

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 764 273**

51 Int. Cl.:

C12N 9/10 (2006.01)

C12P 7/64 (2006.01)

C12R 1/89 (2006.01)

C12P 33/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **10.07.2015 PCT/US2015/039951**

87 Fecha y número de publicación internacional: **14.01.2016 WO16007862**

96 Fecha de presentación y número de la solicitud europea: **10.07.2015 E 15747630 (0)**

97 Fecha y número de publicación de la concesión europea: **09.10.2019 EP 3167053**

54 Título: **Nuevos genes de cetoacil ACP sintasa y uso de los mismos**

30 Prioridad:

10.07.2014 US 201462023112 P
18.11.2014 US 201462081143 P

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
02.06.2020

73 Titular/es:

CORBION BIOTECH, INC. (100.0%)
One Tower Place, Suite 600
South San Francisco, CA 94080, US

72 Inventor/es:

DAVIS, DAVID;
RUDEKO, GEORGE;
SOMANCHI, ARAVIND;
CASOLARI, JASON;
FRANKLIN, SCOTT y
EWING, AREN

74 Agente/Representante:

TOMAS GIL, Tesifonte Enrique

ES 2 764 273 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

Nuevos genes de cetoacil ACP sintasa y uso de los mismos

5 CAMPO TÉCNICO

[0001] La presente invención se refiere a genes de β -cetoacil ACP sintasa nuevos y métodos para usar los genes que incluyen la expresión de los genes en células hospedadoras oleaginosas para producir triglicéridos con perfiles de ácidos grasos alterados.

10 ANTECEDENTES

[0002] Algunos organismos que incluyen plantas y algunas microalgas usan una vía biosintética de ácidos grasos de tipo II, caracterizada por el uso de enzimas discretas en un complejo multimérico para la síntesis de ácidos grasos. En cambio, los mamíferos y los hongos usan una única proteína grande multifuncional.

[0003] En los organismos que usan una vía biosintética de ácidos grasos de tipo II, la β -cetoacil-ACP sintasa I (KAS I, EC 2.3.1.41) es una de las enzimas responsables del alargamiento de acil-ACP graso de cadena media creciente de 4 a 16 átomos de carbono de longitud. La KAS I usa acil-ACP C2-C14 como sustratos para la condensación con una unidad C2 derivada de malonil-ACP. KASIV es una enzima relacionada que tiene una función de alargamiento similar. Así, KASI y KASIV pueden considerarse ambas enzimas tipo KASI.

[0004] Se ha introducido tales genes en plantas usando tecnología del ADN recombinante. Véanse por ejemplo las patentes US7301070, US6348642, US6660849, US6770465 y US2006/0094088. En las células plastídicas, tales como las procedentes de plantas, macroalgas y microalgas, se localizan enzimas tipo KAS I en los cloroplastos u otros plástidos junto con otra enzima de la vía de síntesis de ácidos grasos (FAS).

[0005] Las publicaciones PCT WO2010/063032, WO2011/150411, WO2012/106560 y WO2013/158938 revelan ingeniería genética de microalgas oleaginosas que incluye dirigir productos génicos de FAS exógenos al plástido de microalga.

[0006] Las patentes US 2013/004646 y WO 2010/063031 revelan células de microalga que contienen genes exógenos que codifican, por ejemplo, una lipasa, un transportador de sacarosa, una invertasa de sacarosa, una fructoquinasa, una enzima de degradación de polisacárido, una enzima de cetoacil-ACP sintasa, una acil-ACP tioesterasa grasa, una acil-CoA/aldehído reductasa grasa, una acil-CoA reductasa grasa, una aldehído reductasa grasa, una aldehído-descarbonilasa grasa y/o una proteína portadora de acilo.

[0007] La base de datos Geneseq [en línea] 14 de octubre de 2010 (2010-10-04), "Palmitic acid production-related gene, SEQ:20024.", recuperado del n.º de acceso a EBI GSN:AXE01814 Número de acceso a la base de datos AXE01814 divulga una secuencia de polinucleótidos que tiene un 93,6 % de identidad con la SEQ ID NO 21. La base de datos Geneseq [en línea] 15 junio 2007 (2007-06-15), "Oil-associated gene related protein #865.", recuperado del n.º de acceso a EBI GSP:ADJ49365 Número de acceso a la base de datos ADJ49365 divulga una secuencia polipeptídica que tiene un 90,2 % de identidad con la SEQ ID NO 2. La base de datos Geneseq [en línea] 29 octubre 2009 (2009-10-29), "Nucleotide sequence SEQ ID 134280, recuperado del n.º de acceso a EBI GSN:AWK61076 Número de acceso a la base de datos AWK61076 divulga una secuencia de polinucleótidos que tiene un 93,55 % de identidad con la SEQ ID NO 21. La base de datos Geneseq [en línea] 20 junio 2013 (2013-06-20), "Plasmid pSZ1491 DNA construct, SEQ ID 232.", recuperado de n.º de acceso a EBI GSN:BAN54762 Número de acceso a la base de datos BAN54762 divulga una secuencia de polinucleótidos que tiene un 92,44 % de identidad con la SEQ ID NO 39.

50 RESUMEN

[0008] En un aspecto, la invención proporciona un polinucleótido con al menos un 95 % de identidad de secuencia con la SEQ ID NO: 21 o la SEQ ID NO: 39, o un polinucleótido que codifica una proteína de tipo KASI que tiene al menos un 95 % de identidad de secuencia de aminoácidos con la SEQ ID NO: 2, o una proteína madura producida a partir del mismo, donde dicho polinucleótido, cuando se coexpresa con un gen exógeno de acil ACP tioesterasa FATA o FATB en una célula hospedadora de microalga oleaginosa, es capaz de enriquecer el contenido C14:0 del aceite celular producido por la célula hospedadora, con respecto a un aceite producido por una célula del mismo tipo que no expresa dicho polinucleótido.

[0009] En otro aspecto, formas de realización de la invención incluyen un vector de transformación que comprende un polinucleótido tal y como se ha mencionado anteriormente. En algunos casos, el vector comprende secuencias promotoras y 3'UTR en conexión funcional con el polinucleótido y, opcionalmente, una secuencia flanqueante para la recombinación homóloga. Las secuencias promotoras o 3'UTR son secuencias de nucleótidos heterólogas. Las secuencias promotoras heterólogas o 3'UTR heterólogas pueden proceder de un organismo diferente del organismo a partir del cual se obtuvo en primer lugar la codificación de secuencias de nucleótidos KAS.

[0010] En un aspecto, el vector de transformación comprende una secuencia promotora heteróloga o 3'UTR heteróloga obtenida a partir del mismo organismo del que se aisló en primer lugar el gen KAS. Cuando la secuencia promotora, la secuencia 3'UTR y las secuencias de nucleótidos KAS son del mismo organismo, el promotor heterólogo no guía naturalmente la expresión de KAS y la 3'UTR no se produce naturalmente aguas abajo de las secuencias de nucleótidos de KAS en el organismo de origen.

[0011] En otro aspecto, el vector de transformación se utiliza para expresar el gen KAS en el organismo del que se aisló en primer lugar el gen KAS. Cuando el gen KAS se expresa recombinantemente en el organismo del que se aisló en primer lugar el gen KAS, el gen se expresa en un locus cromosómico diferente del locus cromosómico natural del gen KAS. De forma alternativa, el gen KAS se expresa en el citoplasma.

[0012] En otro aspecto, formas de realización de la invención incluyen una célula hospedadora que es una célula oleaginosa plástica que tiene una vía de biosíntesis de ácidos grasos de tipo II que comprende el polinucleótido y/o el vector mencionado anteriormente, y que expresa una proteína de KAS funcional codificada por el polinucleótido. En algunos casos, la célula hospedadora comprende además un gen exógeno que codifica una acil-ACP tioesterasa FATA o acil-ACP tioesterasa FATB funcional. En un aspecto, la acil-ACP tioesterasa FATB tiene al menos un 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98 o 99 % de identidad de secuencia de aminoácidos con cualquiera de las SEQ ID NO: 1 o SEQ ID NO: 57. En algunos casos, la célula hospedadora produce un aceite celular caracterizado por un perfil de ácidos grasos con (i) al menos un 30, 40, 50 o 55 % de C14:0, (ii) al menos un 7, 8, 9, 10, 11, 12, 13 o 14 % de C8:0, (iii) al menos un 10, 15, 20, 25, 30 o 35 % en área para la suma de C8:0 y C10:0, o (iv) una relación C8/C10 en el rango de 2,2-2,5, 2,5-3,0 o 3,0-3,4. En algunos casos, la célula hospedadora es una microalga. En algunos casos, la célula hospedadora es de *Trebouxioophyceae*, y opcionalmente del género *Chlorella* o *Prototheca*. En algunos casos, la microalga es de la especie *Prototheca moriformis*.

[0013] En otro aspecto, formas de realización de la invención incluyen un método para hacer un aceite celular, donde el método comprende cultivar una célula hospedadora tal y como se ha mencionado anteriormente para producir el aceite celular, donde el aceite comprende triglicéridos y esteroides de microalga. En algunos casos, el aceite celular comprende esteroides caracterizados por un perfil de esteroles y el perfil de esteroles tiene un exceso de ergosterol con respecto al β -sitosterol y/o la presencia de 22,23-dihidrobrassicasterol, poriferasterol o clonasterol.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

[0014] La fig. 1 muestra un árbol filogenético para genes de tipo KAS1 en relación con el ejemplo 3.

DESCRIPCIÓN DETALLADA DE LAS FORMAS DE REALIZACIÓN DE LA INVENCION

[0015] Tal y como se usa con respecto a los ácidos nucleicos, el término "aislado" se refiere a un ácido nucleico que está libre de al menos otro componente que está típicamente presente con el ácido nucleico de origen natural. De este modo, un ácido nucleico de origen natural está aislado si se ha purificado apartado de al menos otro componente de origen natural con el ácido nucleico.

[0016] Un "aceite celular" o "grasa celular" significará un aceite predominantemente de triglicérido obtenido a partir de un organismo, donde el aceite no se ha sometido a mezcla con otro aceite natural o sintético, o a fraccionamiento para alterar sustancialmente el perfil de ácidos grasos del triglicérido. En relación con un aceite que comprende triglicéridos de una regioespecificidad particular, el aceite celular o grasa celular no ha sido sometido a interesterificación u otro proceso sintético para obtener tal perfil triglicérido regioespecífico, más bien la regioespecificidad se produce naturalmente, por una célula o una población de células. Para un aceite celular o grasa celular producido por una célula, el perfil de esteroles del aceite se determina generalmente por los esteroides producidos por la célula, no por reconstitución artificial del aceite por adición esteroides para imitar el aceite celular. En relación con un aceite celular o grasa celular, y tal y como se utiliza generalmente a lo largo de la presente descripción, los términos aceite y grasa se usan de forma intercambiable, excepto donde se indique lo contrario. De este modo, un "aceite" o una "grasa" puede ser líquido, sólido o parcialmente sólido a temperatura ambiente, dependiendo de la composición de la sustancia y otras condiciones. Aquí, el término "fraccionamiento" significa eliminar material del aceite de manera que cambie su perfil de ácidos grasos con respecto al perfil producido por el organismo, independientemente de cómo se lleve a cabo. Los términos "aceite celular" y "grasa celular" engloban tales aceites obtenidos a partir de un organismo, donde el aceite ha sido sometido a un procesamiento mínimo, que incluye refinación, blanqueo y/o desgomado, que no cambia sustancialmente su perfil de triglicéridos. Un aceite celular también puede ser un "aceite celular no interesterificado", lo que significa que el aceite celular no ha sido sometido a un proceso en el que los ácidos grasos se han redistribuido en sus enlaces de acilo con el glicerol y permanece esencialmente con la misma configuración que cuando se extrajo del organismo.

[0017] "Gen exógeno" designará un ácido nucleico que codifica para la expresión de un ARN y/o proteína que se ha introducido en una célula (por ejemplo mediante transformación/transfección) y también se denomina "transgén". Una célula que comprende un gen exógeno se puede denominar célula recombinante, en la que se

pueden introducir uno o más genes exógenos adicionales. El gen exógeno puede proceder de una especie diferente (y por tanto heteróloga), o de la misma especie (y por lo tanto homóloga), con respecto a la célula que se está transformando. De este modo, un gen exógeno puede incluir un gen homólogo que ocupa una ubicación diferente en el genoma de la célula o está bajo control diferente, con respecto a la copia endógena del gen. Un gen exógeno puede estar presente en más de una copia en la célula. Un gen exógeno se puede mantener en una célula, por ejemplo, como una inserción en el genoma (nuclear o plástido) o como una molécula episómica.

[0018] "Ácidos grasos" designará ácidos grasos libres, sales de ácidos grasos o fracciones de acilo graso en un glicerolípido. Se entenderá que los grupos acilo grasos de glicerolípidos se pueden describir en términos del ácido carboxílico o anión de un ácido carboxílico que se produce cuando el triglicérido se hidroliza o saponifica.

[0019] "Microalgas" son organismos microbianos que contienen un cloroplasto u otro plástido, y opcionalmente que son capaces de realizar fotosíntesis, o un organismo microbiano procariótico capaz de realizar fotosíntesis. Las microalgas incluyen fotoautótrofos obligados, que no pueden metabolizar una fuente de carbono fijo como energía, así como heterótrofos, que pueden vivir exclusivamente de una fuente de carbono fija. Las microalgas incluyen organismos unicelulares que se separan de las células hermanas poco después de la división celular, tales como las *Chlamydomonas*, al igual que los microbios tales como, por ejemplo, *Volvox*, que es un microbio fotosintético multicelular simple de dos tipos de célula diferentes. Las microalgas incluyen células tales como *Chlorella*, *Dunaliella* y *Prototheca*. Las microalgas incluyen también otros organismos fotosintéticos microbianos que presentan una adhesión célula-célula, tales como *Agmenellum*, *Anabaena* y *Pyrobotrys*. Las microalgas incluyen también microorganismos heterótrofos obligados que han perdido la capacidad de realizar fotosíntesis, tales como algunas especies de algas dinoflageladas y especies del género *Prototheca*.

[0020] Una célula "oleaginosa" es una célula capaz de producir al menos un 20 % de lípido por peso de célula en seco, naturalmente o a través de la mejora de cepa recombinante o tradicional. Un "microbio oleaginoso" o "microorganismo oleaginoso" es un microbio, que incluye una microalga, que es oleaginoso.

[0021] El término "porcentaje de identidad de secuencia", en el contexto de dos o más secuencias de aminoácidos o de ácidos nucleicos, se refiere a dos o más secuencias o subsecuencias que son iguales o tienen un porcentaje específico de residuos de aminoácidos o nucleótidos idéntico, cuando se comparan y alinean para una correspondencia máxima, medida usando un algoritmo de comparación de secuencias o mediante inspección visual. Para la comparación de secuencias para determinar el porcentaje de nucleótidos o de identidad de aminoácido, típicamente una secuencia actúa como una secuencia de referencia, con la que se comparan secuencias de prueba. Cuando se usa un algoritmo de comparación de secuencias, las secuencias de prueba y referencia se introducen en un ordenador, se designan las coordenadas de la subsecuencia, si es necesario, y se designan los parámetros del programa de algoritmo de secuencias. El algoritmo de comparación de secuencias calcula entonces el porcentaje de identidad de secuencia para la o las secuencias de prueba con respecto a la secuencia de referencia, basándose en los parámetros del programa designado. Se puede llevar a cabo un alineamiento óptimo de secuencias para la comparación usando el software NCBI BLAST (ncbi.nlm.nih.gov/BLAST/) configurado en los parámetros por defecto. Por ejemplo, para comparar dos secuencias de ácidos nucleicos, se puede usar blastn con la herramienta "BLAST 2 Sequences" versión 2.0.12 (21 de abril de 2000) configurada en los parámetros por defecto siguientes: Matriz: BLOSUM62; Recompensa por coincidencia: 1; Penalización por incompatibilidad: -2; Hueco abierto: 5 y Hueco de extensión: 2 penalizaciones; Hueco x reducción: 50; Probabilidad: 10; Tamaño de palabra: 11; Filtro: activado. Para una comparación de parejas de dos secuencias de aminoácidos, se puede usar la herramienta "BLAST 2 Sequences" versión 2.0.12 (21 de abril de 2000) con blastp configurado, por ejemplo, en los siguientes parámetros por defecto: Matriz: BLOSUM62; Hueco abierto: 11 y Hueco de extensión: 1 penalizaciones; Hueco x reducción 50; Probabilidad: 10; Tamaño de palabra: 3; filtro: activado.

[0022] Donde se proporcionan múltiples identidades de secuencia para una cepa con un par de genes exógenos, esto engloba todas las combinaciones de identidades de secuencia. Por ejemplo, la coexpresión de un primer gen que codifica una primera proteína que tiene al menos un 85, 90, 91, 92, 93, 94, 95, 96, 97, 98 o 99 % con el gen A y un segundo gen que codifica una segunda proteína que tiene al menos un 85, 90, 91, 92, 93, 94, 95, 96, 97, 98 o 99 % con el gen A debe entenderse englobando (i) al menos un 85 % de identidad con el gen A y al menos un 85 % de identidad con el gen B, (ii) al menos un 85 % de identidad con el gen A y al menos un 99 % de identidad con el gen B, (iii) al menos un 92 % de identidad con el gen A y al menos un 95 % de identidad con el gen B y todas las demás combinaciones.

[0023] En relación con un aceite celular, un "perfil" es la distribución de especies particulares de triglicéridos o grupos acilo grasos en el aceite. Un "perfil de ácidos grasos" es la distribución de grupos acilo grasos en los triglicéridos del aceite sin referencia a la fijación a una estructura de glicerol. Los perfiles de ácidos grasos se determinan típicamente por conversión a un éster metílico de ácidos grasos (FAME), seguida de un análisis de cromatografía de gases (GC) con detección de ionización de llama (FID). El perfil de ácidos grasos se puede expresar como uno o más por ciento de un ácido graso en la señal de ácidos grasos totales determinada a partir del área bajo la curva para dicho ácido graso. La medición FAME-GC-FID se aproxima a los porcentajes en peso de los ácidos grasos.

[0024] Tal y como se utiliza en este caso, se dice que un aceite está "enriquecido" en uno o más ácidos grasos particulares si hay al menos un 10 % de aumento en la masa de dicho ácido graso en el aceite con respecto al aceite no enriquecido. Por ejemplo, en el caso de una célula que expresa un gen FatB heterólogo descrito aquí, se dice que el aceite producido por la célula está enriquecido con, por ejemplo, ácidos grasos C8 y C16 si la masa de estos ácidos grasos en el aceite es al menos un 10 % superior a la del aceite producido por una célula del mismo tipo que no expresa el gen FatB heterólogo (por ejemplo aceite de tipo salvaje).

[0025] "Recombinante" es una célula, ácido nucleico, proteína o vector que ha sido modificado debido a la introducción de un ácido nucleico exógeno o a la alteración de un ácido nucleico nativo. Así, por ejemplo, las células (hospedadoras) recombinantes pueden expresar genes que no se encuentran en la forma nativa (no recombinante) de la célula o expresar genes nativos de forma diferente a como se expresan tales genes por una célula no recombinante. Las células recombinantes pueden, sin limitación, incluir ácidos nucleicos recombinantes que codifican un producto génico o elementos de supresión tales como mutaciones, inactivaciones, antisentido, ARN (ARNi) o ARNbc interferente que reducen los niveles de producto génico activo en una célula. Un "ácido nucleico recombinante" es un ácido nucleico formado originalmente *in vitro*, en general por la manipulación del ácido nucleico, por ejemplo usando polimerasas, ligasas, exonucleasas y endonucleasas, usando síntesis química o, de otro modo, está en una forma que no se encuentra normalmente en la naturaleza. Se pueden producir ácidos nucleicos recombinantes, por ejemplo, para colocar dos o más ácidos nucleicos en el enlace funcional. Así, tanto un ácido nucleico aislado como un vector de expresión formados *in vitro* por nucleico mediante ligamiento de moléculas de ADN que no se unen normalmente en la naturaleza, se consideran recombinantes para los fines de esta invención. También se pueden producir ácidos nucleicos recombinantes de otras formas; por ejemplo usando síntesis de ADN química. Una vez que se hace un ácido nucleico recombinante y se introduce en una célula u organismo hospedador, se puede replicar usando la maquinaria celular *in vivo* de la célula hospedadora; sin embargo, tales ácidos nucleicos, una vez que se han producido de forma recombinante, aunque posteriormente se repliquen intracelularmente, todavía se consideran recombinantes para fines de la presente invención. De forma similar, una "proteína recombinante" es una proteína hecha usando técnicas recombinantes, es decir a través de la expresión de un ácido nucleico recombinante.

[0026] Un gen o enzima "de tipo KAS I" designará un gen o enzima KAS I o KAS IV.

[0027] Las formas de realización de la presente invención se refieren al uso de genes de tipo KASI tal y como se describe en las reivindicaciones, aislados de plantas, que se pueden expresar en una célula hospedadora transgénica para alterar el perfil de ácidos grasos de un aceite celular producido por la célula hospedadora. Aunque la *Prototheca moriformis* de microalga se ha usado para cribar los genes por capacidad para alterar el perfil de ácidos grasos, los genes descubiertos son útiles en una amplia variedad de células hospedadoras para las que son conocidas técnicas de transformación genética. Por ejemplo, los genes se pueden expresar en bacterias, cianobacterias, otras microalgas eucarióticas o plantas superiores. Los genes se pueden expresar en plantas superiores según los métodos descritos en las patentes de EE. UU. de números: US7301070, US6348642, US6660849 y US6770465. Se ha descubierto que los transgenes de tipo KASI se pueden usar solos o en combinación con un transgén FatB (que codifica una acil-ACP tioesterasa activa) para estimular los niveles de ácidos grasos de cadena media (por ejemplo, ácidos cápricos, caprílicos, láuricos, mirísticos o palmíticos) en el perfil de ácidos grasos del aceite celular. La combinación de un gen de tipo KASI exógeno con un gen FATA o FATB exógeno en una célula hospedadora puede proporcionar niveles de ácidos grasos de cadena media y/o ácidos grasos de cadena larga (por ejemplo esteárico u oleico) superiores a un gen exógeno solo. Los ácidos grasos del aceite celular se pueden convertir además en triglicéridos, aldehídos grasos, alcoholes grasos y otros oleoquímicos ya sea sintética o biosintéticamente.

[0028] En formas de realización específicas, los triglicéridos se producen por una célula hospedadora que expresa un gen de tipo KASI nuevo tal y como se describe en las reivindicaciones (a partir de un ADNc nuevo y/o bajo control de un promotor heterólogo). Se puede extraer un aceite celular de la célula hospedadora. Generalmente, el aceite celular comprende principalmente triglicéridos y esteroides. El aceite celular se puede refinar, desgomar, blanquear y/o desodorizar. El aceite, en su forma sin procesar o procesada, se puede usar para alimentos, productos químicos, combustibles, cosméticos, plásticos y otros usos.

[0029] Los genes KAS se pueden usar en una variedad de construcciones genéticas que incluyen plásmidos u otros vectores para una expresión o recombinación en una célula hospedadora. Los genes pueden presentar una optimización de codones para la expresión en una célula hospedadora diana. Los genes se pueden incluir en un casete de expresión que incluye un promotor (por ejemplo un promotor heterólogo) y elemento regulador aguas abajo. El vector puede incluir secuencias flanqueantes para la recombinación homóloga. Por ejemplo, el vector puede causar la inserción en un cromosoma de la célula hospedadora, donde se puede expresar de forma estable. Las proteínas producidas por los genes se pueden usar *in vivo* o en forma purificada. En una forma de realización, un casete de expresión comprende un promotor homólogo, una CDS (secuencia codificante, por sus siglas en inglés) operable para expresar una enzima de tipo KASI de la tabla 1 y una 3'UTR. La 3'UTR puede comprender un sitio de poliadenilación.

[0030] Tal y como se describe en los ejemplos más adelante, los genes KAS nuevo se han descubierto a partir de ADNc producido a partir de transcritos de ARNm de semillas vegetales. Por consiguiente, las secuencias de gen no son naturales porque carecen de intrones que están presentes en los genes vegetales y transcritos de ARNm de los genes antes del empalme de ARNm. Por consiguiente, la invención comprende un gen de tipo KASI no natural aislado en la tabla 1. Otra desviación del gen natural está en el uso de elementos reguladores heterólogos y expresión en células hospedadoras para las que tales genes no se producen de forma natural.

[0031] Por ejemplo, el gen se puede preparar en un vector de expresión que incluye un promotor y 5'UTR enlazados operativamente. Donde una célula plastídica se usa como el hospedador, un péptido dirigido a plástido adecuadamente activo (designado también más adelante como un "péptido de tránsito") se puede fusionar al gen de tipo KASI, como en los ejemplos más adelante. Los genes descritos comprenden una secuencia de identificación de plástido N-terminal hidrofóbica, que se puede sustituir por una secuencia de identificación alternativa y variar en longitud. Variar el péptido dirigido a plástido puede mejorar la localización celular y la actividad enzimática para un tipo de célula hospedadora determinada. De este modo, la invención contempla delecciones y proteínas de fusión para optimizar la actividad enzimática en una célula hospedadora determinada. Por ejemplo, se puede usar un péptido de tránsito del hospedador o especies relacionadas en vez del de los genes vegetales recién descubiertos descritos aquí. Las delecciones adicionales terminales o internas pueden hacerse siempre que se retenga la actividad enzimática. El péptido dirigido se puede escindir por la célula hospedadora para producir una proteína de tipo KASI madura que carece del péptido dirigido.

[0032] Un gen marcador seleccionable puede estar incluido en el vector para ayudar en el aislamiento de una célula transformada. Ejemplos de marcadores seleccionables útiles en microalgas incluyen invertasa de sacarosa, alfa galactosidasa (para la selección en melibiosa) y genes de resistencia antibiótica.

[0033] Las secuencias de gen descritas también se pueden usar para preparar ARN antisentido o inhibitorio (por ejemplo, ARNi o ARN en horquilla) para inhibir genes complementarios en una planta u otro organismo. Por ejemplo, con la ayuda del conocimiento de una secuencia de genes de la tabla 1, se puede obtener por ingeniería una planta con el mismo gen de tipo KASI o similar para expresar un constructo de ARNi, un gen inactivado, una mutación puntual o similar y así reducir la actividad de KASI o KASIV de la semilla de la planta. Como resultado de ello, la planta puede producir un aceite con un perfil de ácidos grasos alterado donde se disminuye o aumenta la longitud de cadena media, dependiendo de la presencia de otros genes de síntesis de ácidos grasos.

[0034] Los genes/proteínas de tipo KASI que se han revelado como útiles para producir perfiles de ácidos grasos deseados en una célula se resumen más adelante en la tabla 1 y las proteínas relativas descubiertas mediante secuenciación de transcripción (como en los ejemplos 1 y 2) se muestran en la tabla 1a. Se pueden usar ácidos nucleicos o proteínas que tienen la secuencia de SEQ ID NOS: 2 y 39 para alterar el perfil de ácidos grasos de una célula recombinante. También se pueden usar variantes de ácidos nucleicos; por ejemplo variantes que tengan al menos un 95 % de identidad de secuencia con las SEQ ID NOS: 21 o 39. Se contempla la optimización de codones de los genes para una variedad de organismos hospedadores, como lo es el uso de fragmentos génicos. Los codones preferidos para las cepas de *Prototheca* y para *Chlorella protothecoides* se muestran a continuación en las tablas 2 y 3, respectivamente. El uso de codones para *Cuphea wrightii* se muestra en la tabla 4. El uso de codones para *Arabidopsis* se muestra en la tabla 5; por ejemplo, se puede seleccionar el codón más preferido para cada aminoácido. En la técnica se conocen tablas de codones para otros organismos que incluyen microalgas y plantas superiores. En algunas formas de realización, los primeros y/o segundos codones de *Prototheca* más preferidos se emplean para la optimización de codones. En formas realización específicas, las secuencias de aminoácidos nuevas contenidas en los listados de secuencias más adelante se convierten en secuencias de ácidos nucleicos según el uso de codón más preferido en *Prototheca*, *Chlorella*, *Cuphea wrightii* o *Arabidopsis* tal y como se expone en las tablas 2 a 3b o secuencias de ácidos nucleicos con al menos un 95 % de identidad de secuencia con estas secuencias de ácidos nucleicos derivadas. Por ejemplo, el gen de tipo KASI puede presentar optimización de codones para *Prototheca moriformis* sustituyendo los codones más preferidos según la tabla 2 por al menos un 10, 20, 30, 40, 50, 60, 70, 80 o 90 % de todos los codones. Asimismo, el gen de tipo KASI puede presentar optimización de codones para *Chlorella protothecoides* sustituyendo los codones más preferidos según la tabla 3 por al menos un 10, 20, 30, 40, 50, 60, 70, 80 o 90 % de todos los codones. De forma alternativa, el gen de tipo KASI puede presentar optimización de codones para *Chlorella protothecoides* o *Prototheca moriformis* sustituyendo primeros o segundos codones más preferidos según la tabla 2 o 3 por al menos un 10, 20, 30, 40, 50, 60, 70, 80 o 90 % de todos los codones. Los genes de codones optimizados son de origen no natural porque se optimizan para la expresión en un organismo hospedador.

[0035] En algunas formas de realización, el porcentaje de identidad de secuencia para las variantes de los ácidos nucleicos o proteínas mencionados anteriormente se puede calcular usando la secuencia de ácidos nucleicos en toda su longitud (por ejemplo una de la SEQ ID NOS: 21 y 39 o la secuencia de aminoácidos en toda su longitud (por ejemplo la SEQ ID NO: 2) como la secuencia de referencia y comparando la secuencia de prueba en toda su longitud con esta secuencia de referencia. Para los fragmentos, el porcentaje de identidad de secuencia para las variantes de ácido nucleico o fragmentos de proteína se puede calcular a lo largo de toda la longitud del fragmento. En algunas formas de realización, hay un ácido nucleico o fragmento de proteína con al menos un 96, 97, 98 o 99 % de identidad de secuencia con una de las SEQ ID NOS: 21, 39 o 2.

[0036] Opcionalmente, el péptido dirigido a plástido se puede intercambiar por otro péptido que funciona para guiar la enzima tipo KASI a un plástido sintetizador de ácidos grasos de una célula hospedadora plastídica. Por consiguiente, en varios ejemplos de la divulgación, una célula de transgén o transgénica hospedadora comprende un nucleótido o fusión peptídica correspondiente de una secuencia dirigida a plástido y una secuencia de dominio de enzima (la secuencia restante después de la eliminación del péptido de tránsito), donde la proteína madura tiene al menos un 70, 80, 85, 90, 95, 96, 97, 98 o 99 % de identidad de secuencia con una secuencia de proteína madura enumerada en la tabla 1 o la tabla 1a. Los péptidos de tránsito/dirigidos a plástido están subrayados en el listado de secuencias informal anexo. Ejemplos de péptidos dirigidos incluyen aquellos de la tabla 1 y otros conocidos en la técnica, especialmente en relación con la dirección de los productos génicos de KAS I, KAS II, KAS III, FATA, FATB y SAD (estearoil-ACP desaturasa) a cloroplastos u otros plástidos de plantas y microalgas. Véanse los ejemplos de *Chlorophyta* proporcionados en las publicaciones PCT WO2010/063032, WO2011/150411, WO2012/106560 y WO2013/158938. Opcionalmente, los genes tipo KASI codifican 450, 475 o 500 aminoácidos o más (con o sin el péptido de tránsito), o aproximadamente 555 residuos (con el péptido de tránsito), a diferencia de las secuencias truncadas conocidas.

Tabla 1. Genes tipo KASI: el casete de expresión usado para evaluar los genes en combinación con un transgén FATB se proporciona en la SEQ ID NO: 38 (es decir, sustituyendo la secuencia codificante de KASIV de *CpaI* de la SEQ ID NO: 38 con varias otras secuencias codificantes de la tabla 1), excepto por el hecho de que el KASIV de *Cuphea hookeriana* se evaluó usando el casete de expresión de la SEQ ID NO: 61. Véanse los ejemplos 1-4.

Especie	Nombre del gen	Secuencia de aminoácidos	secuencia codificante de nucleótido (de ADNc producido a partir de ARNm de semilla, optimización de codones)	Secuencia de nucleótidos con optimización de codones de <i>Prototheca moriformis</i>
<i>Cuphea palustris</i>	KASIV	2	21	39
<i>Cinnamomum camphora</i>	KASIV	3	22	40
<i>Cinnamomum camphora</i>	KASI	4	23	41
<i>Umbellularia californica</i>	KASI	5	24	42
<i>U. Californica</i>	KASIV	6	25	43
<i>Cuphea. Wrightii</i>	KASAI	7	26	44
<i>Cuphea avigera</i>	KASIVb	8	27	45
<i>Cuphea paucipetala</i>	KASIVb	9	28	46
<i>C. ignea</i>	KASIVb	10	29	47
<i>Cuphea procumbens</i>	KASIV	11	30	48
<i>C. paucipetala</i>	KASIVa	12	31	49
<i>Cuphea painteri</i>	KASIV	13	32	50
<i>C. avigera</i>	KASIVa	14	33	51
<i>C. ignea</i>	KASIVa	15	34	52
<i>C. avigera</i>	KASIIa	16	35	53
<i>C. pulcherrima</i>	KASI	17	36	54
<i>C. avigera</i>	KAS mitochondrial	18	37	55
<i>Cuphea hookeriana</i>	KASIV	59		60, 61

Tabla 1a. Proteínas adicionales codificadas por ADNc descubiertas mediante análisis por perfil de transcripción de las semillas. Las secuencias codificantes se pueden derivar de tablas de codones para varias células hospedadoras.

Especie	Nombre del gen	Secuencia de aminoácidos
Varios (Clado 1)	Secuencia de consenso KASIV	69, 71
Varios (Clado 2)	Secuencia de consenso KASIV	70, 72
<i>Cuphea aequipetala</i>	KASIV	62
<i>Cuphea glassostoma</i>	KASIV	63
<i>Cuphea hookeriana</i>	KASIV	64
<i>Cuphea glassostoma</i>	KASIV	65

Especie	Nombre del gen	Secuencia de aminoácidos
<i>Cuphea carthagenesis</i>	<i>KASIV</i>	66, 67
<i>C. pulcherrima</i>	<i>KASIV</i>	68

Tabla 2. Uso de codones en cepas de *Prototheca*.

Ala	GCG	345 (0,36)	Asn	AAT	8 (0,04)
	GCA	66 (0,07)		AAC	201 (0,96)
	GCT	101 (0,11)			
	GCC	442 (0,46)	Pro	CCG	161 (0,29)
Cys				CCA	49 (0,09)
	TGT	12 (0,10)		CCT	71 (0,13)
	TGC	105 (0,90)		CCC	267 (0,49)
Asp	GAT	43 (0,12)	Gln	CAG	226 (0,82)
	GAC	316 (0,88)		CAA	48 (0,18)
Glu	GAG	377 (0,96)	Arg	AGG	33 (0,06)
	GAA	14 (0,04)		AGA	14 (0,02)
				CGG	102 (0,18)
				CGA	49 (0,08)
Phe	TTT	89 (0,29)		CGT	51 (0,09)
	TTC	216 (0,71)		CGC	331 (0,57)
Gly	GGG	92 (0,12)	Ser	AGT	16 (0,03)
	GGA	56 (0,07)		AGC	123 (0,22)
	GGT	76 (0,10)		TCG	152 (0,28)
	GGC	559 (0,71)		TCA	31 (0,06)
His	CAT	42 (0,21)		TCT	55 (0,10)
	CAC	154 (0,79)		TCC	173 (0,31)
Ile	ATA	4(0,01)	Thr	ACG	184 (0,38)
	ATT	30 (0,08)		ACA	24 (0,05)
	ATC	338 (0,91)		ACT	21 (0,05)
				ACC	249 (0,52)
Lys	AAG	284 (0,98)	Val	GTG	308 (0,50)
	AAA	7 (0,02)		GTA	9 (0,01)
				GTT	35 (0,06)
Leu	TTG	26 (0,04)		GTC	262 (0,43)
	TTA	3 (0,00)			
	CTG	447 (0,61)	Trp	TGG	107 (1,00)
	CTA	20 (0,03)			
	CTT	45 (0,06)			
	CTC	190 (0,26)	Tyr	TAT	10 (0,05)
Met				TAC	180 (0,95)
	ATG	191 (1,00)			

Parada TGA/TAG/TAA

Tabla 3. Uso preferido de codones en *Chlorella protothecoides*.

TTC (Phe)	TAC (Tyr)	TGC (Cys)	TGA (parada)
TGG (Trp)	CCC (pro)	CAC (His)	CGC (Arg)
CTG (Leu)	CAG (Gln)	ATC (Ile)	ACC (Thr)
GAC (Asp)	TCC (Ser)	ATG (Met)	AAG (Lys)
GCC (Ala)	AAC (Asn)	GGC (Gly)	GTG (Val)
GAG (Glu)			

Tabla 4: uso de codones para *Cuphea wrightii* (codón, aminoácido, frecuencia, por mil, número)

UUU F 0,48 19,5 (52)	UCU S 0,21 19,5 (52)	UAU Y 0,45 6,4 (17)	UGU C 0,41 10,5 (28)
UUC F 0,52 21,3 (57)	UCC S 0,26 23,6 (63)	UAC Y 0,55 7,9 (21)	UGC C 0,59 15,0 (40)
UUA L 0,07 5,2 (14)	UCA S 0,18 16,8 (45)	UAA * 0,33 0,7 (2)	UGA * 0,33 0,7 (2)
UUG L 0,19 14,6(39)	UCG S 0,11 9,7 (26)	UAG * 0,33 0,7 (2)	UGG W 1,00 15,4 (41)
CUU L 0,27 21,0 (56)	CCU P 0,48 21,7 (58)	CAU H 0,60 11,2 (30)	CGU R 0,09 5,6 (15)
CUC L 0,22 17,2 (46)	CCC P 0,16 7,1 (19)	CAC H 0,40 7,5 (20)	CGC R 0,13 7,9 (21)
CUA L 0,13 10,1 (27)	CCA P 0,21 9,7(26)	CAA Q 0,31 8,6 (23)	CGA R 0,11 6,7 (18)
CUG L 0,12 9,7 (26)	CCG P 0,16 7,1 (19)	CAG Q 0,69 19,5 (52)	CGG R 0,16 9,4 (25)
AUU 10,44 22,8 (61)	ACU T 0,33 16,8 (45)	AAU N 0,66 31,4 (84)	AGU S 0,18 16,1 (43)
AUC yo 0,29 15,4 (41)	ACC T 0,27 13,9 (37)	AAC N 0,34 16,5 (44)	AGC S 0,07 6,0 (16)
AUA yo 0,27 13,9 (37)	ACA T 0,26 13,5 (36)	K AAA 0,42 21,0 (56)	AGA R 0,24 14,2 (38)
AUG M 1,00 28,1 (75)	ACG T 0,14 7,1 (19)	AAG K 0,58 29,2 (78)	AGG R 0,27 16,1 (43)
GUU V 0,28 19,8 (53)	GCU A 0,35 31,4 (84)	GAU D 0,63 35,9 (96)	GGU G 0,29 26,6 (71)
GUC V 0,21 15,0 (40)	GCC A 0,20 18,0 (48)	GAC D 0,37 21,0 (56)	GGC G 0,20 18,0 (48)
GUA V 0,14 10,1 (27)	GCA A 0,33 29,6 (79)	GAA E 0,41 18,3 (49)	GGA G 0,35 31,4 (84)
GUG V 0,36 25,1 (67)	GCG A 0,11 9,7 (26)	GAG E 0,59 26,2 (70)	GGG G 0,16 14,2 (38)

Tabla 5: uso de codones para *Arabidopsis* (codón, aminoácido, frecuencia, por mil)

UUU F 0,51 21,8	UCU S 0,28 25,2	UAU Y 0,52 14,6	UGU C 0,60 10,5
UUC F 0,49 20,7	UCC S 0,13 11,2	UAC Y 0,48 13,7	UGC C 0,40 7,2
UUA L 0,14 12,7	UCA S 0,20 18,3	UAA * 0,36 0,9	UGA * 0,44 1,2
UUG L 0,22 20,9	UCG S 0,10 9,3	UAG * 0,20 0,5	UGG W 1,00 12,5
CUU L 0,26 24,1	CCU P 0,38 18,7	CAU H 0,61 13,8	CGU R 0,17 9,0
CUC L 0,17 16,1	CCC P 0,11 5,3	CAC H 0,39 8,7	CGC R 0,07 3,8
CUA L 0,11 9,9	CCA P 0,33 16,1	CAA Q 0,56 19,4	CGA R 0,12 6,3
CUG L 0,11 9,8	CCG P 0,18 8,6	CAG Q 0,44 15,2	CGG R 0,09 4,9
AUU I 0,41 21,5	ACU T 0,34 17,5	AAU N 0,52 22,3	AGU S 0,16 14,0
AUC I 0,35 18,5	ACC T 0,20 10,3	AAC N 0,48 20,9	AGC S 0,13 11,3
AUA I 0,24 12,6	ACA T 0,31 15,7	K AAA 0,49 30,8	AGA R 0,3519,0
AUG M 1,00 24,5	ACG T 0,15 7,7	AAG K 0,51 32,7	AGG R 0,20 11,0
GUU V 0,40 27,2	GCU A 0,43 28,3	GAU D 0,68 36,6	GGU G 0,34 22,2
GUC V 0,19 12,8	GCC A 0,16 10,3	GAC D 0,32 17,2	GGC G 0,14 9,2
GUA V 0,15 9,9	GCA A 0,27 17,5	GAA E 0,52 34,3	GGA G 0,37 24,2
GUG V 0,26 17,4	GCG A 0,14 9,0	GAG E 0,48 32,2	GGG G 0,16 10,2

5

Combinaciones de genes

[0037] En una forma de realización, un gen/producto génico tal y como se describe en las reivindicaciones se coexpresa en una célula hospedadora con un gen de acil-ACP tioesterasa FATA o FATB exógeno. En una forma de realización específica, el producto génico FATB tiene al menos un 85, 90, 91, 92, 93, 94, 95, 95,5, 96, 96,5 97, 97,5, 98, 98,5 o 99 % de identidad de secuencia de aminoácidos con el FATB2 de *Cuphea palustris* ("Cpal FATB2", acceso AAC49180, SEQ ID NO: 1) o el FATB2 de *C. hookeriana* ("Ch FATB2", acceso U39834, SEQ ID NO: 57) o fragmento de los mismos. Opcionalmente, el producto génico FATB tiene al menos un 85, 90, 91, 92, 93, 94, 95, 95,5, 96, 96,5, 97, 97,5, 98, 98,5 o 99 % de identidad de secuencia de aminoácidos con el dominio de péptido de no tránsito de FATB2 de *Cuphea palustris* ("Cpal FATB2", acceso AAC49180, SEQ ID NO: 1) o FATB2 de *C. Hookeriana* ("Ch FATB2", acceso U39834 SEQ ID NO: 57).

10

15

[0038] Los genes FATA codifican enzimas que preferentemente, pero no exclusivamente, hidrolizan ácidos grasos de cadena larga con una actividad máxima hacia C18:1. Los genes FATB codifican un grupo de enzimas con especificidades de sustrato más heterogéneas pero muestran generalmente actividad más alta hacia los ácidos grasos saturados. Las especificidades de sustrato de las enzimas de FATB son bastante heterogéneas; hay un número de enzimas de FATB que muestran una alta actividad hacia C18:0 y C18:1. Las enzimas FATA y FATB terminan la síntesis de ácidos grasos hidrolizando el enlace de tioéster entre la fracción de acilo y la proteína portadora de acilo (ACP).

[0039] En una forma de realización, una célula hospedadora se transforma para expresar tanto un transgén FATA o FATB y tipo KASI. La célula hospedadora produce un aceite celular. Juntos, los genes FATA o FATB y tipo KASI se expresan para producir sus productos génicos respectivos y alteran así el perfil de ácidos grasos del aceite celular. Los dos genes funcionan por adición o sinérgicamente con respecto a las cepas de control que carecen de uno de los dos genes. Opcionalmente, la célula hospedadora es oleaginosa y pueden ser microalgas eucarióticas oleaginosas tales como las descritas previamente o más adelante. El perfil de ácidos grasos del aceite celular puede estar enriquecido (con respecto a un control apropiado) en C14:0 (mirístico), C8:0, C10:0 o una combinación de C8/C10.

[0040] En una forma de realización, el perfil de ácidos grasos de la célula está enriquecido con ácidos grasos C14:0. En esta forma de realización, el gen FATB expresa una enzima de acil-ACP tioesterasa que tiene al menos un 85, 90, 91, 92, 93, 94, 95, 95,5, 96, 96,5, 97, 97,5, 98, 98,5 o 99 % de identidad de secuencia de aminoácidos en porcentaje de identidad de aminoácidos con la enzima de la SEQ ID NO: 1. El gen de tipo KASI coexpresado codifica una beta-cetoacil ACP sintasa que tiene al menos un 95, 95,5, 96, 96,5, 97, 97,5, 98, 98,5 o 99 % de identidad de secuencia de aminoácidos en porcentaje de identidad de aminoácidos con la enzima de la SEQ ID NO: 2. También se describe que el gen de tipo KASI coexpresado codifica una beta-cetoacil ACP sintasa que tiene al menos un 85, 90, 91, 92, 93, 94, 95, 95,5, 96, 96,5, 97, 97,5, 98, 98,5 o 99 % de identidad de secuencia de aminoácidos en porcentaje de identidad de aminoácidos con la enzima de la SEQ ID NO: 7. Opcionalmente, el aceite celular tiene un perfil de ácidos grasos caracterizado por al menos un 10 %, 20 %, 30 %, 40 %, 50 % o al menos el 55 % C14:0 (% en área por FAME-GC-FID).

[0041] En otra forma de realización, el perfil de ácidos grasos de la célula está enriquecido con ácidos grasos C8:0 y/o C10:0. En esta forma de realización, el gen FATB expresa una enzima de acil-ACP tioesterasa que tiene al menos un 85, 90, 91, 92, 93, 94, 95, 95,5, 96, 96,5, 97, 97,5, 98, 98,5 o 99 % de identidad de secuencia de aminoácidos en porcentaje de identidad de aminoácidos con la enzima de la SEQ ID NO: 57. El gen de tipo KASI coexpresado codifica una beta-cetoacil ACP sintasa que tiene al menos un 95, 95,5, 96, 96,5, 97, 97,5, 98, 98,5 o 99 % de identidad de secuencia de aminoácidos en porcentaje de identidad de aminoácidos con una enzima de la SEQ ID NOS: 2. Opcionalmente, el aceite celular tiene un perfil de ácidos grasos caracterizado por al menos un 7, 8, 9, 10, 11, 12, 13 o 14 % en área C8:0 (por FAME-GC-FID). Opcionalmente, el aceite celular tiene un perfil de ácidos grasos caracterizado por al menos un 10, 15, 20, 25, 30 o 35 % en área para la suma de ácidos grasos C8:0 y C10:0 (por FAME-GC-FID). Opcionalmente, la proporción C8/C10 del aceite celular está en el rango de 2,2-2,5, 2,5-3,0 o 3,0-3,4.

[0042] Opcionalmente, los aceites producidos por estos métodos pueden tener un perfil de esteroles de acuerdo con aquellos descritos más adelante.

Células hospedadoras

[0043] La célula hospedadora puede ser una única célula (por ejemplo microalga, bacterias o levadura) o parte de un organismo multicelular tal como una planta o un hongo. Se proporcionan métodos para expresar genes tipo KASI en una planta en las patentes US7301070, US6348642, US6660849 y US6770465 o se pueden llevar a cabo usando otras técnicas conocidas generalmente en biotecnología vegetal. Se describe ingeniería de microbios oleaginosos eucarióticos que incluyen microalgas eucarióticas (por ejemplo de *Chlorophyta*) en las patentes WO2010/063032, WO2011/150411 y WO2012/106560 y en los ejemplos más adelante.

[0044] Las células hospedadoras oleaginosas plastídicas que tienen una vía biosintética de ácidos grasos de tipo II incluyen células oleaginosas plastídicas tales como las de algas oleaginosas. Ejemplos específicos de células de microalga incluyen microalgas eucarióticas heterótrofas o heterótrofas obligadas del filo *Chlorophyta*, la clase *Trebouxiophytae*, el orden *Chlorellales* o la familia *Chlorellaceae*. Se proporcionan ejemplos de células hospedadoras de microalgas oleaginosas eucarióticas en las solicitudes de patente de PCT publicadas WO2008/151149, WO2010/06032, WO2011/150410 y WO2011/150411, que incluyen especies de *Chlorella* y *Prototheca*, un género que comprende heterótrofos obligados. Las células oleaginosas pueden ser, por ejemplo, capaces de producir un 25, 30, 40, 50, 60, 70, 80, 85 o aproximadamente un 90 % de aceite por peso celular, ± 5 %. Opcionalmente, los aceites producidos pueden ser bajos en ácidos grasos DHA o EPA. Por ejemplo, los aceites pueden comprender menos del 5 %, 2 % o 1 % de DHA y/o EPA. Las publicaciones anteriormente mencionadas revelan también métodos para cultivar tales células y extraer aceite, especialmente a partir de células de microalga; tales métodos son aplicables a las células descritas aquí. Cuando se usan células de microalga, se pueden cultivar de forma autótrofa (salvo un heterótrofo obligado) o a oscuras usando un azúcar (por ejemplo

glucosa, fructosa y/o sacarosa). Cuando se cultivan heterótrofamente, las células y el aceite celular comprenden menos de 200 ppm, 20 ppm o 2 ppm de impurezas generadoras de color o de clorofila. En cualquiera de las formas de realización descritas aquí, las células pueden ser células heterótrofas que incluyen un gen de invertasa exógeno para permitir que las células produzcan aceite a partir de una materia prima de sacarosa. De forma alternativa, o adicional, las células pueden metabolizar xilosa a partir de materias primas celulósicas. Por ejemplo, las células pueden ser modificadas por ingeniería genética para que expresen uno o más genes de metabolismo de xilosa tales como los que codifican un transportador de xilosa activa, un transportador de xilulosa-5-fosfato, una xilosa isomerasa, una xiluloquinasa, una xilitol dehidrogenasa y una reductasa de xilosa - véase la patente WO2012/154626. Las células se pueden cultivar en una materia prima celulósica despolimerizada tal como bagazo hidrolizado con ácido o enzima, pulpa de remolacha azucarera, hojas y troncos de maíz, astillas de madera, serrín o pasto varilla. Opcionalmente, las células se pueden cultivar en una materia prima celulósica despolimerizada que comprende glucosa y al menos un 5, 10, 20, 30 o 40 % de xilosa, mientras que produce al menos un 20 % de lípido por peso en seco. Opcionalmente, el lípido comprende triglicéridos que tienen un perfil de ácidos grasos caracterizado por al menos un 10, 15 o 20 % de C12:0.

[0045] Opcionalmente, la célula hospedadora comprende ARNr 23S con al menos un 65, 70, 75, 80, 85, 90 o 95 % de identidad de secuencia de nucleótidos con la SEQ ID NO: 58.

Aceites y productos relacionados

[0046] Las células oleaginosas expresan uno o más genes exógenos que codifican enzimas de biosíntesis de ácidos grasos. Como resultado de ello, algunas formas de realización incluyen aceites celulares que no se pueden obtener a partir de un aceite que no es vegetal o no es de semilla, o no se pueden obtener en absoluto.

[0047] Las células oleaginosas producen un aceite de almacenamiento, que es principalmente un triglicérido y se puede almacenar en cuerpos de almacenamiento de la célula. Un aceite crudo se puede obtener a partir de las células rompiendo las células y aislando el aceite. Las patentes WO2008/151149, WO2010/06032, WO2011/150410 y WO2011/1504 revelan un cultivo heterótrofo y técnicas de aislamiento del aceite. Por ejemplo, se puede obtener aceite cultivando, secando y prensando las células. Los aceites celulares producidos se pueden refinar, blanquear y desodorizar (RBD) tal y como se conoce en la técnica del aceite de semilla o como se describe en la patente WO2010/120939. La etapa de refinación puede comprender el desgomado. Los aceites crudos, refinados o RBD se pueden usar en una variedad de alimentos, sustancias químicas y productos o procesos industriales. Después de recuperar el aceite, permanece una biomasa residual valiosa. Usos para la biomasa residual incluyen la producción de papel, plásticos, absorbentes, adsorbentes, como pienso para animales, para nutrición humana o para fertilizante.

[0048] Donde se proporcione aquí un perfil de ácidos grasos de un aceite celular de triglicérido (también denominado un "triglicérido" o "TAG"), se entenderá que esto hace referencia a una muestra no fraccionada del aceite de almacenamiento extraído de la célula analizada bajo condiciones en las que se han eliminado fosfolípidos o con un método de análisis que es sustancialmente insensible a los ácidos grasos de los fosfolípidos (por ejemplo usando cromatografía y espectrometría de masa). El aceite se puede someter a un proceso RBD para eliminar fosfolípidos, ácidos grasos libres y olores y aún así tener solo cambios menores o insignificantes en el perfil de ácidos grasos de los triglicéridos del aceite. Dado que las células son oleaginosas, en algunos casos el aceite de almacenamiento constituirá el volumen de todos los TAG de la célula.

[0049] El valor de isótopo de carbono estable $\delta^{13}\text{C}$ es una expresión de la proporción de $^{13}\text{C}/^{12}\text{C}$ con respecto a un estándar (por ejemplo PDB, carbonita de esqueleto fósil de *Belemnite americana* de la Formación Peedee de Carolina del Sur). El valor de isótopo de carbono estable $\delta^{13}\text{C}$ (0/00) de los aceites se puede relacionar con el valor $\delta^{13}\text{C}$ de la materia prima usada. En algunas formas de realización, los aceites son derivados de organismos oleaginosos cultivados heterótrofamente en azúcar derivados a partir de una planta C4 tal como maíz o caña de azúcar. En algunas formas de realización, el $\delta^{13}\text{C}$ (0/00) del aceite es de -10 a -17 0/00 o de -13 a -16 0/00.

[0050] Los aceites producidos según los métodos anteriores en algunos casos se han elaborado usando una célula hospedadora de microalga. Tal y como se ha descrito anteriormente, la microalga puede ser, sin limitación, una microalga eucariótica que recae en la clasificación de la *Chlorophyta*, *Trebouxiophyceae*, *Chlorellales*, *Chlorellaceae* o *Chlorophyceae*. Se ha descubierto que las microalgas de *Trebouxiophyceae* se pueden distinguir de aceites vegetales en función de sus perfiles de esterol. Se descubrió que el aceite producido por la *Chlorella protothecoides* (un pariente cercano de la *Prototheca moriformis*) producía esteroides que parecían ser brasicasterol, ergosterol, campesterol, estigmasterol y beta-sitosterol, cuando se detectan mediante GC-MS. Sin embargo, se cree que todos los esteroides producidos por la *Chlorella* tienen estereoquímica C24 β . De este modo, se cree que las moléculas detectadas como campesterol, estigmasterol y beta-sitosterol son en realidad 22,23-dihidrobrasicasterol, proferasterol y clionasterol, respectivamente. De este modo, los aceites producidos por las microalgas anteriormente descritas se pueden distinguir de los aceites vegetales por la presencia de esteroides con estereoquímica C24 α y la ausencia de estereoquímica C24 α en los esteroides presentes. Por ejemplo, los aceites producidos pueden contener 22,23-dihidrobrasicasterol mientras que carecen de campesterol; contener clionasterol, mientras que carecen de beta-sitosterol, y/o contener poriferasterol mientras que carecen de

estigmasterol. Alternativa o adicionalmente, los aceites pueden contener cantidades significativas de Δ^7 -poriferasterol.

[0051] En una forma de realización, los aceites proporcionados aquí no son aceites vegetales. Los aceites vegetales son aceites extraídos de plantas y semillas de plantas. Los aceites vegetales se pueden distinguir de los aceites no vegetales proporcionados provistos aquí basándose en su contenido de aceite. Una variedad de métodos para analizar el contenido de aceite se puede emplear para determinar la fuente del aceite o si se ha producido una adulteración de un aceite proporcionado aquí con un aceite de origen diferente (por ejemplo vegetal). La determinación se puede hacer basándose en uno o una combinación de los métodos analíticos. Estas pruebas incluyen, pero no se limitan a, el análisis de uno o más ácidos grasos libres, perfil de ácidos grasos, contenido de triacilglicerol total, contenido de diacilglicerol, valores de peróxido, propiedades espectroscópicas (por ejemplo absorción de UV), perfil de esterol, productos de degradación de esterol, antioxidantes (por ejemplo tocoferoles), pigmentos (por ejemplo clorofila), valores d13C y análisis sensorial (por ejemplo sabor, olor y textura en la boca). Muchas de estas pruebas se han estandarizado para aceites comerciales tales como las normas del código Codex Alimentarius para grasas y aceites comestibles.

[0052] El análisis del perfil del esterol es un método particularmente conocido para determinar la fuente biológica de materia orgánica. Campesterol, β -sitosterol y stigmasterol son esteroides vegetales comunes, donde el β -sitosterol es un esterol vegetal de principio. Por ejemplo, se descubrió que el β -sitosterol estaba en mayor cantidad en un análisis de determinados aceites de semilla, aproximadamente un 64 % en el maíz, 29 % en la colza, 64 % en el girasol, 74 % en la semilla de algodón, 26 % en la soja y 79 % en el aceite de oliva (Gul *et al.* J. Cell and Molecular Biology 5:71-79, 2006).

[0053] El aceite aislado de cepa UTEX1435 de *Prototheca moriformis* se clarificó (CL), refinó y blanqueó (RB), o refinó, blanqueó y desodorizó (RBD) por separado y se evaluó para el contenido de esterol según el procedimiento descrito en JAOCs vol. 60, n.º 8, agosto de 1983. Los resultados del análisis se muestran a continuación (unidades en mg/100g) en la tabla 6.

Tabla 6. Esteroides en aceite de microalga.

	Esterol	Crudo	Clarificado	Refinado y blanqueado	Refinado, blanqueado y desodorizado
1	Ergosterol	384 (56 %)	398 (55 %)	293 (50 %)	302 (50 %)
2	5,22-colestadien-24-metil-3-ol (brasicasterol)	14,6 (2,1 %)	18,8 (2,6 %)	14 (2,4 %)	15,2 (2,5 %)
3	24-metilcolest-5-en-3-ol (campesterol o 22,23-dihidrobrasicasterol)	10,7 (1,6 %)	11,9 (1,6 %)	10,9 (1,8 %)	10,8 (1,8 %)
4	5,22-colestadien-24-etil-3-ol (stigmasterol o poriferasterol)	57,7 (8,4 %)	59,2 (8,2 %)	46,8 (7,9 %)	49,9 (8,3 %)
5	24-etilcolest-5-en-3-ol (β -sitosterol o clionasterol)	9,64 (1,4 %)	9,92 (1,4 %)	9,26 (1,6 %)	10,2 (1,7 %)
6	Otros esteroides	209	221	216	213
	Esteroides totales	685,64	718,82	589,96	601,1

[0054] Estos resultados muestran tres características llamativas. En primer lugar, se descubrió que el ergosterol era el más abundante de todos los esteroides, contabilizando aproximadamente un 50 % o más de los esteroides totales. La cantidad de ergosterol es mayor que la de campesterol, beta-sitosterol y stigmasterol combinados. El ergosterol es un esteroide comúnmente hallado en hongos y que no se encuentra habitualmente en plantas y su presencia, particularmente en cantidades significativas, sirve como un marcador útil para los aceites no vegetales. En segundo lugar, se descubrió que el aceite contenía brasicasterol. A excepción del aceite de colza, el brasicasterol no se encuentra comúnmente en aceites vegetales. En tercer lugar, se descubrió que estaba presente menos del 2 % de beta-sitosterol. El beta-sitosterol es un prominente esterol vegetal que no se encuentra comúnmente en las microalgas y su presencia particularmente en cantidades significativas sirve como un marcador útil para los aceites de origen vegetal. En resumen, se ha descubierto que la cepa UTEX1435 de *Prototheca moriformis* contiene tanto cantidades significativas de ergosterol como solo pequeñas cantidades de beta-sitosterol como un porcentaje de contenido de esterol total. Por consiguiente, la relación de ergosterol: beta-sitosterol o en combinación con la presencia de brasicasterol se puede usar para distinguir este aceite de los aceites vegetales.

[0055] En algunas partes de la divulgación, el contenido de aceite de un aceite proporcionado aquí contiene, como un porcentaje de esteroides totales, menos del 20 %, 15 %, 10 %, 5 %, 4 %, 3 %, 2 % o 1 % de beta-sitosterol. En otras formas de realización el aceite carece de beta-sitosterol.

5 [0056] En algunas partes de la divulgación, el aceite carece de uno o más de beta-sitosterol, campesterol o estigmasterol. En algunas partes de la divulgación, el aceite carece de beta-sitosterol, campesterol y estigmasterol. En algunas partes de la divulgación, el aceite carece de campesterol. En algunas partes de la divulgación, el aceite carece de estigmasterol.

10 [0057] En algunas partes de la divulgación, el contenido de aceite de un aceite proporcionado aquí comprende, como un porcentaje de esteroides totales, menos del 20 %, 15 %, 10 %, 5 %, 4 %, 3 %, 2 % o 1 % de 24-etilcolest-5-en-3-ol. En algunas partes de la divulgación, el 24-etilcolest-5-en-3-ol es clionasterol. En algunas partes de la divulgación, el contenido de aceite de un aceite proporcionado aquí comprende, como un porcentaje de esteroides totales, al menos el 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 % o 10 % de clionasterol.

15 [0058] En algunas partes de la divulgación, el contenido de aceite de un aceite proporcionado aquí contiene, como un porcentaje de esteroides totales, menos del 20 %, 15 %, 10 %, 5 %, 4 %, 3 %, 2 % o 1 % de 24-metilcolest-5-en-3-ol. En algunas partes de la divulgación, el 24-metilcolest-5-en-3-ol es 22,23-dihidrobrasicasterol. En algunas partes de la divulgación, el contenido de aceite de un aceite proporcionado aquí comprende, como un porcentaje de esteroides totales, al menos el 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 % o 10 % de 22,23-dihidrobrasicasterol.

20 [0059] En algunas partes de la divulgación, el contenido de aceite de un aceite proporcionado aquí contiene, como un porcentaje de esteroides totales, menos del 20 %, 15 %, 10 %, 5 %, 4 %, 3 %, 2 % o 1 % de 5,22-colestadien-24-etil-3-ol. En algunas partes de la divulgación, el 5,22-colestadien-24-etil-3-ol es poriferasterol. En algunas partes de la divulgación, el contenido de aceite de un aceite proporcionado aquí comprende, como un porcentaje de esteroides totales, al menos el 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 % o 10 % de poriferasterol.

25 [0060] En algunas partes de la divulgación, el contenido de aceite de un aceite proporcionado aquí contiene ergosterol o brasicasterol o una combinación de los dos. En algunas partes de la divulgación, el contenido de aceite contiene, como un porcentaje de esteroides totales, al menos el 5 %, 10 %, 20 %, 25 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 % o 65 % de ergosterol. En algunas partes de la divulgación, el contenido de aceite contiene, como un porcentaje de esteroides totales, al menos el 25 % de ergosterol. En algunas partes de la divulgación, el contenido de aceite contiene, como un porcentaje de esteroides totales, al menos el 40 % de ergosterol. En algunas partes de la divulgación, el contenido de aceite contiene, como un porcentaje de esteroides totales, al menos el 5 %, 10 %, 20 %, 25 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 % o 65 % de una combinación de ergosterol y brasicasterol.

30 [0061] En algunas partes de la divulgación, el contenido de aceite contiene, como un porcentaje de esteroides totales, al menos el 1 %, 2 %, 3 %, 4 % o 5 % de brasicasterol. En algunas partes de la divulgación, el contenido de aceite contiene, como un porcentaje de esteroides totales menos del 10 %, 9 %, 8 %, 7 %, 6 % o 5 % de brasicasterol.

[0062] En algunas partes de la divulgación la proporción de ergosterol a brasicasterol es de al menos 5:1, 10:1, 15:1 o 20:1.

45 [0063] En algunas partes de la divulgación, el contenido de aceite contiene, como un porcentaje de esteroides totales, al menos el 5 %, 10 %, 20 %, 25 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 % o 65 % de ergosterol y menos del 20 %, 15 %, 10 %, 5 %, 4 %, 3 %, 2 % o 1 % de beta-sitosterol. En algunas partes de la divulgación, el contenido de aceite contiene, como un porcentaje de esteroides totales, al menos el 25 % de ergosterol y menos del 5 % de beta-sitosterol. En algunas partes de la divulgación, el contenido de aceite comprende además brasicasterol. Para cualquiera de los aceites o aceites de célula descritos en esta aplicación, el aceite puede tener el perfil de esteroide de cualquier columna de la tabla 6 anterior con una variación esteroide por esteroide del 30 %, 20 %, 10 % o menos.

50 [0064] Los esteroides contienen de 27 a 29 átomos de carbono (C27 a C29) y se encuentran en todas las eucariotas. Los animales hacen exclusivamente esteroides C27 ya que carecen de la capacidad de modificar adicionalmente los esteroides C27 para producir esteroides C28 y C29. Las plantas, sin embargo, son capaces de sintetizar esteroides C28 y C29, y los esteroides vegetales C28/C29 se denominan a menudo fitoesteroides. El perfil de esteroide de una planta determinada es alto en esteroides C29 y los esteroides primarios en las plantas son generalmente los esteroides C29 beta-Sitosterol y estigmasterol. En cambio, el perfil de esteroide de los organismos no vegetales contiene porcentajes superiores de esteroides C27 y C28. Por ejemplo, los esteroides en los hongos y en muchas microalgas son principalmente esteroides C28. El perfil de esteroide y particularmente el notable predominio de los esteroides C29 sobre los esteroides C28 en las plantas ha sido aprovechado para determinar la proporción de material vegetal y marino en muestras de suelo (Huang, Wen-Yen, Meinschein W. G., "Sterols as ecological indicators"; *Geochimica et Cosmochimica Acta*. Vol 43. Pp. 739-745).

65 [0065] En algunas partes de la divulgación, los esteroides primarios en los aceites de microalga proporcionados aquí son esteroides diferentes del beta-sitosterol y estigmasterol. En algunas partes de la divulgación de los aceites

de microalga, los esteroides C29 componen menos del 50 %, 40 %, 30 %, 20 %, 10 % o 5 % en peso del contenido de esteroide total.

[0066] En algunas partes de la divulgación, los aceites de microalga proporcionados aquí contienen esteroides C28 que exceden los esteroides C29. En algunas partes de la divulgación de los aceites de microalga, los esteroides C28 constituyen más del 50 %, 60 %, 70 %, 80 %, 90 % o 95 % en peso del contenido de esteroide total. En algunas partes de la divulgación, el esteroide C28 es ergosterol. En algunas partes de la divulgación el esteroide C28 es brasicasterol.

[0067] En ejemplos de la presente descripción, las células oleaginosas que expresan uno o más de los genes de la tabla 1 pueden producir un aceite con al menos el 20, 40, 60 o 70 % de ácidos grasos C8, C10, C12, C14 o C16. En una forma de realización específica, el nivel de miristato (C14:0) en el aceite es superior al 30 %.

[0068] De este modo, en ejemplos de la divulgación, hay un proceso para producir un aceite, triglicérido, ácidos grasos o derivado de cualquiera de estos, que comprende transformar una célula con cualquiera de los ácidos nucleicos tratados aquí. En otra forma de realización, la célula transformada se cultiva para producir un aceite y, opcionalmente, se extrae el aceite. El aceite extraído de esta manera se puede usar para producir alimentos, oleoquímicos u otros productos.

[0069] Los aceites mencionados anteriormente solos o en combinación son útiles en la producción de alimentos, combustibles y productos químicos (incluyendo plásticos, espumas, películas, detergentes, jabones, etc.). Los aceites, triglicéridos y ácidos grasos de los aceites se pueden someter a activación C-H, hidroaminometilación, metoxycarbonatación, ozonólisis, transformaciones enzimáticas, epoxidación, metilación, dimerización, tiolación, metátesis, hidroalquilación, lactonización u otros procesos químicos.

[0070] Después de la extracción del aceite, puede quedar una biomasa residual, que puede tener uso como un combustible, como un pienso para animales o como un ingrediente en papel, plástico u otro producto. Por ejemplo, la biomasa residual de algas heterótrofas se puede usar en dichos productos.

Ejemplos

Ejemplo 1: cribado de genes KAS en combinación con acil-ACP tioesterasa de FATB2 de *Cuphea palustris*.

[0071] Se elaboró una cepa de *Prototheca moriformis* que expresa FATB2 de *Cuphea palustris* (Cpal) con optimización de codones tal y como se describe en la patente WO2013/158938, ejemplo 53 (p. 231). La secuencia de aminoácidos del gen FATB2 de Cpal se proporciona en la SEQ ID NO: 1. Esta cepa (S6336) produjo un aceite celular caracterizado por un perfil de ácidos grasos que tiene aproximadamente un 38 % de ácido mirístico (C14:0).

[0072] Se clonaron seis genes tipo KASI a partir de genomas de aceite de semilla. El ARN total se extrajo de semillas maduras secas usando un mortero y mano de mortero de nitrógeno líquido refrigerado para romper las paredes de semilla. A continuación, se precipitó el ARN con una solución de 8M urea, 3M LiCl seguido de una extracción de fenol-cloroformo. Se generó un banco de ADNc con iniciadores oligo dT usando el ARN purificado y sometido a secuenciación de nueva generación. Los genes KAS nuevos se identificaron a partir del transcriptoma ensamblado usando BLAST con genes KAS conocidos como cebo. Las secuencias de gen KAS identificadas se optimizaron en codones para la expresión en *Prototheca* y se sintetizaron para la incorporación en un casete de expresión.

[0073] Para evaluar el impacto en la acumulación de miristato, se transformó S6336 con un plásmido linealizado diseñado para recombinación homóloga en el locus de bucle p y para expresar los genes tipo KASI con coexpresión de un marcador de selección (véase la patente WO2013/1589380). El vector se describe en la SEQ ID NO 38, los genes KAS con optimización de codones restantes se sustituyeron en el segmento de la CDS de KAS de este vector antes de la transformación. Tal y como se muestra en la tabla 7, se observaron aumentos en los niveles de C14:0 en el aceite celular extraído con la expresión de los genes KASIV de *C. camphora* (D3147), KASI de *C. camphora* (D3148), KASI de *U. californica* (D3150) o KASVI de *U. californica* (D3152) en S6336. Se produjeron aumentos todavía mayores en los niveles de C14:0 como resultado de la expresión del gen KASI de KASIV de *C. palustris* (D3145) o KASAI de *C. wrightii* (D3153), con algunas líneas individuales que producen >50 % o >55 % de C14:0. La producción de C14 superaba con mucho la pequeña cantidad hallada en el aceite de tipo salvaje (véase la tabla 7a).

Tabla 7. Genes KAS que efectúan un aumento en los ácidos grasos C14 en el aceite de microalga eucariota.

Gen (ID transformante)	SEQ ID NOs:	C14:0 (% en área. Media de 4 transformantes)	C14:0 máximo observado
KASIV de <i>C. camphora</i>	3, 22, 40	38,0	40,3
KASI de <i>C. camphora</i>	4, 23, 41	33,8	39,3

Gen transformante (ID)	SEQ ID NOS:	C14:0 (% en área. Media de 4 transformantes)	C14:0 máximo observado
KASI de <i>U. californica</i>	5, 24, 42	37,4	42,3
KASVI de <i>U. californica</i>	6, 25, 43	38,4	41,6
KASIV de <i>C. palustris</i>	2, 21, 39	45,4	58,4
KASAI de <i>C. wrightii</i>	7, 26, 44	43,2	53,6

Tabla 7a. Perfil de ácidos grasos de aceite de *Prototheca moriformis* de tipo salvaje (% en área).

C8:0	C10:0	C12:0	C14:0	C16:0	C18:0	C18:1	C18:2	C18:3
0	0	0	2	38	4	48	5	1

Ejemplo 2: cribado de genes KAS en combinación con acil-ACP tioesterasa de FATB de *Cuphea hookeriana*.

[0074] Se construyeron cepas de *P. moriformis* que expresan acil-ACP tioesterasa de ChFATB2 junto con un gen KAS seleccionado entre diez KASI, un KASIII y un KAS mitocondrial fueron clonados a partir de genomas de aceite de semilla, optimizados en codones e introducidos en la *Prototheca* tal y como se ha descrito en el ejemplo 1. Los genes KAS se fusionaron a un epítipo HA TAG en el extremo C-terminal de cada KAS para permitir la confirmación de expresión de proteína.

Tabla 8. Perfiles de ácidos grasos C8:0-C10:0 medios derivados de la transformación de la cepa de microalga que expresa FATB2 con genes tipo KAS/ aislados de genomas de aceite de semilla.

Gen KAS	SEQ ID NOS: (aminoácido, CDS, optimización de codones)	C8:0 (% en área medio)	C10:0 (% en área medio)	Suma C8:0 + C10:0	Relación C10/C8
KASla de <i>C. avigera</i>	16, 35, 53	8,0	21,4	29,3	2,7
KASI de <i>C. pulcherrima</i>	17, 36, 54	7,7	20,3	28,0	2,6
KAS mitocondrial de <i>C. avigera</i>	NE, 37, 55	7,8	20,4	28,2	2,6
KAS III de <i>C. avigera</i>	19, NE, 56	9,5	22,8	32,3	2,4
KASIVb de <i>C. paucipetala</i>	9, 28, 46	7,9	22,5	30,3	2,9
KASIVb <i>C. ignea</i>	10, 29, 47	6,6	18,7	25,4	2,8
KASIV de <i>C. painteri</i>	13, 32, 50	9,0	22,4	31,4	2,5
KASIVa de <i>C. palustris</i>	2, 21, 38	8,6	21,6	30,4	2,5
KASIVb de <i>C. avigera</i>	8, 27, 45	11,0	23,8	34,8	2,2
KASIV de <i>C. procumbens</i>	11, 30, 48	8,2	25,8	34,0	3,2
KASIVa de <i>C. paucipetala</i>	12, 31, 49	8,8	29,9	39,4	3,4
KASIVa de <i>C. ignea</i>	15, 34, 52	8,6	25,8	34,4	3,0
KASIVa de <i>C. avigera</i>	14, 33, 51	10,0	23,0	32,9	2,3
KASIV de <i>C. hookeriana</i>	59, NE, 61	14,5	27,81	42,6	3,0

[0075] La cepa parental es una cepa de microalga estable que expresa el FATB2 de *C. hookeriana* bajo el control del promotor *PmUAPA1* compatible con pH5. La cepa parental acumula el 27,8 % de C8:0-C10:0 con una relación C10/C8 de 2,6. Todos los transformantes se derivan de integraciones de los transgenes KAS/ en el locus de *bucle p* de la cepa parental. Las medias se calculan desde al menos 19 transformantes individuales para cada transgén KAS (NE = no enumerado).

[0076] Como se puede observar en la tabla 8, la expresión de los siguientes genes KAS aumentó significativamente los niveles de C8:0-C10:0: *KASIVb* de *C. avigera* (D3287), *KASIV* de *C. procumbens* (D3290), *KASIVa* de *C. paucipetala* (D3291), *KASIVa* de *C. avigera* (D3293) y *KASIVa* de *C. ignea* (D3294). De manera importante, la expresión de *KASIVb* de *C. avigera* (D3287) aumentó la acumulación de ácidos grasos tanto de C8:0 como de C10:0, mientras que solo aumentaron los niveles de C10:0 tras la expresión de D3290, D3291, D3293 y D3294. En algunos casos, la suma de ácidos grasos C8:0 y C10:0 en el perfil de ácidos grasos fue de al menos el 30 %, o al menos el 35 % (% en área por FAME-GC-FID). La producción de cadena media excedía por mucho la cantidad insignificante encontrada en el aceite de tipo salvaje (véase la tabla 7a).

[0077] Las relaciones C8/C10 medias de la tabla 8 se encontraban en el rango de 2,2 a 3,4. La suma de C8 y C10 medio se encontraba en el rango de 25,4 a 39,4.

[0078] La cepa de producción C8:0 máxima hallada fue D3287, que combinaba *KASIV* de *C. avigera* con *FATB2* de *C. hookeriana*. La media fue del 11,0 % de C8:0 con un rango de 12,4 a 14,8. De este modo, se produjo un aceite celular con un perfil de ácidos grasos superior al 14 % de C8. Además, la relación C10/C8 fue inferior a 2,5.

Ejemplo 3: identificación de clados y secuencias de consenso de KAS

[0079] Las secuencias recién identificadas de genes tipo KAS se compararon con aquellas de la base de datos ThYme de enzimas activas de tioéster mantenidas por la universidad Iowa State University (enzyme.cbirc.iastate.edu) usando el algoritmo BLAST y se extrajeron los primeros resultados. Se descargaron los primeros 50 resultados de BLAST y se creó un alineamiento múltiple usando el algoritmo de alineamiento ClustalW y se creó un árbol filogenético (Fig. 1) usando dicho alineamiento con el método Jukes-Cantor de unión de vecinos. Los genes *KASIV* nuevos se agruparon junto con solo 4 genes *KAS* de ThYme internos a dicho grupo de los 50 posibles. Las secuencias *KAS* de ThYme totales se redujeron a 12 porque casi todas las *KAS* de ThYme se agruparon fuera de las secuencias *KAS* nuevas. Las secuencias ThYme son solo de 222 residuos mientras los nuevos *KASIV* son de aproximadamente 555 residuos de longitud incluyendo el péptido diana.

[0080] Se identificaron dos clados nuevos Clado 1 y Clado 2, caracterizados por las secuencias de consenso SEQ ID NO: 69 y SEQ ID NO:70, que incluyen péptidos de tránsito. Los clados también se pueden caracterizar por las secuencias de las proteínas de consenso maduras SEQ ID NO: 71 y SEQ ID NO: 72, respectivamente. Los genes *KAS* del Clado 1 se asocian con la producción de ácidos grasos C8 y C10 elevados basados en transformaciones en *P. moriformis* en combinación con un *FATB* de acil-ACP tioesterasa como en el ejemplo 2. Los genes *KAS* del Clado 2 se asocian con la producción de ácidos grasos C10 elevados basados en transformaciones en *P. moriformis* en combinación con un *FATB* de acil-ACP tioesterasa como en el ejemplo 2.

[0081] Aunque la discusión anterior divulga varias formas de realización ejemplares de la invención, debería desprenderse del ella que los expertos en la técnica pueden realizar varias modificaciones que conseguirán algunas de las ventajas de la invención sin apartarse del alcance de las reivindicaciones.

LISTADO DE SECUENCIAS

[0082]

SEQ ID NO: 1

Secuencia de aminoácidos de *FATB2* de *Cuphea palustris* (n.º de acceso Genbank AAC49180.1)

```
MVAAAASAAFFSVATPRTNISPSSLVSPFKPKSNHNGGFQVKANASAHPKANGSAVS
LKSGSLETQEDKTSSSSPPPRFTINQLPVWSMLLSAVTTVFGVAEKQWPMLDRKSKR
PDMLVEPLGVDRIVYDGVSFQRQSFSIRS YEIGADRTAS IETLMNMFQETSLNHCKIIGL
LNDGFGRTPEMCKRDLIWVVT KMQIEVNRYPTWGD TIEVNTWVSASGKHGMGRD
WLISDCHTGEILIRATSVWAMMNQKTRRLSKIPYEV RQEIEPQFVDSAPVIVDDRKFH
KLDLKTGDSICNGLTPRWTDLVDNQHVN NVKYIGWILQSVPTVEVFETQELCGLTLEY
RRECGRDSVLESVTAMDPSKEGDRSLYQHLLRLEDGADIVKGRTEWRPKNAGAKG
AILTGKTSNGNSIS
```

SEQ ID NO: 2

Secuencia de aminoácidos de *KASIV* de *C. palustris* (D3145 y D3295, pSZ4312). El péptido de tránsito de alga está subrayado.

MASAAFTMSACPAMTGRAPGARSRGPVATRLRGSTFQCLVTSYIDPCNQFSSSASL
SFLGDNGFASLFGSKPFRSNRGHRRLLGRASHSGEAMAVALEPAQEVATKKKPLVKQ
RRVVVTGMGVVTPLGHEPDVYYNNLLDGVSGISEIEAFDCTQFPTRIAGEIKSFSTDG
WVAPKLSKRMDKFMPLYLLTAGKKALADGGITDDVMKELDKRKCGVLIGSGLGGM
KLFSDSIEALRISYKKMNPFCVPFATTNMGSAAMLAMD LGWMGPNYSISTACATSNFC
ILNSANHIVRGEADMMLCGGSDAVIPIGLGGFVACRALSQRNNDPTKASRPWDSNR
DGFVMGEGAGVLLLEELEHAKKRGATIIYAEFLGGSFTCDAYHMTPEHPGAGVILCI
EKALAQAGVSRREDVNYINAHATSTPAGDIKEYQALAHCFGQNSELRVNSTKSMIGHL
IGAAGGVEAVTVVQAIRTGWHPNLNLEDPKAVDAKVLVGPKKERLNVKVGLSNS
FGFGGHNSILFAPYN

SEQ ID NO: 3

Secuencia de aminoácidos de KASIV de *C. camphora* (D3147, pSZ4338).

5

MAMMAGSCSNLVIGNRELGGNGPSLLHYNGLRPLENIQTASAVKKPNGLFASSTAR
KSKAVRAMVLPTVTAPKREKDPKKRIVITGMGLVSVFGNDIDTFYSKLLEGESGIGPI
DRFDASSFSVRFAGQIHNFSSKGYIDGKNDRRLDDCWRYCLVAGRRALDANLGPE
VLEKMDRSRIGVLIGTGMGGLSAFNGVESLIQKGYKKITPFFIPYSITNMGSAALLAID
TGVMGPNYSISTACATANYCFHAAANHIIRGEAEIMVTGGTEAAVSATGVGGFIACR
ALSHRNDEPQTASRPWDKDRDGFVMGEGAGVLVMESLHHARKRGANIIAEYLGA
VTCDAAHMTDPRADGLGVSSCITKSLEDAGVSPEEVNYVNAHATSTLAGDLAEVNA
IKKVFKDTSEMKMNGTKSMIGHCLGAAGGLEAIATIKAIN TGWLHPTINQFNIEPAVT
IDTVPNVKKKHDIHVGISNSFGFGGHNSVVVFAPFMP

SEQ ID NO: 4 Secuencia de aminoácidos de KASI de *C. camphora* (D3148, pSZ4339).

MQILQTPSSSSSLRMSSMESLSLTPKSLPLKTLPLRPRPKNLSRRKSQNPRPISSSSSP
ERETDPKKRVVITGMGLVSVFGNDVDAYYDRLLSGESGIAPIDRFDASKFPTRFAGQI
RGFTSDGYIDGKNDRRLDDCLRYCIVSGKKALENAGLGPMLMDGKIDKERAGVLVG
TGMGGLTVFSNGVQTLHEKGYRKMT PFFIPYAITNMGSAALLAIELGFMGPNYSISTAC
ATSNYCFYAAANHIIRGEADLMLAGGTEAAIPIGLGGFVACRALSQRNDDPQTASR
PWDKDRDGFVMGEGAGVLVMESLEHAMKRDAPIIAEYLGGAVNCDAYHMTDPR

DGLGVSTCIERSLEDAGVAPEEVNYINAHATSTLAGDLAEVNAIKKVFTNTSEIKINA
TKSMIGHCLGAAGGLEAIATIKAIN TGWLHPSINQFNPEPSVEFDTVANKKQQHEVN
VAISNSFGFGGHNSVVVFSAFKP

10

SEQ ID NO: 5

Secuencia de aminoácidos de KASI de *U. californica* (D3150, pSZ4341).

MESLSLTPKSLPLKTLPLRPRPKNLSRRKSQNPKPISSSSSPERETDPKKRVVITGMGL
VSVFGNDVDAYYDRLLSGESGIAPIDRFDASKFPTRFAGQIRGFTSDGYIDGKNDRRL
DDCLRYCIVSGKKALENAGLGPDLMDGKIDKERAGVLVGTGMGGLTVFSNGVQTL
HEKGYRKMT PFFIPYAITNMGSAALLAIDLGMGPNYSISTACATSNYCFYAAANHIIR
GEADVMLAGGTEAAIPIGLGGFVACRALSQRNDDPQTASRPWDKDRDGFVMGEA
GVLVMESLEHAMKRDAPIIAEYLGGAVNCDAYHMTDPRADGLGVSTCIERSLEDAG
VAPEEVNYINAHATSTLAGDLAEVNAIKKVFTNTSEIKINATKSMIGHCLGAAGGLE
AIATIKAIN TGWLHPSINQFNPEPSVEFDTVANKKQQHEVNVAISNSFGFGGHNSVVV
FSAFKP

15

SEQ ID NO: 6

Secuencia de aminoácidos de KASIV de *U. californica* (D3152, pSZ4343).

MTQTLICPSSMETLSLTKQSHFRLRLPTPPHIRRGGGHRHPPPIISASAAPRRETDPKK
RVVITGMGLVSVFGTNVDVYYDRLLAGESGVGTIDRFDASMFPTRFGGQIRFTSEG
YIDGKNDRRLDDYLRVCLVSGKKAIESAGFDLHNTNKIDKERAGILVSGSMGGLKV
FSDGVESLIEKGYRKISPFIPYMIIPNMGSALLGIDLGFMGPNYSISTACATSNYCIYAA
ANHIRQGDADLMVAGGTEAPIPIGLGGFVACRALSTRNDDPQTASRPWDIDRDGFV
MGEGAGILVLESLEHAMKRDAPILAEYLGGAVNCDAAHHMTDPRADGLGVSTCIESS
LEDAGVAAEEVNYINAHATSTPTGDLAEMKAIKNVFRNTSEIKINATKSMIGHCLGA
SGGLEAIATLKAITTGWLHPTINQFNPEPSVDFDTVAKKKKQHEVNVAISNSFGFGGH
NSVLVFSAPKP

SEQ ID NO: 7

Secuencia de aminoácidos de KASAI de *C. wrightii* (D3153, pSZ4379). El péptido de tránsito de alga está subrayado.

5

MASAAFTMSACPAMTGRAPGARRSGRPVATRLRYVFQCLVASCIDPCDQYRSSASL
SFLGDNGFASLFGSKPFMSNRGHRRLRRASHSGEAMAVALQPAQEAGTKKKPVIKQ
RRVVVTGMGVVTPLGHEPDVFYNNLLDGVSGISEIETFDCTQFPTRIAGEIKSFSTDG
WVAPKLSKRMDKFMLYLLTAGKKALADGGITDEVMMKELDKRKCGVLIGSGMGGM
KVFNDIAEALRVSYKKMNPFCVPFATINMGSAAMLAMD LGWMGPNYSISTACATSN
FCILNAANHIIRGEADMMLCGGSDAVIPIGLGGFVACRALSQNRSDPTKASRPWDSN
RDGFVMGEGAGVLLLEEEHAKKRGATIIAEFLGGSFTCDAYHMTPEHPGAGVIL
CIEKALAQAGVSKEDVNYINAHATSTSAGDIKEYQALARCFGQNSLRVNSTKSMIG
HLLGAAGGVEAVTVVQAIRTGWIHPNLNLEDPKAVDAKLLVGPKKERLNVKVGL
SNSFGFGGHNSSILFAPCNV

SEQ ID NO: 8

Secuencia de aminoácidos de KASIVb de *C. avigera* (D3287, pSZ4453).

10

MASAAFTMSACPAMTGRAPGARRSGRPVATRLRGSTFQCYIGDNGFGSKPPRSNRG
HLRLGRTSHSGEVMAMQSAQEVSTKEKPA TKQRRVVVTGMGVVTALGHDPDV

YYNNLLDGVSGISEIENFDCSQLPTRIAGEIKSF SADGWVAPKFSRRMDKFMLYILTA
GKKALVDGGITEDVMKELDKRKCGVLIGSGLGGMKVFESEIALRTSYKKISPFVFPF
STTNMGSAAMLAMD LGWMGPNYSISTACATSNFCILNAANHIIRGEADMMLCGGSDS
VILPIGMGGFVACRALSQNRNDPTKASRPWDSNRDGFVMGEGAGVLLLEEEHAKK
RGATIIAEFLGGSFTCDAYHMTPEHPGAGVILCIEKALAQSGVSREDVNYINAHATS
TPAGDIKEYQALAHCFGQNSLRVNSTKSMIGHLLGAGGVEAVTVVQAIRTGWIHP
NINLDDPDEGVDAKLLVGPKKEKLKVKVGLSNSFGFGGHNSSILFAPCN

SEQ ID NO: 9

Secuencia de aminoácidos de KASIVb de *C. paucipetala* (D3288, pSZ4454).

15

MASAAFTMSACPAMTGRAPGARRSGRPVATRLRGSTFQCLGDIGFASLIGSKPPRSN
RNIHRLGRTSHSGEVMAMQPAHEASTKNKPVTQRRVVVTGMGVATPLGHDP
DVYYNNLLDGVSGISQIENFDCTQFPTRIAGEIKSFSTEGYVIPKFAKRMDKFMLYLL
TAGKKALEDDGGITEDVMKELDKRKCGVLIGSGMGGMKIINDSIAALNVSYKKMTPF
CVPFSTTNMGSAAMLAMD LGWMGPNYSISTACATSNYCIILNAANHIIRGEADMMLCG
GSDAVIIPVGLGGFVACRALSQNRNDPTKASRPWDSNRDGFVMGEGAGVLLLEEE
HAKKRGATIIAEFLGGSFTCDAYHMTPEHPDGAGVILCIEKALAQSGVSREDVNYIN
AHATSTPAGDIKEYQALAHCFGQNSLRVNSTKSMIGHLLGAAGGVEAVTVVQAIR
TGWIHPNINLENPDEAVDAKLLVGPKKEKLKVKVGLSNSFGFGGHNSSILFAPYN

SEQ ID NO: 10

Secuencia de aminoácidos de KASIVb de *C. ignea* (D3289, pSZ4455). El péptido de tránsito de alga está subrayado.

20

MASAAFTMSACPAMTGRAPGARRSGRPVATRLRGSTSQCLVTSYIDPCNKYCSSASL
SFLGDNGFASLFGSKPFRSNRGHRRLLGRASHSGEAMAVALQPAQEVTTKKKPKVIKQR
RVVVTGMGVVTPLGHEPDVYYNNLLDGVSGISEIETFDCTQFPTRIAGEIKSFSTDGW
VAPKLSKRMDKFMFLYLLTAGKKALADGGITDDVMKELDKRKCGVLIGSGMGGMK
LFNDSIEALRISYKKMNPFCVPPFATTNMGSAAMLAMD LGWMGPNYSISTACATSNFCI
LNASNHIVRGEADMMLCGGSDSVTVPLGVGGFVACRALSQRNNDPTKASRPWDSN
RDGFVMGEGAGVLLLEEEHAKKRGATIYAEFLGGSFTSDAYHMTPEHPGAGVILC
IEKALAQSGVSREDVNYINAHATSTPAGDIKEYQALARCFGQNSELRVNSTKSMIGH
LLGAAGGVEAVAVIQAIRTGWIHPNINLEDPDEAVDPKLLVGPKEKLVKVALSNS
FGFGGHNSSILFAPCN

SEQ ID NO: 11

Secuencia de aminoácidos de KASIV de *C. procumbens* (D3290, pSZ4456). El péptido de tránsito de alga está subrayado.

MASAAFTMSACPAMTGRAPGARRSGRPVATRLRGSTFQCLVTSNHDPCNQYCSSAS
LSFLGDNGFSGSKPFRSNRGHRRLLGRASHSGEAMAVALQPAQEVATKKKPKAMKQRR
VVVTGMGVVTPLGHEPDVYYNNLLDGVSGISEIETFDCTQFPTRIAGEIKSFSTDGWV
APKLSKRMDKFMFLYLLTAGKKALADGGITDDVMKELDKRKCGVLIGSGMGGMKLF
NDSIEALRVSYKKMNPFCVPPFATTNMGSAAMLAMD LGWMGPNYSISTACATSNFCIL
NAANHIVRGEADMMLCGGSDAVIPIGLGGFVACRALSQRNNDPTKASRPWDSNRD
GFVMGEGAGVLLLEEEHAKKRGATIYAEFLGGSFTCDAYHMTPEHPGAGVILCIE
KALAQSGVSREDVNYINAHATSTPAGDIKEYQALAHCFGQNSELRVNSTKSMIGHLL
GAAGGVEAVTVIQAIRTGWIHPNINLEDPDKAVDAKFLVGPKEKRLNVKVGLSNSF
FGFGGHNSSILFAPCN

SEQ ID NO: 12

Secuencia de aminoácidos de KASIVa de *C. paucipetala* (D3291, pSZ4457). El péptido de tránsito de alga está subrayado.

MASAAFTMSACPAMTGRAPGARRSGRPVATRLRGSTFQCLVNSHIDPCNQNVSSAS
LSFLGDNGFSGNPFRSNRGHRRLLGRASHSGEAMAVALQPAQEVATKKKPKAIKQRRV
VVTGMGVVTPLGHEPDVFYNNLLDGVSGISEIETFDCTQFPTRIAGEIKSFSTDGWVA
PKLSKRMDKFMFLYLLTAGKKALADAGITEDVMKELDKRKCGVLIGSGMGGMKLFN
DSIEALRVSYKKMNPFCVPPFATTNMGSAAMLAMD LGWMGPNYSISTACATSNFCILN
AANHIVRGEADMMLCGGSDAVIPIGLGGFVACRALSQRNNDPTKASRPWDSNRDGF
VMGEGAGVLLLEEEHAKKRGATIYAEFLGGSFTCDAYHMTPEHPDGAGVILCIEKA
LAQSGVSREDVNYINAHATSTPAGDIKEYQALAHCFGQNSELRVNSTKSMIGHLLGA
AGGVEAVTVIQAIRTGWIHPNINLEDPDEAVDAKFLVGPKEKRLNVKVGLSNSFSGFG
GHNSSILFAPYN

SEQ ID NO: 13

Secuencia de aminoácidos de KASIV de *C. painteri* (D3292, pSZ4458). El péptido de tránsito de alga está subrayado.

MASAAFTMSACPAMTGRAPGARRSGRPVATRLRGSTPQCLDPCNQHCFLGDNGFAS
LIGSKPFRSNLGHRLRGRTSHSGEVMAVAQEVSTNKKHATKQRRVVVTGMGVVTPL
GHDPDVYYNNLLEGVSGISEIENFDCSQLPTRIAGEIKSFSTDGLVAPKLSKRMDKFM
LYILTAGKKALADGGITEDVMKELDKRKCGVLIGSGLGGMKVFSDSVEALRISYKKI
SPFCVPPFSTTNMGSAAMLAMD LGWMGPNYSISTACATSNFCILNAANHITKGEADM
LCGGSDAAILPIGMGGFVACRALSQRNNDPTKASRPWDSNRDGFVMGEGAGVLLLE
EEHAKKRGATIYAEFLGGSFTCDAYHMTPEHPDGAGVILCIEKALAQSGVSREEVN
YINAHATSTPAGDIKEYQALAHCFGQNSELRVNSTKSMIGHLLGGAGGVEAVTVVQ
AIRTGWIHPNINLEDPDKGVDAKLLVGPKEKLVKVGGLSNSFSGFGGHNSSILFAPCN

SEQ ID NO: 14

Secuencia de aminoácidos de KASIVa de *C. avigera* (D3293, pSZ4459). El péptido de tránsito de alga está subrayado.

MASAAFTMSACPAMTGRAPGARRSGRPVATRLRGSTFQCLVTSYNDPCEQYRSSAS
LSFLGDNGFASLFGSKPFRSNRGHRRLLGRASHSGEAMAVALQPAQEVGTKKKPKVIKQ
 RRVVVTGMSGVVTPLGHEPDVYYNNLLDGVSGISEIETFDCTQFPTRIAGEIKSFSTDG
 WVAPKLSKRMDKFMLYLLTAGKKALADGGITDDVMKELDKRKCGVLIGSGLGGM
 KVFSEIEALRTSYKKISPFVFPSTTNMGSAILAMD LGWMGPNYSISTACATSNFCIL
 NAANHITKGEADMMLCGGSDSVILPIGMGGFVACRALSQRNNDPTKASRPWDSNRD
 GFVMGEGAGVLLLEELEHAKKRGATTYAEFLGGSFTCDAYHMTPEHPGAGVILCIE
 KALAQSGVSRDVDVNYINAHATSTPAGDIKEYQALAHCFGQNSLRVNSTKSMIGHLL
 GGAGGVEAVTVVQAI RTGWHPNINLDDPDEGVDAKLLVGPKKEKLKVKVGLSNSF
 FGFGHNSSILFAPCN

SEQ ID NO: 15

- 5 Secuencia de aminoácidos de *KASIVa* de *C. ignea* (D3294, pSZ4460). El péptido de tránsito de alga está subrayado.

MASAAFTMSACPAMTGRAPGARRSGRPVATRLRGSTSQCCLVTSYIDPCNKYCSSL
SFLGDNGFASLFGSKPFRSNRGHRRLLGRASHSGEAMAVALQPAQEVTTKKKPKVIKQR
 RRVVVTGMSGVVTPLGHEPDVYYNNLLDGVSGISEIETFDCTQFPTRIAGEIKSFSTDGW
 VAPKLSKRMDKFMLYLLTAGKKALADGGITDDVMKELDKRKCGVLIGSGMGGMK
 LFNDSEIALRISYKKMNPFCVPFATTNMGSAMLAMD LGWMGPNYSISTACATSNFCI
 LNASNHIVRGEADMMLCGGSDAVIPIGLGGFVACRALSQRNNDPTKASRPWDSNRD
 GFVMGEGAGVLLLEELEHAKKRGATTYAEFLGGSFTCDAYHMTPEHPGAGVILCIE
 KALAQAGVSKEDVNYINAHATSTPAGDIKEYQALAQCFGQNSLRVNSTKSMIGHLL
 LGAAGGVEAVTVVQAI RTGWHPNLNLEDPKAVDAKLLVGPKKERLNVKVGLSNS
 FGFGHNSSILFAPYN

10

SEQ ID NO: 16

Secuencia de aminoácidos de *KASla* de *C. avigera* (D3342, pSZ4511).

MQSLHSPALRASPLDPLRLKSSANGPSSTA AFRPLRRATLPNIRAASPTVSAPKRETD
PKRRVVTG MGLVSVFGSDVDAYYEKLLSGESGISLDRFDASKFPTRFGGQIRGFENAT
 GYIDGKNDRRLDDCLRYCIVAGKKALENSDLGGDSLKIDKERAGVLVGTGMGGLT
 VFSDGVQNLIEKGHRKISPFPIPYAITNMGSALLAIDLGLMGPNYSTACATSNYCFY
 AAANHIRRGEADLMIAGGTEAAIPIGLGGFVACRALSQRNDDPQTASRPWDKDRDG
 FVMGEGAGVLVMESEHAMKRGAPIIAEYLGGAVNCDAYHMTDPRADGLGVSSCIE
 SSLEDAGVSPEEVNYINAHATSTLAGDLAEINAIKKVFKNTKDIKINATKSMIGHCLG
 ASGGLEAIATIKGITTGWLHPSINQFNPEPSVEFDTVANKKQQHEVNVAISNSFGFGG
 HNSVVAFAFKP

15

SEQ ID NO: 17

Secuencia de aminoácidos de *KASI* de *C. pulcherima* (D3343, pSZ4512).

MHSLQSPSLRASPLDPFRPKSSTVRPLHRASIPNVRAASPTVSAPKRETDPKRRVVTG
MGLVSVFGSDVDAYYDKLLSGESGIGPIDRFDASKFPTRFGGQIRGFNSMGYIDGKN
 DRRLDDCLRYCIVAGKKSLEDADLGADRLSKIDKERAGVLVGTGMGGLTVFSDGVQ
 SLIEKGHRKITPFPIPYAITNMGSALLAIELGLMGPNYSTACATSNYCFHAAANHRR
 GEADLMIAGGTEAAIPIGLGGFVACRALSQRNDDPQTASRPWDKDRDGFVMGEA
 GVLVLESLEHAMKRGAPIIAEYLGGAINCDAYHMTDPRADGLGVSSCIESSLEDAGV
 SPEEVNYINAHATSTLAGDLAEINAIKKVFKNTKDIKINATKSMIGHCLGASGGLEAI
 ATIKGINTGWLHPSINQFNPEPSVEFDTVANKKQQHEVNVAISNSFGFGGHNSVVAFA
 FFKP

20

SEQ ID NO: 18

Secuencia de aminoácidos de *KAS mitocondrial* de *C. avigera* (D3344, pSZ4513).

MVFLPWRKMLCPSQYRFLRPLSSSTTFDPRRVVVTGLGMVTPLGCGVNTTWKQLE
GKCGIRAIISLEDLKMDAFDIDTQAYVFDQLTSKVAATVPTGVNPGEFNEDLWFNQKE
HRAIARFIAYALCAADEALKDANWEPTPEEREMTGVSIGGGTGSISDVLDAGRMIC
EKKLRLSPFFIPRILINMASGHVSMKYGFQGNHAAVTACATGAHSIGDAARMIQF
GDADVMVAGGTESSIDALSIAGFCSRALTTKYNSCPQEASRPFDTRDGRGVIGEGSG
VLVLEELDHARKRGAKMYAEFCGYGMSGDAHHTQPHSDGRGAILAMTRALKQSN
LHPDQVDYVNAHATSTSLGDAIEAKAIKTVFSDHAMSGSLALSSTKGAIGHLLGAAG
AVEAIFSI LAIKNGLAPLTLNVARPDPVFTERFVPLTASKEMHVR AALSNSFGFGGTN
TLLFTSPPQN

SEQ ID NO: 19

Secuencia de aminoácidos de KASIII de *C. avigera* (D3345, pSZ4514).

5

MANAYGEVGS SVPTVGRAAQFQOMGSGFCSVDFISKRVFCCSAVQGADKPASGDSR
AEYRTPRLVSRGCKLIGSGSAIPTLQVSNDDLAKIVDTNDEWISVRTGIRNRRVLTGK
DSLTLNATEAARKALEMAQVDAEDVDMVLMCTSTPEDLFGSAPQIQKALGCKKNPL
SYDITAACSGFVLGLVSAACHIRGGGFNNVLVIGADSLRYVDWTDRTGTCILFGDAA
GAVLVQSCDAEEDGLFAFDLHSDGDGQRHLRAVITENETDHAVGTNGSVSDFPFRRS
SYSCIQMNGKEVFRFACRSVPQSIELALGKAGLNGSNIDWLLHQNQRIDAVATRL
EVPQERVISNLANYGNTS AASIPLALDEAVRGGKVKPGHLIATAGFGAGLTWGS AIV
RWG

SEQ ID NO: 20

Secuencia de aminoácidos TAG del epítipo HA

10

TMYPYDVPDYA

SEQ ID NO: 21

CDS de KASIV de *C. palustris*

ATGGCGGCCCGCCGCTTCCATGGTTGCGTCCCCACTCTGTACGTGGCTCGTAGCCG
CTTG CATGTCCACTTCC TCGACAACGACCCACGTTCCCCGTCCATCAAGCGTCTC
CCCCGCCGGAGGAGGACTCTCTCCCAATCCTCCCTCCGCGGCGGATCCACCTTCC
AATGCCTCGTCACTCATAATCGACCCTTGCAATCAGTTCTCCTCCTCCGCCTCC
CTTAGCTTCTCGGGGATAACGGATTCGCATCCCTTTTCGGATCCAAGCCTTTCCG
GTCCAATCGCGGCCACCGGAGGCTCGGCCGTGCTTCCCATTCCGGGGAGGCCATG
GCCGTGGCTTTGGAACCTGCACAGGAAGTCGCCACGAAGAAGAAACCTCTTGTC
AAGCAAAGGCGAGTAGTTGTTACAGGAATGGGCGTGGTGACTCCTCTAGGCCAT
GAACCTGATGTTTACTACAACAATCTCCTAGATGGAGTAAGCGGCATAAGTGAG
ATAGAGGCCTTCGACTGCACTCAGTTTCCCACGAGAATTGCCGGAGAGATCAAGT
CTTTTTCCACAGATGGATGGGTGGCCCCAAAGCTCTCCAAGAGGATGGACAAGT
CATGCTTTACTTGTGACTGCTGGCAAGAAAGCATTAGCGGATGGTGGAATCACC
GATGATGTGATGAAAGAGCTTGATAAAAGAAAGTGTGGAGTTCTCATTGGCTCC
GGATTGGGCGGCATGAAGCTGTTCAGTGATTCATTGAAGCTCTGAGGATTTTAT
ATAAGAAGATGAATCCCTTTTGTGTACCTTTTGTCTACTACAAATATGGGATCAGC
TATGCTTGCAATGGACTTGGGATGGATGGGTCTTAACCTACTCGATATCAACTGCC
TGTGCTACAAGTAATTTCTGTATACTGAATTTCTGCAAATCACATAGTCAGAGGCG
AAGCTGACATGATGCTTTGTGGTGGCTCGGATGCGGTCACTTATACTATTGGTTT
GGGAGGTTTGTGGCGTGCCGAGCTTTGTACAGAGGAATAATGACCCTACCAA
AGCTTCGAGACCATGGGACAGTAATCGTGATGGATTGTAAATGGGCGAAGGAGC
TGGAGTGTTACTTCTCGAGGAGTTAGAGCATGCAAAGAAAAGAGGTGCCACCAT
TTATGCGGAATTTTATAGGGGGCAGTTTCACTTGGGATGCCTACCATATGACCGAG
CCTCACCTGAAGGTGCTGGAGTGATCCTCTGCATAGAGAAGGCCCTTGGCTCAGG
CCGGAGTCTCTAGAGAAGACGTAAATACATAAATGCGCATGCAACTTCCACTCC
TGCTGGAGATATCAAGGAATACCAAGCTCTCGCACACTGCTTCGGCCAAAACAG
TGAGCTGAGAGTGAATTCCTACTAAATCGATGATCGGTCATCTTATGGAGCAGCT
GGTGGTGTAGAAGCAGTTACCGTAGTTTCAGGCGATAAGGACTGGGTGGATCCAT
CCAAATCTTAATTTGGAGGACCCGGACAAAAGCCGTGGATGCAAAAAGTGCTCGTA
GGACCTAAGAAGGAGAGACTAAATGTCAAGGTGGGTTTGTCCAATTCATTTGGGT
TCGGTGGTCACTAAGTCCGTCCTACTCTTCGCCCTTACAATTAG

15

SEQ ID NO: 22

CDS de KASIV de *C. camphora*

ATGGCAATGATGGCAGGTTCTTGTTCCAATTTGGTGATTGGAAACAGAGAATTGG
GTGGGAATGGGCCTTCTTTGCTTCACTACAATGGCCTCAGACCATTTGGAAAATAT
TCAAACAGCCTCAGCTGTGAAAAAGCCAAATGGGTTATTTGCATCTTCTACAGCT
CGAAAATCCAAAGCTGTTCAGAGCCATGGTATTGCCCACTGTAACAGCTCCAAAA
CGCGAAAAAGATCCCAAGAAGCGGATTGTAATAACAGGAATGGGCCTGGTTTCC
GTCTTTGGAAATGACATTGATACATTTTATAGTAAACTACTGGAAGGAGAGAGCG
GGATTGGCCCAATCGACAGATTTGATGCTTCTTCTCAGTGAGATTTGCTGGT
CAGATTCACAATTTCTCATCCAAAGGATACATTGATGGGAAGAATGATCGTCGGC
TAGATGACTGCTGGAGGTATTGCCCTTGTGGCTGGAAGAAGAGCCCTTGAAGATG
CCAATCTTGGACCAGAGGTATTGGAAAAAATGGACCGATCTCGAATAGGGGTGC
TGATAGGGACAGGAATGGGTGGGTGTGACGCTTTAGCAATGGAGTTGAGTCTCT
GATCCAGAAGGGCTACAAGAAAATCACTCCATTTTTTATTCCTTACTCCATCACC
AATATGGGCTCTGCTCTTTTAGCAATCGACACGGGCGTAATGGGACCAAACACTACT
CCATTTCAACAGCATGTGCAACCGCAAACCTATTGCTTCCATGCTGCTGCAAATCA
TATAAGAAGGGGTGAAGCTGAAATCATGGTGACTGGAGGGACAGAGGCAGCAG
TCTCAGCTACTGGAGTTGGCGGATTTCATAGCATGTAGAGCCTTATCGCACAGGAA
TGATGAGCCCCAGACGGCCTCGAGACCATGGGATAAAGATCGGGATGGTTTCGT
CATGGGCGAAGGCGCTGGTGTGCTGGTGATGGAGAGCTTGCATCATGCAAGAAA
GAGAGGAGCAAACATAATTGCAGAGTATTTAGGAGGAGCAGTAACATGTGATGC
ACATCACATGACAGATCCTCGAGCTGATGGTCTCGGGGTTTCTTCTTGATAACC
AAGAGCTTAGAAGATGCAGGAGTCTCCCCAGAAGAGGTGAACTATGTGAATGCT
CATGCAACATCAACACTTGCAGGAGATTTAGCAGAGGTTAATGCCATAAAGAAG
GTCTTCAAGGACACATCTGAAATGAAAATGAATGGAACCTAAGTCAATGATTGGA
CACTGTCTTGGAGCAGCTGGTGGATTAGAAGCCATTGCGACCATCAAAGCTATCA
ATACTGGCTGGCTACATCCAACCATCAATCAATTTAACATAGAACCAGCGGTAAC
TATCGACACGGTCCCAAATGTGAAGAAAAAGCATGATATCCATGTTGGCATCTCT
AACTCATTGTGGCTTTGGTGGGCACAACCTCGGTGGTCGTTTTTGCTCCCTTCATGCC
ATGA

5

SEQ ID NO: 23

CDS de KASI de *C. camphora*

ATGCAAATCCTCCAAACCCCATCATCATCATCGTCTTCTCTCCGCATGTCGTCCAT
GGAATCTCTCTCTCTCACCCCTAAATCTCTCCCTCTCAAAACCCCTTCTTCCCCTTC
GTCTTCGCCCTAAAAACCTCTCCAGACGCAAATCCCCAAAACCCCTAGACCCATCTC
CTCCTCTTCTTCCCCCGAGAGAGAGACGGATCCCAAGAAGCGAGTCGTATCACC
GGGATGGGCCTCGTCTCCGTCTTCGGCAACGATGTCGATGCCTACTACGACCGCC
TCCTCTCGGGAGAGAGCGGCATCGCCCCCATCGATCGCTTCGACGCTCCAAAGTT
CCCCACCAGATTGCGCCGGTCAGATCCGAGGGTTCACCTCCGACGGCTACATTGAC
GGGAAGAACGACCGCCGGTTAGACGATTGTCTCAGATACTGTATTGTTAGTGGG
AAGAAGGCGCTCGAGAATGCCGGCCTCGGACCCCATCTCATGGACGGAAAGATT
GACAAGGAGAGAGCTGGTGTGCTTGTGCGGGACAGGCATGGGTGGTCTTACAGTT
TTCTCTAATGGGGTCCAGACTCTACATGAGAAAGGTTACAGGAAAATGACTCCGT

10

TTTTCATCCCTTATGCCATAACAAACATGGGTTCTGCCTTGCTTGCAATTGAACTT
GGTTTTATGGGCCCCAACTATTCTATCTCAACTGCATGTGCTACCTCCAATTATTG
CTTTTATGCTGCTGCTAACCATATACGGAGAGGTGAGGCTGATCTGATGCTTGCT
GGTGGAACTGAAGCTGCAATTATTCCTATTGGATTAGGAGGCTTTGTTGCATGTA
GAGCTTTATCACAGAGAAATGATGACCCCCAGACAGCTTCAAGACCATGGGACA
AAGATCGAGACGGTTTTTGTATGGGTGAAGGTGCTGGAGTATTGGTAATGGAGA
GCTTGGAGCATGCTATGAAACGTGATGCACCAATTATTGCTGAGTATTTAGGAGG
TGCAGTGAACGTGTGATGCGTATCATATGACGGATCCTAGAGCTGATGGGCTCGGG
GTTTCAACATGCATAGAAAGAAGTCTTGAAGATGCTGGTGTGGCACCTGAAGAG
GTTAACTACATAAATGCACATGCAACTTCCACTCTTGCAGGAGACCTGGCTGAGG
TGAATGCGATCAAAAAGGTTTTTACAAACACTTCAGAGATCAAAATCAATGCAA
CCAAGTCTATGATAGGGCACTGCCTTGGAGCGGCCGGGGGGTTAGAAGCCATTG
CCACAATCAAAGCAATAAATACTGGTTGGCTGCACCCTTCTATAAACCAATTTAA
TCCAGAGCCCTCTGTGAGTTTGACACTGTAGCAAATAAAAAGCAGCAGCATGA
AGTGAATGTTGCCATTTCCTCAACTCTTTCGGGTTTGGCGGACACAACCTCAGTCGTG
GTGTTTTTCGGCATTCAAGCCTTGA

SEQ ID NO: 24

CDS de KASI de *Umbellularia californica*

5

ATGGAATCTCTCTCTCTCACCCCTAAATCTCTCCCTCTCAAAACCCCTTCTTCCCTTT
CGTCCCTCGCCCTAAAAACCTCTCCAGACGCAAATCCCAAAACCCCTAAACCCATCT
CCTCCTCTTCCCTCCCCGGAGAGAGAGACGGATCCCAAGAAGCGAGTCGTCATCAC
CGGGATGGGCCCTCGTCTCCGTCTTCGGCAACGACGTGATGCCTACTACGACCGC
CTCCTCTCCGGAGAGAGCGGCATCGCCCCATCGATCGCTTCGACGCCTCCAAGT
TCCCCACCAGATTCGCCGGTCAGATCCGAGGGTTCACCTCCGACGGCTACATTGA
CGGGAAGAACGACCGCCGGTTAGACGATTGTCTCAGATACTGTATCGTTAGTGG
GAAGAAGGCGCTCGAGAATGCCGGCCTCGGACCCGATCTCATGGACGGAAAGAT
TGACAAGGAGCGAGCTGGTGTGCTTGTCTGGGACAGGCATGGGTGGTCTTACAGT
TTTCTCTAATGGGGTTTCAGACTCTCCATGAGAAAGGTTACAGGAAAATGACTCCG
TTTTTCATCCCTTATGCCATAACAAACATGGGTTCTGCCTTGCTTGCAATTGACCT
TGGTTTTATGGGCCCCAACTATTCTATCTCAACTGCATGTGCTACCTCCAATTATT
GCTTTTATGCTGCTGCTAACCATATACGGAGAGGTGAGGCTGATGTGATGCTTGC
TGGTGGAACTGAAGCTGCAATTATTCCTATTGGCTTAGGAGGCTTTGTTGCATGT
AGAGCTTTATCACAGCGAAATGATGACCCCCAGACAGCTTCAAGACCATGGGAC
AAAGATCGAGACGGTTTTTGTATGGGTGAAGGTGCTGGAGTATTGGTAATGGAG
AGCTTGGAGCATGCTATGAAACGTGATGCACCAATTATTGCTGAGTATTTAGGAG
GTGCAGTGAACGTGTGATGCGTATCATATGACGGATCCTAGAGCTGATGGGCTCGG
GGTTTCAACATGCATAGAAAGAAGTCTTGAAGATGCTGGTGTGGCACCTGAAGA
GGTTAACTACATAAATGCACATGCAACTTCCACACTTGCAGGTGACCTGGCCGAG
GTGAATGCCATCAAAAAGGTTTTTACAAACACTTCAGAGATCAAAATCAATGCA
ACCAAGTCTATGATAGGGCACTGCCTTGGAGCGGCCGGGGGGTTAGAAGCCATT
GCCACAATCAAAGCAATAAATACTGGTTGGCTGCACCCTTCTATAAACCAATTTA
ATCCAGAGCCCTCTGTGAGTTTGACACTGTAGCAAATAAAAAGCAGCAGCATG
AAGTGAATGTTGCCATTTCCTCAACTCTTTCGGGTTTGGTGGACACAACCTCGGTCGT
GGTGTTCGGCATTCAAGCCTTGA

SEQ ID NO: 25

CDS de KASIV de *Umbellularia californica*

10

ATGACGCAAACCCCTCATCTGCCCATCCTCCATGGAAACCCCTCTCTCTTACCAAAC
 AATCCCATTTCAGACTCAGGCTACCCACTCCTCCTCACATCAGACGCGGCGGCGG
 CCATCGCCATCCTCCTCCCTTCATCTCCGCCCTCCGCCGCCCTAGGAGAGAGACC
 GATCCGAAGAAGAGAGATCGTCATCACGGGAATGGGCCTCGTCTCCGTCTTCGGC
 ACCAACGTCGATGTCTACTACGATCGCCTCCTCGCCGGCGAGAGCGGCGTTGGCA
 CTATCGATCGCTTCGACGCGTCGATGTTCCCGACGAGATTTCGGCGGCCAGATCCG
 GAGGTTACGTCGGAGGGGTACATCGACGGGAAGAACGACCGGCGGCTGGATGA
 CTACCTCCGGTACTGCCTCGTCAGCGGGAAGAAGGCGATCGAGAGTGCTGGCTTC
 GATCTCCATAACATCACCAACAAGATTGACAAGGAGCGAGCTGGGATACTTGTT
 GGGTCAGGCATGGGCGGTCTTAAAGTTTTCTCTGATGGTGTTGAGTCTCTTATCG
 AGAAAGGTTACAGGAAAATAAGTCCATTTTTTCATCCCTTATATGATACCAAACAT
 GGGTTCTGCTTTGCTTGGAATTGACCTTGTTTTTCATGGGACCAAACACTCAATTT
 CAACTGCTTGCTACGTCAAATTATTGCATTTATGCTGCTGCAAATCATATCCGA
 CAAGGTGATGCCGACCTAATGGTTGCTGGTGGAACCTGAGGCTCCAATTATTCCAA
 TTGGCTTAGGGGGCTTTGTAGCATGTAGAGCTTTGTCAACAAGAAATGATGATCC
 CCAGACAGCTTCAAGGCCATGGGACATAGACCGAGATGGTTTTTGTATGGGCGA
 AGGAGCTGGAATATTGGTATTGGAGAGCTTGGAACATGCAATGAAACGTGATGC
 ACCAATTCTTGCTGAGTATTTAGGAGGTGCAGTTAACTGTGATGCTCATCATATG
 ACAGATCCTCGAGCTGATGGGCTTGGGGTTTCAACATGCATTGAAAGCAGTCTTG
 AAGATGCCGGCGTGGCAGCAGAAGAGGTTAACTATATAAATGCACACGCGACTT
 CAACACCTACAGGTGACCTGGCTGAGATGAAGGCTATAAAAAATGTATTTAGGA
 ACACCTCTGAGATCAAAATCAATGCAACCAAGTCTATGATTGGGCATTGCCTTGG
 AGCGTCTGGGGGGCTAGAAGCCATTGCCACATTGAAAGCGATTACAACTGGTTG
 GCTTCATCCAACATAAAACCAATTTAATCCAGAGCCTTCTGTTGACTTTGATACG
 GTGGCAAAGAAAAAGAAGCAGCATGAAGTTAATGTTGCCATTTCAAACCTTTTG
 GATTTCGGAGGACACAACCTCAGTGTTGGTGTTTTTCGGCATTCAAAGCCTTGA

SEQ ID NO: 26
 CDS de KASAI de *C. wrightii* (D3153, pSZ4379)

atgcttccgcggcattcaccattcggcgctgccccgcgatgactggcaggccccctggggcacgtcgtcctccggacggccagtcgc
caccgccctgagggtacgtattccagtgcctgggtggccagctgcacgcaccctgcgaccagtaccgcagcagcggccagcctgagctt
cctggggcgacaacggcttcgccagcctgttcggcagcaagcccttcattgagcaaccgcggccaccggccctgcgccgcgccagc
cacagcggcgaggccatggccgtggccctgcagccccccccaggaggccggcaccagaagaagccccgtatcaagcagcggcc
gcgtgggtgtgaccggcatggcggtgtgtagccccctggggccacgagcccgacgtgttctacaacaacctgtctggacggcggtgag
cggcaltcagcgagatcgagaccttcgactgcaccacgttccccaccggcatcgccggcgagatcaagagcttcagcaccgacggct
gggtggcccccaagctgagcaagcgcatggacaagttcatgctgtacctgctgaccgcggcgcaagaaggccccctggccgacggcgg
catcaccgacgaggtgtatgaaggagctggacaagcgcaagtgccggcgtgctgatcgccagcggcatggcgggcatgaaggtgttc
aacgacgccatcgaggccctgcgcgtgagctacaagaagatgaaccccttctcgtgccccttcgccaccaccaacatgggcagcgc
catgtgtggccatggacctgggctggatggggccccaaactacagcatcagcaccgctgcgccaccagcaacttctgcatctgaacgc
cgccaaccacatcatccggcgcgaggcccgacatgatgctgtgcggcgggcagcgacggcgtgatcatccccatcgccctggggcggc
ttcgtggcctgcgcgcgccctgagccagcgcaacagcgacccccaccaaggccagccgccctgggacagcaaccgcgacggcttc
gtgatggggcgagggcgccggcggtgctgctgctggaggagctggagcacgccaaagaagcggcgccaccatctacgcgagttc
ctggcgggcagcttcacctgcgacgctaccacatgaccgagccccacccgaggggcgccggcggtgatcctgtgcatcgagaagg
ccctggcccaggccggcggtgagcaaggaggacgtgaactacatcaacgccccagccaccagcaccagcggcgggacatcaag
gagttaccaggccctggcccgtgcttcggccagaacagcgagctgcgcgtgaacagcaccagaagagcatgatcgccacctgctgg
gcggcgccggcgcggtggaggccgtgacctgtgtgtagggccatccgcaccggctggattcacccaacctgaacctggaggacc
ccgacaaggccgtgtagcccaagctgctgtgtggcccccaagaaggagcgccctgaacgtgaaggtggcctgagcaacagcttcg
gttcggggggccacaacagcagcatctgttcggccccctgcaacgtgtga

SEQ ID NO: 27

5 CDS de KASIVb de *C. avigera*

ATGGCGGCCGCTTCTTGTCATGGCTGCGTCCCTTTCTGTACGTGCTCGTGGGTGCG
ATGCATGTCGACTTCATCCGACAACGACCCATGTCCCTTTCCCGCCGCGGATCC
ACCTTCCAATGCTACATCGGGGATAACGGATTTCGGATCGAAGCCTCCCCGTTCAA
ATCGTGGCCACCTGAGGCTCGGCCGCACTTCACATTCGGAGAGGTGATGGCTGT
GGCTATGCAATCTGCACAAGAAGTCTCCACAAAGGAGAAAACCTGCTACCAAGCA
AAGGCGAGTTGTTGTACGGGTATGGGTGTGGTGA CTGCTCTAGGCCATGACCC
GATGTTTACTACAACAATCTCCTAGACGGAGTAAGCGGCATAAGCGAGATAGAA
AACTTTGACTGTTCTCAGCTTCCACGAGAATTGCCGGAGAGATCAAGTCTTTTT
CTGCAGATGGGTGGGTGGCCCCGAAGTTCTCCAGGAGGATGGACAAGTTTATGC
TTTACATTC TGA CTGCAGGCAAGAAAGCATTAGTAGATGGTGGAATCACTGAAG
ATGTGATGAAAGAGCTCGATAAAAGAAAGTGTGGAGTTCTCATTGGCTCCGGAT
TGGGCGGTATGAAGGTATTTAGCGAGTCCATTGAAGCTCTGAGGACTTCATATAA
GAAGATCAGTCCCTTTTGTGTACCTTTTTCTACCACGAATATGGGATCCGCTATTC
TTGCAATGGACTTGGGATGGATGGGCCCTAACTATTCGATATCGACTGCCTGTGC
AACAAGTAACTTCTGTATACTGAATGCTGCGAACCACATAACCAAAGGCGAAGC
AGACATGATGCTTTGTGGTGGCTCGGATTCGGTCA TTTTACCTATTGGTATGGGA
GGTTTCGTAGCATGCCGAGCTTTGTACAGAGGAATAATGACCCTACCAAAGCTT
CGAGACCATGGGACAGTAATCGTGATGGATTTGTGATGGGAGAAGGTGCTGGAG
TTTTACTTCTCGAGGAGTTAGAGCATGCAAAGAAAAGAGGGCGCAACCATTTATGC
GGAATTTCTTGGTGGGAGTTTCACTTGCGATGCCTACCACATGACCGAGCCTCAC
CCTGAAGGAGCTGGAGTGATCCTCTGCATAGAGAAGGCCCTTGGCTCAGTCCGGA
GTCTCGAGGGAAGACGTAAATTACATAAATGCGCATGCAACTTCCACTCCCGCTG
GAGATATCAAAGAATACCAAGCTCTCGCCCACTGTTTTCGGCCAAAACAGTGAGTT
AAGAGTGAATTCACCAAGTCGATGATCGGTCACCTTCTTGGAGGAGCCGGTGG
CGTAGAAGCAGTTACAGTCGTTCAAGCAATAAGGACTGGATGGATCCATCCAAA
TATTAATTTGGACGACCCGGACGAAGGCGTGGATGCAAAACTGCTCGTCGGCCCT
AAGAAGGAGAAACTGAAGGTCAAGGTCGGTTTGTCCAATTCATTCGGGTTCGGC
GGCCATAA CTCA TCCATACTCTTTGCCCATGCAATTAG

SEQ ID NO: 28

CDS de KASIVb de *C. paucipetala*

ATGGCGGGCGCTTCATCAATGGTTGCCTCCCCATTCTCTACGTCCCTCGTAGCCGC
CTGCATGTCCACTTCATTCGACAACGACCCACGTTCCCTTTCCCACAACCGCATCC
GCCTCCGCGGATCCACCTTCCAATGCCTCGGGGATATCGGATTTCGCTTCCCTCAT
CGGATCCAAGCCTCCGCGTTCAAATCGCAACCACCGGAGGCTCGGCCGCACTTCC
CATTCGGGGAGGTTCATGGCTGTGGCTATGCAACCTGCACATGAAGCTTCCACAA
AGAATAAACCTGTTACCAAGCAAAGGCGAGTAGTTGTGACAGGTATGGGCGTGG
CGACTCCTCTAGGCCATGACCCCGATGTTTACTACAACAATCTCCTAGACGGAGT
AAGTGGCATAAGTCAGATAGAGAACTTCGACTGCACTCAGTTTCCCACGAGAATT
GCCGGAGAGATCAAGTCTTTCTCCACAGAAGGGTATGTGATCCCGAAGTTCGCCA
AGAGGATGGACAAGTTCATGCTTTACTTGTGACTGCAAGGCAAGAAAGCATTAG
AAGATGGTGGAAATCACTGAAGATGTGATGAAAGAGCTCGATAAAAGAAAGTGTG
GAGTTCTCATTGGCTCCGGAATGGGCGGTATGAAGATAATCAACGATTCCATTGC
AGCTCTGAATGTTTCATATAAGAAGATGACTCCCTTTTGTGTACCTTTTCCACCA
CAAATATGGGATCCGCTATGCTTGCATAGACTTGGGATGGATGGGCCCGAACT
ATTCGATATCAACTGCCTGTGCAACAAGTAAGTACTGTATACTGAATGCTGCGAA
CCACATAGTCAGAGGCGAAGCAGATATGATGCTTTGTGGTGGCTCGGATGCGGT
CATTATACCTGTTGGTTTGGGAGGTTTCGTAGCATGCCGAGCTTTGTACAGAGG
AACAAATGACCCTACCAAAGCTTCGAGACCTTGGGACAGTAACCGTGATGGATTT
GTGATGGGAGAAGGAGCCGGAGTGTTACTTCTCGAGGAGTTAGAGCATGCAAAG
AAAAGAGGTGCAACCATTATGCGGAATTTCTAGGTGGGAGTTTCACTTGCATG
CCTACCACATGACCGAGCCTCACCTGATGGAGCTGGAGTGATCCTCTGCATAGA
GAAGGCTTTGGCACAGTCCGGAGTCTCGAGGGAAGACGTCAATTACATAAATGC
GCATGCAACTTCTACTCCTGCTGGAGATATCAAGGAATACCAAGCTCTCGCCAC
TGTTTCGGCCAAAACAGTGAGTTAAGAGTGAATTCACCAAATCGATGATCGGTC
ACCTTCTTGGAGCTGCTGGTGGCGTAGAAGCAGTTACAGTAGTTCAGGCAATAAG
GACTGGGTGGATCCATCCAAATATTAATTTGGAAAACCCGGACGAAGCTGTGGA
TGCAAAATTGCTCGTCGGCCCTAAGAAGGAGAACTGAAGGTCAAGGTTCGGTTT
GTCCAATTCATTGGGTTCGGTGGGCATAACTCATCCATACTCTTCGCCCCTTACA
ATTAG

5

SEQ ID NO: 29

CDS de KASIVb de *C. ignea*

ATGGCGGGCGGGCGCTTCCATGTTTACGTCCCCACTCTGTACGTGGCTCGTAGCCT
CTTGCAATGTGCACTTCCTTCGACAACGACCCACGTTTCGCCGTCCGTCAAGCGTCT
CCCCCGCCGGAGGAGGATTTCTCTCCCAATGCTCCCTCCGCGGATCCACCTCCCAA
TGCTTCGTACCTCATACATCGACCTTGCAATAAGTACTGCTCCTCCGCCTCCCT
TAGCTTCCTCGGGGATAACGGATTTCGCATCCCTTTTCGGATCTAAGCCATTCCGG
TCCAATCGCGGCCACCGGAGGCTCGGCCGTGCTTCCCATTCGGGGGAGGCCATGG
CTGTGGCTCTGCAACCTGCACAGGAAGTCACCACGAAGAAGAAACCTGTGATCA
AGCAAAGGCGAGTAGTTGTTACAGGAATGGGCGTGGTGAATCCTCTAGGCCATG
AACCTGATGTTTACTACAACAATCTCCTAGATGGAGTAAGCGGCATAAGTGAGAT
AGAGACCTTCGACTGCACTCAGTTTCCCACGAGAATCGCCGGAGAGATCAAGTCT
TTTTCACAGATGGGTGGGTGGCCCCAAAGCTCTCCAAGAGGATGGACAAGTTC

10

ATGCTTTACTTGTGACTGCTGGCAAGAAAGCATTAGCAGATGGTGGAATCACCG
 ATGATGTGATGAAAGAGCTTGATAAAAGAAAGTGTGGGGTTCTCATTTGGCTCTG
 GAATGGGCGGCATGAAGTTGTTCACGATTCATTTGAAGCTCTGAGGATTTTATA
 TAAAAAGATGAATCCCTTTTGTGTACCTTTTGTCTACCACAAATATGGGATCAGCT
 ATGCTTGCAATGGACTTGGGATGGATGGGTCTTAAGTACTCGATATCAACTGCCT
 GTGCAACAAGTAATTTCTGTATACTGAATGCTTCAAACACATAGTCAGAGGCGA
 AGCTGACATGATGCTTTGTGGTGGCTCGGATTCTGTCACTGTACCTTTAGGTGTG
 GGAGGTTTCGTAGCATGCCGAGCTTTGTTCACAGAGGAATAATGACCTTACCAAA
 GCTTCGAGACCTTGGGACAGTAATCGGGATGGATTTGTGATGGGAGAAGGAGCT
 GGAGTGTACTTCTTGAGGAGTTAGAGCATGCAAAGAAAAGAGGTGCAACCATT
 TATGCGGAATTTCTCGGTGGGAGCTTTACTTCTGATGCCTACCACATGACCGAGC
 CTCACCCCGAAGGAGCTGGAGTGATTCTCTGCATTGAGAAGGCCTTGGCTCAGTC
 CGGAGTCTCGAGGGAAGACGTGAATTATATAAATGCGCATGCAACTTCCACTCCT
 GCTGGTGATATAAAGGAATACCAAGCTCTCGCCCGCTGTTTCGGCCAAAACAGTG
 AGTTAAGAGTGAATTCCACCAAATCGATGATCGGTACCTTCTTGGAGCAGCTGG
 TGGCGTAGAAGCAGTTGCAGTAATTCAGGCAATAAGGACTGGATGGATCCATCC
 AAATATTAATTTGGAAGACCCCGACGAAGCCGTGGATCCAAAATTGCTCGTCGG
 CCTAAGAAGGAGAACTGAAGGTCAAGGTAGCTTTGTCCAATTCATTCGGGTTC
 GGCGGGCATAACTCATCCATACTCTTTGCCCTTGCAATTAG

SEQ ID NO: 30

CDS de KASIV de *C. procumbens*

ATGGCGGCGGCGCCCTCTTCCCCACTCTGTACGTGGCTCGTAGCCGCTTGCATGT
 CCACTTCCTTCGACAACAACCCACGTTTCGCCCTCCATCAAGCGTCTCCCCCGCCG
 GAGGAGGGTTCTCTCCCAATGCTCCCTCCGTGGATCCACCTTCCAATGCCCTCGTC
 ACCTCACACAACGACCTTTGCAATCAGTACTGCTCCCTCCGCCCTCCCTTAGCTTCCT
 CGGGGATAACGGATTTCGGATCCAAGCCATTCCGGTCCAATCGCGGCCACCGGAG
 GCTCGGCCGTGCTTCGCATTCGCGGGAGGCCATGGCTGTGGCCTTGCAACCTGCA
 CAGGAAGTCGCCACGAAGAAGAAACCTGCTATGAAGCAAAGGCGAGTAGTTGTT
 ACAGGAATGGGCGTGGTGACTCCTCTGGGCCATGAACCTGATGTTTACTACAACA
 ATCTCCTAGATGGAGTAAGCGGCATAAGTGAGATAGAGACCTTCGACTGCATCTC
 AGTTTCCACGAGAATCGCCGGAGAGATCAAGTCTTTTTCCACAGATGGATGGGT
 GGCCCCAAAGCTCTCCAAGAGGATGGACAAGTTCATGCTTTACTTGTGACTGCT
 GGCAAGAAAGCATTAGCAGATGGTGGAATCACTGATGATGTGATGAAAGAGCTT
 GATAAAAGAAAGTGTGGAGTTCTCATTTGGCTCTGGAATGGGCGGCATGAAGTTG
 TTCAACGATTCCATTGAAGCTCTGAGAGTTTCATATAAGAAGATGAATCCCTTTT
 GTGTACCTTTTGTCTACCACAAATATGGGATCAGCTATGCTTGCAATGGACTTGGG
 ATGGATGGGTCTTAAGTACTCGATATCAACTGCCTGTGCAACAAGTAATTTCTGT
 TACTGAATGCTGCAAACACATAGTCAGAGGCGAAGCTGACATGATGCTTTGT
 GGTGGCTCGGATGCGGTCATTATACCTATTGGTTTGGGAGGTTTTGTGGCGTGCC
 GAGCTTTGTACAGAGGAATAATGACCCTACCAAGGCTTCGAGACCATGGGATA
 GTAATCGTGATGGATTTGTAATGGGCGAAGGAGCTGGAGTGTACTTCTCGAGGA
 GTTAGAGCATGCAAAGAAAAGAGGTGCAACCATTATGCGGAATTTTAGGGGG

5

10

CAGTTTCACCTTGCGATGCCTACCATATGACCGAGCCTCACCCCTGAAGGAGCTGGA
GTGATCCTCTGCATAGAGAAGGCCTTGGCTCAGTCCGGAGTCTCTAGAGAAGAC
GTAAATTACATAAATGCGCATGCAACTTCCACTCCTGCTGGAGATATCAAAGAAT
ACCAAGCTCTCGCCCACTGTTTCGGCCAAAACAGTGAGCTGAGAGTGAATTCAC
TAAATCGATGATCGGTCATCTTCTTGGAGCAGCTGGTGGTGTAGAAGCAGTTACC
GTAATTCAGGCGATAAAGGACTGGGTGGATCCATCCAAATCTTAATTTGGAAGACC
CGGACAAAGCCGTGGATGCAAAATTTCTCGTGGGACCTAAGAAGGAGAGACTGA
ATGTCAAGGTCGGTTTGTCCAATTCATTTGGGTTCGGGGGGCATAACTCATCCAT
ACTCTTTGCCCCTTGCAATTAG

SEQ ID NO: 31

CDS de KASIVa de *C. paucipetala*

5

ATGGCGGCGGCGGCCTCTTCCCCACTCTGCACATGGCTCGTAGCCGCTTGCAATGT
CCACTTCATTCGACAACAACCCACGTTCCGCTCCATCAAGCGTCTCCCCCGCCG
GAGGAGGGTTCTCTCCCAATGCTCCCTCCGCGGATCCACCTTCCAATGCCTCGTC
AACTCACACATCGACCTTGCAATCAGAACGTCCTCCGCTCCCTTAGCTTCCT
CGGGGATAACGGATTTCGGATCCAATCCATTCCGGTCCAATCGCGGCCACCGGAG
GCTCGGCCGGGCTTCCCATTCGGGGGAGGCCATGGCTGTTGCTCTGCAACCTGCA
CAGGAAGTCGCCACGAAGAAGAAACCTGCTATCAAGCAAAGGCGAGTAGTTGTT
ACAGGAATGGGCGTGGTGACTCCTCTAGGCCATGAGCCTGATGTTTTCTACAACA
ATCTCCTAGATGGAGTAAGCGGCATAAGTGAGATAGAGACCTTCGACTGCACTC
AGTTTCCCACGAGAATTGCCGGAGAGATCAAGTCTTTTTCCACAGATGGGTGGGT
GGCCCCAAAGCTCTCCAAGAGGATGGACAAGTTCATGCTTTACTTGTGACTGCT
GGCAAGAAAGCATTAGCAGATGCTGGAATTACCGAGGATGTGATGAAAGAGCTT
GATAAAAGAAAGTGTGGAGTTCTCATTGGCTCCGGAATGGGCGGCATGAAGTTG
TTCAACGATTCCATTGAAGCTCTGAGGGTTTCATATAAGAAGATGAATCCCTTTT
GTGTACCTTTTGCTACCACAAATATGGGATCAGCTATGCTTGCAATGGACTTGGG
ATGGATGGGTCCCTAACTACTCGATATCGACTGCCTGTGCAACAAGTAATTTCTGT
ATACTGAATGCTGCAAACACATAATCAGAGGCGAAGCTGACATGATGCTTTGT
GGTGGTTTCGGATGCGGTCATTATACCTATTGGTTTGGGAGGTTTTGTGGCGTGCC
GAGCTTTGTACAGAGGAATAGTGACCCTACCAAAGCTTCGAGACCATGGGATA
GTAATCGTGATGGATTGTGAATGGGCGAAGGAGCTGGAGTGTACTTCTCGAGGA
GTTAGAGCATGCAAAGAAAAGAGGTGCAACCATTTATGCGGAATTTTATAGGGGG
CAGCTTCACCTTGCGATGCCTACCACATGACCGAGCCTCACCCCTGATGGAGCTGGA
GTGATCCTCTGCATAGAGAAGGCTTTGGCACAGTCCGGAGTCTCGAGGGAAGAC
GTCAATTACATAAATGCGCATGCAACTTCTACTCCTGCTGGAGATATCAAGGAAT
ACCAAGCTCTCGCCCACTGTTTCGGCCAAAACAGTGAGCTGAGAGTGAATTCAC
TAAATCGATGATCGGTCATCTTCTTGGTGCAGCTGGTGGTGTAGAAGCTGTTACT
GTAATTCAGGCGATAAAGGACTGGGTGGATTTCATCCAAATCTTAATTTGGAAGACC
CGGACGAAGCCGTGGATGCAAAATTTCTCGTGGGACCTAAGAAGGAGAGATTGA
ATGTCAAGGTCGGTTTGTCCAATTCATTTGGGTTCGGTGGGCATAACTCATCCAT
ACTCTTCGCCCCTTACAATTAG

SEQ ID NO: 32

CDS de KASIV de *C. painteri*

10

ATGGCGGCCCTCCTCTTGCATGGTTGCGTCCCCGTTCTGTACGTGGCTCGTATCCGC
 ATGCATGTCTACTTCATTCGACAACGACCCACGTTCCCTTTCCACAAAGCGGCTC
 CGCCTCTCCCGTCGCCGGAGGCCCTCTCTCCTCTCATTTGCTCCCTCCGCGGATCCAC
 TCCCCAATGCCTCGACCTTGCAATCAGCACTGCTTCCTCGGGGATAACGGATTCTC
 GCTTCCCTCATCGGATCCAAGCCTCCCGTTCCAATCTCGGCCACCTGAGGCTCG
 GCCGCACTTCCCATTCGGGGGAGGTCATGGCTGTGGCACAGGAAGTCTCCACAA
 ATAAGAAACATGCTACCAAGCAAAGGCGAGTAGTTGTGACAGGTATGGGCGTGG
 TGACTCCTCTAGGCCATGACCCCGATGTTTACTACAACAATCTCCTAGAAGGAGT
 AAGTGGCATCAGTGAGATAGAGAACTTCGACTGCTCTCAGCTTCCCACGAGAATT
 GCCGGAGAGATCAAGTCTTTTCCACAGATGGGTTGGTGGCCCCGAAGCTCTCCA
 AGAGGATGGACAAGTTCATGCTTTACATCCTGACTGCAGGCAAGAAAGCATTAG
 CAGATGGTGGAACTCACTGAAGATGTGATGAAAGAGCTCGATAAAAGAAAGTGTG
 GAGTCTCATTTGGCTCCGGATTGGGCGGTATGAAGGTATTCAGCGACTCCGTTGA
 AGCTCTGAGGATTTTCATATAAGAAGATCAGTCCCTTTTGTGTACCTTTTTCTACCA
 CAAATATGGGATCCGCTATGCTTGCAATGGACTTGGGATGGATGGGCCCTAACTA
 TTCGATATCAACTGCCTGTGCAACAAGTAACCTCTGTATACTGAATGCTGCGAAC
 CACATAACCAAAGGCGAAGCTGACATGATGCTTTGTGGTGGCTCGGATGCGGCC
 ATTTTACCTATTTGGTATGGGAGGTTTCGTGGCATGCCGAGCTTTGTACAGAGGA
 ATAATGACCTACCAAAGCTTCGAGACCATGGGACAGTAATCGTGATGGATTTGT
 GATGGGAGAAGGAGCTGGAGTGTTACTTCTCGAGGAGTTAGAGCATGCAAAGAA
 AAGAGGTGCAACCATTTATGCGGAATTTCTAGGTGGGAGTTTCACTTGGCATGCC
 TACCACATGACCGAGCCTCACCTGATGGAGCTGGAGTGATCCTCTGCATAGAGA
 AGGCCCTTGGCTCAGTCCGGAGTCTCGAGGGAAGAAGTAAATTACATAAATGCGC
 ATGCAACTTCCACTCCTGCTGGAGATATCAAGGAATACCAAGCTCTCGCCCATTTG
 TTTCGGCCAAAACAGTGAGTTAAGAGTGAATTCACCAAATCGATGATCGGTCAC
 CTTCTTGGAGGAGCTGGTGGCGTAGAAGCAGTTACAGTAGTTACAGGCAATAAGG
 ACTGGATGGATCCATCCAAATATTAATTTGGAAGACCCGGACAAAGGCGTGGAT
 GCAAACTGCTCGTCGGCCCTAAGAAGGAGAACTGAAGGTCAAGGTCGGTTTG
 TCCAATTCATTTGGGTTTCGGCGGCCATAACTCATCCATACTCTTTGCCCATGCAA
 TTAG

SEQ ID NO: 33
 CDS de KASIVa de *C. avigera*

5

ATGGCGGCCCGCCGCTTCCATGGTTGCGTCCCCATTCTGTACGTGGCTCGTAGCCG
 CTTGCATGTCCACTTCCGTCGACAAAGACCCACGTTCCGCGTCTATCAAGCGTCT
 CCCCCGCCGGAAGAGGATTCAATCCCAATGCTCCCTCCGCGGATCCACCTTCCAA
 TGCTTCGTCACCTCATACAACGACCCCTTGCGAACAATAACCGCTCATCCGCCCTCC
 TTAGCTTCCTCGGGGATAACGGATTCCGATCCCTTTTCGGATCCAAGCCATTCGG
 GTCCAATCGCGGCCACCGGAGGCTCGGCCGTGCTTCCCATTCGGGGGAGGCCATG
 GCCGTGGCACTGCAACCTGCACAGGAAGTTGGCACGAAGAAGAAACCTGTTATC
 AAGCAAAGGCGAGTAGTTGTTACAGGAATGGGCGTGGTGAATCTCTAGGCCAT
 GAACCTGATGTTTACTACAACAATCTCCTAGACGGAGTAAGCGGCATAAGTGAG

10

ATAGAGACCTTCGACTGCACTCAGTTTCCACGAGAATTGCCGGAGAGATCAAGT
 CTTTTTCCACAGATGGGTGGGTGGCTCCAAAGCTCTCTAAGAGGATGGACAAGTT
 CATGCTTTACTTGTGACTGCTGGCAAGAAAGCATTGGCAGATGGTGGAATCACC
 GATGATGTGATGAAAGAGCTTGATAAAAGAAAGTGTGGAGTTCTCATTGGCTCC
 GGATTGGGCGGTATGAAGGTATTTAGCGAGTCCATTGAAGCTCTGAGGACTTCAT
 ATAAGAAGATCAGTCCCTTTTGTGTACCTTTTTCTACCACGAATATGGGATCCGCT
 ATTCTTGCAATGGACTTGGGATGGATGGGCCCTAACTATTTCGATATCGACTGCCT
 GTGCAACAAGTAACCTTCTGTATACTGAATGCTGCGAACCACATAACCAAAGGCG
 AAGCAGACATGATGCTTTGTGGTGGCTCGGATTCCGGTCATTTTACCTATTGGTAT
 GGGAGGTTTCGTAGCATGCCGAGCTTTGTTCACAGAGGAATAATGACCCTACCAA
 AGCTTCGAGACCATGGGACAGTAATCGTGATGGATTTGTGATGGGAGAAGGTGC
 TGGAGTTTTACTTCTCGAGGAGTTAGAGCATGCAAAAGAAAAGAGGCGCAACCAT
 TTATGCGGAATTTCTTGGTGGGAGTTTCACTTTCGATGCCTACCACATGACCGAG
 CCTCACCTTGAAGGAGCTGGAGTGATCCTCTGCATAGAGAAGGCCTTGGCTCAGT
 CCGGAGTCTCGAGGGAAGACGTAATAATACATAAATGCGCATGCAACTTCCACTC
 CCGCTGGAGATATCAAAGAATACCAAGCTCTCGCCCACTGTTTCGGCCAAAACA
 GTGAGTTAAGAGTGAATTCACCAAGTCGATGATCGGTCACCTTCTTGGAGGAGC
 CGGTGGCGTAGAAGCAGTTACAGTCGTTTCAGGCAATAAGGACTGGATGGATCCA
 TCCAAATATTAATTTGGACGACCCGGACGAAGGCGTGGATGCAAAACTGCTCGT
 CGGCCCTAAGAAGGAGAACTGAAGGTCAAGGTTCGGTTTGTCCAATTTCATTCGG
 GTTCGGCGGCCATAACTCATCCATACTCTTTGCCCCATGCAATTAG

SEQ ID NO: 34

CDS de KASIVa de *C. ignea*

ATGGCGGCGGCCGCTTCCATGTTTACGTCCCCACTCTGTACGTGGCTCGTAGCCT
 CTTGCATGTGCACTTCCTTCGACAACGACCCACGTTTCGCCGTCCGTCAAGCGTCT
 CCCCCGCCGGAGGAGGATTCTCTCCCAATGCTCCCTCCGCGGATCCACCTCCCAA
 TGCTTCGTACCTCATACTGACCCCTTGCAATAAGTACTGCTCCTCCGCCCTCCCT
 TAGCTTCCTCGGGGATAACGGATTTCGCATCCCTTTTCGGATCTAAGCCATTCCGG
 TCCAATCGCGGCCACCGGAGGCTCGGCCGTGCTTCCCATTCGGGGAGGCCATGG
 CTGTGGCTCTGCAACCTGCACAGGAAGTCACCACGAAGAAGAAACCTGTGATCA
 AGCAAAGGCGAGTAGTTGTTACAGGAATGGGCGTGGTGACTCCTCTAGGCCATG
 AACCTGATGTTTACTACAACAATCTCCTAGATGGAGTAAGCGGCATAAGTGAGAT
 AGAGACCTTCGACTGCACTCAGTTTCCACGAGAATCGCCGGAGAGATCAAGTCT
 TTTTCCACAGATGGGTGGGTGGCCCCAAAGCTCTCCAAGAGGATGGACAAGTTC
 ATGCTTTACTTGTGACTGCTGGCAAGAAAGCATTAGCAGATGGTGGAATCACCG
 ATGATGTGATGAAAGAGCTTGATAAAAGAAAGTGTGGGGTTCTCATTGGCTCTG
 GAATGGGCGGCATGAAGTTGTTCAACGATTCCATTGAAGCTCTGAGGATTTCATA
 TAAAAAGATGAATCCCTTTTGTGTACCTTTTGTCTACCACAAATATGGGATCAGCT
 ATGCTTGCAATGGACTTGGGATGGATGGGTCTTAATACTCGATATCAACTGCCT
 GTGCAACAAGTAATTTCTGTATACTGAATGCTTCAAACCACATAGTCAGAGGCGA
 AGCTGACATGATGCTTTGTGGTGGCTCGGATGCGGTTATTATACCTATTGGTTTG
 GGAGGTTTTGTGGCGTGCCGAGCTTTGTTCACAGAGGAATAATGACCCTACCAAA

5

10

GCTTCGAGGCCATGGGATAGTAATCGTGATGGATTTGTAATGGGCGAAGGAGCT
GGAGTGTTACTTCTCGAGGAGTTAGAGCATGCAAAGAAAAGAGGTGCAACCATT
TATGCGGAATTTTTAGGGGGCAGTTTCACTTGCGATGCCTACCACATGACCGAGC
CTCACCTGAAGGAGCTGGAGTGATCCTCTGCATAGAGAAGGCCTTGGCTCAGG
CCGGAGTCTCTAAAGAAGATGTAAATTACATAAATGCGCATGCAACTTCTACTCC
TGCTGGAGATATCAAGGAATACCAAGCTCTCGCCCAATGTTTCGGCCAAAACAGT
GAGCTGAGAGTGAATCCACTAAATCGATGATCGGTCATCTTCTTGGAGCAGCTG
GTGGTGTAGAAGCAGTTACTGTGGTTCAGGCGATAAGGACTGGGTGGATCCATC
CAAATCTTAATTTGGAAGACCCGGACAAAGCCGTGGATGCAAAGTTGCTCGTGG
GACCTAAGAAGGAGAGACTGAATGTCAAGGTCGGTTTGTCCAATTCATTTGGGTT
CGGTGGGCATAATTCGTCCATACTCTTCGCCCCTTACAATTAG

SEQ ID NO: 35

CDS de KASla de *C. avigera*

5

ATGCAATCCCTCCATTCCCCTGCCCTCCGGGCCTCCCCTCTCGACCCCTCTCCGACT
CAAATCCTCCGCCAATGGCCCCCTCTTCCACCGCCGCTTTCCGTCCCCCTCCGCCGCG
CCACCCCTCCCCAACATTCCGGGCCGCTCCCCACCGTCTCCGCCCCCAAGCGCGA
GACCGACCCCAAGAAGCGTGTCTCATCACCGGCATGGGCCTCGTCTCCGTCTTC
GGCTCCGATGTCGACGCTTATTACGAAAAGCTCCTCTCCGGCGAGAGCGGGATCA
GCTTAATCGACCGCTTCGACGCTTCCAAGTTCCCCACGAGGTTCCGGCGGCCAGAT
CCGGGGATTC AACGCCACGGGATACATCGACGGCAAAAACGACAGGAGGCTCGA
CGATTGCCTCCGCTACTGCATTGTGCGCCGGGAAGAAGGCTCTCGAAAATTCCGAT
CTCGGCGGCGATAGTCTCTCAAAGATTGATAAGGAGAGAGCTGGAGTGCTAGTT
GGAAC TG CATGGGTGGCCTAACCGTCTTCTCTGACGGGGTTCAGAATCTAATCG
AGAAAGGTCACCGGAAGATCTCCCCGTTTTTCATTCCATATGCCATTACAAACAT
GGGGTCTGCCCTGCTTGCCATCGATTTGGGTCTGATGGGCCCCAAATTATTCGATTT
CAACTGCGATGTGCTACTTCCAAC TACTGCTTTTATGCTGCTGCTAATCATATCCGC
CGAGGCGAGGCTGACCTCATGATTGCTGGAGGAACTGAGGCTGCAATCATTCCA
ATTGGGTTAGGAGGATTCGTTGCTTGCAGGGCTTTATCTCAAAGGAATGATGACC
CTCAGACTGCCTCAAGGCCGTGGGATAAGGACCGTGATGGTTTTGTGATGGGTGA
AGGGGCTGGAGTATTGGTTATGGAGAGCTTAGAACATGCAATGAAACGAGGAGC
GCCGATTATTGCAGAATATTTGGGAGGTGCAGTCAACTGTGATGCTTATCATATG
ACTGATCCAAGGGCTGATGGGCTTGGTGTCTCCTCGTGCATTGAGAGCAGTCTCG
AAGATGCCGGGGTCTCACCTGAAGAGGTCAATTACATAAATGCTCATGCGACTTC
TACTCTTGCTGGGGATCTTGCCGAGATAAATGCCATCAAGAAGGTTTTCAAGAAC
ACCAAGGATATCAAAATCAATGCAACTAAGTCGATGATTGGACACTGTCTTGGA
GCATCAGGGGGTCTTGAAGCCATTGCGACAAATTAAGGGAATAACCACTGGCTGG
CTTCATCCCAGCATAAACCAATTCAATCCCGAGCCATCAGTGGAATTTGACACTG
TTGCCAACAAGAAGCAGCAACATGAAGTCAATGTTGCTATCTCAAATTCATTCGG
ATTCGGAGGCCACAAC TCAGTTGTAGCTTTCTCAGCTTTCAAGCCATGA

SEQ ID NO: 36

CDS de KASI de *C. pulcherrima*

10

15

ATGCATTCCCTCCAGTCACCCCTCCCTTCGGGCCTCCCCGCTCGACCCCTTCCGCCC
CAAATCATCCACCGTCCGCCCCCTCCACCGAGCATCAATTCCCAACGTCCGGGCC
GCTTCCCCCACCCTCTCCGCTCCCAAGCGCGAGACCGACCCCAAGAAGCGCGTCG
TGATCACCGGAATGGGCCTTGCTCTCCGTTTTCGGGCTCCGACGTGATGCGTACTA
CGACAAGCTCCTGTCAGGCGAGAGCGGGATCGGCCCAATCGACCGCTTCGACGC
CTCCAAGTTCCCCACCAGGTTTCGGCGGCCAGATTTCGTGGCTTCAACTCCATGGGA
TACATTGACGGCAAAAACGACAGGCGGCTTGATGATTGCCTTCGCTACTGCATTG
TCGCCGGGAAGAAGTCTCTTGAGGACGCCGATCTCGGTGCCGACCGCCTCTCCAA
GATCGACAAGGAGAGAGCCGGAGTGCTGGTTGGGACAGGAATGGGTGGTCTGAC
TGTCTTCTCTGACGGGGTTCAATCTCTTATCGAGAAGGGTCACCGGAAAATCACC
CCTTCTTCATCCCCTATGCCATTACAAACATGGGGTCTGCCCTGCTCGCTATTGA
ACTCGGTCTGATGGGCCCCAACTATTCAATTTCCACTGCATGTGCCACTTCCAAC
TACTGCTTCCATGCTGCTGCTAATCATATCCGCCGTGGTGAGGCTGATCTTATGAT
TGCTGGAGGCACTGAGGCCGCAATCATTCCAATTTGGGTTGGGAGGCTTTGTGGCT
TGCAGGGCTCTGTCTCAAAGGAACGATGACCTCAGACTGCCTCTAGGCCCTGGG
ATAAAGACCGTGATGGTTTTGTGATGGGTGAAGGTGCTGGAGTGTGGTGCTGGA
GAGCTTGGAACATGCAATGAAACGAGGAGCACCTATTATTGCAGAGTATTTGGG
AGGTGCAATCAACTGTGATGCTTATCACATGACTGACCCAAGGGCTGATGGTCTC
GGTGCTCTCTTGCATTGAGAGTAGCCTTGAAGATGCTGGCGTCTCACCTGAAG
AGGTCAATTACATAAATGCTCATGCGACTTCTACTCTAGCTGGGGATCTCGCCGA
GATAAATGCCATCAAGAAGGTTTTCAAGAACACAAAGGATATCAAAATTAATGC
AACTAAGTCAATGATCGGACACTGTCTTGGAGCCTCTGGAGGTCTTGAAGCTATA
GCGACTATTAAGGGAATAAACACCGGCTGGCTTCATCCCAGCATTAAATCAATTCA
ATCCTGAGCCATCCGTGGAGTTCGACACTGTTGCCAACAAGAAGCAGCAACACG
AAGTTAATGTTGCGATCTCGAATTCATTGGATTTCGGAGGCCACAACCTCAGTCGT
GGCTTTCTCGGCTTTCAAGCCATGA

SEQ ID NO: 37

CDS de KAS mitochondrial de *C. aviga*

ATGGTGTTCCTTCCTTGGCGAAAAATGCTCTGTCCATCTCAATACCGTTTTTTGCG
GCCCTTATCTTCATCTACAACCTTTTGATCCTCGTAGGGTTGTTGTTACAGGCCTGG
GTATGGTGACTCCATTAGGATGCGGGGTGAACACCACATGGAAACAACTCATAG
AGGGGAAATGTGGGATAAGAGCAATATCCCTTGAAGACCTAAAGATGGATGCTT
TTGATATTGATACTCAGGCCTATGTATTTGATCAGCTGACCTCGAAGGTCGCTGC
CACCGTGCCACCGGAGTGAATCCCGGAGAATTTAATGAAGATTTATGGTTCAAT
CAGAAGGAGCACCGTGCTATTGCAAGGTTTCATAGCTTATGCACTCTGTGCAGCTG
ATGAAGCTCTTAAAGATGCAAAATTTGGGAACCTACTGAACCTGAAGAGAGAGAAA
TGACGGGTGTCTCCATTGGTGGAGGGACTGGAAGCATTAGCGATGTATTAGATGC
TGGTCGGATGATTTGTGAGAAGAAATTCGCTCGCCTAAGTCCATTCTTCATTCCA
CGCATATTGATAAATATGGCCTCTGGTCATGTGAGCATGAAATATGGTTTCCAGG
GACCCAACCATGCTGCTGTGACAGCTTGTGCAACAGGGGCTCATTCGATAGGTGA
TGCTGCAAGGATGATACAGTTTGGAGATGCAGATGTCATGGTCGCTGGAGGCAC
AGAATCTAGCATAGACGCCTTATCCATTGCAGGATTTTGCAGGTCAAGGGCTCTT

ACAACAAAGTATAATTCTTGCCCAACAAGAAGCTTCACGACCCTTTGATACCGATA
GAGATGGGTTTGTAAATAGGTGAAGGGTCTGGCGTCTTGGTATTGGAGGAACTAG
ATCATGCAAGAAAACGTGGTGCAAAGATGTATGCCGAGTTCTGTGGATATGGAA
TGTCTGGTGTATGCGCATCATATAACCCAACCTCATAGCGATGGAAGAGGTGCCAT
TTAGCAATGACCCGTGCATTGAAGCAGTCAAATCTACATCCGGATCAGGTGGAT
TATGTAAATGCTCACGCTACGTCTACTTCTTTAGGTGATGCAATTGAAGCTAAGG
CGATTAAAACAGTTTTCTCGGATCATGCGATGTCAGGTTGCTCGCCCTTTCTCTCC
ACCAAGGGAGCTATTGGGCATCTCCTCGGAGCAGCGGGTGCTGTGGAAGCCATT
TTCTCCATTCTGGCTATAAAAAACGGACTTTCGCCCTTTGACGCTAAATGTCGCAA
GACCAGACCCTGTGTTTACCGAGCGGTTTGTGCCTTTGACTGCTTCAAAAGAGAT
GCATGTAAGGGCGGCGTTGTCAAACCTCTTTTGGCTTTGGAGGTACAAATACTACA
CTTCTTTTCACTTCACCTCCTCAAAACTAA

SEQ ID NO: 38

KAS IV de *Cuphea palustris* con optimización de codones para *Prototheca* con secuencia y etiquetas de clonación.
5 Secuencia de nucleótidos del vector de expresión de KASIV de *C. palustris* (D3145 y D3295, pSZ4312). Los brazos de homología 5 y 3 que permiten la integración dirigida en el locus de bucle p están anotados en minúsculas; el promotor PmHXT1-2 está anotado en mayúsculas en cursiva que guía la expresión del marcador de selección ScMelibiasa anotado en minúscula en cursiva seguido del terminador PmPGK 3'UTR subrayado en mayúsculas. El promotor PmACP (anotado en **negrita**) guía la expresión de KASIV de *Cpal* con optimización de codones (anotado en minúsculas en **negrita**) y se termina con la 3'UTR C_vNR anotada en minúsculas en **negrita** subrayadas. Los lugares de clonación de restricción y los fragmentos de DNA espaciador están anotados subrayados, en mayúsculas y con letras sin formato.

aacggaggctgtcaccaaatggacccgtctattgcgggaaccacggcgatggcacgtttcaaaacttgatgaaatacaatattcag
tatgtcggggcgacggcgagctgatgtcgcgtgggtattgcttaategccagcttcgccccgtcttgccgcgagggcgt
gaacaagccgacccgatgtgacagagcaaatcctgacactagaagggctgactegccggcagcgctgaattacacaggttgcaaa
aataaccagaatttgacgcaccgtattcgcgggtattttgttgacagtgaatagcgatgcggcaatggcttggtggttagaagggtcga
cgaaggtggtgccaccactgtgccagccagtcctggcggtccccagggcccgatcaagagccaggacatccaaactaccacag
catcaacgccccggcctatactcgaaacccacttgacactctgcaatggatgggaaccacggggcagctctgtgtgggtcgcgcctat
cgcggctcggcggaagaccgggaaGGTACCCCGCTCCCGTCTGGTCTCACGTTCTGTGTACGGCCT
GGATCCCGGAAAGGGCGGATGCACGTGGTGTGCCCCGCCATTGGCGCCACGTTTC
AAAGTCCCCGGCCAGAAATGCACAGGACCGGCCCGGCTCGCACAGGCCATGACGAAT
GCCCAGATTTGACAGCAAAACAATCTGGAATAATCGCAACCATTCGCGTTTTGAACGA
AACGAAAAGACGCTGTTTAGCACGTTTCCGATATCGTGGGGGCCGAAGCATGATTGGG
GGGAGGAAAGCGTGGCCCCAAGGTAGCCATTCTGTGCCACACGCCGACGAGGACCA
ATCCCCGGCATCAGCCTTCATCGACGGCTGCGCCGCACATATAAAGCCGGACGCCTTC
CCGACACGTTCAAACAGTTTTATTTCTCTCCACTTCCTGAATCAAACAAATCTTCAAGGAA
GATCCTGCTCTTGAGCAACTAGTatgttcggttctacttcctgacggcctgcctctccctgaagggcggtgttcggc
gtctccccctctacaacggcctgggcctgacgccccagatgggctgggacaactggaacacgttcgctgcgacgtctccgagc
agctgctgtggacacggccgacccgatctccgacctgggcctgaaggacatgggctacaagtacatccttgacgactgct
ggctctccggccgactccgacggccttcctggctgccgacgagcagaagttccccaacggcatgggcccacgtcggcaccacc
tgcacaacaactcttctgttcggcatgtactctccgcggcgagtacacgtgcggcgtaccccggtccctgggcccgcgag

gaggaggacgccagttctcgcgaacaaccgcgtggactacctgaagtagacaactgctacaacaagggccagttcggcac
 gcccagatctctaccaccgctacaaaggccatgtccgacgccctgaacaagacgggcccacccatcttctactccctgtgcaac
 tggggccaggacctgaccttctacttggggctccggcctcgcgaactcctggcgcatgtccggcgacgtcacggcgagttcacgc
 gccccgactccccgcigccccigcgacggcgacgagtagactgcaagtagcggcgttccactgciccatcatgaacatcctgaa
 caagggcggccccatggggccagaacgcggcgctcggcggtggaacgacctggacaacctggagggtcggcgctggcaacct
 gacggacgacgaggagaaggcgcacttctccatgtggggccatggtgaagtcccccctgacatcggcggaacgtgaacaacc
 tgaaggcctctctactctactctactccagggcgtccgtcgcctcaaccaggactccaacggcaccgccacgcgcgtc
 tggcgctactacgtgtccgacacggacgagtagcggcaggcgagatccagatgtgttccggccccctggacaacggcgacca
 ggtcgtggcgctgtgaacggcggtcgtgtccggccccatgaacacgacctggaggagatcttcttgactccaaacctgggct
 ccaagaagctgacctccacctgggacatctacgacctgtggggcgaaccgcgtcgacaactccacggcgctccgcatcttgggccc
 gcaacaagaccgccaccggcactcgtgacaacgccaccgagcagtcctacaaggacggcctgtccaagaacgacacccgct
 gttcggccagaagatcggcctcctgtcccccaacgcgactcgtgaacacgacgcgtcccccacggcactcgcgttctaccgcctg
 cggccctctctctgaTACAACCTTATTACGTATTCTGACCGGCGCTGATGTGGCGCGGACGC
 CGTCGTAATCTTTTACAGACTTTACTCTTGAGGAATTGAACCTTTCTCGCTTGCTGGC
 ATGTAACATTTGGCGCAATTAATTGTGTGATGAAGAAAGGGTGGCACAAGATGG
 ATCGCGAATGTACGAGATCGACAACGATGGTGATTGTTATGAGGGGCCAAACCT
 GGCTCAATCTTGTGCGATGTCCGGCGCAATGTGATCCAGCGGCGTGACTCTCGCA
 ACCTGGTAGTGTGTGCGCACCGGGTCGCTTTGATTAAACTGATCGCATTGCCAT
 CCCGTCAACTCACAAGCCTACTCTAGCTCCCATTTGCGCACTCGGGCGCCCGGCTC
 GATCAATGTTCTGAGCGGAGGGCGAAGCGTCAGGAAATCGTCTCGGCAGCTGGA
 AGCGCATGGAATGCGGAGCGGAGATCGAATCAGGATCCCGCGTCTCGAACAGAG
 CGCGCAGAGGAACGCTGAAGGTCTCGCCTCTGTGCGCACCTCAGCGCGGCATACA
 CCACAATAACCACCTGACGAATGCGCTTGGTTCTTTCGTCCATTAGCGAAGCGTCC
 GGTTACACACAGTGCACAGTTGGCGAGGTGGCAGGTGACAATGATCGGTGGAGC
 TGATGGTTCGAAACGTTACAGCCTAGGGATATCGCCTGCTCAAGCGGGCGCTC
 AACATGCAGAGCGTCAGCGAGACGGGCTGTGGCGATCGCGAGACGGACGA
 GGCCGCCTCTGCCCTGTTTGAACCTGAGCGTCAGCGCTGGCTAAGGGGAGGG
 AGACTCATCCCCAGGCTCGCGCCAGGGCTCTGATCCCGTCTCGGGCGGTGA
 TCGGCGCGCATGACTACGACCCAACGACGTACGAGACTGATGTCGGTCCCCG
 ACGAGGAGCGCCGCGAGGCACTCCCGGGCCACCGACCATGTTTACACCGAC
 CGAAAGCACTCGCTCGTATCCATTCGCTGCGCCCGCACATGCATCATCTTTT
 GGTACCGACTTCGGTCTTGTTTTACCCCTACGACCTGCCTTCCAAGGTGTGA
 GCAACTCGCCCCGGACATGACCGAGGGTGATCATCCGGATCCCCAGGCCCA
 GCAGCCCCCTGCCAGAATGGCTCGCGCTTTCCAGCCTGCAGGCCCGTCTCCC
 AGGTCGACGCAACCTACATGACCACCCCAATCTGTCCAGACCCCAAAACACC
 CTCTTCCCTGCTTCTCTGTGATCGCTGATCAGCAACACATatggettccggcgattca
ccatgtcggcggtcccccgcatgactggcagggccccctggggcagctcgtcggcagggcagtcgccaccgccctgagggg
ctccaccttcagtgctggtagctctctacatcgacctgcaaccagttctctctcgcctccctgtcttctggggcgaca
acggcttcgctccctgttgggtccaaagcccttcgctccaaaccggcgccaccgcgctggggcgcgctcccaactcggcg
aggccatggcggtggccctggagcccgccaggaggtggccaccaagaagaagccccctggtgaagcagcgccgctggtg
gtgaccggcatggcggtgtagccccctggggcagagcccgacgtgtaactacaacaacctgtggacggcggtgtcggga
tctccagatcgaggccttcgactgcaaccagttccccaccgcacgcggcgagatcaagtccttctccaccgacggctggg
tggcccccaagctgtccaagcgatggacaagttcatgtgtacctgtgacccgggcaagaagggcctggcgacggcg
catcaccgacgacgtgatgaaggagctggacaagcgcaagtgcggcggtgctgatcggtccggcctggggcgcatgaagct

gttctccgaactccatcgaggccctgccaatctcctacaagaagatgaaccccttctgctgccccttccgacaccaccaacatggg
ctccgcaatgctggccatggacctgggctggatgggcccactactccatctccaccgctgcccacctccaacttctgcatc
ctgaactccgccaaccaatcgctgcccggcgaggccgacatgatgctgtgcccggctccgacgcccgtgacatccccatcgg
cctggggcggttctgctggcctgcccgcctgtccagcgcaacaacgaccccaccaagggctcccgcccctgggaactccaacc
ggcagggcttctgctgctggcgaggcgccggcgctgctgctgctggaggagctggagcagccaagaagcgccggcgccacc
atctacgcccagttctggggcggtcttcaactgcccagcctaccacatgaccgagccccaccccaggggcgccggcgctgat
cctgtgcatcgagaaggccctggcccaggccggcgctgctccgagaggacgtgaactacatcaaccccacgcccacctccacc
cccggcgcgacatcaaggagtaccaggccctggcccactgcttggccagaaactccgagctgcgctgaactccaacagt
ccatgatcgccacatgatcgccgcccggcgccggcgctggaggccgtgaccgtgggtgacggccatccgacccggctggatcca
ccccaaactgaacctggaggaccccagacaaggccgtggagcgaagggtgctgggtgggcccgaagaaggagcgctgaacg
tgaagggtgggctgtccaaactccttggcttggcgccacaaactcctccatcctgttgccecttacaacacatgtaccctta
cgacgtgcccgaactacgctgaTATCGAGGcagcagcagctcgatagatcgacacactctggacgctggctgctga
tggactgttccgcccacacttctgctgcttgaactgtgaataccttgcgcttttatcaaacagcctcagtgcttctgctgct
gtacgcgcttttgcgagttgctagctgcttctgctatcttgcgaataccacccccagcctcccttccctgcttcatatcgcttgc
ccccacccgaacttatctacgctgcttctgctatccctcagcgctgctcctgctcctgctcactgcccctcgcacagccttgggttgg
gctccgctgtattctctgctgactgcaacctgtaaaccagcactgcaatgctgatgcacgggaagtagtgggatgggaacac
aatggaAAGCTTGAGCTCagcgccgacggctcgtgctaccgtacgacgttggcgacgcccataaagtgttataccga
gctgttgagcgaactgaagcgccgctcaaggatactgaactcctggattgatatcggtccaataatggatggaaaatccgaacctc
gtgcaagaactgagcaaacctcgttaccatggatgcacagtcgccagtcgaatgaacattgaagtgcgaactgttgccttccgtggc
agtactactcaagaatgagctgctgttaaaaatgcactctcgttctctcaagtgcagtgagtgagatgagtgctacgcccgtgaccttgcctg
cccgtgctatgcctgcccccaaaatttgaaaaaaggatgagattattgggcaatggacgacgtcgtcgtccgggagtcaggac
cgggggaaaataagaggcaacacactccgttctta

SEQ ID NO: 39

KAS IV de *Cuphea palustris* con optimización de codones para *Prototheca*

5

atggcttccgcgccallcaccatgctggcgtgccccgcgatgactggcagggcccttggggcacgtcgtccggacggccagtcgc
caccggcctgaggggctccaccttccagtgccctggtgacctctacatcgacccctgcaaccagttctctctccgctccctgctctt
cctggcgacaaaggcttgcctccctgttggctccaagcccttccgctccaaccggcgccaccgcccgtggggcgccctccc
actccggcgaggccatggcgtgcccctggagcccggcaggaggtggccaccaagaagaagccctggtgaagcagcgccgc
gtggtggtgacggccatggcggtggtgacccccctgggccaagagcccagcgtgactacaacaacctgctggacggcggtgctcg
gcatctccgagatcgaggccttgcactgcaccagttcccaccggcagtcgcggcgagatcaagtcttctccaccgacggctggg
tggccccaaagctgtccaagcgcatggacaagttcatgctgtacctgctgaccgcccggcaagaaggccctggccgacggcgccatc
accgacgacgtgatgaaggagctggacaagcgcaagtgccgctgctgacggctccggcctggcgccatgaagctgttctccg
actccatcgaggccctgcgcatctctacaagaagatgaaccccttctgctgccccttcccaccaccaacatgggctccgcatgct
ggccatggacctgggctggatgggccccaaactactccatctccaccgctgcccacactccaacttctgcatctgaactccgccaac
cacatgctgcggcgaggccgacatgatgctgtgcggcggtccgacgcccgtgatcccccacggcctggcgcgcttctgtggc
ctgcggcgccctgtccagcgcaacaacgaccccaccaaggcctccggcccctgggactccaaccgacggcttctgtatgggc
gaggcgccggcgctgctgctgctggaggagctggagcagccaagaagcgcgggcgccaccatctacgccgagttcctggcgccg
tcttcaactgagacgctaccacatgaccgagccccaccccaggggcgccggcgctgacctgtgcatcgagaaggccctggccca
ggccggcgctgtcccgcgaggacgtgaactacatcaacgcccacgcccactccacccccggcgacatcaaggagtagccagggc
cctggcccactgttctggccagaactccgagctgcgctgaactccaccaagttccatgatcgccacctgatcgccgcccggcg
gctggaggccgtgacctggtgagggccatccgacccggctggatccacccaacctgaacctggaggacccccgacaaggccg

tggacgccaaggtgctggtgggccccgaagaaggagcgctgaacgtgaaggtgggctgtccaactccttccgcttccggcgccac
aacctccatctgttgccecttacaacacatgtacctacgacgtgcccgaactacgctga

SEQ ID NO: 40

KASIV de *C. camphora* con optimización de codones para *Prototheca*. Secuencia de nucleótidos del vector de expresión de KASIV de *C. camphora* (D3147, pSZ4338). Solo se muestra la secuencia de KASIV de *C. camphora*

10

con optimización de codones; el promotor, la 3'UTR, el marcador de selección y los brazos dirigidos son iguales que en la SEQ ID NO: 38.

atggccatgatggcggctcctgctccaacctgggtgatcggaaccgcgagctggcggaacggccctccctgctgactacaa
cgccctgcgccccctggagaacatccagaccgcctccgcccgtgaagaagcccaacggcctgttcgctcctccaccgcccgaagt
ccaagggcgtgcgcccctgggtgctgcccaccgtgaccgcccccaagcgcgagaaggaccccaagaagcgcatcgatcaccg
gcatggcctggtgtccgtgttcggcaacgacatcgacaccttctactccaagctgctggaggcgagtcggcatcgcccccatcg
accgcttcgacgcctcctccttctccgtgcgcttcgcccggccagatccacaacttctcctccaagggtacatcgacggcaagaacga
ccgccgctggagcactgctggcgctactgctgtggcgccggccgcccctggaggacgccaacctgggccccgaggtgct
ggagaagatggaccgctcccgatcggtgctgacggcaccggcatggcgggcctgtccgcttctccaacggcggtggagtcce
tgatccagaagggtacagaagatcccccttcttctactccctactccatccaacatgggtccgcccctgctggccatcgacacc
ggcggtgatggccccaaactactccatcctaccgctgcccaccgccaactactgcttccacggccggcccaaccacatccgcgc
ggcgaggccgagatcatggtgaccggcgccaccgaggccgctgtccgccaccggcggtggcggttcacgctgcccgcgc
ctgtcccaccgcaacgacgagccccagaccgctcccgcccctgggacaaggaccgagcggttcgtgatggcgagggcgcc
ggcggtgctgggtgatggagtcctgcaccacggccgaagcgccggcccaacatcatcgccgagtagctggcgccgcccgtgacct
gcgacgcccaccacatgaccgacccccgcgcgacggcctggcggtgtctcctcctgcatcacaagtcctggaggacggcggt
gtcccccgaggaggtgaactacgtgaacggccacggccacctccacctggcgccgacgtggcgaggtgaacggcatcaagaa
ggtgttcaggacacctccgagatgaagatgaacggcaccagtcctgatcgccactgcctggcgccggcgccggtgag
ggcatcgccaccatcaaggccatcaacaccggctggtgcacccaccatcaaccagttcaacatcgagcccgccgtgaccatcga
caccgtgcccacgtgaagaagaagcagacatccacgtggcgatcctccaactccttggcttcggcgcccaactccgtgggtgt
gttcgcccccttcatgcccaccatgtacccctacgacgtgcccactacgcctga

5

SEQ ID NO: 41

KASI de *C. camphora* (D3148; pSZ4339) con optimización de codones para *Prototheca*

atgcagatcctgcagacccccctcctcctcctcctcctcctgcgatgtcctccatggagtcctgtccctgacccccaaagtcctgccc
ctgaagacctgtgccccctgcgccccgccccagaacctgtcccgccgaagtcccagaacccccgccccatcctcctcctc
tccccgagcgcgagaccgaccccaagaagcgctggtgatcaccggcatggcctggtgttcgtgttcggcaacgacgtggacg
cctactacgaccgctgctgtccggcgagtcggcatcgccccatcgaccgttcgacgcctccaagttccccaccgcttcgcg
ggcagatccgcggttcacctccgacggctacatcgacggcaagaacgacggccgctggacgactgctgctactgcatcggt
tcgggcaagaaggccctggagaacggcgccctgggccccacctgatggacggcaagatcgacaaggagcgcgccggcggtgct
ggtgggacggcgatggcgccgctgaccgtgttctccaacggcggtgcagacctgcacgagaagggtaccgcaagatgaccccc
tttctacccccctacgccatccaacatgggtccgcccctgctggccatcgagctgggttcctgggccccaaactactccatcctacc
gctgcccacctccaactactgcttctacggcgccgccaaccacatccggcgccgagggcgacctgatgctggcgccgggca
ccgagggcccatcatccccatcgccctggcgccgttcgtgacctgcgcgcccctgtcccaaggcaacgacgacccccagaccgc
ctcccgcccctgggacaaggaccgacgggttcgtgatggcgagggcgccggcggtgctgggtgatggagtcctggagcagc
catgaagcgcgacgccccatcatcgccgagtagctggcgccgctgaactgcgacgctaccatgacggacccccgcgc
gacggcctggcggtgtccacctgcatcgagcgtccctggaggacggcggtggcccccgaggaggtgaactacatcaacggcc
acggcacctccacctggcgccgacgtggcgaggtgaacggcatcaagaagggtgttcaccaacacctccgagatcaagatcaac

gccaccaagtcctatgatcgccactgcctggcgccgcccggcgccctggaggccatcgccaccatcaaggccatcaacaccggct
gggtgcacccccctcatcaaccagttcaaccccgagccccctgtggagttcgacacctgggcaacaagaagcagcagcaggggtg
aacgtggccatcctcaactccttcggcttcggcgcccaactccgtggtggtgttctccgcttcaagcccaccatgtaccctacga
cgtgccccactacgcctga

10

SEQ ID NO: 42

KASI de *U. californica*

KASI de *U. californica* (D3150, pSZ4341) con optimización de codones para *Prototheca*

atggagtcctgtccctgacccccaaagtcctgccccgaagacctgtgccccccgccccgccccaaagacctgtcccgccgc
aagtcaccagaacccccagcccatctcctcctcctcctccccgagcgcgagaccgacccccagaagcgctggatgacacggcat
gggctgtgtgtccgtgttcggcaacgacgtggacgctactacgacgcctgtgtccggcgagtcggcatgcccccatgacc
gttcgacgctccaagtccccaccgcttcgcccggccagatccggcgcttccctccgacggctacatcgacggcaagaacgac
gcccctggacgactgctgctactgcatcgtgtccggcaagaagggcctggagaacgcccgtggccccgacctgatgga
cggcaagatcgacaaggagcgccggcggtgtgtgtggcgacggcgatggcgggcctgacctgtttcccaacggcggtgcagac
cctgcacgagaagggtaccgcaagatgaccccccttttcatccctacgccatcaccacatgggctccgcccctgctggccatcgac
ctgggttcctggtggccccaaactactccatctccaccgctgcccacctccaactactgtttctacggcgcccaaccacatccgc
ggcgcgaggccgacgtgatgtgtggcgccggcgacccgagggcccatcatccccatcgccctggcgccgttcgtggcctgcccgc
ccctgtcccagcgcaacgacgacccccagaccgctcccggccctgggacaaggaccgagcgacggcttcgtgatggcgaggcg
ccggcggtgtgtgtgatggagtcctggagcacgccatgaagcgcgacgccccatcatcgccgagtagctggcgccgctgaa
ctgcgacgctaccacatgaccgacccccgcccgcacggcctggcggtgttccacctgcatcgagcgctccctggaggacggcg
gtggccccgaggaggtgaactacatcaacgccacgccacctccacctggcgccgacctggccgaggtgaacgccatcaag
aggtgttcaccaaacctccgagatcaagatcaacgccaccaagtcctgatcgccactgctggcgccgcccggcgccctggag
gccatcgccacatcaaggccatcaacaccggctggctgcacctccatcaaccagttcaaccccgagccctcgtggagttcgac
acctgtggcaacaagaagcagcagcagaggtgaacgtggccatctccaactccttcggcttcggcgccacaactccgtgggtgt
gttctccgcttcaagcccacatgtacccctacgacgtgcccactacgcctga

SEQ ID NO: 43

KASIV de *U. californica* (D3152, pSZ4343) con optimización de codones para *Prototheca*

5

atgaccagacctgatctgccccctcctccatggagacctgtccctgaccaagcagtccttccgctgcccctgcccaccccc
cccacatccgcccggcgccggccaccgccaccccccccccttcatctccgctccgcccggcgagaccgacccca
agaagcgctggatgacacggcatgggctgtgttcggcaccacgtggacgtgtactacgacgcctgtgtggcgcc
gagtcggcggtggcgaccatcgaccgttcgacgctccatgttccccaccgcttcggcgccagatccgcttccctccgag
ggctacatcgacggcaagaacgaccggcctggacgactacctgcgctactgcctgggttcggcaagaagccatcgatccg
ccggcttcgacctgcaaacatcaccacaagatcgacaaggagcgccggcgcacccctgtgtgggtccggcatggcgccctgaa
ggtgttctccgacggcggtgagtcctgatcgagaagggtaccgcaagatcccccttttcatccctacatgatccccaacatgg
gtccgcccctgtgggcatcgacctgggcttcctgggccccaaactactcctccaccgctgcccacctccaactactgcatctac
ggcgccgccaaccacatccgccaggcgacggcgacctgatgtgtggcgccggcgaccgagggccccatcatcccatcgccctg
ggcggttcgtggcctgcccggccctgtccacccgcaacgacgacccccagaccgctcccggccctgggacatcgaccgcgacg
gttcgtgatggcgaggcgccggcctcctgtgtgtggagtcctggagcagccatgaagcgcgacgccccatcctggccga
gtacctggcgccggcggtgaactgcgacgccaccacatgaccgacccccgcccgacggcctggcggtgtccacctgcatcgag
tctccctggaggacggcgcggtggcgccgagggaggtgaactacatcaacgccacgccacctccacccccaccggcgacctgg
ccgagatgaaggccatcaagaacgtgttccgcaaacctccgagatcaagatcaacgccaccaagtcctgatcgccactgctgg
ggcctccggcgccctggaggccatcgccacctgaaggccatcaccaccggctggctgcaccccaccatcaaccagttcaaccc

cgagccctccgtggacttcgacacgtggccaagaagaagcagcagcagaggtgaacgtggccatctccaactccttcggcttcg
ggcgccacaactccgtgtgtgtgttccgcttcaagcccaccatgtacccctacgacgtgcccactacgcctga

SEQ ID NO: 44

KASAI de *C. wrightii* (D3153, pSZ4379) con optimización de codones para *Prototheca*

10

15

20

atggcttccgcggcattaccattgctggcgctgccccgcgatgactggcaggggccccctggggcacgtcgtccggagcgccagtcgc
caccgcctgaggtacgtattccagtgccctgggtggccagctgcatcgacccctgcgaccagtlaccgcagcagcgccagcctgagctt
cctggggcgacaacggcttcgccagcctgttcggcagcaagccccctcatgagcaaccgcggccaccggccctgcggcgcccgacc
cacagcggcgagggccatggcgtggtgcccgtgcagccccccaggaggccggcaccagaagaagccccgtgatcaagcagcgcc
gcgtggtggtgaccggcatggggcggtggtgacccccctggggccacgagcccgacgtgttctacaacaacctgctggacggcggtgag
cgggcatcagcgagatcgagaccttcgactgcaaccagttccccaccggcatcgccggcgagatcaagagcttcagcaccgacggct
gggtggcccccaagctgagcaagcgcatggacaagttcatgctgtacctgctgaccgcgggcaagaaggccctggccgacggcg
cataccgacgaggtgatgaaggagctggacaagcgcaagtgccggctgctgcatcgccagcgggcatggcgcatgaaggtgttc
aacgacgccatcgaggccctgcggctgagctacaagaagatgaaccccccttgcgtgccccttcgccaccaccaacatgggcagcgc
catgttggccatggacctgggctggatggggccccaaclacagcatcagcacccgctgcggccaccagcaacttctgcatctgaacg
cgccaaccacatcatccgcggcgagggcgacatgatgtgtgcggcggcagcgacggcgtgatcatccccatggcctggggggc
ttcgtggcctgcgcgcccgtgagccagcgcaacagcgacccccaccaaggccagccgcccctgggacagcaaccgcgacggcttc
gtgatggggcgagggcgccggcggtgctgctgctggaggagctggagcacgccaaagaagcgcgggcgccaccatctacggcgagttc
ctggggcggcagcttcacctgcgacgcctaccacatgaccgagccccacccgaggggcgccggcggtgatcctgtgcatcgagaagg
ccctggcccaggccggcggtgagcaaggaggacgtgaactacatcaacgccaccgcaccagcaccagcgccggcgacatcaag
gagtaccaggccctggcccgtgcttcggccagaaacagcgagctgcgcgtgaacagcaccagaagagcatgatcgccacctgctgg
gcggcgccggcgcggtggaggccgtgacctgggtgcaggccatccgcaccggctggattaccccaacctgaacctggaggacc
ccgacaaggccgtggacgccaaagctgctggtggggccccaaagaaggagcgccctgaacgtgaaggtgggcctgagcaacagcttcg
gttcggcgcccaacagcagcatcctgttcgccccctgcaacgtgtga

SEQ ID NO: 45

KASIVb de *C. avigera* (D3287, pSZ4453) con optimización de codones para *Prototheca*

5

atggtctccgcggcattaccatgttcggcgtgccccgcgatgactggcagggccccctggggcacgtctgcctcggacggccagtcgc
caccgcctgagggggtccaccttcacgtgtctacatcggcgacaacggcttcggctccaagcccccccgctccaaccggcgccacc
tgcccttggggccgcacctcccactccggcgaggtgatggcgttggcatgcagtcggccaggaggtgtccaccaaggagaagcc
cgccaccaagcagcgcccgctgtgtgtgaccggcatggcggtgtgtgaccgccttgggccacgaccccgacgtgtactacaacaa
cctgtctggacggcggttccggcatctccgagatcgagaattcgcactgtcccgacttggccaccggcctcggcgagatcaagtc
cttctccgccgacggctgggtggcccccaagtctcccgccgcattggacaagttcatgtgtacatcctgaccggcggaagaaggc
cctgttggacggcgcatcaccgaggacgtgatgaaggagctggacaagcgcaagtgcggcggtgtgctgacggctccggcctggg
cggcataaggtgttctccgagtcacatcgaggcccttgcgcacctctacaagaagatcccccttctgcgtgccccttctccaccacca
acatgggctccggcctcgtggccatggacctgggcttgatgggccccaaactactccatctccaccgcctgcgccacctccaactctg
catctgaacggcgccaaccacatcaccaaggcgaggccgacatgatgtgtgcggcggtctccgactccgtgatcttgcctatcg
gcatggcggttctgtgcttgcgcgccctgtcccagcgcaacaacgacccccaccaaggcctccggccccctgggactccaaccg
cgacggcttctgtatggcgaggggcgccggcggtgtgtgtgtgtggaggagctggagcacgccaagaagcgcgggcgccaccatcta
cgccgagttcctggggcggtcttccacttgcgacgcctaccacatgaccgagccccaccccgagggcgccggcggtgatcctgtgca
tcgagaaggcccttggcccagtcggcggtgtcccgcgaggacgtgaactacatcaacggccacggccacctccacccccggcgcgga
catcaaggagtaccaggcccttggccactgttctggccagaactccgagctgcgcgtgaactccaccaagtcctgatcggccacct
gtctggcgcgccggcgccggtgtggaggccgtgacctgtgtgtgacggccatccgcaccggctgtgatcccccacatcaacctgga

cgaccccgacgagggcgtggacgccaagctgctgggtgggcccgaagaaggagaagctgaaggfgaaggtgggcctgtccaactc
cttcggcttcggcgccacaaactctccatctcttcgccccctgcaacaccatgtacctctacgactgccccactacgcctga

SEQ ID NO: 46

KASIVb de *C. paucipetala* con optimización de codones para *Prototheca*

10

15

atggcttccgcggcattaccatgtcggcgtgccccgcgatgactggcagggccccctggggcacgtcgtccggacggccagtcgc
caccggcctgaggggtccacctccagtgcttggcgacatggcttccctcctgatcggctccaagccccccgtccaaccg
caaccaccggcgtggggcgcacctcccaactccggcgagggtgatggcgtggccatgcagccggccacgagggcctccacca
gaacaagcccgtagcaagcagcggcgtggtggtgaccggcatggcggtggccacccccctggggcacgaccccgacgtgta
ctacaacaacctgctggacggcgtgctcggcatctccagatcgagaacttcgactgcacccagttccccacccgcatcgccggcga
gatcaagtcttctccaccgggggtacgtgacccccaaagtgcgaagcgcagtgacaagtcatgctgtacctgctgaccggcggc
aagaaggccctggaggacggcgccatcaccgaggacgtgatgaaggagctggacaagcgcaagtgcggcgtgctgaltcggctcc
ggcatggggcgcatgaagatcatcaacgactccatcgccgcctgaacgtgtctacaagaagatgaccccccttctgctgccccttct
ccaccaccaacatgggctccggcatgctggccatcgacctgggctggtatgggccccaaactactccatctccaccgctgcgccacct
ccaactactgcatctgaacggcgccaaccacatcgtgcggcgaggccgacatgatgctgtcggcggtccgacggcgtgac
atccccgtgggctggcggttctggtgctgcccgcgcctgtcccagcgcaacaacgaccccccaaggcctcccgccccctggg
actccaaccggcagcgttctgtagggcgaggcgccggcgtgctgctgctggaggagctggagcacgccaagaagcgcgggcg
ccaccatctacggcagttcttggcggtccttccatcgacgcctaccacatgaccgagccccaccccgacggcgccggcgtg
atcctgtgcatcgagaaggccctggcccagtcggcggtgctccggcgaggacgtgaactacatcaacggccacgcccactccacccc
cgccggcgacatcaaggagtaccagccctggcccactgcttcggccagaactccgagctgcgcgtgaactccaccaagtcctga
tcggccacctgctggcgccggcgccggcggtgaggccgtgaccgtggtgcaggccatccgcaccggctggatccacccaacat
caacctggagaaccccagcaggcggtggacgccaagctgctggtgggccccagaaggagaagctgaaggtgaaggtggcgct
gtccaactcttctggcttggcgggccacaactctccatctgttgcgccctacaacaccatgacctctacgacgtgcccgaactacg
cctga

SEQ ID NO: 47

KASIVb de *C. ignea* (D3289, pSZ4455) con optimización de codones para *Prototheca*

5

atggcttccgcggcattaccatgtcggcgtgccccgcgatgactggcagggccccctggggcacgtcgtccggacggccagtcgc
caccggcctgaggggtccacctccagtgcttggcgacatggcttccctcctgatcggctccaagccccccgtccaaccg
tcttggggcgaacaaggcttccctcctgttgcggctccaaagcccttccgctccaaccggcgccaccggcctggggcgccctccc
actccggcgaggccatggcgtgcccctgcagccggccaggaggtgaccaccaagaagaagcccgtagcaagcagcggcg
gtggtggtgaccggcatggcggtggtgacccccctgggcaagcggcggcgtgactacaacaacctgctggacggcgtgtccg
gcatctccgagatcgagaccttcgactgcacccagttccccacccgcatcgccggcgagatcaagtccttctccaccgacggctgggt
ggcccccaagctgtccaagcgcagtgacaagttcatgctgtacctgctgaccggcggaagaaggccctggccgacggcggtc
accgacgacgtgatgaaggagctggacaagcgcaagtgcggcggtgctgacggctccggcatggcggtgatgaagctgttcaacg
actccatcgaggccctgcgcctcctacaagaagatgaaccccttctgctgccccttgcgccaccaccaacatgggctccgcatgct
ggccatggacctggcgtggtgggccccaaactactccatctccaccgctgcgccacctccaacttctgcatctgaacgcctccaac
cacatcgtgcggcgaggccgacatgatgctgtgcggcgctccgactccgtgaccgtgccccctggcggtggcggttctggtg
ctgcggcgccctgtcccagcgcaacaacgacccccaccaaggcctcccgccccctgggactccaaccggcagggcttctgtagggc
gagggcgccggcgtgctgctgctggaggagctggagcacgccaagaagcgcgccgccaccatctacggcagttcttggcgggc
tcttccactccgacgctaccacatgaccgagccccaccccgaggcgccggcggtgaltctgtgcatcgagaaggccctggccca
gtccggcggtgtcccgaggacgtgaactacatcaacggccacgcccactccacccccggcgacatcaaggagtaccaggcc
ctggccccgtgttctggccagaactccgagctgcgcgtgaactccaccaagttcatgatggccacctgtggggcgccggcgcg

cgtggaggccgtggcgtgatccaggccatccgacccggctggatccaccccaacatcaacctggaggaccccgacgaggccgtg
gaccccaagctgtgtggggcccaagaaggagaagctgaaggtgaaggtggccctgtccaactcttctggcttggcgggccacaa
ctctccatctgttgcgccctgcaacaccatgtacctctacgacgtgcccgaactacgcctga

SEQ ID NO: 48

KASIV de *Cuphea procumbens* (D3290, pSZ4456) con optimización de codones para *Prototheca*

10

atggcttccgcggcattaccatgtcggcgtgccccgcgatgactggcagggccccctggggcacgtcgtctccggacggccagtcgc
caccgcctgaggggtccaccttccagtgcctggtgacctcccaaacgaccttgcaccagtaactgtctctccgctccctgtcc
ttctggggcgacaacggcttcggctccaagcccttccgctccaaccggccacggccgctggggcggcctcccaactccggcga
ggccatggcgtggccctgcagcccgccaggaggtggccaccaagaagaagcccgccatgaagcagcggcggcgtggtggtga
ccggcatggcggtggtgacccccctgggccacgagcccgacgtgtactacaacaacctgctggacggcgtgtccggcatctccgag
atcgagaccttcgactgcaccagttccccaccggcgcggcgagatcaagtccttccaccgacggctgggtggcccccaag
ctgtccaagcgcattggacaagttcatgtgtacctgtgaccggcgcaagaaggccccctggccgacggcgccatcaccgacgacgt
gatgaaggagctggacaagcgaagtgcggcggtgtgacggctccggcatggcgggcatgaagctgttcaacgactccatcgag
gccttgcgcgtgtcttacaagaagatgaaccttcttgcgtgccccttgcaccaccaacatgggctccgcatgtgtgcatggacc
tgggctggatggggcccccaactactccatctccaccggcctgcccacctccaaacttctgcatctgaacggcgccaaccacatcgtcgc
cggcgagggcgacatgatgtgtgcggcggtccgacggcgtgacatccccatggcctggcggttctgtggcctgccgcgcc
tgtccacgcgaacaacgacccaccaaggcctcccgccccctgggactccaaccggacggcttctgtatggcgaggcgccg
gctgtctgtctggaggagctggagcacgccaagaagcgcgggcgccaccatctacggcagttcttggcggtcccttccactgc
gacgctaccacatgaccgagccccaccggagggcgccggcgtgacatccccatggcctggcggttctgtggcctgccgcgcc
tgtccacgcgaacaacgacccaccaaggcctcccgccccctgggactccaaccggacggcttctgtatggcgaggcgccg
gctgtctgtctggaggagctggagcacgccaagaagcgcgggcgccaccatctacggcagttcttggcggtcccttccactgc
gacgctaccacatgaccgagccccaccggagggcgccggcgtgacatccccatggcctggcggttctgtggcctgccgcgcc
cttccggcagaactccgagctgcgcgtgaactccaccaagtcctgatcgccacctgctggcgccggcgccggcggtggaggcc
gtgacgctgatccagggcatccgacggcgtggatccaccaacctgaacctggaggaccccgacaaggcgtggacgccaagt
tcttgggtggggcccccaagaaggagcgctgaacgtgaagggtggcctgtccaactcttctggcttccggcgccacaactctccatcc
tgttcccccttgaacaccatgtaccttaccgacgtgcccgaactacgctga

SEQ ID NO: 49

KASIVa de *C. paucipetala* (D3291, pSZ4457) con optimización de codones para *Prototheca*

5

atggcttccgcggcattaccatgtcggcgtgccccgcgatgactggcagggccccctggggcacgtcgtctccggacggccagtcgc
caccgcctgaggggtccaccttccagtgcctggtgacctcccaacatgaccttgcaccagaacgtgtctctccgctccctgtcc
ttctggggcgacaacggcttcggctccaaccccttccgctccaaccggccacggccgctggggcggcctcccaactccggcga
ggccatggcgtggccctgcagcccgccaggaggtggccaccaagaagaagcccgccatcaagcagcggcggcgtggtggtga
ccggcatggcggtggtgacccccctgggccacgagcccgacgtgttctacaacaacctgctggacggcgtgtccggcatctccgag
atcgagaccttcgactgcaccagttccccaccggcgcggcgagatcaagtccttccaccgacggctgggtggcccccaag
ctgtccaagcgcattggacaagttcatgtgtacctgtgaccggcggaagaaggccccctggccgacggcgccatcaccgagacgt
gatgaaggagctggacaagcgaagtgcggcggtgtgacggctccggcatggcgggcatgaagctgttcaacgactccatcgag
gccttgcgcgtgtcttacaagaagatgaaccttcttgcgtgccccttgcaccaccaacatgggctccgcatgtgtgcatggacc
tgggctggatggggcccccaactactccatctccaccggcctgcccacctccaaacttctgcatctgaacggcgccaaccacatccg
cggcgagggcgacatgatgtgtgcggcggtccgacggcgtgacatccccatggcctggcggttctgtggcctgccgcgcc
tgtccacgcgaactccgacccaccaaggcctcccgccccctgggactccaaccggacggcttctgtatggcgaggcgccg
cgtgtctgtctggaggagctggagcacgccaagaagcgcgggcgccaccatctacggcagttcttggcggtcccttccactgc
acgctaccacatgaccgagccccaccggacggcgccggcgtgacatccccatggcctggcggttctgtggcctgccgcgcc
ccggcaggacgtgaactacatcaacgcccacgcccaccccccggcgacatcaaggagtaaccaggccccctggccccactgc

ttcggccagaactccgagctgcgcgtgaactccaccaagtcctgatcgccacctgctggcgccggcgccggcggtggaggccg
tgaccgtgatccagggcatccgacccggctggatccaccaacctgaacctggaggaccccgacgagggcgtggacgccaagtt
cctggtggggcccccaagaaggagcgctgaacgtgaagggtggcctgtccaactcttctggcttccggcgccacaactctccatcc
gttcccccttacaacaccatgtaccttaccgacgtgcccgaactacgctga

SEQ ID NO: 50

KASIV de *Cuphea painteri* (D3292, pSZ4458) con optimización de codones para *Prototheca*

10

15

atggttccgcggcattcaccatgtcggcgtgccccgcgatgactggcaggccccctggggcacgtcgtccggacggccagtcgc
caccgcctgaggggtccacccccagtgccctggaccctgcaaccagcactgcttctgggcgacaacggcttcgctccctgat
cggctccaagccccccgctccaacctggggccacctggcctggggcgacctcccactccggcgagggtgatggcctggcccag
gagggtgtccaccaacaagaagcacgccaccaagcagcgccgcgtggtggtgaccggcatggcggtggtgacccccctggggcac
gaccccgacgtgactacaacaacctgctggaggcggtgtcggcgtatccgagatcgagaacttcactgctccagctgccacc
cgcacgcggcgagatcaagtcttctccaccgacggcctggtggcccccaagctgtccaagcgcagtgacaagttcatgctgtac
atcctgaccggcggaagaagggcctggccgacggcgccatcaccgaggacgtgatgaaggagctggacaagcgcaagtggg
cgtgctgatcggctccggcctggggcgcatgaaggtgttctccgactcctggaggccctgcgcacgtccclacaagaagatcccc
ttctgctgccccttctccaccacaacatgggctccgcatgtcggcatggacctgggctggatgggcccactactccatccac
cggctggccacctccaacttctgcatcctgaacggcgccaaccacatcacaaggcgaggccgacatgatgctgtggcgggct
ccgacggcgccatcctgccatcggcgatggggcggttcgtggcctgcgcgcctgtcccagcgcaaacgaccccaccaaggc
ctccccccctgggactccaaccgcgacggcttcgtgatggcgaggcgccggcgctgctgctgctggaggagctggagcacgcc
aagaagcgcgggccaccatctacccgagttcctggcggtccttccactgcgacgctaccacatgaccgagccccaccgga
cggcgccggcgtgatcctgtgcatcgagaaggccctggcccagtcggcgctgtcccgcgaggaggtgaactacatcaacggccac
ggcactccacccccggcgacatcaaggagatcaccagccctggcccactgcttcggccagaactccgagctgcgcgtgaactc
caccagtcctgatcggccacctgtcggcgggcgccggcggtggaggccgtgaccgtggtgagggccatccgcaccggctg
gatecccccacatcaacctggaggaccccgacaaggcggtggacgccaagctgctggtgggcccccaagaaggagaagctgaa
ggtgaaggtggcctgtccaactccttggcttcggcgccacaactctccatcctgttcgccccctgcaacaccatglaccctacg
acgtgccccgactacgctga

SEQ ID NO: 51

KASIVa de *C. avigera* (D3293, pSZ4459) con optimización de codones para *Prototheca*

5

atggttccgcggcattcaccatgtcggcgtgccccgcgatgactggcaggccccctggggcacgtcgtccggacggccagtcgc
caccgcctgaggggtccaccttccagtgccctggtgacctctacaacgacctctgcgagcagtlaccgctcctccgctccctgtcc
ttctgggggacaacggcttcgctccctgttcggctccaagccttccgctccaaccggggccaccggcctggggcgccctcc
cactccggcgaggccatggcctggtccctgcagcccgccaggaggtgggccaagaagaagcccgatcaagcagcgccg
cgtggtggtgacggcgatggcggtggtgacccccctggggcagagcccgacgtgactacaacaacctgtcggacggcggttccg
gcatctccgagatcgagaccttgcactgcacccagttccccaccgcatcgccggcgagatcaagttcttctccaccgacggctgggt
ggcccccaagctgtccaagcgcagtgacaagttcatgctgtacctgtgaccggcggaagaaggccctggccgacggcgcatc
accgacgacgtgatgaaggagctggacaagcgcaagtgcggcggtgctgatcggctccggcctggcgcgcatgaaggtgttctccg
agtcacatgaggccctgcgcacctctacaagaagatctcccccttctgctgcttctccaccaccaacatgggctccgccatcctg
gccatggacctgggtggtggtggcccccaactactccatctccaccgctgcgccactccaacttctgcatcctgaacggcgccaacc
acatcacaaggcgaggccgacatgatgctgtgcccgggctccgactccgtgatcctgcccatcggcagtgggcggttcgtggcct
ggcgcgccctgtcccagcgcaacaacgaccccaccaaggcctcccgccccctgggactccaaccgcgacggcttcgtgatggcgga
ggggcgccggcggtgctgctgctggaggagctggagcacgccaagaagcgcgggcgccaccatctacggcagttcctggggcgctc
cttcacctgcgacgctaccacatgaccgagccccccccgaggggcgccggcggtgatcctgtgcatcgagaaggccctggcccagt

ccggcggttcccgcgaggacgtgaactacatcaacggcccacggccactccacccccggcgcgacatcaaggagtlaccaggccct
ggcccactgcttcggccagaactccgagctgcgcgtgaactccaccaagtcctatgatcgccacctgctggggcgccggcgggc
gtggaggccgtgacgtggtgagggccatccgcaccggctggatccaccccaacatcaacctggacgaccccgacgaggcggtg
gacgccaagctgctggtggggcccccaagaaggagaagctgaaggtgaaggtgggctgtccaactccttccggttcggcgccaca
actctccatcctgttcgccccctgcaacaccatgtaccctacgacgtgcccactacgctga

SEQ ID NO: 52

KASIVa de *C. ignea* (D3294, pSZ4460) con optimización de codones para *Prototheca*

10

atggcttcccgccgcatccaccatgctggcgctgccccgcgatgactggcagggcccccctggggcacgctcgcctccgacggccagtcgc
caccggcctgaggggctccacctcccagtgccctggtgacctctacalcagccccctgcaacaagtactgctctccggcctccctgctct
tcttggggcgacaacggcttgcctccctgttcggctccaagccctccgctccaaccgcggccaccggcctggggccggcgcctccc
actccggcgaggccatggccgctggccctgcagccccgccaggaggtgaccaccaagaagaagccccgtgatcaagcagcgccgc
gtggtggtgaccggcatggcgctggtgacccccctggggccacgagcccgactgttactacaacaacctgctggacggcggtgtccg
gcatctccgagatcgagaccttcgactgcacccagttccccaccggcatcgccggcgagatcaagtccttctccaccgacggctgggt
ggcccccgaagctgtccaagcgcatggacaagttcatgctgtacctgtgaccggccggcaagaaggccctggccgacggcgccgcatc
accgacgacgtgatgaaggagctggacaagcgcaagtgccggcgtgctgaltcggctccggcatggggcgccatgaagctgttcaacg
actccatcgaggccctgcgcaltctctacaagaagatgaaccccttctgcgtgccccttcggccaccaccaacatggggtccggcatgct
ggccatggacctggcgctggatggggcccaactactccatctccaccgctgcggccacctccaacttctgcatctgtaacgcttccaac
cacatcgtgcggcgagggccgacatgatgctgtgcggcggtccgacggcgtgatcatccccatcgccctggcgggcttcgtggc
ctggcgccctgtcccagcgcaacaacgacccccaccaaggcctccggccctgggactccaaccgcgacggcttcctgatgggc
gagggcgccggcgctgctgctgctggaggagctggagcacgccaaagcgcggcgccaccatctacccgagttcctggggcgcc
tcttcacctgcgacgctaccacatgaccgagccccaccccgaggggcgccggcgctgactctgtgcatcgagaaaggccctggccca
ggccggcggtgtccaaggaggacgtgaactacatcaacgccacgccacctccaccccgccggcgacatcaaggagtagaccaggc
cctggcccagtgcttcggccagaactccgagctgcgcgtgaactccaccaagttccatgacggccacctgctgggcggcgccggcg
gcgtggaggccgtgaccgtggtgcaggccatccgcaccggctggatccacccaacctgaacctggaggacccccacaaggccg
tggacgccaagctgctgtgtggggcccaagaaggagcgccctgaacgtgaaggtgggcctgtccaactccttcggcttcggcgccac
aactcctccatctgttcggcccttacaacaccatgtacctctacgacgtgcccgactacgctga

SEQ ID NO: 53

KASla de *C. avigera* (D3342, pSZ4511) con optimización de codones para *Prototheca*

5

atgacgtccctgcactccccgccttgcgcgcctccccctggacccttgcgcctgaagtccctcccaacggccccctctccacc
gcccgccttccgccccctgcgcgcgcaccctgcccaacatccgcgcgcctccccaccgtgtccgcccccaagcgcgcgaccg
accccaagaagcgcgtgtgtatcacccggcatgggcctgtgtgtctcgtgttcggctccgacgtggacgcctactacgagaagctgtgt
ccggcgagtccggcatctccctgatcgaccgttcgacgcctccaagttccccaccgccttcggcgccagatccgcggcttcaacg
ccaccggctacatcgacggcaagaacgaccgcgccttggacgactgcctcgcgtactcgcategtggccggcaagaagggccctgga
gaactccgacctgggcggcgactccctgtccaagatcgacaaggagcgcgcggcgctgtgtgtgggcaccggcatggcgccct
gaccgtgttctccgacggcgctgcagaacctgatcgagaaggccaccgcaagatcccccttcttcccccctacgccataccaac
atgggtccgcccctgtgtggccatcgacctgggcctgatgggccccaaactactccatctccaccgcctgcgccacctccaactactgtt
ctacgcgcgcgccaaaccacatccgcgcggcgaggccgacctgatgatcgcggcgccaccgaggccgccatcatccccatcgg
cctggcgccgttcgtgtgcctgccgcgcctgtccccagcgcaacgacgacccccagaccgcctccccccccctgggacaaggaccg
cgacggcttcgtgatggcgaggggcgccggcgctgtgtgtgatggagtccttggagcagccatgaagcgcggcgcccccatc
ggcgagtacctggcgggcgccgtgaactcgacgcctaccacatgaccgacccccgcgcgcagggcctggcgctgtcctcctgca
tcgagtcctccctggaggacggcgcgctgtcccccgaggaggtgaactacatcaacgccccaccctccaccctggccggcga

cttggccgagatcaacgccatcaagaaggtgtcaagaacaccaaggacatcaagatcaacgccaccaagttcatgctggccact
gcttggggcgctccggcgccctggaggccatgccaccatcaaggccatcaccaccggctggtctgacccctccatcaaccagttc
aaccccgagcgctccgtggagttcgacaccgtggccaacaagaagcagcagcagagggtgaacgtggccatctccaaactcttcgg
cttcggcgccacaaactcgtgttggtctctccgcttcaaggccaccatgtacctctacgagctgcccgaactacgcttga

SEQ ID NO: 54

KASI de *C. pulcherrima* (D3343, pSZ4512) con optimización de codones para *Prototheca*

10

atgcactccctgcagtcctccctgcgcgcctccccctggacccttccgccccagtcctccaccgtgcgccccctgcaccgc
gcctccatccccaacgtgcgcgcgcctccccaccgtgtccgcccccaagcgcgagaccgaccccaagaagcgcgtggtgatca
ccggcatgggctggtgtccgtgttcggctccgacgtggacgctactacgacaagctgtgtccggcgagtcggcatcgcccca
tcgaccgcttcgacgctcccaagtccccaccgccttcggcgccagatccgcggttcaactccatgggctacatcgacggcaaga
acgaccgcccctggacgactgctgtgctactgcatcgtggcgccgaagaagtcctggaggacggcgacctggcgccgacc
gcctgtccaagatcgacaaggagcgccggcggtgctgtgggacccggcatggcgccgtgacctgttctccgacggcggtgca
gtccctgatcgagaagggccaccgcaagatcaccccttcttcatccctacgccatcaccaacatgggctccgcccctgctggccatc
gagctgggcctgatgggccccaaactactccatccaccgctgcgccacctccaactactgttccacggcgccgccaaccacatc
cgccgcccggagggcgacctgatgacggcgccgaccgaggcgccatcccccacggcctggggcgcttcgtggcctgcc
ggccctgtcccagcgcaacgacgacccccagaccgctcccgccctgggacaaggaccgacggccttcgtgatggcgagg
ggcgccggcggtgctggtgctgagtcctggagcagccatgaagcgccggcccccatcagccgagtaacctggcgccgccc
caactgcgacgctaccacatgaccgacccccggcgccgacggcctggcggtgtctctcctgcatcagtcctccctggaggacggcg
gctgtcccccgaggaggtgaactacatcaacgcccacgcaacctccacctggcgccgacctggccgagatcaacgccatcaa
gaaggtgttcaagaacaccaaggacatcaagatcaacgccaccaagtcctgatcgccacctgctggcgccctccggcgccctgg
aggccatcgccacatcaaggcgatcaacaccggctggctgcacccctccatcaaccagttaacccccgagccctccgtggagttcg
acaccgtggccaacaagaagcagcagcagaggtgaacgtggccatctcaactccttcggcttcggcgcccaactccgtgtg
gccttctccgcttcaagcccaccatgacccctacgacgtgcccgaactacgctga

SEQ ID NO: 55

KAS mitochondrial de *C. avigera* (D3344, pSZ4513) con optimización de codones para *Prototheca*

5

atgggtgttctgcccctggcgcaagatgctgtgccccctccagtcaccgcttctgcgccccctgtctctccaccacttcgacccccg
ccgctggtgtggtgaccggcctgggcatggtgacccccctgggtgctggcggtgaacaccacctggaagcagctgatcaggggcaag
tgcggcatccggccatctccctggaggacctgaagatggacgcttcgacatcgacacccaggcctacgtgttcgaccagctgacc
tcaaggtggcgccaccgtgcccaccggcggtgaacccccggcgagttcaacgaggacctgtgttcaaccagaaggagcaccggc
ccatcgcccgttcacgctacgcccgtgtgcgcggcgacgagggcctgaaggacgccaactgggagcccaccgagcccagg
agcgcgagatgaccggcggtgcccacggcgccggcgacccggctccatctccgacgtgctggacggcgccgcatgatctgcgagaa
gaagctgcgcgcctgtcccccttcttcatcccccgatcctgatcaacatggcctccggccacgtgtccatgaagtacggcttcagg
gccccaaaccacgcccgtgaccgctgcgcgaccggcgcccactccatcgcgacggcgcccgcatgatccagttcggcgacg
ccgacgtgatggtggcgcgccgacgagtcctccatcgacgcccgttccatcgccggcttctgcgctcccgcgcctgaccacc
aagtacaactcctgccccaggaggcctcccgcccttcgacaccgaccgcgacggcttcgtatcgcgagggtcctggcggtgt
ggtgctggaggagctggaccacggccgcaagcgcgccgccaagatgtacggcgagttctcgggctacggcatgcccggcgacgc
ccaccacatcaccagccccactccgacggcgccggcgccatcctggccatgaccggcgccctgaagcagttcaacctgcacccc
gaccaggtggactacgtgaacggccacggccactccactcctggcgacggccatcgaggccaaggccatcaagacctgttctc
cgaccacggcatgctcggtccttggccctgtctccaccaaggcgccatcgccacactgctggcgccgcccggcgccgtggag
gccaatcttccatcctggccatcaagaacggcctggccccctgacctgaacgtggccccgccccgaccccggtgttcaccgagcgc

ttcgtgccccctgaccgctccaaggagatgcagtgctgcgcgcgcctgtccaactccttcggcttcggcgccaccaacaccacctg
ctgttaccctccccccccagaacaccatgtaccctacgacgtgcccgaactacgctga

SEQ ID NO: 56

KASIII de *C. avigera* (D3345, pSZ4514) con optimización de codones para *Prototheca*.

10

atggccaacgcctacggcttcgtgggtctctccgtgccaccgtggggccgcgcgcagttccagcagatgggctccggcttctgc
tccgtggacttcattccaagegcgtgttctgtctccgccgtgcaggggccgacaagcccgctccggcgactcccgcccgag
taccgcaccccccgctgggtgctcccgccgtgcaagctgatccggctccggctccgcacccccaccctgcaggtgtccaacgacga
cctggccaagatcgtggacaccaacgacgagtgatccgtgcgcaccggcaccgcaaccgcgcgtgctgacgggcaaggac
tcctgaccaacctggccaaccgagggcccgccgcaaggccctggagatggccaggtggacccgaggacgtggacatggtgctg
atgtgcacctccacccccgaggacctgttcggctccggccccagatccagaaggccctgggctgcaagaagaacccccctgctctac
gacatcaccgcgcctgctccggcttcgtgtgggctgtgtccgccctgcccacatccgcggcgccggttcaacaacgtgctg
gtgatccggcgccgactccctgtcccgctacgtggactggaccgaccgcggcactgcatctgttcggcgacgcgcggcgccgct
gctgtgagctcctgcgacgcgagggagggacggcctgttcgcttcgacctgcactccgacggcgacggccagcgccacctgcgc
gccgtgatcaccgagaacgagaccgaccacgcctgggcaaccaacggctccgtgtccgacttcccccccgccgctcctctactc
ctgcatccagatgaacggcaaggaggtgttcgcttcgctccgctccgtgccccagtcacagctggccctgggcaaggccg
gectgaacggctccaacatcgactggctgctgctgcaccaggccaaccagcgcatcagacgcgcgtggccaccccgctggaggtg
ccccaggagcgctgatcccaacctggccaactacggcaacacctccgcgcctccatccccctggccctggacgagggcgctg
cgggcgcaaggtgaagcccgccacctgatccaccgcgcggcttcggcgccgacctgacctggggctccggcaccgtgctg
gggcaccatglaccctacgacgtgcccgactacgctga

SEQ ID NO: 57

FATB2 de *C. hookeriana* ("Ch FATB2")

5

MVAAAASSAFFPVPAPGASPKPGKFGNWSSLSPSFKPKSIPNGGFQVKANDSAHPK
ANGSAVSLKSGSLNTQEDTSSSPPPRTFLHQLPDWSRLLTAITTVFVKSKRPDMHDRK
SKRPDMLVDSFGLSTVQDGLVFRQSFSIRSIEIGTDRTASIELTMNHLQETSLNHCK
STGILLDGFGRITLEMCKRDLIWVVIKMQIKVNRYPAWGDTVEINTRFSLGKIGMR
DWLISDCNTGEILVRATSAYAMMNQKTRRLSKLPYEVHQEIVPLFVDSPIEDSDLK
VHKFKVKTGDSIQKGLTPGWNLDVNQHVS NVKYIGWILESMPTVEVLETQELCSLA
LEYRRECGRDSVLESVTAMDP SKVGVR SQYQHLLRLEDGTAIVNGATEWRPNKAGA
NGAISTGKTSNGNSVS

SEQ ID NO: 58

23S ARNr para UTEX 1439, UTEX 1441, UTEX 1435, UTEX 1437 *Prototheca moriformis*

10

TGTTGAAGAATGAGCCGGCGACTTAAAATAAATGGCAGGCTAAGAGAATTAATA
ACTCGAAACCTAAGCGAAAGCAAGTCTTAATAGGGCGCTAATTTAACAAAACAT
TAAATAAAATCTAAAGTCATTTATTTTAGACCCGAACCTGAGTGATCTAACCATG
GTCAGGATGAACTTGGGTGACACCAAGTGGAAGTCCGAACCGACCGATGTTGA
AAAATCGGCGGATGAACTGTGGTTAGTGGTGAAATACCAGTCGAACTCAGAGCT
AGCTGGTTCTCCCCGAAATGCGTTGAGGCGCAGCAATATATCTCGTCTATCTAGG
GGTAAAGCACTGTTTCGGTGCGGGCTATGAAAATGGTACCAAATCGTGGCAAAC
TCTGAATACTAGAAATGACGATATATTAGTGAGACTATGGGGGATAAGCTCCAT
AGTCGAGAGGGAAACAGCCCAGACCACCAGTTAAGGCCCAAAATGATAATGAA

GTGGTAAAGGAGGTGAAAATGCAAATACAACCAGGAGGTTGGCTTAGAAGCAGC
CATCCTTTAAAGAGTGCGTAATAGCTCACTG

SEQ ID NO: 59

Secuencia de aminoácidos de la KASIV de *C. hookeriana* (D3668, pSZ4756). El péptido de tránsito de alga está subrayado.

15

MASAAFTMSACPAMTGRAPGARRSGRPVATRLRGSTFQCLDPCNQQRFLGDNGFAS
LFGSKPLRSNRGHLRLGRTSHSGEVMAVAMQPAQEVSTNKKPATKQRRVVVTGMG
VVTPLGHDPDVYYNNLLDGISGISEIENFDCSQFPTRIAGEIKSFSTDGWVAPKFSERM
DKFMLYMLTAGKKALADGGITEDAMKELNKRKCGVLIGSGLGGMKVFSDSIEALRT
SYKKISPFVCFSTTNMGSAILAMD LGWMGPNYSISTACATSNFCILNAANHIIKGEA
DMMLCGGSDAAVLPVGLGGFVACRALSQRNNDPTKASRPWDSNRDGFVMGEGAG
VLLLEELEHAKKRGATTYAEFLGGSFTCDAYHMTPEHPGAGVILCIEKALAQSGVSR
EDVNYINAHATSTPAGDIKEYQALAHCFGQNSELRVNSTKSMIGHILLGGAGGVEAV
AVVQAIRTGWIIHPNINLEDPDEGVDAKLLVGPKKEKLKVKVGLSNSFGFGGHNSIL
FAPCN

SEQ ID NO: 60

- 5 Secuencia de nucleótidos del vector de expresión de *KASIV* de *C. hookeriana* (D3668, pSZ4756). Los brazos de homología 5 y 3 que permiten la integración dirigida en el locus SAD2-1 están anotados en minúsculas. El promotor endógeno SAD2-1 (presente en el brazo de homología dirigido 5) guía la expresión de *KASIV* de *Ch* con optimización de codones (anotada en minúsculas en **negrita**) y se termina con la 3'UTR *PmHSP90* anotada en minúsculas en **negrita** subrayadas. El promotor *PmHXT1* está anotado en mayúsculas en cursiva que guía expresión del marcador de selección de *ScMelibiasa* anotado en minúsculas en cursiva seguido del terminador de 3'UTR *PmPGK* en **mayúsculas** subrayadas. Los lugares de clonación de restricción y los fragmentos de DNA espaciador están anotados subrayados, en mayúsculas y con letras sin formato.
- 10

gcccggcaccacccgcagctcgtactacagcgcacgcaccgcttcgtgaccaccgggtgaacglagtcctcgacggaaacatctg
gttcggggcctcctgcttgcactcccgcacatgccgacaacctttctgctgttaccacgaccacacaatgcaacgcgacacgaccgtgtg
ggactgacgggttcactgcacctgcacatgtcacaagcgttactccaattgtattcgttttttctgggagcagttgctcgaccgc
ccgcgtcccgagggcagcgatgacgtgtgctgtggcctgggtgttctgctgaaaggccagcaaacctaaatgcagggcgatccggag
attgggatctgatccgagtttggaccagatccgccccgatgcggcaggggaactgcacgactcggcgcggaacccagcttctgtaa
atgccagattgggtgctcgataacctggatttccatcagcgaacaagacttcagcagcgagcgtatttggcgggctgctaccaggggt
gcatacattgccatttctgtctgaccgcttactggcgagaggggtgagttgatgggttggcagggcatcgaaacgcgcgtgcatgg
tgtgctgtctgttttggctgcacgaattcaatagtcggatggcgacggtagaattgggtgtggcgctcgctgcatgcctcgcccc
gtcgggtgcatgaccgggactggaatccccctcgcgaccatttctgaacgctcccgactcctccgaccgcgcgaggatagact
cttgttcaaccaatcgacaGGTACCatggcttcggcggaattcaacatgtcggcgtgccccgcgatgactggcagggccccct
ggggcagctcgttcggcagggcagtcgccacccgcctgaggggcagcaccttcagtgccctggacccctgcaaccagcagc
gttctctggggcagcaaacggcttcgctgctgttctggctccaagccctgcgcagcaaacgcggccacctgcgctggggcgc
acctgcactccggcgaggtgatggcctgcgctgacgccccccaggaggtgagcaccacaagaagcccgcgaccaa
gcagcgccgctggctgtgacggcatggcgctgtgacccccctggggccacgaccccgacgtgtattataacaacctgttgg
acggcatctgggcatctccgagatcgagaacttcgactgcagccagttccccacccgcacgcggcgagatcaagtcgttc
tcaccgacggctgggtcgcgccaagtccagcgagcgatggacaagttcatgctgtatgtgaccccgccgaagaagg
cgctggccgacggcgcatcaccgaggacgcgatgaaggagctgaacaagcgcaagtgcggcgtgctgatcggtcgggc

ctggggggcatgaaggcttctccgaacagcagcagggccctgcgcacctcgtataagaagatccccctctcggtgcccttc
 agcaccaccaacatgggctcggcgatcctggcgatgggacctgggctggatgggccccaaactattccatcagcaccgcgtgcg
 ccacctcgaacttctgcatcctgaacgcggccaaccacatcatcaagggcgaggcggacatgatgctgtgcggcggtccga
 cggcgcggtgctgcccgtcgccctgggcggttctggtgctgcccgcgctgagccagcgcaacaacgacccccaccaaggcc
 tgcgccccctgggactccaaccgcgacggcttcgtcatgggcgaggcgcgggcggtgctgctgctgaggagctggagcacg
 ccaagaagcgcgcgcgacacatctatgcccagttctcggggcgagcttcacctgcgacgcgtatcacatgaccgagcccca
 ccccgagggcgcgggcgctcatcctgtgcatcgagaaggcgctggccagtcggggcggtgccgaggacgtgaactatalc
 aacgcgcacgccaccagcacccccggggcgacatcaaggagatcaggccctggcgactgcttcggccagaactcggag
 ctgcggctcaactccaccaagagcatgatcgccacctgctggcgcgcgcgggcggtggaggcggtgcgctggtccagg
 cgtatcgacacggctggatccaccccaacatcaacctggaggaccccgacgagggcggtggacgccaagctgctggtcgcc
 ccaagaaggagaagctgaaggtaaggctggcctgtcgaactccttcggcttcggcgccacaacagctcgatcctgttcgc
 gccctgcaactgaCTCGAGacagacgaccttggcagcgctcggtaggaggtggtggtgagtcctcgtatgccatc
gcacgcatccaacgacctatagcgcctcgaatgacctgcgtgctcctctctgctcctccgttttggagatgctcaggcttgg
gcactcctcggtggcgagccacgttgcgcctcgtcgtcgttgcctctcttgcgcctctggtactggaaaatcatcgagcc
cgttttttgcctccatttcttccctacatcttgaagcaaacgacaaacgaagcagcaagcaagagacacgaggacgggtg
aacaagtctctcactgtatacatctatttcccccgggcgacctactctctctcctcccccgagagtcagctgccttacgtg
acCCTAGGTGCGGTGAGAATCGAAAATGCATCGTTTCTAGGTTTCGGAGACGGTCAATT
 CCTGCTCCGGCGAATCTGTGCGTCAAGCTGGCCAGTGGACAATGTTGCTATGGCAGC
 CCGCGCACATGGGCCCTCCCGACGCGGCCATCAGGAGCCCAAACAGCGTGTACGGGT
 ATGTGAAACTCAAGAGGTCCCTGCTGGGCACTCCGGCCCCACTCCGGGGGCGGGAC
 GCCAGGCATTTCGCGGTCCGTCCCGCGCGACGAGCGAAATGATGATTTCGGTTACGAGA
 CCAGGACGTCTGTCGAGGTGAGAGGCAGCCTCGGACACGTCTCGCTAGGGCAACGC
 CCCGAGTCCCCGCGAGGGCCGTAACAATTGTTTCTGGGTGTCGGAGTGGGCATTTTG
 GGCCCGATCCAATCGCCTCATGCCGCTCTCGTCTGGTCTCACGTTTCGCGTACGGCCT
 GGATCCCCGAAAGGGCGGATGCACGTGGTGTGCCCCGCCATTGGCGCCACGTTTC
 AAAGTCCCCGGCCAGAAATGCACAGGACCGGCCCGGCTCGCACAGGCCATGCTGAAC
 GCCCAGATTTTCGACAGCAACACCATCTAGAATAATCGCAACCATCCGCGTTTTGAACGA
 AACGAAACGGCGCTGTTTAGCATGTTTCCGACATCGTGGGGGCCGAAGCATGCTCCG
 GGGGGAGGAAAGCGTGGCACAGCGGTAGCCATTCTGTGCCACACGCCGACGAGGA
 CCAATCCCCGGCATCAGCCTTCATCGACGGCTGCGCCGCACATATAAAGCCGGACGC
 CTAACCGGTTTCGTGGTTATGACTAGTatgttcgcgttacttctgacggcctgcatcctccctgaaggcggtg
 ttggcgctccccctctacaaacggcctgggctgacgccccagatgggctgggacaaactggacacgttcgctgcgacgtctc
 cgagcagctgctgtggacacggcgaccgcactctccgacctgggctgaaggacatgggctacaagtacatcactctggacga
 ctgctggctcctcgccgcgactccgacggcttctggtcgcgcagagcagaagttcccaacggcatggggccacgtcggcga
 ccacctgcacaacaactccttctgttcggcatgactcctccgcggcgaglacacgtgcgcggctacccccggctccctgggce
 gcgaggaggaggacgcccagttctcgcgaacaaccgcgtggactacctgaagtacgacaactgclacaacaagggccagttc
 ggcacgcccagatctcctaccaccgctacaaggccatgtccgacgcccgaacaagacggcgccccacttctactcctgt
 gcaactggggccaggacctgaccttacttgggctccggcctcgcgaactcctggcgcatgtccggcgacgtcagggcgagtt
 cacgcgccccgactcccgctgccccgcgacggcgacgagtagctgcaagtagccggcttccactgctccatcatgaacalc
 ctgaacaaggccgccccatgggccaagacgcgggcgtcggcggtggaacgacctggacaacctggaggtcggcgctggc
 aacctgacggacgacgaggagaaggcgcaacttctccatgtgggcatgtggaagtcccccctgatcatcggcgcaacgtgaa
 caacctgaaggccctcctctactccatctactcccaggcgctccgcatcgcacatcaaccaggactccaacggcatccccgccacg
 cgcgtctggcgctactacgtgtccgacacggacgagtagccaggcgagatccagatgtgttcggccccctggacaacgg
 cgaccaggctgtggcgctgtgaacggcggtcctgttccccgcccatgaacacgacctggaggagatcttctgactccaac

ctgggtccaagaagctgacctccacctgggacatctacgacctgtggggaaccgctcgacaactccacggcgtccgccatc
ctgggcccgaacaagaccgccaccggcatcctgtacaacgccaccgagcagtcctacaaggacggcctgtccaagaacgaca
cccgcctgttgcggcagaagatcggtccctgtcccccaacgcgatcctgaacacgaccgtccccgccacggcatcgcgttcta
ccgcctgcggccctctctctgATACAACCTTATTACGTATTCTGACCGGCGCTGATGTGGCGCG
GACGCCGTCGTACTCTTTTACAGACTTTACTCTTTGAGGAATTGAACCTTTCTCGCTTG
CTGGCATGTAAACATTGGCGCAATTAATTGTGTGATGAAGAAAGGGTGGCACAA
GATGGATCGCGAATGTACGAGATCGACAACGATGGTGATTGTTATGAGGGGCCA
AACCTGGCTCAATCTTGTTCGATGTCCGGCGCAATGTGATCCAGCGGCGTGACTC
TCGCAACCTGGTAGTGTGTGCGCACCGGGTCGCTTTGATTAAAACCTGATCGCATT
GCCATCCCGTCAACTCACAAGCCTACTCTAGCTCCCATTTGCGCACTCGGGCGCCC
GGCTCGATCAATGTTCTGAGCGGAGGGCGAAGCGTCAGGAAATCGTCTCGGCAG
CTGGAAGCGCATGGAATGCGGAGCGGAGATCGAATCAGATATCAAGCTCCATCG
AGCTCagccacggcaacaccgcgccttggggcgagcagggcgacaagaacctgagcaagatctcggggtgatcgcc
agcgacgaggcgccgacgagatcgctacacgcgcctgtggagagttcttcgcctcgaccccgaggcgccgctcgccgct
acgccaacatgatggcaagcagatcaccatgcccgcgacctcatggagacatggggcaccggcgaggccaacccggcgccga
acctcttcgcgacttctccggtcgccgagagatcgacgtctacgacgcgaggactactcgcccatcctggagcactcaacg
cgcgctggaaagtggagcagcgccaggtcagcgccagggcgccgagaccaggagtagctctgggctgccccagcgcttc
ggaaactcgccgagaagaccgcgcgaagcgagcgctcgccgagggcgccgcttctcttgatctcggggcgcgaga
tcatggtctaggagcagcaggtgtgctgctggggcgctggcgagtgggacgccctctcgtctctctgttctgaacggacaat
cgccaccccgcgctacgcgcacgcacgagcaacgaagaaacccccgatgataggttgcgggtgctcgccggatagatc
cgccgcacatcaaaaggccctcgcgcagagaagaagctcttccagcagactct

SEQ ID NO: 61

Secuencia de nucleótidos de la CDS de KASIV de *C. hookeriana* con optimización de codones para *P. moriformis*.

5

atggttccgcggcattcaccatgtcggcgtgccccgcgatgactggcagggcccttggggcacgtcgtccggacggccagtcgc
caccggcctgagggcgacaccttccagtgcctggaccctgcaaccagcagcgcttctgggggacaacggcttcgctgctgtt
cggctccaagcccttgcgcagcaaccggggccacctgcgcttggcgccgacctcgcactcggcgaggatggcgtcgcgat
gcagcccgccagggaggtgagcaccacaagaagcccgccagcaagcagcgccgctggtcgtgaccggcatggcgctgga
ccccctggggcacgaccccgacgtgtattataacaacctgctggacggcatctcggcatctccgagatcgagaacttgcactgca
gccagttccccaccgcacgcggcgagatcaagtcgttctccaccgacggctgggtcgcgcccaagttcagcgagcgcatggac
aagttcatgctgtatagctgaccgcggcaagaaggcgtgcccgcacggcgccatcaccgaggacgcgatgaaggagctgaaca
agcgcaagtgcggcgtgctgatcggtcgggctggcgccgatgaaggtcttctccgacagcatcaggccctgcgcacctglat
aagaagatctcccccttctgctgcttccagcaccacaacatgggctcggcgatcctggcgatggacctggctggatggcccc
aactatccatcagcaccgctgcgccacctgcaacttctgcatcctgaacgcggccaaccacatcaaggcgaggcgacatg
atgctgtcggcggtcgcacgcggcggtgctgcccgtcggcctggcggttctggtgctgcccgcgctgagccagcgcaaca
acgacccccaccaaggctcgcgccctgggactccaaccgcgacggcttctgcatggcgaggcgccggcgctgctgctgga
ggagctggagcagccaagaagcgcgccgcgaccatctatgccgagttcctggcgggcagcttcacctgcgacgcgtatcatga
ccgagccccaccccgaggcgccggcgctcatcctgtgcatcgagaaggcgctggccagtcggcggtgtcccgcgaggacgtga
actatatcaacgcgcacgccaccagcaccgccggcgacatcaaggatcaggccctggcgacgtcttccgccaagaactgc
gagctgcgcgtaactccaccaagagcatgacggccacctgctggcgccgcggcgccgctggaggcggtcgccgtgttcag
gcatcgcacggctggtatccaccaacatcaacctggaggaccccgacgaggcggtggacgccaagctgctggtcggcccc
aagaaggagaagctgaaggtgaaggtcggcctgtcgaactcttggcttggcgccacaacagctcagctctgttcgcccctgc
aactga

SEQ ID NO: 62

Secuencia de aminoácidos de KASIV de *C. aequipetala*. El péptido de tránsito de alga está subrayado.
KASIV de C aequ

10

MAAAASMVASPLCTWLVAACMSTSFNDNPRSPSIKRIPRRRRILSQSSLRGSTFQCLV
 TSYIDPCNQFSSSASLSFLGDNGFASLFGSKPFRSIRGHRRLGRASHSGEAMAVALEPA
 QEVATKKKPVVKQRRVVVTGMGVVTPLGHEPDVYYNNLLDGVSGISEIETFDCNQF
 PTRIAGEIKSFSTDGWWVAPKLSKRMDKFMLYLLTAGKKALADGGITDDVMKELDKR
 KCGVLIGSGLGGMKLFSDSIEALRISYKKMNPFCVPFATTNMGSAMLAMD LGWMGP
 NYSISTACATSNFCILNSANHIVRGEADMMLCGGSDAVIPIGLGGFVACRALSQRNN
 DPTKASRPWDSNRDGFVMGEGAGVLLLEEELEHAKKRGATIYAEFLGGSFTCDAYHM
 TEPHPEGAGVILCIEKALAQAGVSREDVNYINAHATSTPAGDIKEYQALAHCFGHNS
 ELRVNSTKSMIGHLIGAAGGVEAVTVVQAI RTGWIIHPNLNLED PDKAVDAKLLVGP
 KKERLNVKVGLSNSFGFGGHNSSILFAPYN

SEQ ID NO: 63

Secuencia de aminoácidos de KASIV de *C. glassostoma*. El péptido de tránsito de alga está subrayado.

5 S07_Cg_Locus_4548_Transcript_4/9_translation

MAAAASSQLCTWLVAACMSTSFNDNPRSPSIKRLPRRRRVLSHC SLRGSTFQCLVTS
 YIDPCNQYC SSASLSFLGDNGFTPLIGSKPFRSNRGHPRLGRASHSGEAMAV ALQPAQ
 EVATKKK PAMKQRRVVVTGMGVVTPLGHEPDVYYNNLLDGVSGISEIETFDC TQFP
 TRIAGEIKSFSTDGWWVAPKLSKRMDKFMLYLLTAGKKALADGGITDDVMKELDKRK
 CGVLIGSGMGMKLFND SIEALRVSYKKMNPFCVPFATTNMGSAMLAMD LGWMGP
 NYSISTACATSNFCILNAANHIVRGEADMMLCGGSDAVIPIGLGGFVACRALSQRNN
 DPTKASRPWDSNRDGFVMGEGAGVLLLEEELEHAKKRGATIYAEFLGGSFTCDAYHM
 TEPHPEGAGVILCIEKALAQAGVSREDVNYINAHATSTPAGDIKEYQALAHCFGQNS
 ELRVNSTKSMIGHLLGAAGGVEAVTVIQAI RTGWIIHPNLNLD DDPDKAVDAKFLVGP
 KKERLNVKVGLSNSFGFGGHNSSILFAPYN

SEQ ID NO: 64

Secuencia de aminoácidos de KASIV de *C. hookeriana*. El péptido de tránsito de alga está subrayado.

10 S26_ChookKASIV_trinity_43853 - translation

MAASSCMVGSPFCTWLVSACMSTSFNDNPRSLSHKRLRLSRRRRRTLSSHCSLRGSTP
 QCLDPCNQHCFLGDNGFASLFGSKPPRSDLGHLRLGRTSHSGEVMAVAQEVSTNKK
 PATKQRRVVVTGMGVVTPLGHD PDVYYNNLLDGVSGISEIETFDC TQFP TRIAGEIKS
 FSTDGLVAPKLSKRMDKFMLYILT AGKKALADGGIT EDVMKELDKRKCGVLIGSGL
 GGMKVFSDSVEALRISYKKISPFCVPFSTTNMGSAILAMD LGWMGPNYSISTACATS
 NFCILNAANHITKGEADMMLCGGSDAAILPIGMGGFVACRALSQRNN DPTKASRPW
 DSNRDGFVMGEGAGVLLLEEELEHAKKRGATIYAEFLGGSFTCDAYHMT EPHPEGAG
 VILCIEKALAQAGVSREDVNYINAHATSTPAGDIKEYQALAHCFGQNS ELRVNSTKS
 MIGHLIGAAGGVEAVTVIQAI RTGWIIHPNLNL ENPDKAVDAKLLVGP KKERLDVKV
 GLSNSFGFGGHNSSILFAPYN

SEQ ID NO: 65

Secuencia de aminoácidos de KASIV de *C. glassostoma*. El péptido de tránsito de alga está subrayado.

15 S07_Cg_Locus_3059_Transcript_2/2_translation

MAAASSMVASSFSTSLVAACMSTSFNDPRFLSHKRIRLSLRGSTFQCLGDNGFAS
LIGSKPPRSNHGHRRLGRTSHSGEAMAVAMQPAQEASTKNKHVTKQRRVVVTGMG
VVTPLGHDVPDYNNLLDGVSGISEIENFDCSQFPTRIAGEIKSFSTEGYVIPKFAKRM
DKFMLYLLTAGKKALEDGGITEDVMKELDKRKCGVLIGSGMGGMKIINDSIAALNV
SYKKMTPFCVPFSTTNMGSAIDLGLWMGPNYSISTACATSNYCILNAANHIIRGE
ANMMLCGGSDAVVIPVGLGGFVACRALSQRNNDPTKASRPWDSNRDGFVMGEA
GVLLLEEELEHAKKRGATIYAEFLGGSFTCDAYHMTEPHPDGAGVILCIEKALAQSGV
SREDVNYINAHATSTPAGDIKEYQALAHCFGQNSELRVNSTKSMIGHLLGAAGGVEA
VSVVQAI RTGWHPNINLEDPDEAVDAKLLVGPKKEKLKVKVGLSNSFGFGGHNSI
LFAPCN

SEQ ID NO: 66

Secuencia de aminoácidos de la KASIV de *C. carthagenesis*. El péptido de tránsito de alga está subrayado.

5 S05_CcrKASIV_17190_Seq_7/7_translation

MAAAAFASPFCITWLVAACMSSASRHDPLPSPSSKPRLRRKILFQCAGRGSSAGSGS
SFHSLVTSYLGCLEPCHEYYTSSSSLGFSLSFGSTPGRTSRRQRRLHRASHSGEAMAV
ALQPAQEVTTKKKPSIKQRRVVVTGMGVVTPLGHDVPDYNNLLDGVSGISEIETFD
CAQFPTRIAGEIKSFSTDGWVAPKLSKRMDKFMILYMLTAGKKALADGGISEDVMKE
LDKRKCGVLIGSAMGGMKVFNDAIEALRISYKKMNPFCVPFATTNMGSAMLAMD
LGLWMGPNYSISTACATSNFCILNAANHITRGEADMMLCGGSDAVIPIGLGGFVACRA
LSQRNNDPTKASRPWDSNRDGFVMGEGAGVLLLEEELEHAKKRGATIYAEFLGGSFT
CDAYHMTEPHPKGAGVILCIERALAQSGVSREDVNYINAHATSTPAGDIKEYQALAH
CFGQNSELRVNSTKSMIGHLLGAAGGVEAVTVVQAI RTGWVHPNINLENPDEGVDA
KLLVGPKKEKLKVKVGLSNSFGFGGHNSILFAPYN

SEQ ID NO: 67

Secuencia de aminoácidos de la KASIV de *C. carthagenesis*. El péptido de tránsito de alga está subrayado.

10 S05_CcrKASIV_17190_Seq_6/7_translation

MAAAASVVASPFCTWLVAACMSASFNDNEPRSLSPKRRRSLSRSSSASLRLGGNGFA
SLFGSDPLRPNRGHRLRHASHSGEAMAVALQPAQEVSTKKKPVTKQRRVVVTGM
GVVTPLGHDVPDYNNLLDGVSGISEIETFDCTQFPTRIAGEIKSFSTDGWVAPKLSK
RMDKFMILYMLTAGKKALADGGITEEVMKELDKRKCGVLIGSGMGGMKLFNDSIEA
LRISYKKMNPFCVPFATTNMGSAMLAMDGLWMGPNYSISTACATSNFCILNAANHIT
RGEADMMLCGGSDAVIPIGLGGFVACRALSQRNNDPTKASRPWDSNRDGFVMGEG
AGVLLLEEELEHAKKRGATIYAEFLGGSFTCDAYHMTEPHPKGAGVILCIERALAQSG
VSREDVNYINAHATSTPAGDIKEYQALAHCFGQNSELRVNSTKSMIGHLLGAAGGVE

AVTVVQAI RTGWVHPNINLENPDEGVDAKLLVGPKKEKLKVKVGLSNSFGFGGHNS
SILFAPYN

SEQ ID NO: 68

Secuencia de aminoácidos de la KASIV de *C. pulcherrima*. El péptido de tránsito de alga está subrayado.

15 pSZ2181 - CpulcKASIV

MPAASSLLASPLCTWLLAACMSTSFHPSDPLPPSISSPRRRRLSRRRRLSQCAPLPSASSA
LRGSSFHITLVTSYLACFEPCHDYYSASLFGSRPIRTTRRHRRRLNRASPSREAMAVAL
QPEQEVTTKKKPSIKQRRVVVTGMGVVTPPLGHDPDVYNNLLDGTSGISEIETFDCA
QFPTRIAGEIKSFSTDGWVAPKLSKRMDKFMLYMLTAGKKALTDGGITEDVMKELD
KRKCGVLIGSAMGGMKVFNDIAEALRISYKKMNPFCVPFATTNMGSAMLAMD LGW
MGPNY SISTACATSNFCIMNAANHIRGEADVMLCGGSDAVIPIGMGGFVACRALSQ
RNSDPTKASRPWDSNRDGFVMGEGAGVLLLEEELEHAKKRGATIYAEFLGGSFTCDA
YHMTPEHPDGAGVILCIEKALAQSGVSREDVNYINAHATSTPAGDIKEYQALIHCFG
QNRELKVNSTKSMIGHILLGAAGGVEAVSVVQAI RTGWIIHPNINLENPDEGVDTKLLV
GPKKERLNVKVGLSNSFGFGGHNSSILFAPYI

SEQ ID NO: 69

Clado 1 C8 y C10 de consenso de KASIV

5

MAAASCMVASPFCTWLVAACMSTSDNDPRSLSHKRLRLSRRRRRLSSHCSLRGSTF
QCLDPCNQHCFLGDNGFASLFGSKPPSRNRGHLRLGRTSHSGEVMAVAXQXAEVS
TNKKPATKQRRVVVTGMGVVTPPLGHDPDVYNNLLDGVSGISEIENFDCSQFPTRIA
GEIKSFSTDGWVAPKLSKRMDKFMLYILT AGKKALADGGITEDVMKELDKRKCGVL
IGSGLGGMKVFSDSIEALRTSYKKISPFVCFSTTNMGSAILAMD LGWMGPNY SISTA
CATSNFCILNAANHITKGEADMMLCGGSDAAILPIGMGGFVACRALSQRNNDPTKAS
RPWDSNRDGFVMGEGAGVLLLEEELEHAKKRGATIYAEFLGGSFTCDAYHMTPEHPE
GAGVILCIEKALAQSGVSREDVNYINAHATSTPAGDIKEYQALAHCFGQNSELRVNS
TKSMIGHILLGGAGGVEAVTVVQAI RTGWIIHPNINLEDPDEGVDAKLLVGPKKEKLV
VKVGLSNSFGFGGHNSSILFAPCN

SEQ ID NO: 70

Clado 2 solo C10 de consenso de KASIV

10

MAAAAASMXXSPLCTWLVAACMSTSFNDNDPRSPSIKRLPRRRRVLSQC SLRGSTFQCL
VTSYIDPCNQYCSSASLSFLGDNGFASLFGSKPFRSNRGHRLGRASHSGEAMAV
QPAQEVATKKKPVIKQRRVVVTGMGVVTPPLGHEPDVYNNLLDGVSGISEIETFDCT
QFPTRIAGEIKSFSTDGWVAPKLSKRMDKFMLYLLTAGKKALADGGITDDVMKELD
KRKCGVLIGSGMGGMKLFNDSIEALRXSYKKMNPFCVPFATTNMGSAMLAMD LGW
MGPNY SISTACATSNFCILNAANHIVRGEADMMLCGGSDAVIPIGLGGFVACRALSQ
RNNDPTKASRPWDSNRDGFVMGEGAGVLLLEEELEHAKKRGATIYAEFLGGSFTCDA
YHMTPEHPEGAGVILCIEKALAQAGVSREDVNYINAHATSTPAGDIKEYQALAHCFG
QNSELRVNSTKSMIGHILLGAAGGVEAVTVXQAI RTGWIIHPNLNLEDPDKAVDAKLL
VGPKKERLNVKVGLSNSFGFGGHNSSILFAPYNV

SEQ ID NO: 71

Clado 1 proteína madura de consenso de KASIV

15

KQRRVVVTGMGVVTPPLGHDPDVYNNLLDGVSGISEIENFDCSQFPTRIAGEIKSFST
DGWVAPKLSKRMDKFMLYILT AGKKALADGGITEDVMKELDKRKCGVLIGSGLGG
MKVFSDSIEALRTSYKKISPFVCFSTTNMGSAILAMD LGWMGPNY SISTACATSNFC
ILNAANHITKGEADMMLCGGSDAAILPIGMGGFVACRALSQRNNDPTKASRPWDSN
RDGFVMGEGAGVLLLEEELEHAKKRGATIYAEFLGGSFTCDAYHMTPEHPEGAGVIL
CIEKALAQSGVSREDVNYINAHATSTPAGDIKEYQALAHCFGQNSELRVNSTKSMIG
HLLGGAGGVEAVTVVQAI RTGWIIHPNINLEDPDEGVDAKLLVGPKKEKLVKVGLS
NSFGFGGHNSSILFAPCN

SEQ ID NO: 72

Clado 2 proteína madura de consenso KASIV

KQRRVVVTGMGVVTPLGHEPDVYYNNLLDGVSGISEIETFDCTQFPTRIAGEIKSFST
DGWVAPKLSKRMDKFMLYLLTAGKKALADGGITDDVMKELDKRKCGVLIGSGMG
GMKLFNDSIEALRXXSYKKMNPFCVFPATTNMGSAAMLAMD LGWMGPNYSISTACAT
SNFCILNAANHIVRGEADMMLCGGSDAVIIPIGLGGFVACRALSQRNNDPTKASRPW
DSNRDGFVMGEGAGVLLLEELEHAKKRGATIYAEFLGGSFTCDAYHMTPEHPGAG
VILCIEKALAQAGVSREDVNYINAHATSTPAGDIKEYQALAHCFGQNSELRVNSTKS
MIGHILLGAAGGVEAVTVXQAI RTGWIHPNLNLEDPKAVDAKLLVGPKKERLNVK
VGLSNSFGFGGHNSSILFAPYNV

5

REIVINDICACIONES

1. Polinucleótido con al menos un 95 % de identidad de secuencia con la SEQ ID NO: 21 o la SEQ ID NO: 39, o un polinucleótido que codifica una proteína de tipo KASI que tiene al menos un 95 % de identidad de secuencia de aminoácidos con la SEQ ID NO: 2, o una proteína madura producida a partir de estos, donde dicho polinucleótido, cuando se coexpresa con un gen de acil-ACP tioesterasa FATA o FATB exógeno en una célula hospedadora de microalga oleaginosa, es capaz de enriquecer el contenido de C14:0 del aceite celular producido por la célula hospedadora, con respecto a un aceite producido por una célula del mismo tipo que no expresa dicho polinucleótido.
2. Vector de transformación que comprende el polinucleótido según la reivindicación 1, que comprende opcionalmente secuencias promotoras y 3'UTR en conexión funcional con el polinucleótido y, opcionalmente, una secuencia flanqueante para una recombinación homóloga.
3. Célula hospedadora oleaginosa plastídica que comprende el polinucleótido según la reivindicación 1 o el vector según la reivindicación 2, donde la célula hospedadora tiene una vía de biosíntesis de ácidos grasos de tipo II.
4. Célula hospedadora oleaginosa plastídica que comprende:
 - a) un polinucleótido según la reivindicación 1; y
 - b) un polinucleótido que codifica una acil-ACP tioesterasa FATA o acil-ACP tioesterasa FATB, donde la célula hospedadora tiene una vía de biosíntesis de ácidos grasos de tipo II.
5. Célula hospedadora según la reivindicación 4, donde la acil-ACP tioesterasa FATA o la acil-ACP tioesterasa FATB es una acil-ACP tioesterasa FATB que tiene al menos un 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98 o 99 % de identidad de secuencia de aminoácidos con la SEQ ID NO: 1 o la SEQ ID NO: 57.
6. Célula hospedadora según la reivindicación 5, donde la célula hospedadora produce un aceite celular **caracterizado por** un perfil de ácidos grasos con (i) al menos un 30, 40, 50 o 55 % de C14:0, (ii) al menos un 7, 8, 9, 10, 11, 12, 13 o 14 % de C8:0, (iii) al menos un 10, 15, 20, 25, 30 o 35 % en área para la suma de C8:0 y C10:0, o (iv) una relación C8/C10 en el rango de 2,2-2,5; 2,5-3,0 o 3,0-3,4.
7. Célula hospedadora según cualquiera de las reivindicaciones 4 a 6, donde uno o más de los polinucleótidos presenta una optimización de codones para la expresión en la célula hospedadora de manera que la secuencia codificante del polinucleótido contenga el codón más preferido o el segundo codón más preferido para al menos el 60 % de los codones de la secuencia codificante de manera que la secuencia de codones optimizados se traduzca de forma más eficiente en la célula hospedadora con respecto a una secuencia no optimizada, opcionalmente donde la secuencia codificante contiene el codón más preferido para al menos el 80 % de los codones de la secuencia codificante.
8. Célula hospedadora según cualquiera de las reivindicaciones 3 a 7, donde la célula hospedadora es una microalga.
9. Célula hospedadora según la reivindicación 8, donde la célula hospedadora es de *Trebouxiophyceae*, y opcionalmente del género *Chlorella* o *Prototheca*, opcionalmente donde, además, la microalga es de la especie *Prototheca moriformis*.
10. Método para hacer un aceite celular, donde el método comprende cultivar una célula hospedadora según cualquiera de las reivindicaciones 3 a 9, para producir el aceite celular, donde el aceite comprende triglicéridos y esteroides de microalga, donde opcionalmente el aceite celular comprende esteroides, **caracterizado por** un perfil de esteroles y el perfil de esteroles tiene un exceso de ergosterol con respecto al β -sitosterol y/o la presencia de 22,23-dihidrobrassicasterol, poriferasterol o clionasterol.

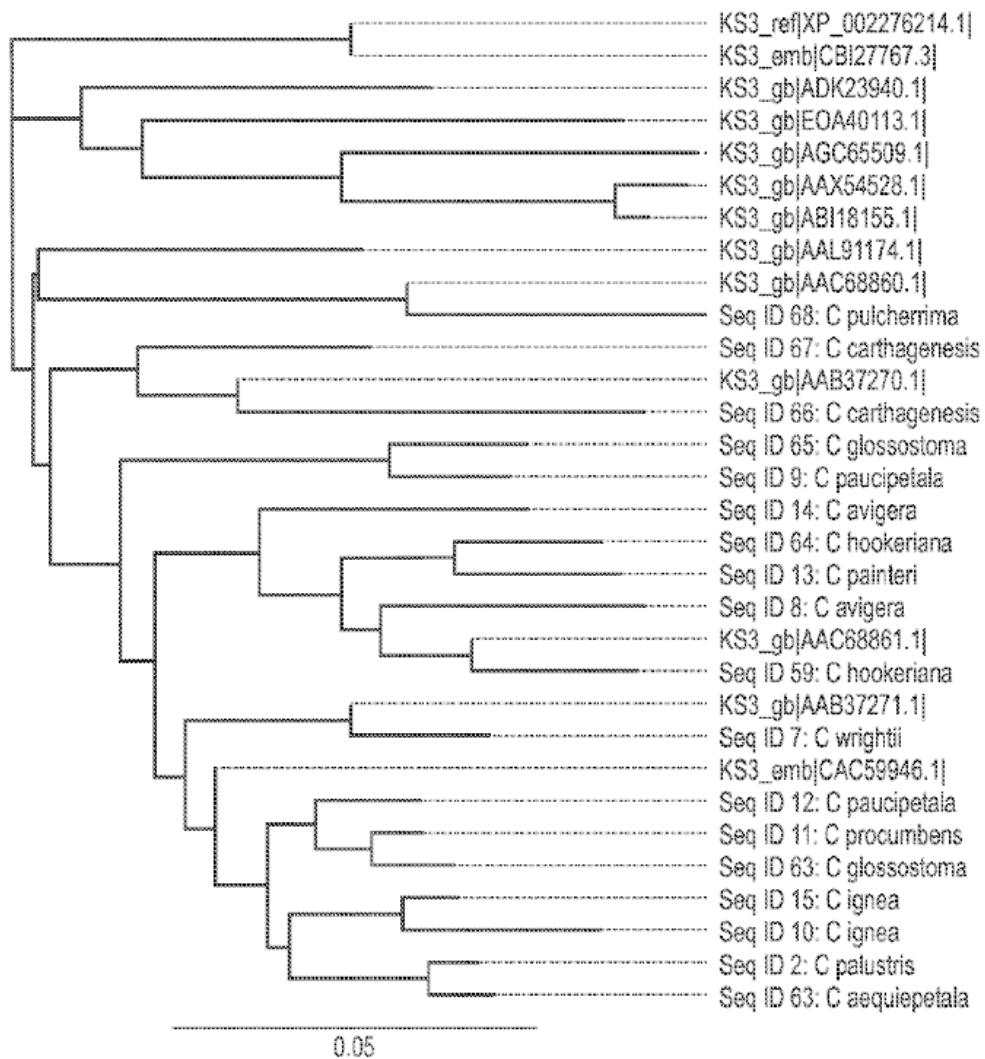


FIG. 1