaagggcga ccgcggcgc ccgcaccaa gaagcgcctg ttccttgagc 2774
gcatgtgcaa tattgagaag ggtgtctatg ctgcgaagaa cgggtgtctgt gtcggcagca 2834
gcagcagcgg cgtctgctcc ctggcgga cgtgtcttcc ccgctaaggg gagcacagca 2894
agaatatcat gtaatgcagc aagagcattc tgagttgaag tatcgatttt cgatgccata 2954
ttgtatgtgt attgtattaa gtgtgtattg tcttaagtgt gtaagagaca tttatttgtg 3014
tcaacaatag cgacgccact gaaaacctca aatatcgtat ttattaatcc ccttcccccc 3074
agcgcagatc gtcccgtcga tttctattgt ttgggcatta tcagcgacgc gacggcgacg 3134

<210> 85

<211> 377

5 <212> PRT

<213> Issatchenka orientalis

<400> 85

Met Leu Ser Lys Thr Ile Thr Ala Ala Leu Arg Gly Asn Thr Thr Arg
1 5 10 15

Thr Ala Phe Arg Ile Asn Ala Ile Arg Ser Leu Ala Ile Pro Ala Ile
20 25 30

Pro Glu Thr Gln Lys Gly Val Ile Phe Tyr Glu Asn Gly Gly Glu Leu
35 40 45

Phe Tyr Lys Asp Ile Pro Val Pro Lys Pro Lys Pro Asn Glu Ile Leu
50 55 60

Val Asn Val Lys Tyr Ser Gly Val Cys His Thr Asp Leu His Ala Trp
65 70 75 80

Lys Gly Asp Trp Pro Leu Ala Thr Lys Leu Pro Leu Val Gly Gly His
85 90 95

Glu Gly Ala Gly Val Val Val Ala Lys Gly Asp Asn Val Thr Asn Phe

ES 2 602 808 T3

100	105	110
Glu Ile Gly Asp Tyr Ala Gly Ile Lys Trp Leu Asn Gly Ser Cys Met		
115	120	125
Gly Cys Glu Phe Cys Gln Gln Gly Ala Glu Pro Asn Cys Pro Gln Ala		
130	135	140
Asp Leu Ser Gly Tyr Thr His Asp Gly Ser Phe Gln Gln Tyr Ala Thr		
145	150	155
Ala Asp Ala Val Gln Ala Ala Lys Ile Pro Gln Gly Thr Asp Leu Ala		
165	170	175
Gln Val Ala Pro Ile Leu Cys Ala Gly Ile Thr Val Tyr Lys Ala Leu		
180	185	190
Lys Thr Ala Glu Leu Arg Pro Gly Gln Trp Val Ala Ile Ser Gly Ala		
195	200	205
Ala Gly Gly Leu Gly Ser Leu Ala Val Gln Tyr Ala Lys Ala Met Gly		
210	215	220
Leu Arg Val Leu Gly Ile Asp Gly Gly Glu Glu Lys Gly Lys Phe Ala		
225	230	235
Lys Ser Leu Gly Ala Glu Val Phe Ile Asp Phe Thr Lys Ser Lys Asp		
245	250	255
Ile Val Lys Asp Ile Gln Glu Ala Thr Asn Gly Gly Pro His Gly Val		
260	265	270
Ile Asn Val Ser Val Ser Pro Ala Ala Ile Ser Gln Ser Thr Gln Tyr		
275	280	285
Val Arg Thr Leu Gly Lys Val Val Leu Val Gly Leu Pro Ala His Ala		
290	295	300
Val Cys Glu Ser Ser Val Phe Asp His Val Val Lys Ser Ile Gln Ile		
305	310	315
Arg Gly Ser Tyr Val Gly Asn Arg Glu Asp Thr Ser Glu Ala Ile Asp		
325	330	335
Phe Phe Thr Arg Gly Leu Val Lys Ser Pro Ile Lys Ile Val Gly Leu		
340	345	350
Ser Glu Leu Pro Lys Ile Tyr Glu Leu Met Glu Gln Gly Lys Ile Leu		
355	360	365
Gly Arg Tyr Val Val Asp Thr Ser Lys		
370	375	

5 <210> 86
<211> 1596

<212> DNA
<213> Bacteroides thetaiotaomicron

<220>

5 <221> CDS

<222> (1)..(1596)

<400> 86

atg aaa tta gat gca aaa tca acc atc gag aca ggt aaa gct atc ctt	48
Met Lys Leu Asp Ala Lys Ser Thr Ile Glu Thr Gly Lys Ala Ile Leu	
1 5 10 15	
ggc ata gaa ctc ggt tct aca cga ata aaa gct gtt ctg att gac cag	96
Gly Ile Glu Leu Gly Ser Thr Arg Ile Lys Ala Val Leu Ile Asp Gln	
20 25 30	
gaa aac aaa cct atc gct caa ggc agc cac aca tgg gaa aat caa ctg	144
Glu Asn Lys Pro Ile Ala Gln Gly Ser His Thr Trp Glu Asn Gln Leu	
35 40 45	
gtc aac gga ctt tgg act tac agc att gat gcc atc tgg tcc gga ctg	192
Val Asn Gly Leu Trp Thr Tyr Ser Ile Asp Ala Ile Trp Ser Gly Leu	
50 55 60	
caa gat tgc tac gcc gac ctc cgc tcc aac gtg aag aaa tta tac gac	240
Gln Asp Cys Tyr Ala Asp Leu Arg Ser Asn Val Lys Lys Leu Tyr Asp	
65 70 75 80	
aca gag atc gaa aca ctg gca gcc atc ggt gtc agc gcc atg atg cat	288
Thr Glu Ile Glu Thr Leu Ala Ala Ile Gly Val Ser Ala Met Met His	
85 90 95	
ggt tac atg cct ttc aat gaa aaa gaa gaa atc ctc gtg cct ttc cgc	336
Gly Tyr Met Pro Phe Asn Glu Lys Glu Glu Ile Leu Val Pro Phe Arg	
100 105 110	
act tgg aga aat acc aat aca ggc cgt gct gcg gca gaa tta tcc gaa	384
Thr Trp Arg Asn Thr Asn Thr Gly Arg Ala Ala Ala Glu Leu Ser Glu	
115 120 125	
tta ttt gtc tat aac atc cct ttg aga tgg agc att tct cat ttg tac	432
Leu Phe Val Tyr Asn Ile Pro Leu Arg Trp Ser Ile Ser His Leu Tyr	
130 135 140	
cag gct att ctg gac aac gaa gcg cac gtc aaa gac atc aag ttc ctg	480
Gln Ala Ile Leu Asp Asn Glu Ala His Val Lys Asp Ile Lys Phe Leu	
145 150 155 160	
aca act ctt gca ggt tat gta cat tgg cag ata aca ggc gaa aag gtg	528
Thr Thr Leu Ala Gly Tyr Val His Trp Gln Ile Thr Gly Glu Lys Val	
165 170 175	

10

ttg ggc att ggt gac gca tgc ggt atg ctc ccc ata gat ccg act acc Leu Gly Ile Gly Asp Ala Ser Gly Met Leu Pro Ile Asp Pro Thr Thr 180 185 190	576
aac aac tat tcc gcc gaa atg gtg gcc aaa ttc aac aat ctg att gct Asn Asn Tyr Ser Ala Glu Met Val Ala Lys Phe Asn Asn Leu Ile Ala 195 200 205	624
tcg aaa gaa tac agt tgg aaa ctg gaa gac att ctg ccc aaa gta ttg Ser Lys Glu Tyr Ser Trp Lys Leu Glu Asp Ile Leu Pro Lys Val Leu 210 215 220	672
tcg gct ggt gaa aat gcc ggt gtc ctc aca ccg gaa ggc tgt aaa aaa Ser Ala Gly Glu Asn Ala Gly Val Leu Thr Pro Glu Gly Cys Lys Lys 225 230 235 240	720
ctc gat gca tcc ggt cat ctg aag gca gga ata ccg gtc tgc cca ccg Leu Asp Ala Ser Gly His Leu Lys Ala Gly Ile Pro Val Cys Pro Pro 245 250 255	768
gaa gga gac gca gcc acc ggc atg gta gca acc aac gcc gtc aag caa Glu Gly Asp Ala Gly Thr Gly Met Val Ala Thr Asn Ala Val Lys Gln 260 265 270	816
cgc acc gcc aac gta tgc gca ggt act tct tct ttc tct atg atc gta Arg Thr Gly Asn Val Ser Ala Gly Thr Ser Ser Phe Ser Met Ile Val 275 280 285	864
ttg gaa aaa gaa ttg tgc aag cca tac gaa atg atc gac atg gtc acc Leu Glu Lys Glu Leu Ser Lys Pro Tyr Glu Met Ile Asp Met Val Thr 290 295 300	912
act ccc gac gga agc ctc gta gcc atg gta cat tgc aac aac tgt act Thr Pro Asp Gly Ser Leu Val Ala Met Val His Cys Asn Asn Cys Thr 305 310 315 320	960
tcg gat ctt aac gca tgg gtc aac ctg ttc aaa gaa tac cag gaa ctt Ser Asp Leu Asn Ala Trp Val Asn Leu Phe Lys Glu Tyr Gln Glu Leu 325 330 335	1008
ctg ggt ata cct gta gat atg gat gaa ctc tat gcc aaa ctt tat aac Leu Gly Ile Pro Val Asp Met Asp Glu Leu Tyr Gly Lys Leu Tyr Asn 340 345 350	1056
att gcc ctt acc ggt gat acc gat tgc ggt ggt ctc ctc tcc tac aac Ile Ala Leu Thr Gly Asp Thr Asp Cys Gly Gly Leu Leu Ser Tyr Asn 355 360 365	1104
tac att tca gcc gaa cct gtt acg gga ctt gcc gag gga aga cct ttg Tyr Ile Ser Gly Glu Pro Val Thr Gly Leu Ala Glu Gly Arg Pro Leu 370 375 380	1152
ttc gta cgt tgc gcc aat gac aag ttc aac ctt gca aac ttt atg cgg Phe Val Arg Ser Ala Asn Asp Lys Phe Asn Leu Ala Asn Phe Met Arg 385 390 395 400	1200
gct cat ttg tac gcc tca gtc gga gtt ctc aag att gcc aac gac atc Ala His Leu Tyr Ala Ser Val Gly Val Leu Lys Ile Gly Asn Asp Ile 405 410 415	1248
ttg ttc aac gaa gaa aag atc aaa gtc gac aga atc aca ggt cac gga Leu Phe Asn Glu Glu Lys Ile Lys Val Asp Arg Ile Thr Gly His Gly 420 425 430	1296

gga ttg ttc aga acc aaa gga gtc ggt caa aga gta ctt gca gca gcc	1344
Gly Leu Phe Arg Thr Lys Gly Val Gly Gln Arg Val Leu Ala Ala Ala	
435 440 445	
atc aac tcg ccc ata tct gtt atg gaa aca gcc ggt gaa ggc ggt gca	1392
Ile Asn Ser Pro Ile Ser Val Met Glu Thr Ala Gly Glu Gly Gly Ala	
450 455 460	
tgg gga att gcc ctg ctg ggt tct tac ctg gta aac aat aaa aag ggt	1440
Trp Gly Ile Ala Leu Leu Gly Ser Tyr Leu Val Asn Asn Lys Lys Gly	
465 470 475 480	
caa tct ctt gcc gat ttc ctg gat gaa agt gta ttt gtc agc gat gct	1488
Gln Ser Leu Ala Asp Phe Leu Asp Glu Ser Val Phe Val Ser Asp Ala	
485 490 495	
ggt gtc gag gta tca ccc aca ccc gaa gat gta gcc ggc ttc aac aca	1536
Gly Val Glu Val Ser Pro Thr Pro Glu Asp Val Ala Gly Phe Asn Thr	
500 505 510	
tac atc gaa agc tac aag gca ggt ttg cct ata gaa gaa gca gcc gtc	1584
Tyr Ile Glu Ser Tyr Lys Ala Gly Leu Pro Ile Glu Glu Ala Ala Val	
515 520 525	
aaa ttc aaa taa	1596
Lys Phe Lys	
530	

<210> 87
 <211> 3347
 5 <212> DNA
 <213> Issatchenkia orientalis

<220>
 <221> CDS
 10 <222> (1001)..(2347)

<400> 87	
gttttgattc atcaaggtct gagaaatctt ctgcatcgac tgtaaacaga aaaaatacga	60
tatcgggggtc tcctaatatc gaagggtccc cacaaacgtc aatgcacagt ggagctctga	120
ttgttgagg atcacaagtt cagcagcccc atattgatcc gaagaagtta cgtaagatta	180
gcgaaactgt tccttatatg ttaagacacg agactatccg tagcaaacaa gctaaattgg	240
aagagttgga acgtcaaagt gccttagaat taagcagaag agcagcggaa cttgaacgca	300
aggccaagga attgaagatg cgtgaaacaa tggtgcgtga acgttttagct aaggccaacg	360
gcagcggcag tctctccttg tcaaacacag ccgacagtaa agggggtct tctatacgtt	420
ccaatgtctc agaaaaccga aagagagaga cttcatacag cgtccaaact accactagt	480
aatacgaaga taccgttgaa gacaacgacg gatctgctgc ttagcttggg atctatgtac	540
tgactttgta tgtacctata aatctacact ctatacctca acttgaacta catgggctgt	600
tttctaggaa acaatcagct atacgcgtgg ctgtcctctt acttaacttc atgtttccct	660

ggtagctgtc ttgttctact gtcccacgga tacaggatgc agacaacaac actcttcttc	720
gggtggtaaa caaccttggt agaacagtgt aagaccagca cagtgcattc ttagtatcat	780
tgcatccgtg tggggaaacg gcaaattcta cattgaagcg ttcctgtttt tttttttgtg	840
gcacaaacaa aggagaaggt gctttttcgt tggttcataa gaacggttat ttctgtggtg	900
aattgccacc tggttttctc ctttagaagc tagctttgtg tgtttccagt atcttgattt	960
tctgtgtaga gacaaacaat agccaagagc cactccaaaa atg tct cct tca caa	1015
Met Ser Pro Ser Gln	
1 5	
att aac gtt gac aac tta tct aat tgg act gaa gaa ttc aaa tct gac	1063
Ile Asn Val Asp Asn Leu Ser Asn Trp Thr Glu Glu Phe Lys Ser Asp	
10 15 20	
gcc aag act caa atc ggg ggt tct gta ttg caa cat tcc aac att gat	1111
Ala Lys Thr Gln Ile Gly Gly Ser Val Leu Gln His Ser Asn Ile Asp	
25 30 35	
gag gtc ttg att aac aga gat gca gaa atc gcc aac aag cat atc ttc	1159
Glu Val Leu Ile Asn Arg Asp Ala Glu Ile Ala Asn Lys His Ile Phe	
40 45 50	
aac cac aag att gaa att gaa ggt cta cct gtc atg gat cag aag gct	1207
Asn His Lys Ile Glu Ile Glu Gly Leu Pro Val Met Asp Gln Lys Ala	
55 60 65	
tct ggt aga tgt tgg ttg ttt gca tcg act aac ttg atg cgt gtt act	1255
Ser Gly Arg Cys Trp Leu Phe Ala Ser Thr Asn Leu Met Arg Val Thr	
70 75 80 85	
gca atg aag aaa tac aat ttg aag gaa atc aag ctt tcc cca tcg tat	1303
Ala Met Lys Lys Tyr Asn Leu Lys Glu Ile Lys Leu Ser Pro Ser Tyr	
90 95 100	
ttg ttt ttc tat gac aaa ttg gaa aga gca aac tat ttc ctt gaa caa	1351
Leu Phe Phe Tyr Asp Lys Leu Glu Arg Ala Asn Tyr Phe Leu Glu Gln	
105 110 115	
atc atc gac act cat aag gaa cca atc gat tca aga ttg gtt caa tat	1399
Ile Ile Asp Thr His Lys Glu Pro Ile Asp Ser Arg Leu Val Gln Tyr	
120 125 130	
ttc ctg acc aat cca gtt gaa gat ggt ggt caa ttc acc atg atg gca	1447
Phe Leu Thr Asn Pro Val Glu Asp Gly Gly Gln Phe Thr Met Met Ala	
135 140 145	
caa att gct acc aaa tac ggt gtt gtt cct gat caa gtc tac cca gat	1495
Gln Ile Ala Thr Lys Tyr Gly Val Val Pro Asp Gln Val Tyr Pro Asp	
150 155 160 165	
tct ttc aac aca acc act tcg agg att atg aac aga tta gtc aac cac	1543
Ser Phe Asn Thr Thr Thr Ser Arg Ile Met Asn Arg Leu Val Asn His	
170 175 180	
aga tta cgt tct tat gca atg act tta cgt aac gct cta gat gaa ggt	1591
Arg Leu Arg Ser Tyr Ala Met Thr Leu Arg Asn Ala Leu Asp Glu Gly	
185 190 195	
aaa gat gta atg tcc ttg aag aat gag atg caa aaa gaa att tat cgt	1639

ES 2 602 808 T3

Lys	Asp	Val	Met	Ser	Leu	Lys	Asn	Glu	Met	Gln	Lys	Glu	Ile	Tyr	Arg	
		200					205					210				
ttg	cta	aca	atg	ttc	ctt	ggt	aac	cca	cca	aag	cca	aac	gaa	gag	ttt	1687
Leu	Leu	Thr	Met	Phe	Leu	Gly	Asn	Pro	Pro	Lys	Pro	Asn	Glu	Glu	Phe	
		215				220					225					
gtc	tggt	gaa	ttc	acc	gat	aaa	gat	ggt	aaa	tat	gaa	tct	att	aaa	act	1735
Val	Trp	Glu	Phe	Thr	Asp	Lys	Asp	Gly	Lys	Tyr	Glu	Ser	Ile	Lys	Thr	
230					235					240					245	
aca	cca	tta	aaa	tat	gca	act	gaa	ggt	ttg	gat	ttc	cat	gct	cca	gaa	1783
Thr	Pro	Leu	Lys	Tyr	Ala	Thr	Glu	Val	Leu	Asp	Phe	His	Ala	Pro	Glu	
				250					255					260		
tat	ggt	tcc	ttg	tta	aat	gac	cca	aga	aat	aag	tat	aac	aag	atg	ggt	1831
Tyr	Val	Ser	Leu	Leu	Asn	Asp	Pro	Arg	Asn	Lys	Tyr	Asn	Lys	Met	Val	
			265					270						275		
caa	ggt	gaa	aga	tta	ggt	aat	ggt	gct	ggt	ggc	gaa	cca	ggt	gca	tac	1879
Gln	Val	Glu	Arg	Leu	Gly	Asn	Val	Ala	Gly	Gly	Glu	Pro	Val	Ala	Tyr	
		280				285						290				
tta	aac	tta	gaa	att	gaa	aag	tta	tct	caa	gct	ggt	ggt	aac	aga	atc	1927
Leu	Asn	Leu	Glu	Ile	Glu	Lys	Leu	Ser	Gln	Ala	Val	Val	Asn	Arg	Ile	
		295				300					305					
aaa	aat	aac	aaa	cca	ggt	ttc	ttt	ggt	acc	gat	aca	cct	aaa	ttt	atg	1975
Lys	Asn	Asn	Lys	Pro	Val	Phe	Phe	Gly	Thr	Asp	Thr	Pro	Lys	Phe	Met	
310					315					320					325	
gat	aaa	agt	aga	ggt	att	atg	gat	atc	aat	tta	tggt	gac	tat	gag	tta	2023
Asp	Lys	Ser	Arg	Gly	Ile	Met	Asp	Ile	Asn	Leu	Trp	Asp	Tyr	Glu	Leu	
				330				335						340		
tta	ggt	tat	gat	gtc	cgt	acc	atg	tca	aag	aag	gaa	aga	ggt	ggt	ttt	2071
Leu	Gly	Tyr	Asp	Val	Arg	Thr	Met	Ser	Lys	Lys	Glu	Arg	Val	Val	Phe	
			345					350					355			
ggt	gat	tct	tta	atg	acc	cac	gct	atg	ttg	att	act	gca	gtg	cac	ggt	2119
Gly	Asp	Ser	Leu	Met	Thr	His	Ala	Met	Leu	Ile	Thr	Ala	Val	His	Val	
		360				365						370				
gat	gaa	aat	ggc	aaa	cct	gtc	aga	tac	aga	gtc	gaa	aac	agt	tggt	ggt	2167
Asp	Glu	Asn	Gly	Lys	Pro	Val	Arg	Tyr	Arg	Val	Glu	Asn	Ser	Trp	Gly	
		375				380					385					
acc	aag	agt	ggt	caa	gaa	ggt	tat	tac	aca	atg	acc	caa	gaa	tat	ttt	2215
Thr	Lys	Ser	Gly	Gln	Glu	Gly	Tyr	Tyr	Thr	Met	Thr	Gln	Glu	Tyr	Phe	
390					395					400					405	
gaa	gag	tac	ggt	tat	caa	gta	gtc	att	gaa	aag	agt	gaa	ttt	gct	gcc	2263
Glu	Glu	Tyr	Val	Tyr	Gln	Val	Val	Ile	Glu	Lys	Ser	Glu	Phe	Ala	Ala	
			410					415					420			
cta	aac	ctc	gat	ggt	tcc	att	ctg	gag	gat	aaa	gaa	cca	gtc	gtc	ttg	2311
Leu	Asn	Leu	Asp	Val	Ser	Ile	Leu	Glu	Asp	Lys	Glu	Pro	Val	Val	Leu	
			425					430					435			
cca	cct	tat	gac	cct	atg	ggt	gca	ctt	gct	tta	taa	attgattttg				2357
Pro	Pro	Tyr	Asp	Pro	Met	Gly	Ala	Leu	Ala	Leu						
		440				445										

taggggggaa aaaaaggaaa aagagacatc acaagtcaat gtgatttagc aagttatddd 2417
tagcttacat tgaagggttat cttcttaaag ctctcccaac gtatgtccaa caagtttggg 2477
tatcaaaact atgtaatcct tttagataaa aaacaatagt actaaatacg cttttattag 2537
accctggcaa gctgtaaact tagcaaaaag agtttagctcc taacaactgt accagcacta 2597
ttactgttgt tgaaaattgc gatctctctt ccaaatttgg cgagcaatga agtttcacca 2657
aatcgtggag ttttaaaaaa aacatttgtga tacggttctg gttttctaca acttttcttt 2717
tcctctctt gactgctttg ctggcactgg ttgacaatct cattctgaga atcggacctg 2777
ttactccttt ctctctgtct aaataaaaaa aagataaggc agaaaattag aaaactagga 2837
agcatgaatt ttacctgggt attcaaaatt ttgctgtgtt tgatattttc aaatcgtatc 2897
aatgccatag gcaataagga ctttgctaac aaagactacc taatggtaga agtaaagggg 2957
gattatgatt tacataacat acaagaatta catccagatt gggaatatga atatgaaatt 3017
gactttctag ataattttca tgtgtttagt atgaagaaag accatcagtt aatagaaaag 3077
ttttcaaagt acacatctat caaagagctc ttgcaagggtg acaaaaaatt aattaaacgt 3137
gaagattatg attttttgag tagtttgaat gataacaatg ttacaggtgt tcatttacta 3197
tcgaggaaac aactagtaaa acgtttccct gttcccgat catatgggaa agacgtacct 3257
gttctacatt ctcgagagaa cacaacggtt gattcaagct tagatgaaat agctcaagtt 3317
gcagaagaat ttggtattaa tgatccaatc 3347

<210> 88

<211> 448

<212> PRT

<213> Issatchenkia orientalis

<400> 88

Met Ser Pro Ser Gln Ile Asn Val Asp Asn Leu Ser Asn Trp Thr Glu
1 5 10 15

Glu Phe Lys Ser Asp Ala Lys Thr Gln Ile Gly Gly Ser Val Leu Gln
20 25 30

His Ser Asn Ile Asp Glu Val Leu Ile Asn Arg Asp Ala Glu Ile Ala
35 40 45

Asn Lys His Ile Phe Asn His Lys Ile Glu Ile Glu Gly Leu Pro Val
50 55 60

Met Asp Gln Lys Ala Ser Gly Arg Cys Trp Leu Phe Ala Ser Thr Asn
65 70 75 80

Leu Met Arg Val Thr Ala Met Lys Lys Tyr Asn Leu Lys Glu Ile Lys

5

10

ES 2 602 808 T3

85										90					95				
Leu	Ser	Pro	Ser	Tyr	Leu	Phe	Phe	Tyr	Asp	Lys	Leu	Glu	Arg	Ala	Asn				
			100					105					110						
Tyr	Phe	Leu	Glu	Gln	Ile	Ile	Asp	Thr	His	Lys	Glu	Pro	Ile	Asp	Ser				
		115					120					125							
Arg	Leu	Val	Gln	Tyr	Phe	Leu	Thr	Asn	Pro	Val	Glu	Asp	Gly	Gly	Gln				
	130					135					140								
Phe	Thr	Met	Met	Ala	Gln	Ile	Ala	Thr	Lys	Tyr	Gly	Val	Val	Pro	Asp				
145					150					155					160				
Gln	Val	Tyr	Pro	Asp	Ser	Phe	Asn	Thr	Thr	Thr	Ser	Arg	Ile	Met	Asn				
				165					170					175					
Arg	Leu	Val	Asn	His	Arg	Leu	Arg	Ser	Tyr	Ala	Met	Thr	Leu	Arg	Asn				
			180					185					190						
Ala	Leu	Asp	Glu	Gly	Lys	Asp	Val	Met	Ser	Leu	Lys	Asn	Glu	Met	Gln				
		195					200					205							
Lys	Glu	Ile	Tyr	Arg	Leu	Leu	Thr	Met	Phe	Leu	Gly	Asn	Pro	Pro	Lys				
	210					215					220								
Pro	Asn	Glu	Glu	Phe	Val	Trp	Glu	Phe	Thr	Asp	Lys	Asp	Gly	Lys	Tyr				
225					230					235					240				
Glu	Ser	Ile	Lys	Thr	Thr	Pro	Leu	Lys	Tyr	Ala	Thr	Glu	Val	Leu	Asp				
				245					250					255					
Phe	His	Ala	Pro	Glu	Tyr	Val	Ser	Leu	Leu	Asn	Asp	Pro	Arg	Asn	Lys				
			260					265					270						
Tyr	Asn	Lys	Met	Val	Gln	Val	Glu	Arg	Leu	Gly	Asn	Val	Ala	Gly	Gly				
		275					280					285							
Glu	Pro	Val	Ala	Tyr	Leu	Asn	Leu	Glu	Ile	Glu	Lys	Leu	Ser	Gln	Ala				
		290				295					300								
Val	Val	Asn	Arg	Ile	Lys	Asn	Asn	Lys	Pro	Val	Phe	Phe	Gly	Thr	Asp				
305					310					315					320				
Thr	Pro	Lys	Phe	Met	Asp	Lys	Ser	Arg	Gly	Ile	Met	Asp	Ile	Asn	Leu				
				325					330					335					

ES 2 602 808 T3

Trp	Asp	Tyr	Glu	Leu	Leu	Gly	Tyr	Asp	Val	Arg	Thr	Met	Ser	Lys	Lys	
			340					345					350			
Glu	Arg	Val	Val	Phe	Gly	Asp	Ser	Leu	Met	Thr	His	Ala	Met	Leu	Ile	
		355					360					365				
Thr	Ala	Val	His	Val	Asp	Glu	Asn	Gly	Lys	Pro	Val	Arg	Tyr	Arg	Val	
	370					375					380					
Glu	Asn	Ser	Trp	Gly	Thr	Lys	Ser	Gly	Gln	Glu	Gly	Tyr	Tyr	Thr	Met	
385					390					395					400	
Thr	Gln	Glu	Tyr	Phe	Glu	Glu	Tyr	Val	Tyr	Gln	Val	Val	Ile	Glu	Lys	
				405					410					415		
Ser	Glu	Phe	Ala	Ala	Leu	Asn	Leu	Asp	Val	Ser	Ile	Leu	Glu	Asp	Lys	
			420					425					430			
Glu	Pro	Val	Val	Leu	Pro	Pro	Tyr	Asp	Pro	Met	Gly	Ala	Leu	Ala	Leu	
	435						440					445				

REIVINDICACIONES

1. Una célula de levadura genéticamente modificada que expresa en exceso un polipéptido transportador de xilosa que comprende una secuencia de aminoácidos con al menos el 90 % de identidad de secuencia con una secuencia de aminoácidos seleccionados del grupo que consiste en SEQ ID NO:2 y SEQ ID NO:4.
- 5 2. La célula de levadura genéticamente modificada de la reivindicación 1 que comprende además una vía de fermentación de arabinosa activa, en donde dicha célula comprende uno o más genes exógenos de la vía de fermentación de arabinosa seleccionados del grupo que consiste en genes Al, RK y RE.
- 10 3. La célula de levadura genéticamente modificada de cualquiera de las reivindicaciones 1 y 2 que comprende además una vía de fermentación de xilosa activa, en donde dicha célula comprende uno o más genes exógenos de la vía de fermentación de xilosa seleccionados del grupo que consiste en genes XI y XK.
4. La célula de levadura genéticamente modificada de cualquiera de las reivindicaciones 2 y 3 que comprende además una vía de pentosa fosfato no oxidativa activa, en donde dicha célula comprende uno o más genes exógenos de la vía de pentosa fosfato no oxidativa seleccionados del grupo que consiste en genes TKL y TAL.
- 15 5. La célula de levadura genéticamente modificada de la reivindicación 4, en donde dicha célula comprende además uno o más genes exógenos de la vía de pentosa fosfato no oxidativa seleccionados del grupo que consiste en genes RPE y RKL.
6. La célula de levadura genéticamente modificada de cualquiera de las reivindicaciones 1, 2, 3, 4 y 5, en donde dicha célula de levadura comprende además una delección o alteración de uno o más genes seleccionados del grupo que consiste en genes AR/XR, arabitol 4-deshidrogenasa, xilulosa reductasa y XDH.
- 20 7. La célula de levadura genéticamente modificada de cualquiera de las reivindicaciones 1, 2, 3, 4, 5 y 6, en donde dicha célula de levadura comprende además una delección o alteración de uno o más genes seleccionados del grupo que consiste en genes ALD y ADH.
8. La célula de levadura genéticamente modificada de cualquiera de las reivindicaciones 1, 2, 3, 4, 5, 6 y 7, en donde dicha célula de levadura pertenece al clado de *I. orientalis*/*P. fermentans*.
- 25 9. La célula de levadura genéticamente modificada de la reivindicación 8, en donde dicha célula de levadura es *I. orientalis*.
10. Un método, que comprende un proceso de fermentación en donde una célula de levadura genéticamente modificada como se cita en cualquiera de las reivindicaciones 1, 2, 3, 4, 5, 6 y 7 se cultiva en medio de fermentación que comprende xilosa.
- 30 11. Un método según la reivindicación 10, de producción de etanol a partir de un medio que contiene xilosa que comprende cultivar una célula de levadura genéticamente modificada como se cita en cualquiera de las reivindicaciones 1, 2, 3, 4, 5, 6 y 7 en un medio que contiene xilosa.
12. El proceso de fermentación de la reivindicación 11, en donde dicho medio comprende al menos 10 g/l de xilosa de un hidrolizado de biomasa vegetal.
- 35 13. Un polinucleótido aislado que codifica un polipéptido que comprende una secuencia de aminoácidos con al menos el 90 % de identidad de secuencia con una secuencia de aminoácidos seleccionada del grupo que consiste en las secuencias de aminoácidos expuestas en SEQ ID NO:2 y SEQ ID NO:4, en donde dicho polipéptido es capaz de transportar xilosa en una célula de levadura.
- 40 14. Un polipéptido aislado que comprende una secuencia de aminoácidos con al menos el 90 % de identidad de secuencia con una secuencia de aminoácidos seleccionada del grupo que consiste en las secuencias de aminoácidos expuestas en SEQ ID NO:2 y SEQ ID NO:4, en donde dicho polipéptido es capaz de transportar xilosa en una célula de levadura.

Figura 1

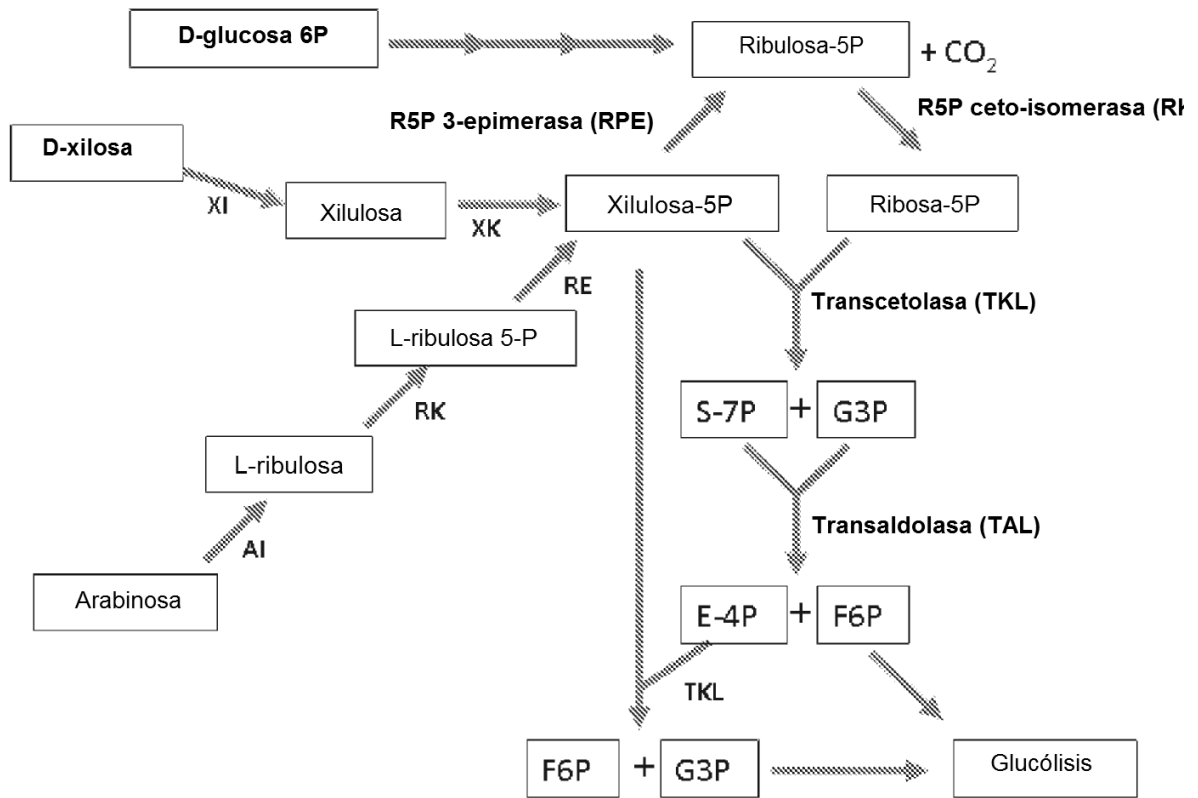


Figura 2

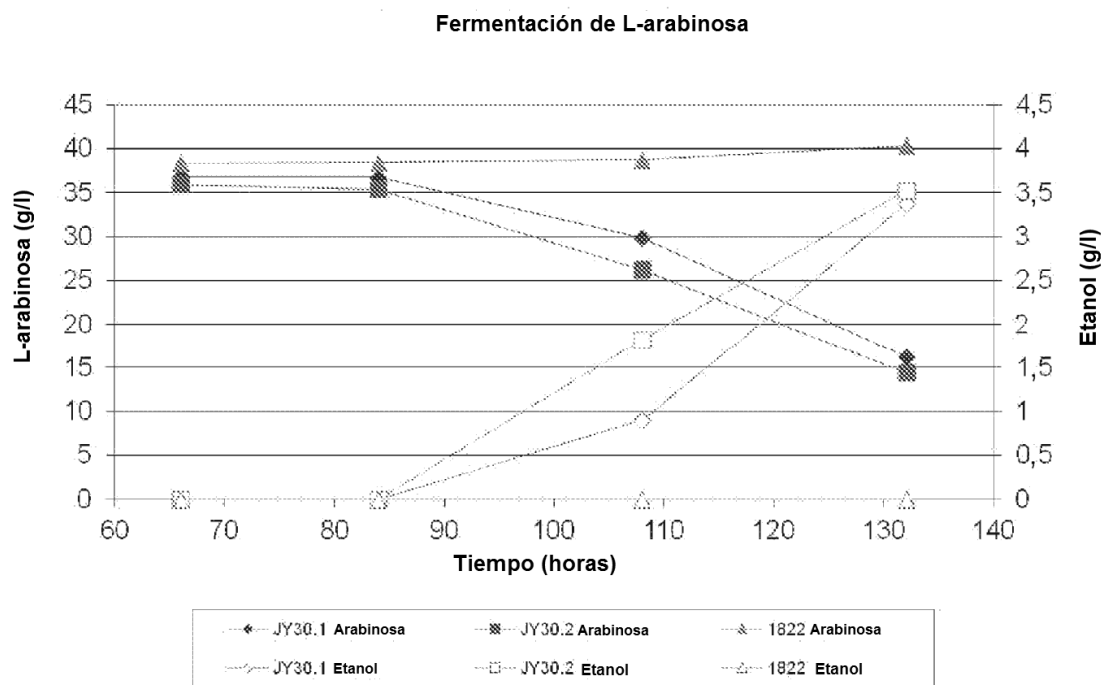


Figura 3

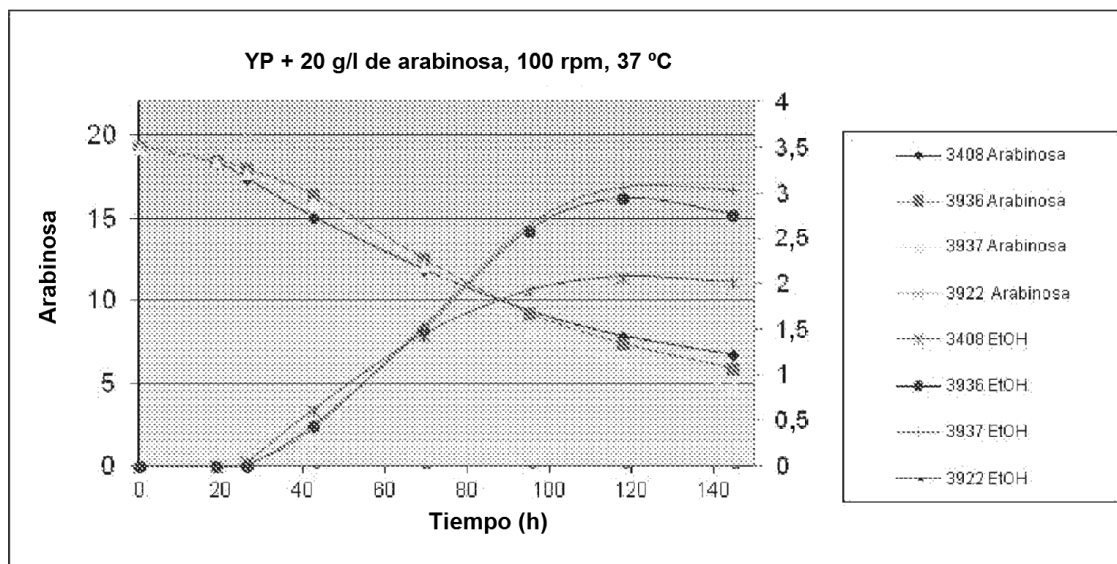


Figura 4

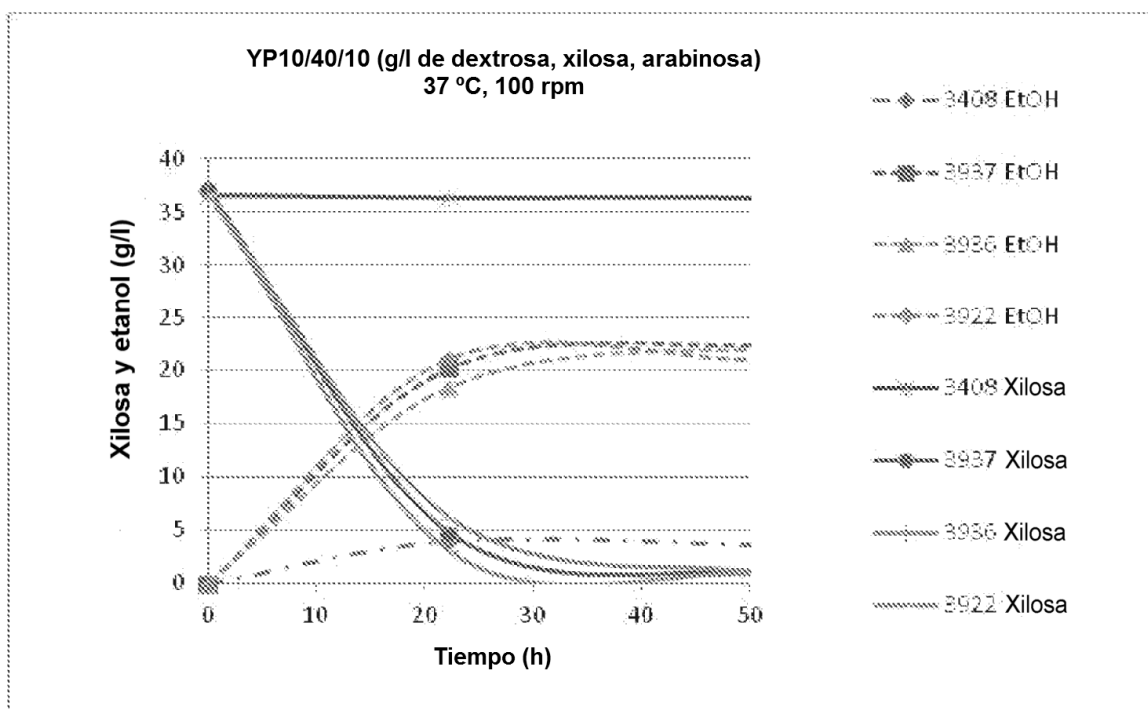


Figura 5

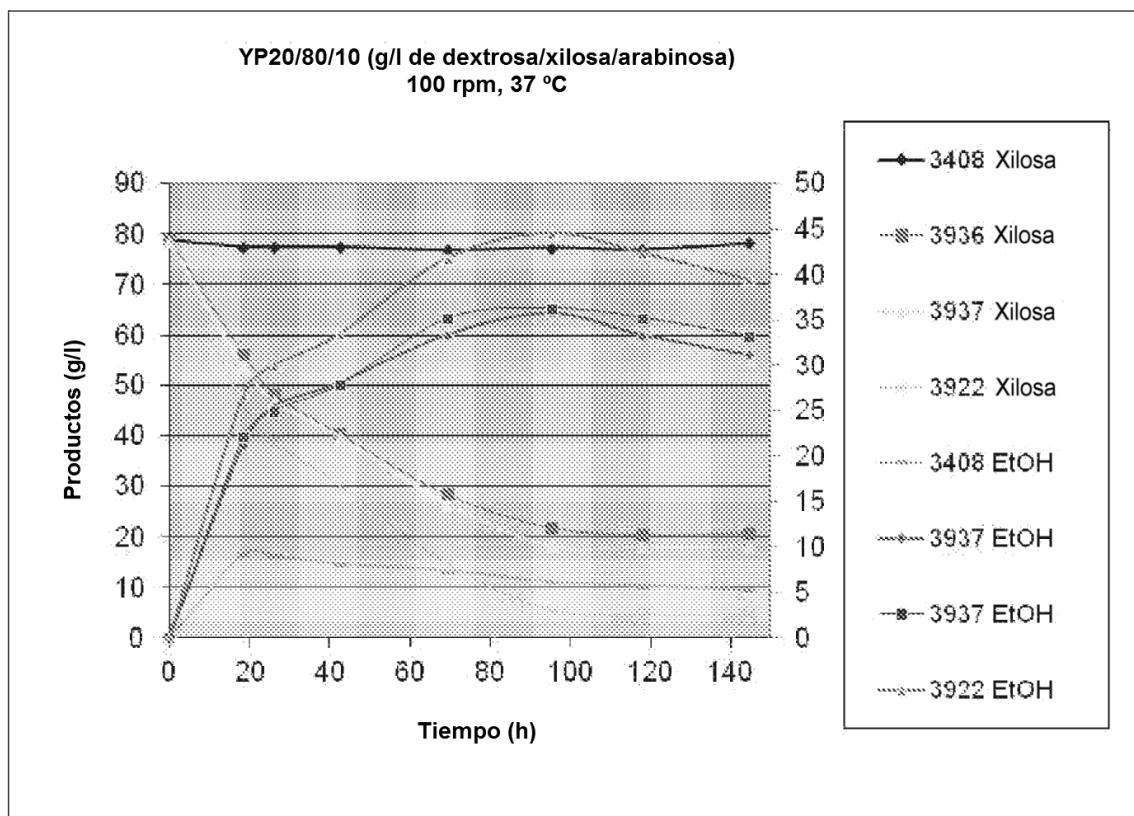


Figura 6

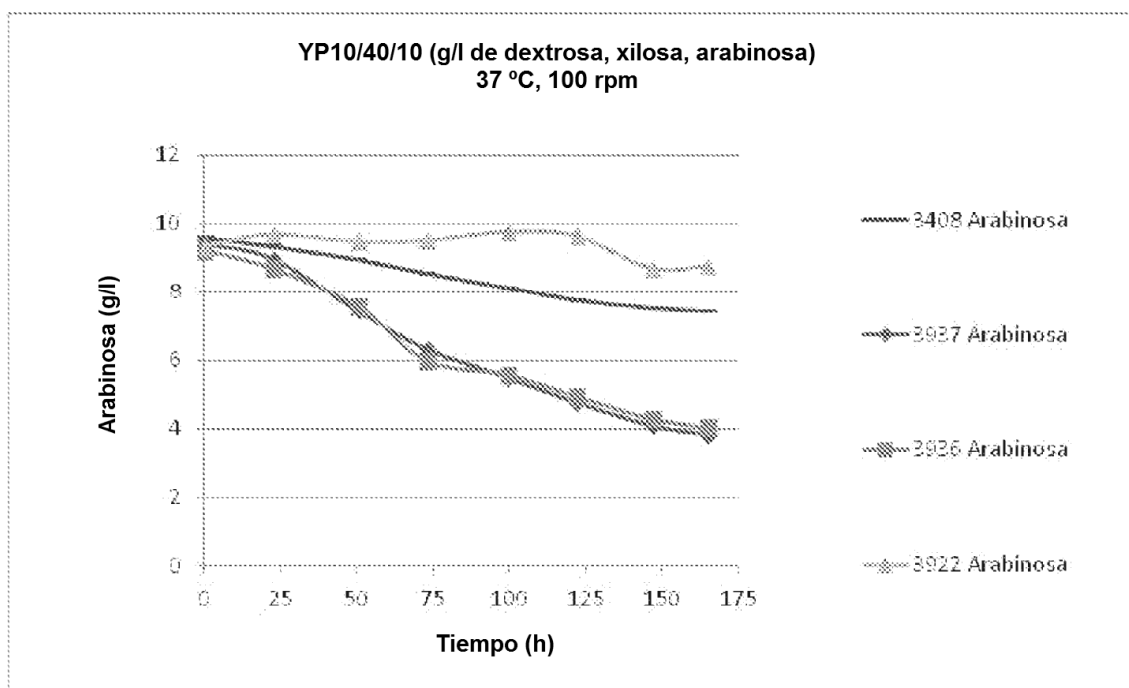


Figura 7

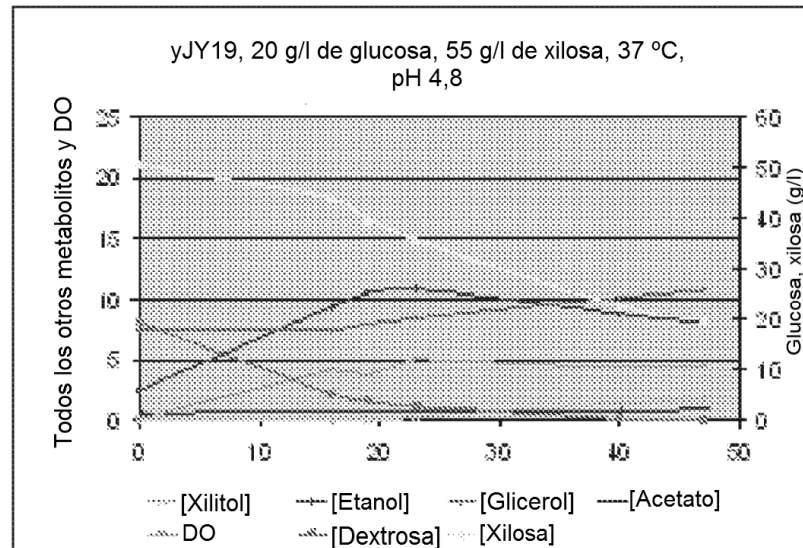


Figura 8

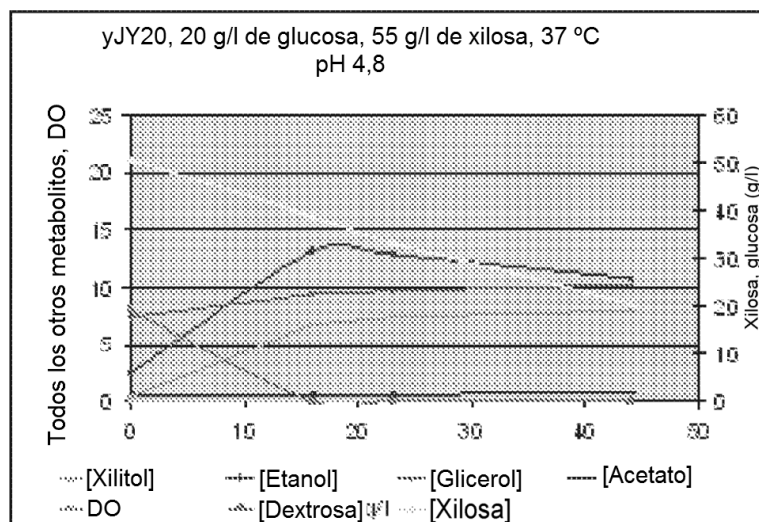


Figura 9

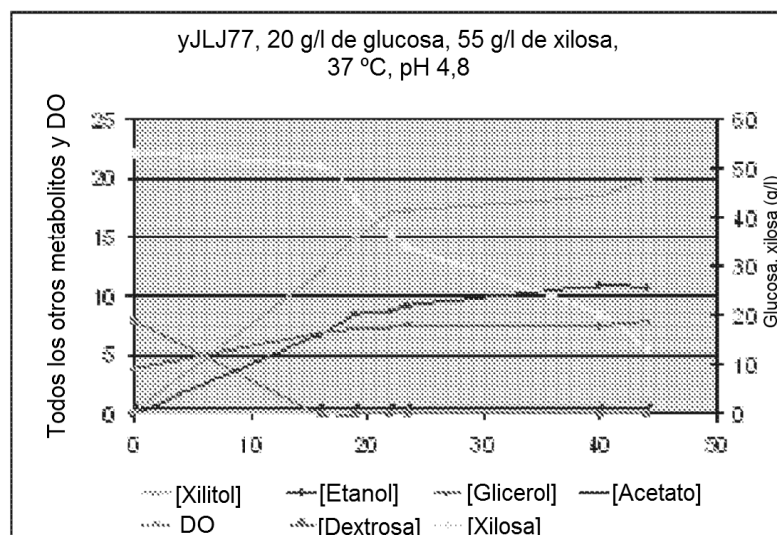


Figura 10

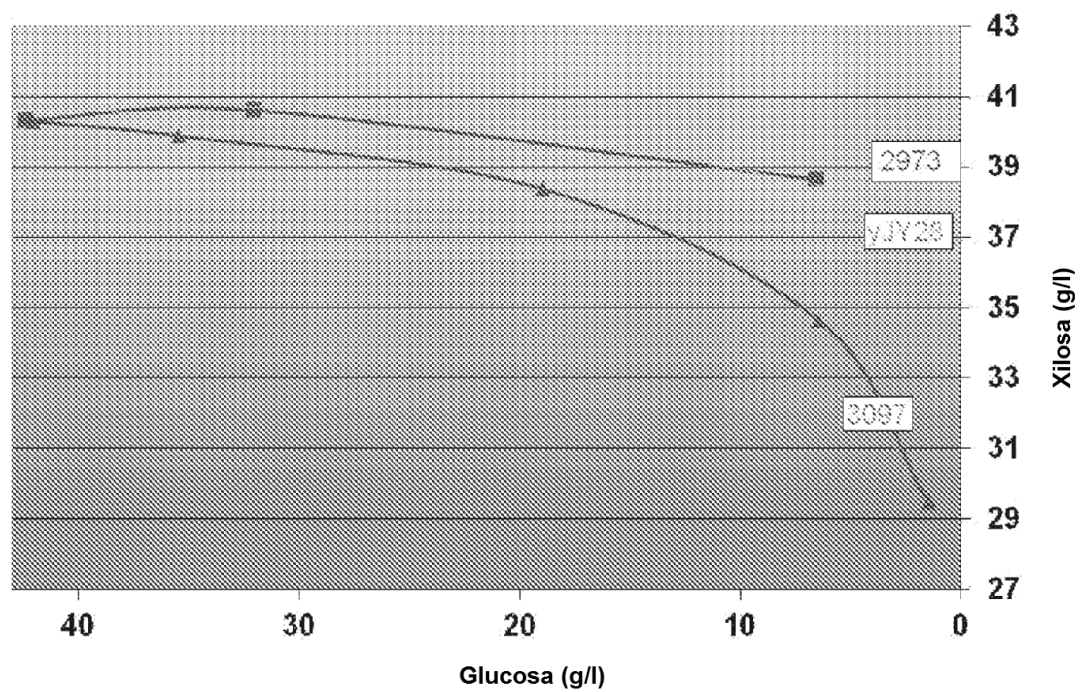


Figura 11

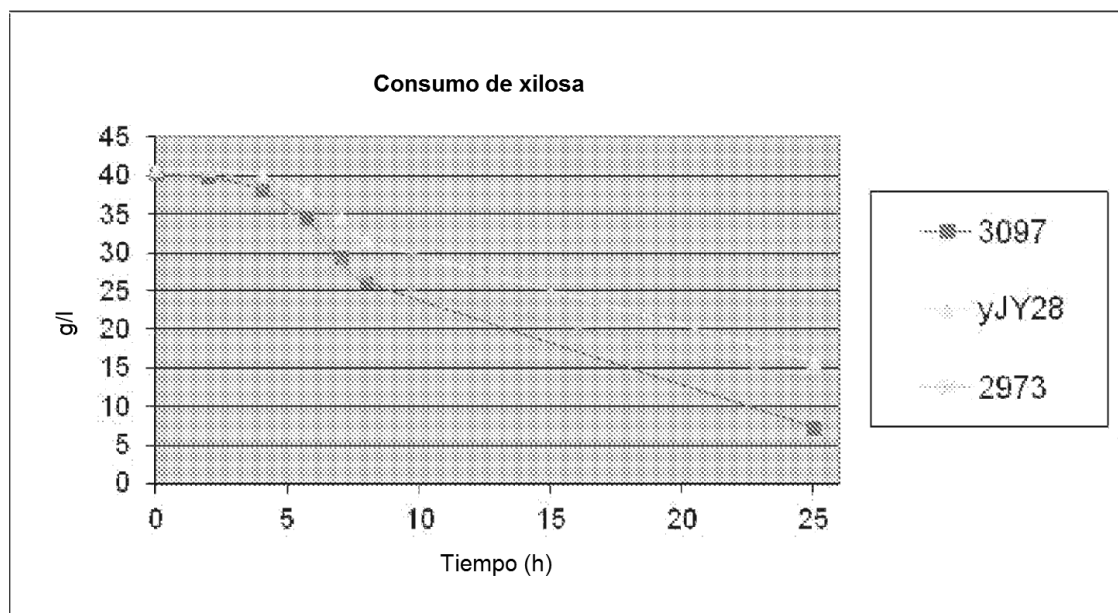


Figura 12

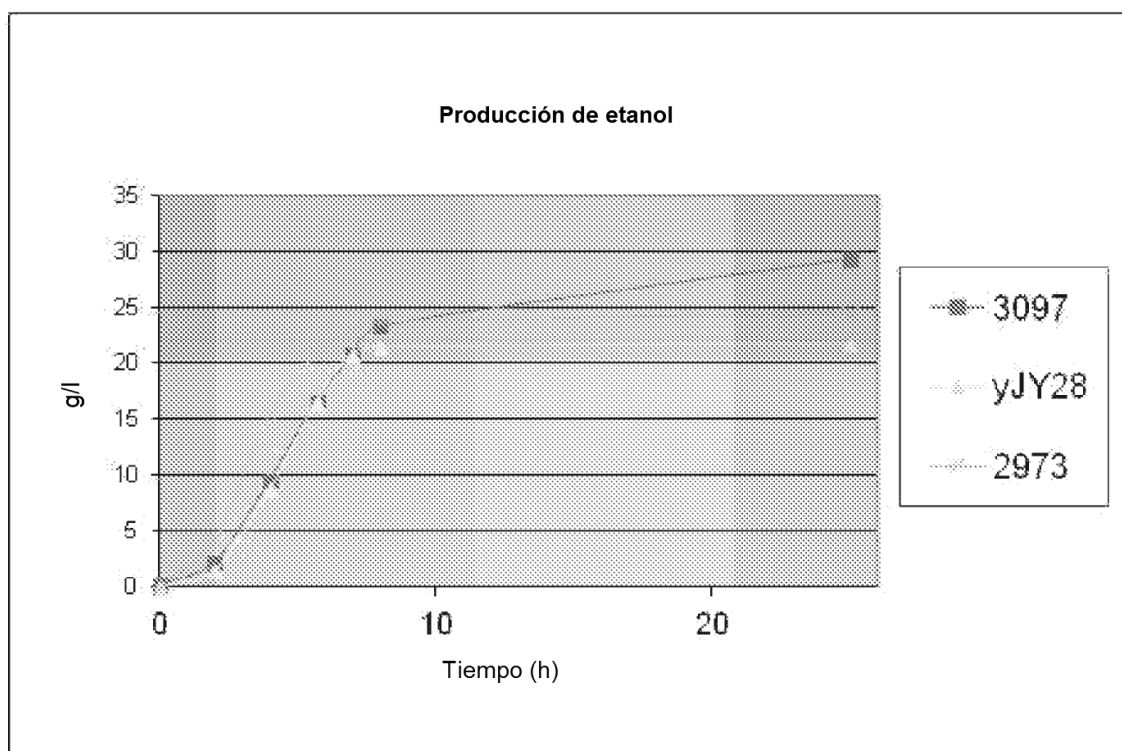


Figura 13

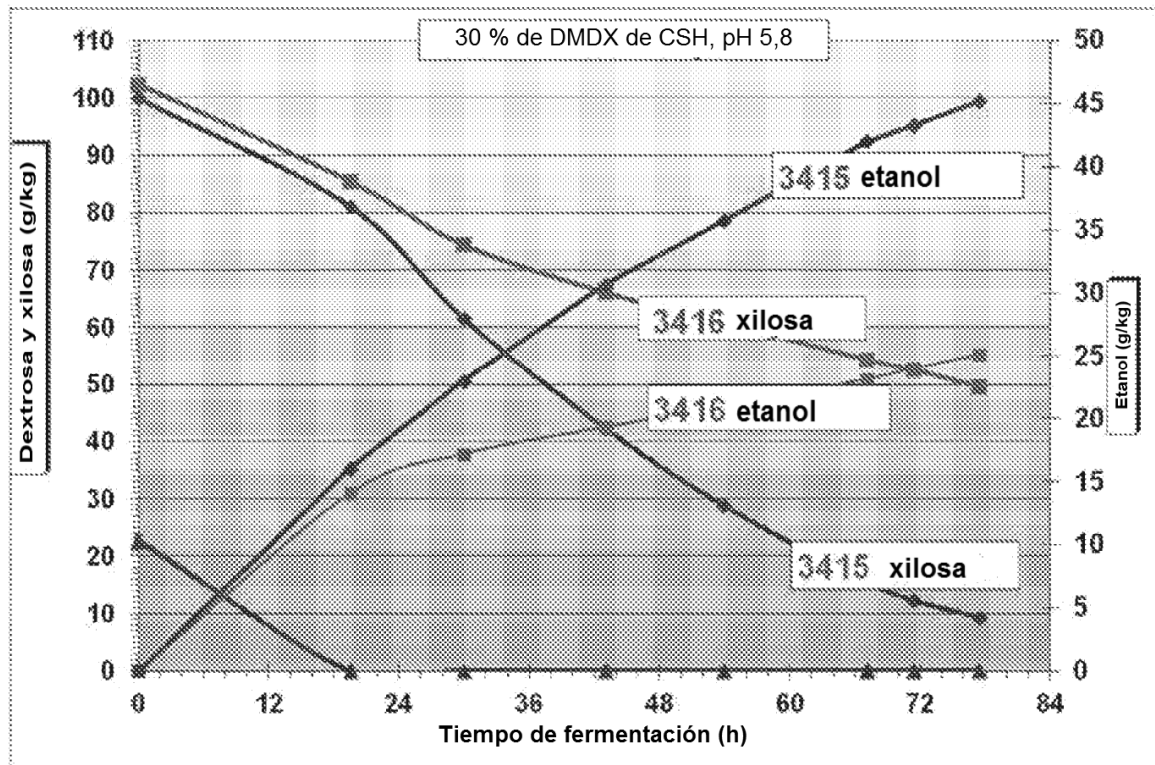


Figura 14

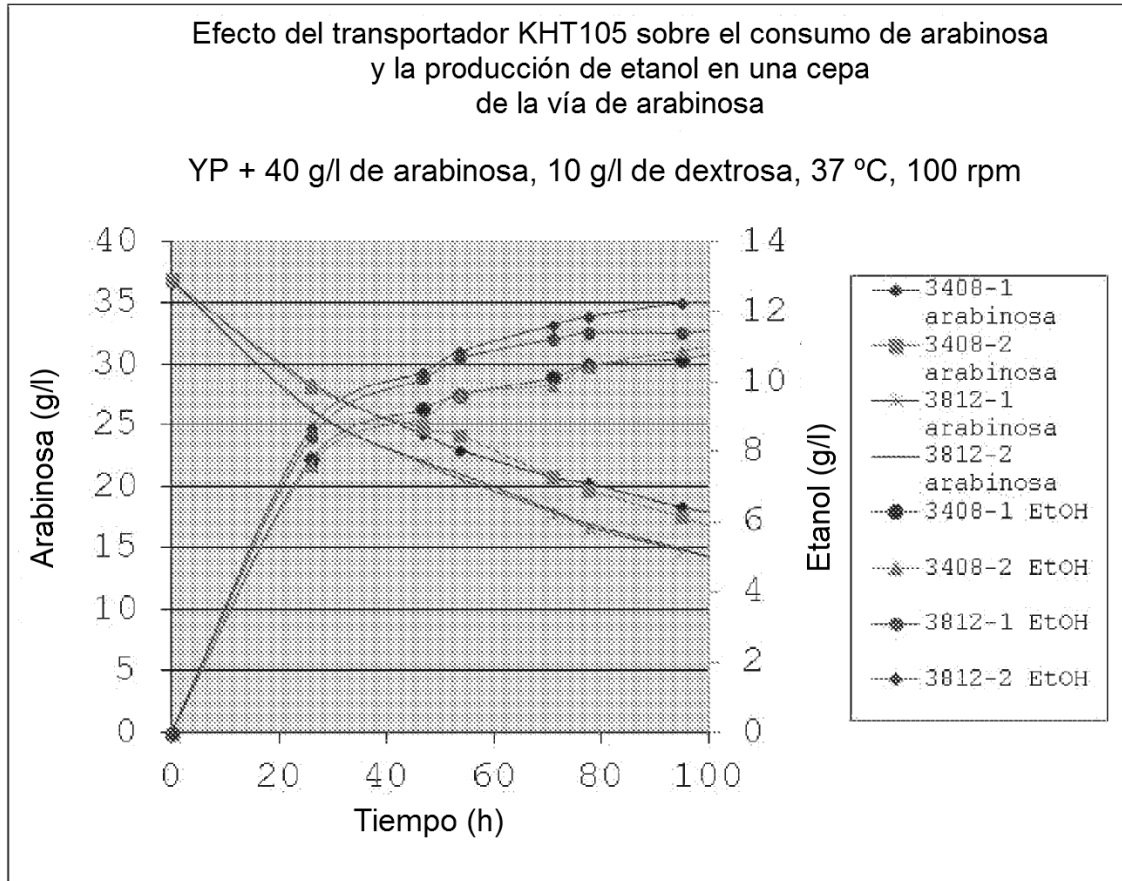


Figura 15

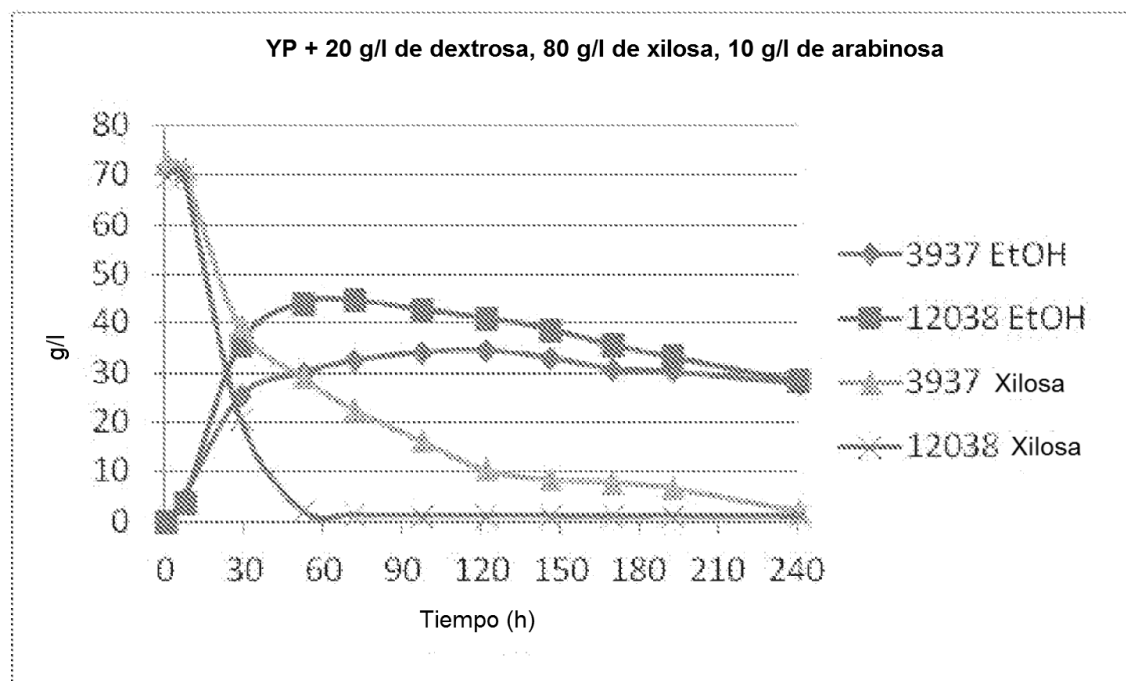


Figura 16

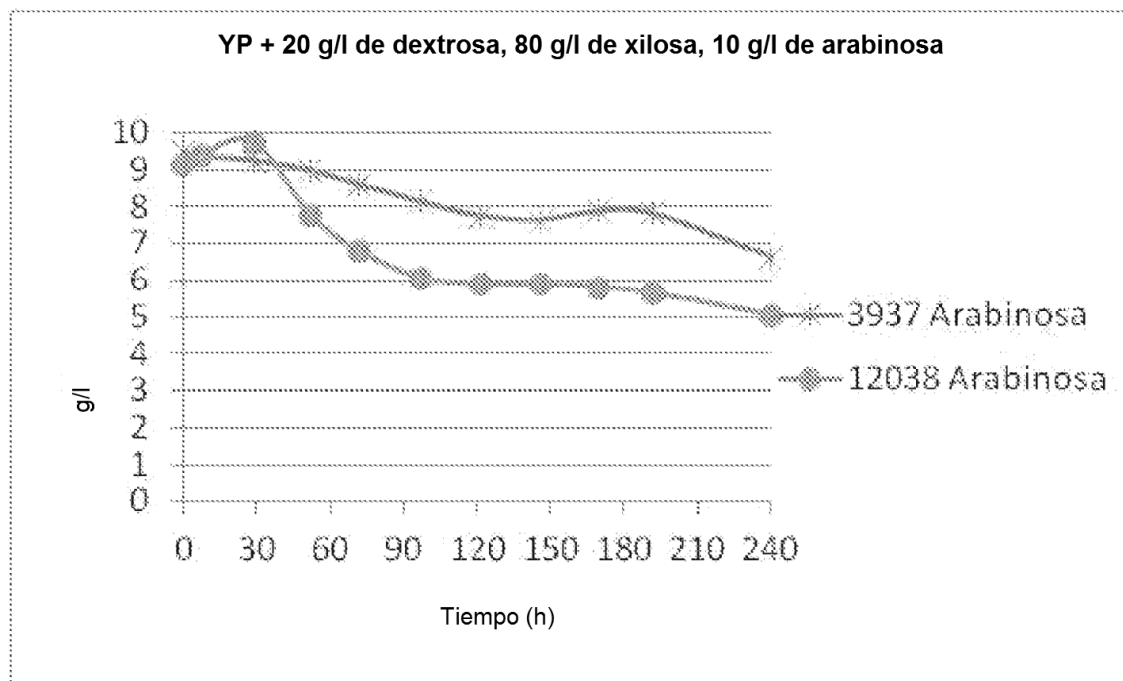


Figura 17

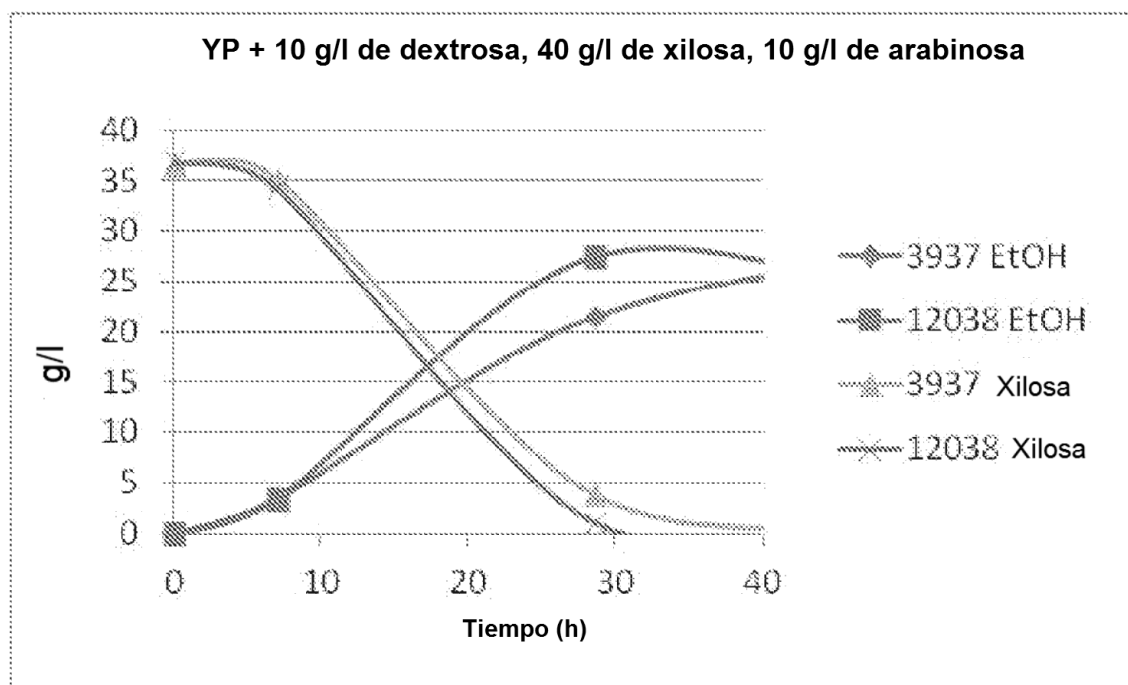


Figura 18

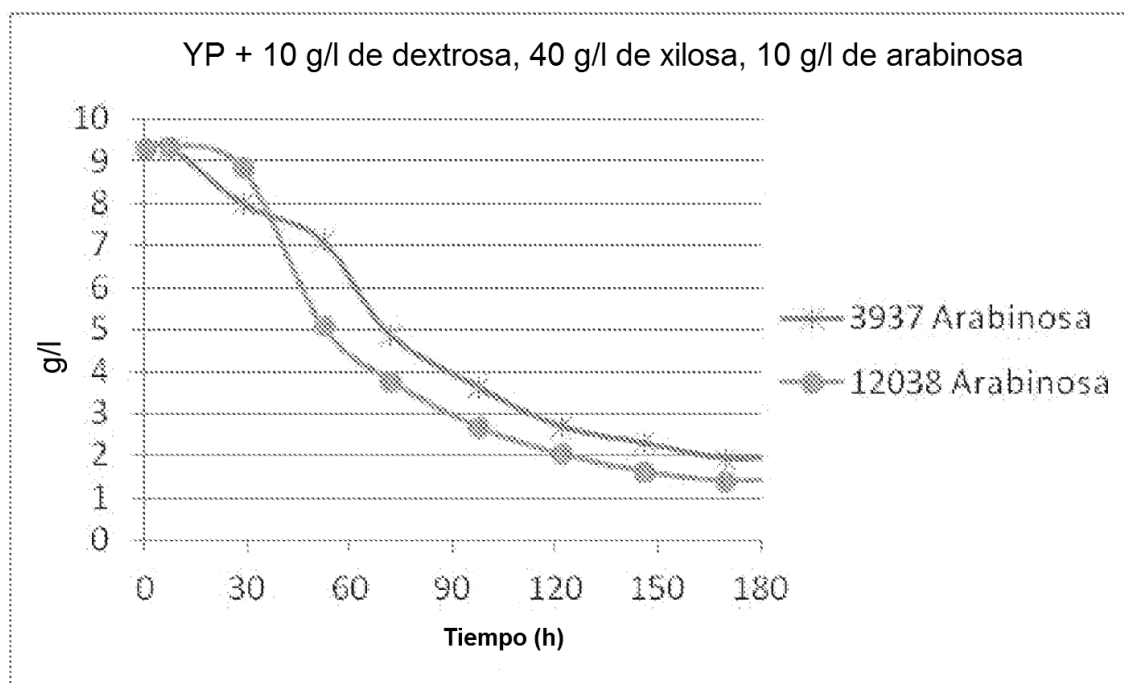


Figura 19

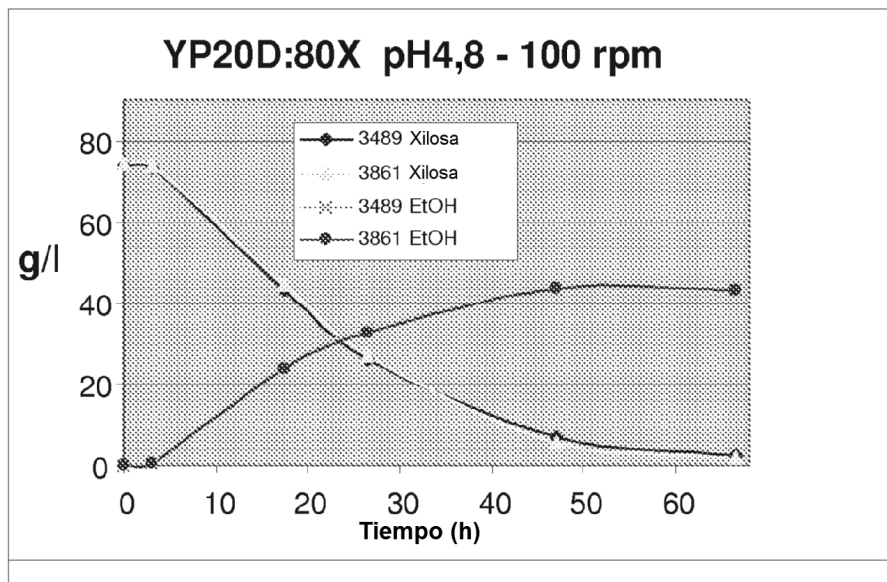


Figura 20

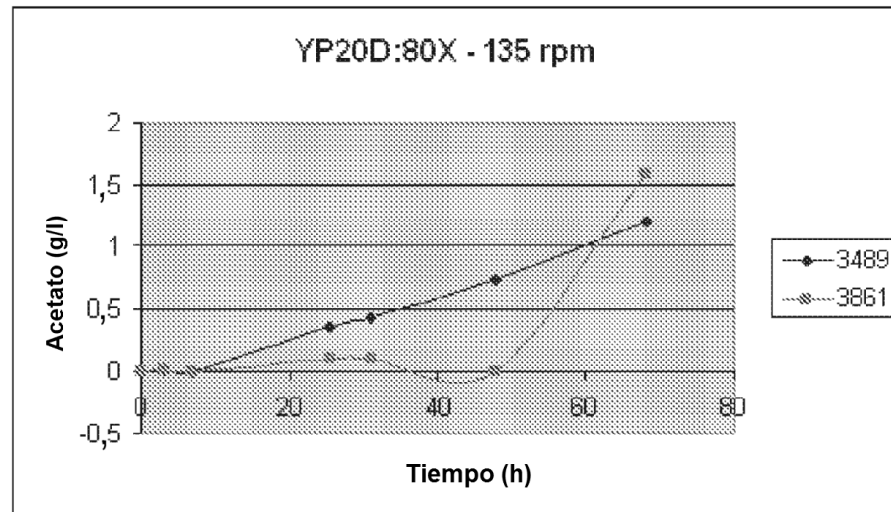


Figura 21

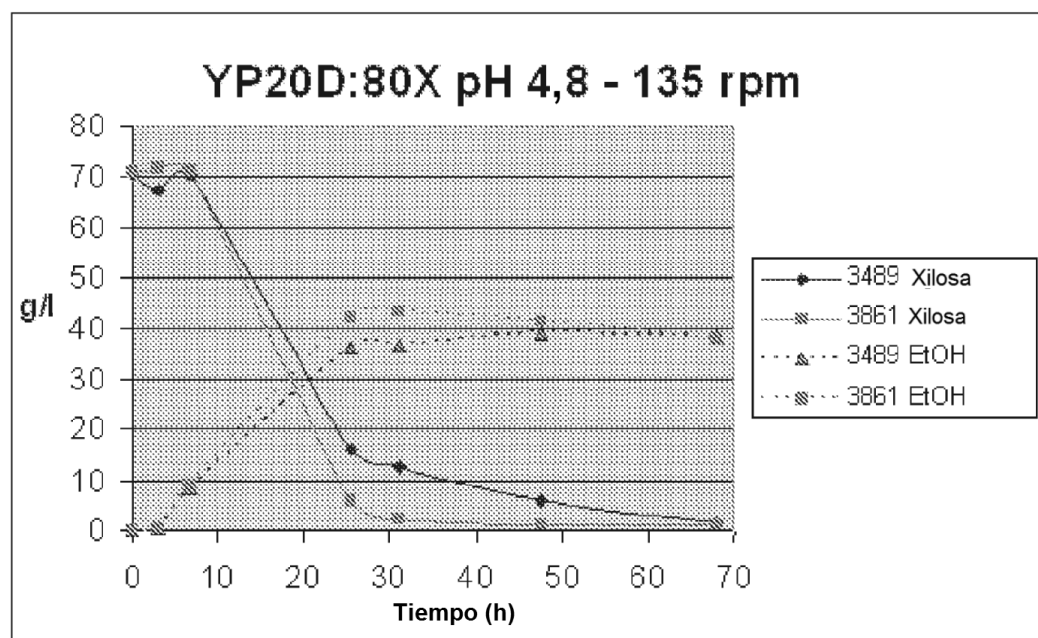


Figura 22

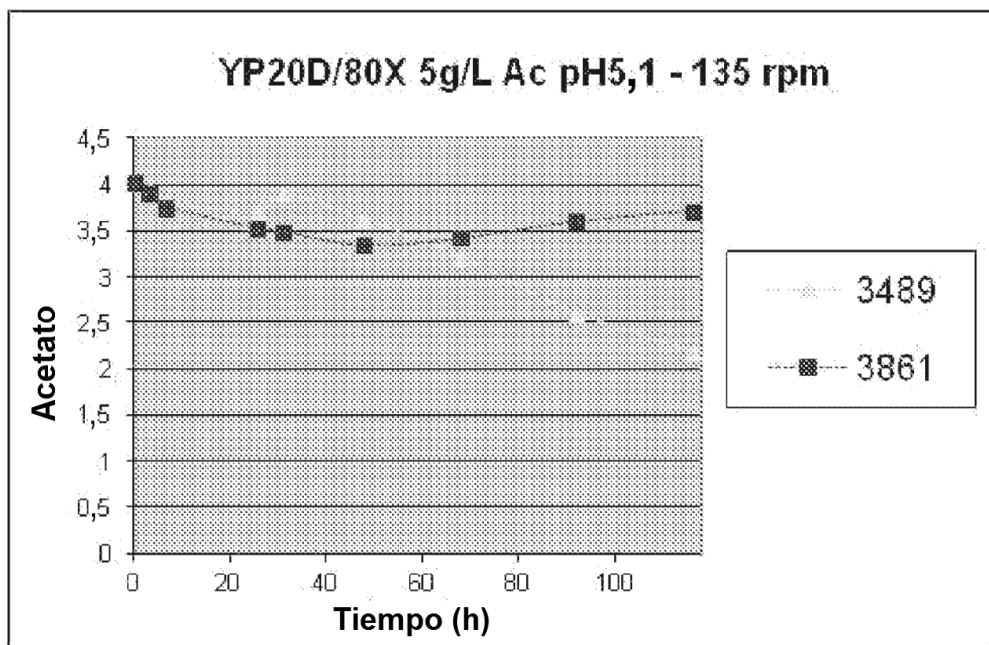


Figura 23

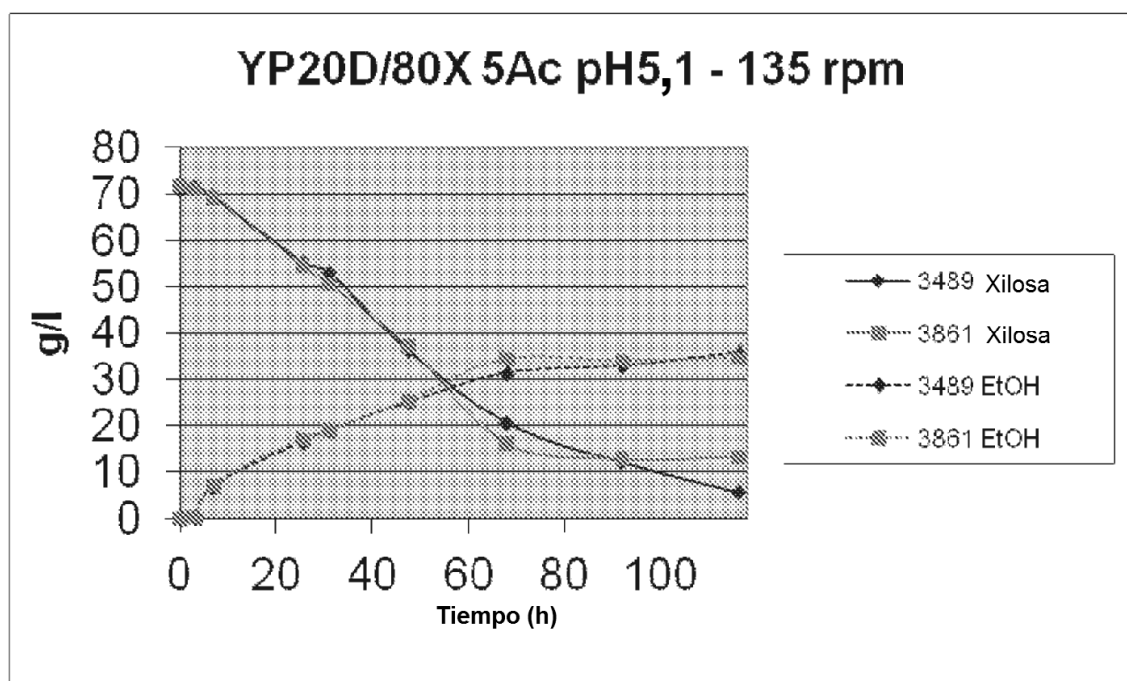


Figura 24

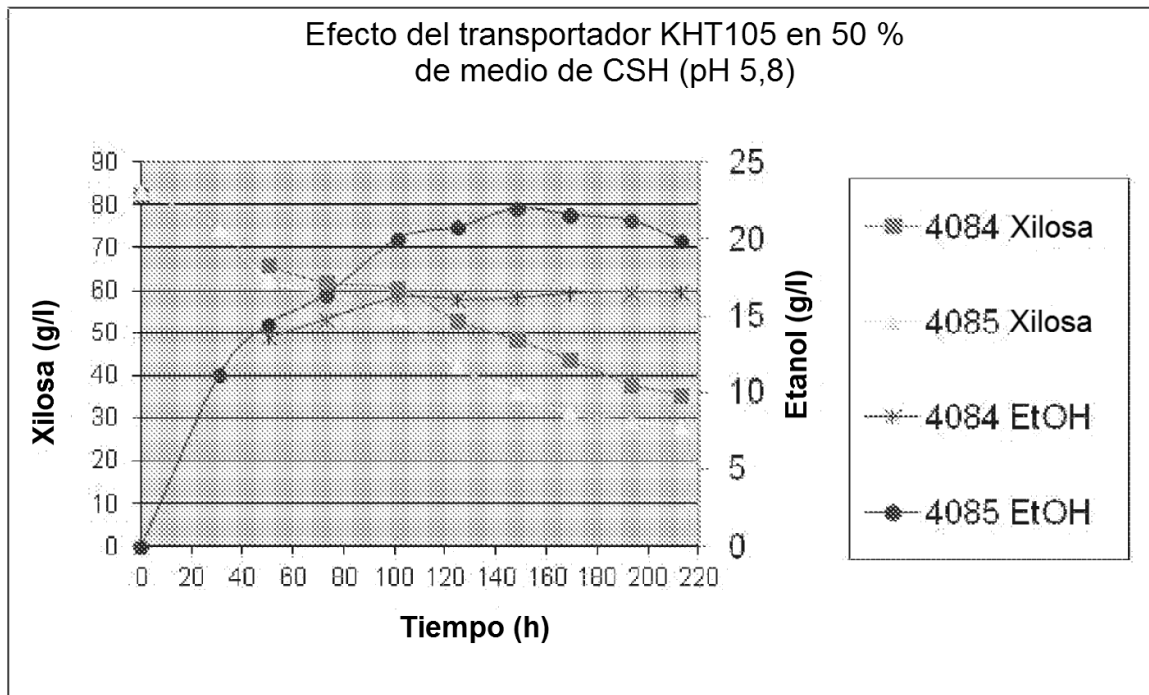


Figura 25

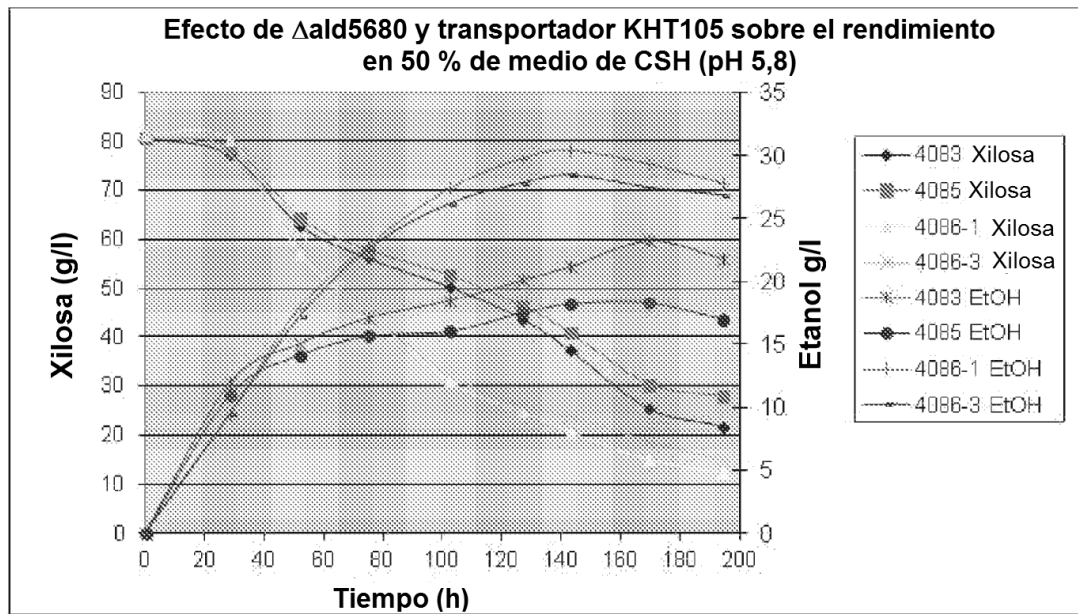


Figura 26

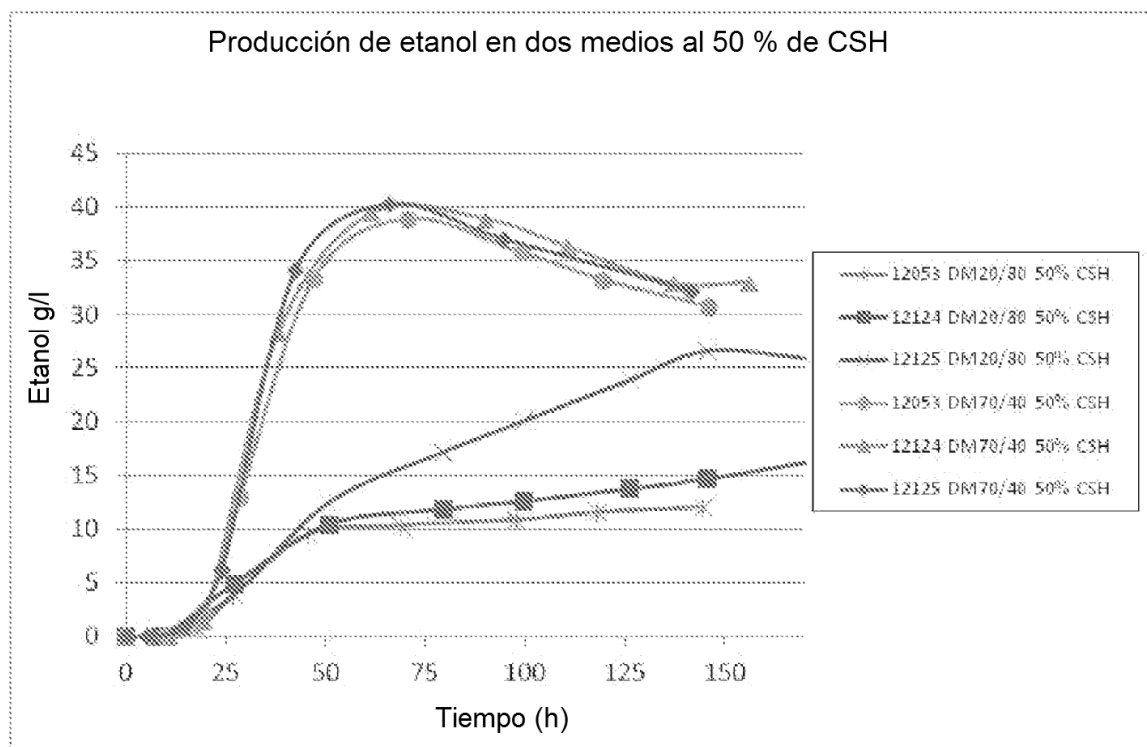


Figura 27

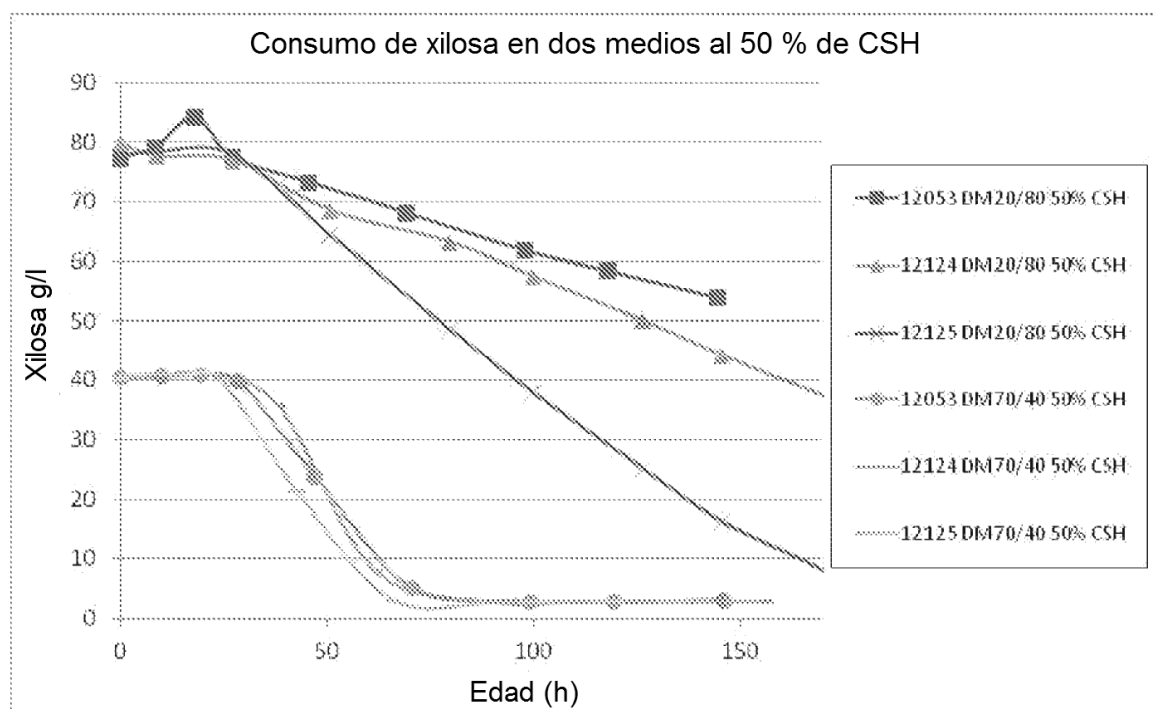


Figura 28

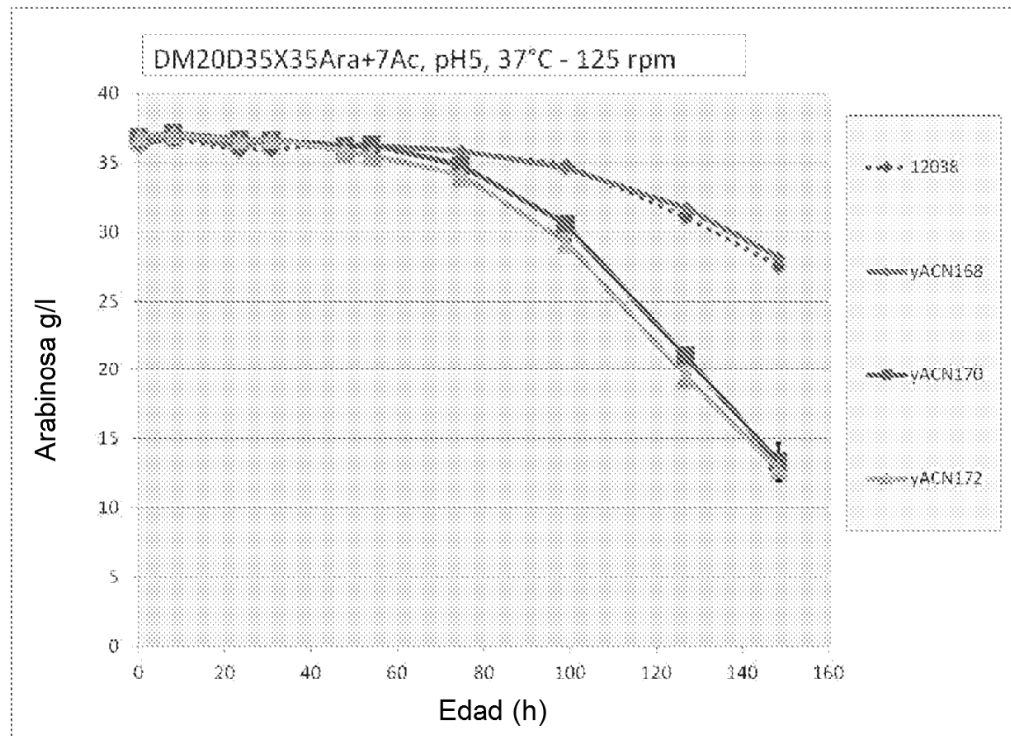


Figura 29

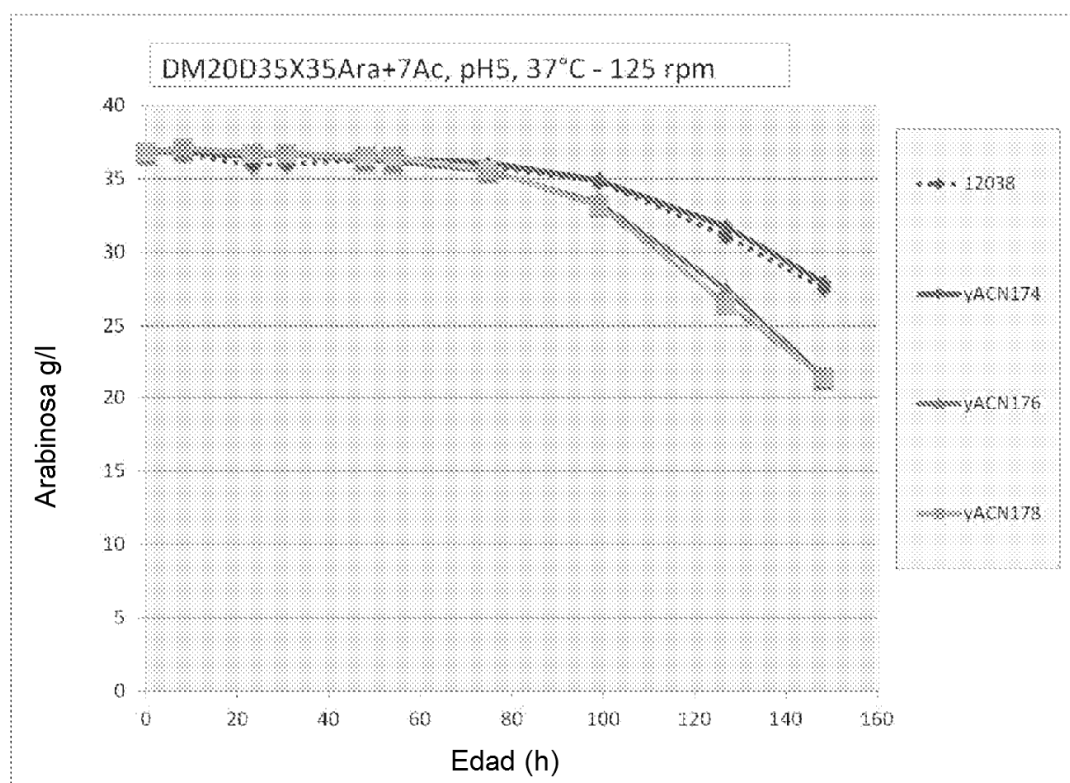


Figura 30

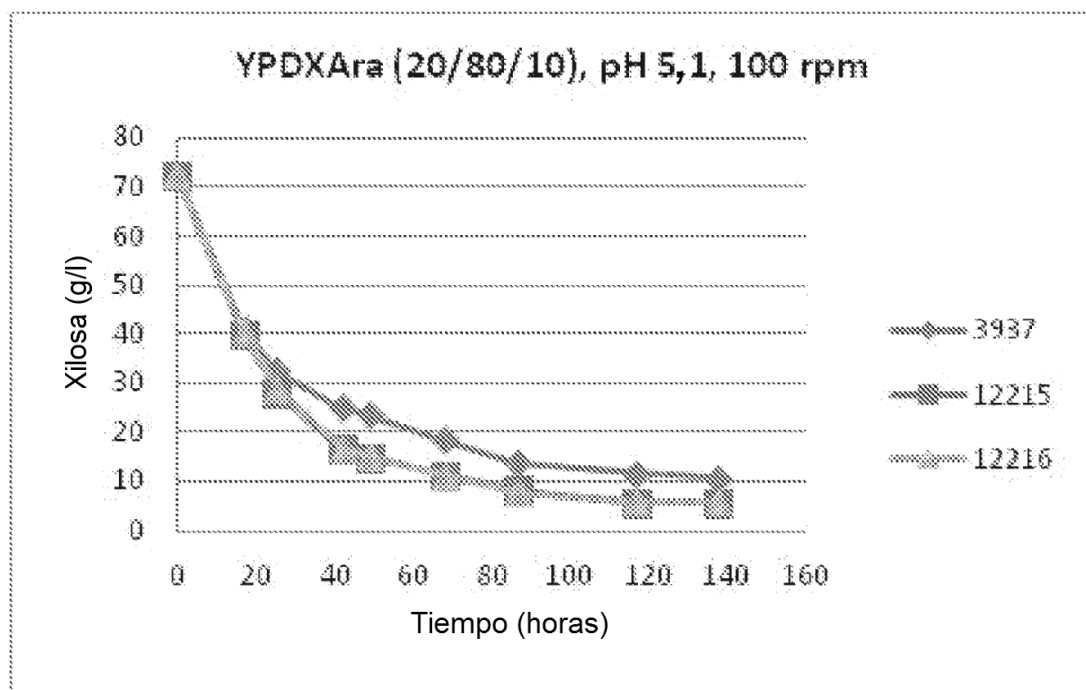


Figura 31

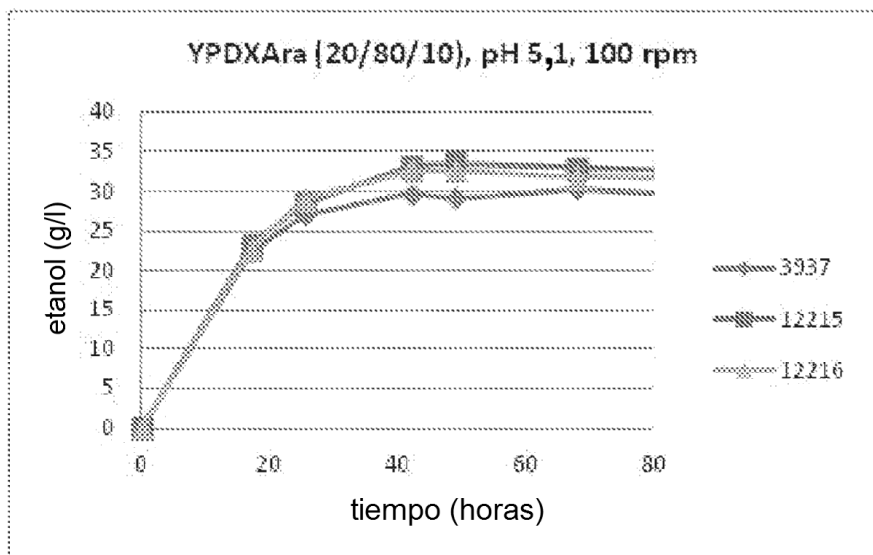


Figura 32

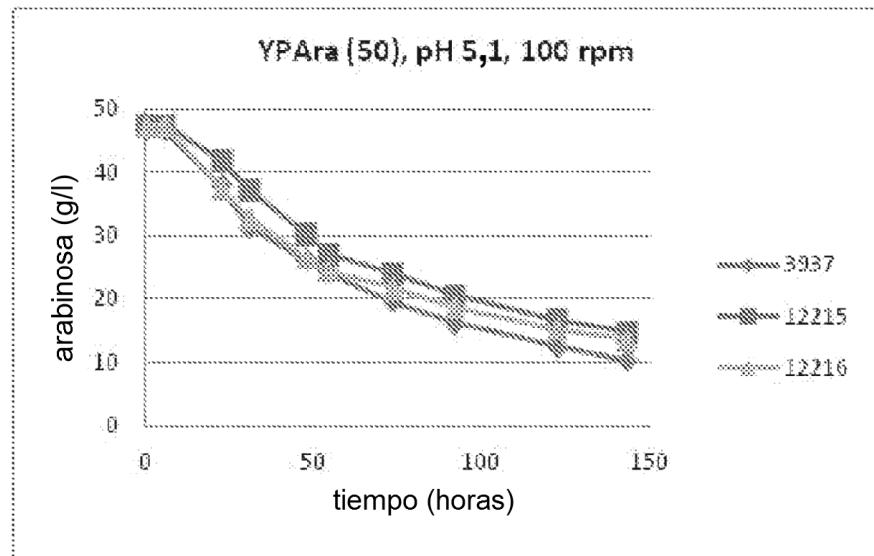


Figura 33

YP con 20 g/l de glucosa y 80 g/l de xilosa, pH 4,8

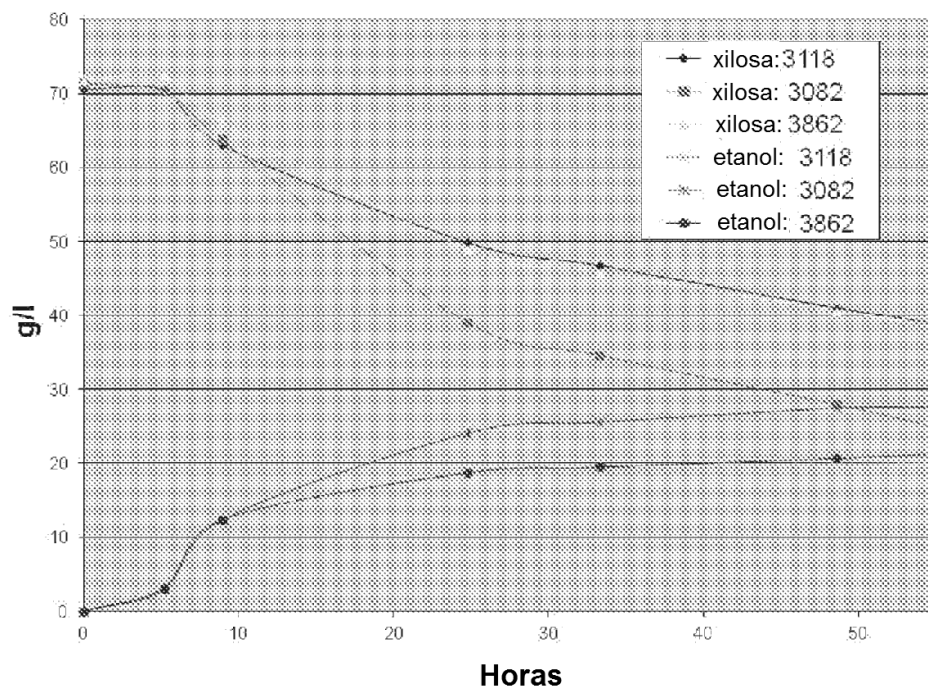


Figura 34

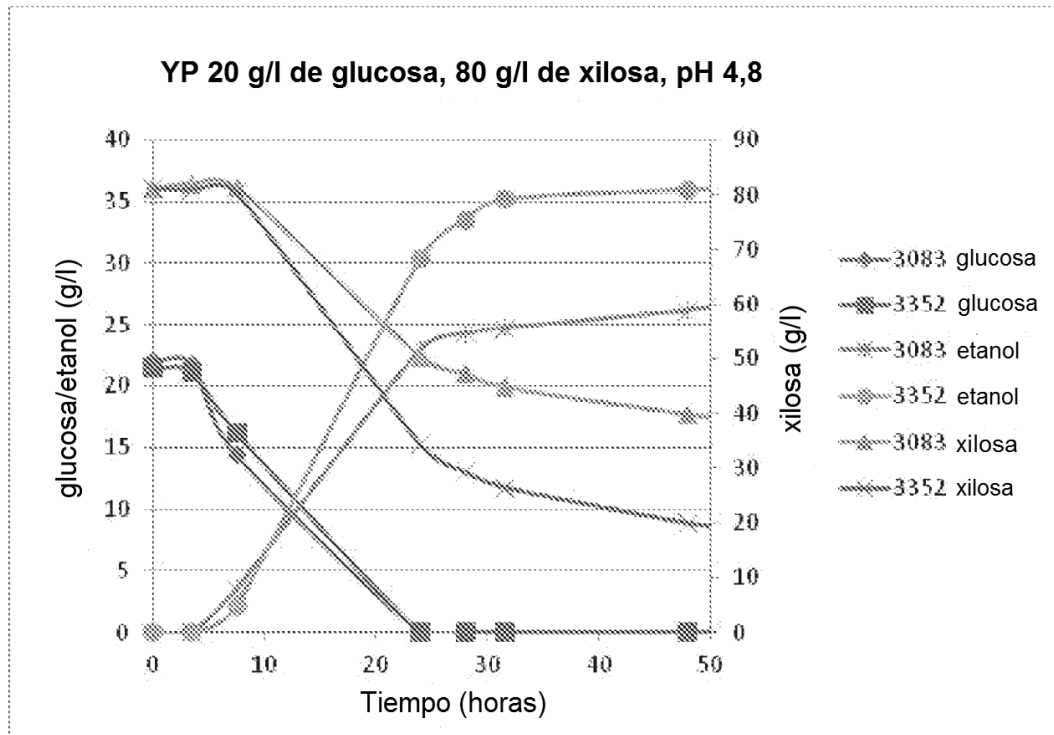


Figura 35

