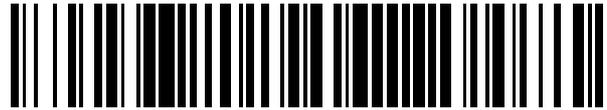


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 550 467**

51 Int. Cl.:

C12N 15/00 (2006.01)

C12P 1/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **07.10.2009 E 09819847 (6)**

97 Fecha y número de publicación de la concesión europea: **02.09.2015 EP 2352824**

54 Título: **Métodos y composiciones para producir aldehídos grasos**

30 Prioridad:

07.10.2008 US 103447 P

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

10.11.2015

73 Titular/es:

**REG LIFE SCIENCES, LLC (100.0%)
600 Gateway Boulevard
South San Francisco, CA 94080, US**

72 Inventor/es:

**ALIBHAI, MURTAZA F. y
HU, ZHIHAO**

74 Agente/Representante:

MILTENYI, Peter

ES 2 550 467 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

Métodos y composiciones para producir aldehídos grasos

Antecedentes de la invención

5 El petróleo es un recurso natural limitado que se encuentra en la Tierra en forma líquida, gaseosa o sólida. El petróleo está compuesto principalmente por hidrocarburos, que están compuestos principalmente por carbono e hidrógeno. También contiene cantidades significativas de otros elementos, tales como nitrógeno, oxígeno o azufre, en diferentes formas.

10 El petróleo es un recurso valioso, pero los productos petrolíferos se desarrollan con costes considerables, tanto desde el punto de vista económico como medioambiental. En primer lugar, deben descubrirse fuentes de petróleo. La exploración petrolífera es una empresa cara y arriesgada. El coste de explorar pozos en aguas profundas puede superar los 100 millones de dólares. Además, no hay ninguna garantía de que estos pozos contendrán petróleo. Se estima que sólo el 40% de los pozos perforados conducen a pozos productivos que generan hidrocarburos comerciales. Además del coste económico, la exploración petrolífera conlleva un alto coste medioambiental. Por ejemplo, la exploración en alta mar altera los entornos marinos circundantes.

15 Tras descubrirse un pozo productivo, debe extraerse el petróleo de la Tierra con un alto coste. Durante la recuperación primaria, la presión natural subterránea es suficiente para extraer aproximadamente el 20% del petróleo en el pozo. A medida que esta presión natural disminuye, se emplean métodos de recuperación secundaria, si resultan económicos. Generalmente, la recuperación secundaria implica aumentar la presión del pozo, por ejemplo, mediante inyección de agua, inyección de gas natural o elevación por gas. Usando métodos de recuperación secundaria, se recupera del 5% al 15% más del petróleo. Una vez agotados los métodos de recuperación secundaria, pueden usarse métodos de recuperación terciaria, si resultan económicos. Los métodos terciarios implican reducir la viscosidad del petróleo para hacer que sea más fácil de extraer. Usando métodos de recuperación terciaria, se recupera del 5% al 15% más del petróleo. Por tanto, incluso en las mejores circunstancias, sólo puede extraerse el 50% del petróleo en un pozo. La extracción petrolífera también conlleva un coste medioambiental. Por ejemplo, la extracción petrolífera puede dar como resultado grandes filtraciones de petróleo que suben hasta la superficie. Además, la perforación en alta mar implica el dragado del fondo marino lo cual altera o destruye el entorno marino circundante.

30 Dado que los depósitos de petróleo no se encuentran de manera uniforme por toda la Tierra, el petróleo debe transportarse grandes distancias desde regiones productoras de petróleo hasta regiones consumidoras de petróleo. Además de los costes de envío, también existe el riesgo medioambiental de devastadores derrames de petróleo.

35 En su forma natural, el petróleo crudo extraído de la Tierra tiene pocos usos comerciales. Es una mezcla de hidrocarburos (por ejemplo, parafinas (o alcanos), olefinas (o alquenos), alquinos, naftenos (o cicloalcanos), compuestos alifáticos, compuestos aromáticos, etc.) de diversa longitud y complejidad. Además, el petróleo crudo contiene otros compuestos orgánicos (por ejemplo, compuestos orgánicos que contienen nitrógeno, oxígeno, azufre, etc.) e impurezas (por ejemplo, azufre, sal, ácido, metales, etc.).

Por tanto, el petróleo crudo debe refinarse y purificarse antes de poder usarse comercialmente. Debido a su alta densidad de energía y su fácil capacidad de transporte, la mayor parte del petróleo se refina para dar combustibles, tales como combustibles para transporte (por ejemplo, gasolina, diésel, combustible para aviación, etc.), aceite para calefacción, gas licuado de petróleo, etc.

40 El petróleo crudo también es una fuente principal de materias primas para producir productos petroquímicos. Las dos clases principales de materias primas derivadas del petróleo son olefinas de cadena corta (por ejemplo, etileno y propileno) y compuestos aromáticos (por ejemplo, benceno e isómeros de xileno). Estas materias primas se derivan de hidrocarburos de cadena más larga en el petróleo crudo mediante craqueo del mismo con un coste considerable usando una variedad de métodos, tales como craqueo catalítico, craqueo por vapor o reformado catalítico. Estas materias primas se usan para producir productos petroquímicos, que no pueden refinarse directamente a partir de petróleo crudo, tales como monómeros, disolventes, detergentes o adhesivos.

50 Un ejemplo de una materia prima derivada de petróleo crudo es etileno. El etileno se usa para producir productos petroquímicos, tales como polietileno, etanol, óxido de etileno, etilenglicol, poliéster, éter de glicol, etoxilato, acetato de vinilo, 1,2-dicloroetano, tricloroetileno, tetracloroetileno, cloruro de vinilo y poli(cloruro de vinilo). Un ejemplo adicional de una materia prima es propileno, que se usa para producir alcohol isopropílico, acrilonitrilo, polipropileno, óxido de propileno, propilenglicol, éteres de glicol, butileno, isobutileno, 1,3-butadieno, elastómeros sintéticos, poliolefinas, alfa-olefinas, alcoholes grasos, ácido acrílico, polímeros acrílicos, cloruro de alilo, epíclorhidrina y resinas epoxídicas.

55 Estos productos petroquímicos pueden usarse entonces para preparar productos químicos especializados, tales como plásticos, resinas, fibras, elastómeros, productos farmacéuticos, lubricantes o geles. Productos químicos especializados particulares que pueden producirse a partir de materias primas petroquímicas son ácidos grasos, hidrocarburos (por ejemplo, de cadena larga, de cadena ramificada, saturados, insaturados, etc.), alcoholes grasos,

ésteres, aldehídos grasos, cetonas, lubricantes, etc.

Los aldehídos se usan para producir muchos productos químicos especializados. Por ejemplo, los aldehídos se usan para producir polímeros, resinas (por ejemplo, baquelita), tintes, saborizantes, plastificantes, perfumes, productos farmacéuticos y otros productos químicos. Algunos se usan como disolventes, conservantes o desinfectantes. Algunos compuestos naturales y sintéticos, tales como vitaminas y hormonas, son aldehídos. Además, muchos azúcares contienen grupos aldehído.

Obtener estos productos químicos especializados a partir del petróleo crudo requiere una inversión económica significativa así como una gran cantidad de energía. También es un procedimiento ineficaz porque con frecuencia los hidrocarburos de cadena larga en el petróleo crudo se craquean para producir monómeros más pequeños. Entonces estos monómeros se usan como materia prima para fabricar los productos químicos especializados más complejos.

Además de los problemas con la exploración, extracción, transporte y refinado del petróleo, el petróleo es un recurso limitado y cada vez más escaso. Una estimación del consumo de petróleo mundial es de 30 mil millones de barriles al año. Según algunas estimaciones, se predice que con los niveles de producción actuales, las reservas mundiales de petróleo se agotarán antes del año 2050.

Finalmente, la quema de combustibles basados en petróleo libera gases de efecto invernadero (por ejemplo, dióxido de carbono) y otras formas de contaminación atmosférica (por ejemplo, monóxido de carbono, dióxido de azufre, etc.). A medida que aumenta la demanda mundial de combustible, también aumenta la emisión de gases de efecto invernadero y otras formas de contaminación atmosférica. La acumulación de gases de efecto invernadero en la atmósfera puede conducir a un aumento del calentamiento global. Por tanto, además de dañar localmente el medio ambiente (por ejemplo, derrames de petróleo, dragado de entornos marinos, etc.), la quema de petróleo también daña el medio ambiente a nivel global.

Debido a los inherentes desafíos planteados por el petróleo, existe una necesidad de una fuente de petróleo renovable que no necesite explorarse, extraerse, transportarse largas distancias o refinarse sustancialmente como el petróleo. También existe una necesidad de una fuente de petróleo renovable que pueda producirse de manera económica sin crear el tipo de daño medioambiental producido por la industria petrolífera y la quema de combustibles basados en petróleo. Por motivos similares, también existe una necesidad de una fuente renovable de productos químicos que se derivan normalmente del petróleo.

Un método de producción de petróleo renovable es modificando por ingeniería microorganismos para producir productos petrolíferos renovables. Algunos microorganismos tienen una capacidad natural para producir productos químicos. Por ejemplo, se han usado levaduras durante siglos para producir etanol (por ejemplo, cerveza, vino, etc.). En los últimos años, mediante el desarrollo de biotecnologías avanzadas, es posible modificar por ingeniería metabólica un organismo para producir bioproductos que nunca antes se habían producido. Los productos, tales como productos químicos, derivados de estas actividades celulares se conocen como bioproductos. Los combustibles producidos por estas actividades celulares se conocen como biocombustibles. Los biocombustibles son un combustible renovable alternativo a combustibles basados en petróleo. Los biocombustibles pueden sustituir a cualquier combustible basado en petróleo (por ejemplo, gasolina, diésel, combustible para aviación, aceite para calefacción, etc.). Los biocombustibles pueden derivarse de fuentes renovables, tales como materia vegetal, materia animal o incluso productos residuales. Estas fuentes renovables se conocen colectivamente como biomasa. Una ventaja de los biocombustibles con respecto a combustibles basados en petróleo es que no requieren exploración o extracción cara y arriesgada. Además, los biocombustibles pueden producirse localmente. Por tanto, no requieren transporte a lo largo de largas distancias. Además, los biocombustibles pueden producirse directamente sin necesidad de refinado caro y que requiere mucha energía tal como se necesita con el refinado de petróleo crudo. En otras circunstancias, el biocombustible puede requerir un nivel limitado y rentable de refinado. Además, el uso de biocombustibles mejora el medio ambiente al reducir la cantidad de emisiones perjudiciales para el medio ambiente (por ejemplo, gases de efecto invernadero, contaminación atmosférica, etc.) liberadas durante la combustión. Por ejemplo, los biocombustibles mantienen un ciclo de carbono equilibrado porque los biocombustibles se producen a partir de biomasa, un recurso natural renovable. Aunque la quema de biocombustibles liberará carbono (por ejemplo, como dióxido de carbono), este carbono se recirculará durante la producción de biomasa (por ejemplo, el cultivo de cosechas), equilibrando así el ciclo de carbono al contrario que los combustibles basados en petróleo.

Por motivos similares, los productos químicos derivados biológicamente ofrecen las mismas ventajas que los biocombustibles con respecto a los combustibles basados en petróleo. Los productos químicos derivados biológicamente son una alternativa renovable a los productos petroquímicos. Los productos químicos derivados biológicamente, tales como hidrocarburos (por ejemplo, alcanos, alquenos o alquinos), alcoholes grasos, ésteres, ácidos grasos, aldehídos grasos y cetonas, son superiores a los productos petroquímicos porque se producen directamente sin refinado caro. Al contrario que los productos petroquímicos, los productos químicos derivados biológicamente no necesitan refinarse como el petróleo crudo para recuperar materias primas que entonces deben procesarse adicionalmente para producir productos petroquímicos más complejos. Los productos químicos derivados biológicamente se convierten directamente a partir de la biomasa en el producto químico deseado.

El n.º de registro de EBI, UNIPROT: A0R484 da a conocer una secuencia de aminoácidos indicada como supuesta ácido graso de cadena larga-CoA ligasa de *Mycobacterium smegmatis* (cepa ATCC 700084 / mc(2)155).

El documento WO 03/089610 describe proteínas de alcohol graso oxidasa (FAO) y métodos de producción de aldehídos durante la segunda etapa de la ruta de omega-oxidación de ácidos grasos.

- 5 El documento US 2004/180400 describe la identificación de la ácido carboxílico reductasa de *Nocardia* sp. NRRL 5646.

El documento WO 2007/136762 describe microorganismos modificados por ingeniería genética para producir productos de la ruta de biosíntesis de ácidos grasos (derivados de ácidos grasos) y métodos de su uso.

- 10 El documento US 2004/197896 describe un método para la identificación y selección de genes esenciales para la supervivencia o virulencia de especies de micobacterias mediante análisis genómico de micobacterias comparativo.

El documento US 2007/281345 describe métodos de preparación enzimática de aldehídos a partir de ácidos grasos usando una ácido carboxílico reductasa para reducir los ácidos grasos para dar sus correspondientes aldehídos.

Sumario de la invención

- 15 La invención se basa, al menos en parte, en la identificación de genes que codifican para polipéptidos de biosíntesis de aldehídos grasos. Por consiguiente, la invención presenta un método de producción de un aldehído graso según la reivindicación 1. El método incluye expresar en una célula huésped un gen que codifica para un polipéptido de biosíntesis de aldehídos grasos que comprende la secuencia de aminoácidos de SEQ ID NO: 22, o una variante de la misma con al menos el 80% de identidad de secuencia, y un gen que codifica para un polipéptido que tiene actividad tioesterasa. En algunas realizaciones, el método incluye además aislar el aldehído graso de la célula huésped. En algunas realizaciones, el aldehído graso está presente en el entorno extracelular. En determinadas realizaciones, el aldehído graso se aísla del entorno extracelular de la célula huésped. En algunas realizaciones, el aldehído graso se secreta a partir de la célula huésped. En realizaciones alternativas, el aldehído graso se transporta al entorno extracelular. En otras realizaciones, el aldehído graso se transporta de manera pasiva al entorno extracelular.

- 25 En algunas realizaciones, el polipéptido de biosíntesis de aldehídos grasos comprende la secuencia de aminoácidos de SEQ ID NO: 22, con una o más sustituciones, adiciones, inserciones o deleciones de aminoácidos, y el polipéptido tiene actividad ácido carboxílico reductasa, y al menos el 80% de identidad de secuencia. El polipéptido puede tener actividad ácido graso reductasa.

- 30 En algunas realizaciones, el polipéptido comprende una o más de las siguientes sustituciones de aminoácidos conservativas: reemplazo de un aminoácido alifático, tal como alanina, valina, leucina e isoleucina, por otro aminoácido alifático; reemplazo de una serina por una treonina; reemplazo de una treonina por una serina; reemplazo de un residuo ácido, tal como ácido aspártico y ácido glutámico, por otro residuo ácido; reemplazo de un residuo que porta un grupo amida, tal como asparagina y glutamina, por otro residuo que porta un grupo amida; intercambio de un residuo básico, tal como lisina y arginina, por otro residuo básico; y reemplazo de un residuo aromático, tal como fenilalanina y tirosina, por otro residuo aromático. En algunas realizaciones, el polipéptido tiene aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100 o más sustituciones, adiciones, inserciones o deleciones de aminoácidos. El polipéptido tiene actividad ácido carboxílico reductasa. El polipéptido puede tener actividad ácido graso reductasa.

- 40 En algunas realizaciones, el método incluye además cultivar la célula huésped en presencia de al menos un sustrato biológico para el polipéptido de biosíntesis de aldehídos grasos.

- 45 En algunas realizaciones, el método incluye además modificar la expresión de un gen que codifica para una ácido graso sintasa en la célula huésped. En determinadas realizaciones, modificar la expresión de un gen que codifica para una ácido graso sintasa incluye expresar un gen que codifica para una ácido graso sintasa en la célula huésped y/o aumentar la expresión o actividad de una ácido graso sintasa endógena en la célula huésped. En realizaciones alternativas, modificar la expresión de un gen que codifica para una ácido graso sintasa incluye atenuar un gen que codifica para una ácido graso sintasa en la célula huésped y/o disminuir la expresión o actividad de una ácido graso sintasa endógena en la célula huésped. En algunas realizaciones, la ácido graso sintasa es una tioesterasa. En realizaciones particulares, la tioesterasa se codifica por *tesA*, *tesA* sin secuencia líder, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA* o *fatA1*.

- 50 En otras realizaciones, la célula huésped se modifica por ingeniería genética para expresar un nivel atenuado de una enzima de degradación de ácidos grasos con respecto a una célula huésped silvestre. En algunas realizaciones, la célula huésped se modifica por ingeniería genética para expresar un nivel atenuado de una acil-CoA sintasa con respecto a una célula huésped silvestre. En realizaciones particulares, la célula huésped expresa un nivel atenuado de una acil-CoA sintasa codificada por *fadD*, *fadK*, *BH3103*, *yhfL*, *Pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa3p* o el gen que codifica para la proteína ZP_01644857. En determinadas realizaciones, la célula huésped modificada por ingeniería genética comprende una desactivación de uno o más genes que codifican

para una enzima de degradación de ácidos grasos, tales como los genes de acil-CoA sintasa mencionados anteriormente.

5 En aún otras realizaciones, la célula huésped se modifica por ingeniería genética para expresar un nivel atenuado de una enzima deshidratasa/isomerasa, tal como una enzima codificada por *fabA* o por un gen indicado en la figura 6. En algunas realizaciones, la célula huésped comprende una desactivación de *fabA* o un gen indicado en la figura 6. En otras realizaciones, la célula huésped se modifica por ingeniería genética para expresar un nivel atenuado de una cetooacil-ACP sintasa, tal como una enzima codificada por *fabB* o por un gen indicado en la figura 7. En otras realizaciones, la célula huésped comprende una desactivación de *fabB* o un gen indicado en la figura 7. En aún otras realizaciones, la célula huésped se modifica por ingeniería genética para expresar un nivel modificado de un gen que codifica para una enzima desaturasa, tal como *desA*.

10 En algunas realizaciones, el polipéptido es de una bacteria, una planta, un insecto, una levadura, un hongo o un mamífero.

15 En determinadas realizaciones, el polipéptido es de una célula de mamífero, célula vegetal, célula de insecto, célula de levadura, célula de hongo, célula de hongo filamentoso, célula bacteriana, o cualquier otro organismo descrito en el presente documento. En algunas realizaciones, la bacteria es una micobacteria seleccionada del grupo que consiste en *Mycobacterium smegmatis*, *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum* y *Mycobacterium ulcerans*. En otras realizaciones, la bacteria es *Nocardia* sp. NRRL 5646, *Nocardia farcinica*, *Streptomyces griseus*, *Salinispora arenicola* o *Clavibacter michiganensis*.

20 En algunas realizaciones, el método incluye además cultivar la célula huésped en presencia de al menos un sustrato biológico para el polipéptido de biosíntesis de aldehídos grasos.

25 La invención presenta un método de producción de un aldehído graso según la reivindicación 1. El método incluye expresar en una célula huésped un gen que codifica para un polipéptido de biosíntesis de aldehídos grasos que comprende una secuencia de aminoácidos que tiene al menos el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 91%, al menos aproximadamente el 92%, al menos aproximadamente el 93%, al menos aproximadamente el 94%, al menos aproximadamente el 95%, al menos aproximadamente el 96%, al menos aproximadamente el 97%, al menos aproximadamente el 98% o al menos aproximadamente el 99% de identidad de secuencia con respecto a la secuencia de aminoácidos de SEQ ID NO: 22. En algunas realizaciones, la secuencia de aminoácidos es la secuencia de aminoácidos de SEQ ID NO: 22.

30 En algunas realizaciones, el método incluye además aislar el aldehído graso de la célula huésped. En algunas realizaciones, el aldehído graso está presente en el entorno extracelular. En determinadas realizaciones, el aldehído graso se aísla del entorno extracelular de la célula huésped. En algunas realizaciones, el aldehído graso se secreta a partir de la célula huésped. En realizaciones alternativas, el aldehído graso se transporta al entorno extracelular. En otras realizaciones, el aldehído graso se transporta de manera pasiva al entorno extracelular.

35 También se da a conocer en el presente documento un método de producción de un aldehído graso, en el que el método incluye expresar en una célula huésped un polinucleótido que se hibrida con un complemento de la secuencia de nucleótidos de SEQ ID NO: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 113, 115, 117, 119 ó 121, o con un fragmento del mismo, en el que el polinucleótido codifica para un polipéptido que tiene actividad ácido carboxílico reductasa. En algunos aspectos, el polipéptido tiene actividad ácido graso reductasa.

40 En algunos aspectos, el polinucleótido se hibrida en condiciones de rigurosidad baja, rigurosidad media, rigurosidad alta o rigurosidad muy alta, con un complemento de la secuencia de nucleótidos de SEQ ID NO: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 113, 115, 117, 119 ó 121, o con un fragmento del mismo.

45 También se da a conocer en el presente documento un método de producción de un aldehído graso, en el que el método incluye expresar en una célula huésped un vector recombinante que comprende una secuencia de nucleótidos de biosíntesis de aldehídos grasos que tiene al menos aproximadamente el 70% de identidad de secuencia con respecto a la secuencia de nucleótidos de SEQ ID NO: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 113, 115, 117, 119 ó 121. En algunos aspectos, la secuencia de nucleótidos tiene al menos aproximadamente el 75%, al menos aproximadamente el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 91%, al menos aproximadamente el 92%, al menos aproximadamente el 93%, al menos aproximadamente el 94%, al menos aproximadamente el 95%, al menos aproximadamente el 96%, al menos aproximadamente el 97%, al menos aproximadamente el 98% o al menos aproximadamente el 99% de identidad de secuencia con respecto a la secuencia de nucleótidos de SEQ ID NO: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 113, 115, 117, 119 ó 121. En algunos aspectos, la secuencia de nucleótidos es la secuencia de nucleótidos de SEQ ID NO: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83,

85, 87, 89, 91, 113, 115, 117, 119 ó 121.

También se describe en el presente documento un método de producción de un aldehído graso, en el que el método incluye expresar en una célula huésped un gen que codifica para un polipéptido de biosíntesis de aldehídos grasos que comprende (i) SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 y SEQ ID NO: 10; (ii) SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13 o SEQ ID NO: 14; y/o (iii) SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 10 y SEQ ID NO: 11; en el que el polipéptido tiene actividad ácido carboxílico reductasa. En algunos aspectos, el polipéptido tiene actividad ácido graso reductasa.

En algunas realizaciones, el polipéptido tiene de aproximadamente 1.000 aminoácidos a aproximadamente 2.000 aminoácidos de longitud. En determinadas realizaciones, el polipéptido tiene aproximadamente 1.000 aminoácidos de longitud, aproximadamente 1.050 aminoácidos de longitud, aproximadamente 1.100 aminoácidos de longitud, aproximadamente 1.150 aminoácidos de longitud, aproximadamente 1.200 aminoácidos de longitud, aproximadamente 1.250 aminoácidos de longitud, aproximadamente 1.300 aminoácidos de longitud, aproximadamente 1.400 aminoácidos de longitud, aproximadamente 1.500 aminoácidos de longitud, aproximadamente 1.600 aminoácidos de longitud, aproximadamente 1.700 aminoácidos de longitud, aproximadamente 1.800 aminoácidos de longitud, aproximadamente 1.900 aminoácidos de longitud o aproximadamente 2.000 aminoácidos de longitud. En otras realizaciones, el polipéptido tiene hasta aproximadamente 1.500 aminoácidos de longitud, hasta aproximadamente 1.400 aminoácidos de longitud, hasta aproximadamente 1.300 aminoácidos de longitud, hasta aproximadamente 1.250 aminoácidos de longitud, hasta aproximadamente 1.200 aminoácidos de longitud, hasta aproximadamente 1.150 aminoácidos de longitud, hasta aproximadamente 1.100 aminoácidos de longitud, hasta aproximadamente 1.050 aminoácidos de longitud o hasta aproximadamente 1.000 aminoácidos de longitud.

En cualquiera de los aspectos de la invención descritos en el presente documento, la célula huésped puede seleccionarse del grupo que consiste en una célula de mamífero, célula vegetal, célula de insecto, célula de levadura, célula de hongo, célula de hongo filamentoso y célula bacteriana.

En algunas realizaciones, la célula huésped es una célula bacteriana Gram-positiva. En otras realizaciones, la célula huésped es una célula bacteriana Gram-negativa.

En algunas realizaciones, la célula huésped se selecciona del género *Escherichia*, *Bacillus*, *Lactobacillus*, *Rhodococcus*, *Pseudomonas*, *Aspergillus*, *Trichoderma*, *Neurospora*, *Fusarium*, *Humicola*, *Rhizomucor*, *Kluyveromyces*, *Pichia*, *Mucor*, *Myceliophthora*, *Penicillium*, *Phanerochaete*, *Pleurotus*, *Trametes*, *Chrysosporium*, *Saccharomyces*, *Stenotrophomonas*, *Schizosaccharomyces*, *Yarrowia* o *Streptomyces*.

En determinadas realizaciones, la célula huésped es una célula *Bacillus lentus*, una célula *Bacillus brevis*, una célula *Bacillus stearothermophilus*, una célula *Bacillus licheniformis*, una célula *Bacillus alkalophilus*, una célula *Bacillus coagulans*, una célula *Bacillus circulans*, una célula *Bacillus pumilis*, una célula *Bacillus thuringiensis*, una célula *Bacillus clausii*, una célula *Bacillus megaterium*, una célula *Bacillus subtilis* o una célula *Bacillus amyloliquefaciens*.

En otras realizaciones, la célula huésped es una célula *Trichoderma koningii*, una célula *Trichoderma viride*, una célula *Trichoderma reesei*, una célula *Trichoderma longibrachiatum*, una célula *Aspergillus awamori*, una célula *Aspergillus fumigates*, una célula *Aspergillus foetidus*, una célula *Aspergillus nidulans*, una célula *Aspergillus niger*, una célula *Aspergillus oryzae*, una célula *Humicola insolens*, una célula *Humicola lanuginosa*, una célula *Rhodococcus opacus*, una célula *Rhizomucor miehei* o una célula *Mucor michei*.

En aún otras realizaciones, la célula huésped es una célula *Streptomyces lividans* o una célula *Streptomyces murinus*.

En aún otras realizaciones, la célula huésped es una célula *Actinomycetes*.

En algunas realizaciones, la célula huésped es una célula *Saccharomyces cerevisiae*.

En realizaciones particulares, la célula huésped es una célula de una planta eucariota, alga, cianobacteria, bacteria verde del azufre, bacteria verde no del azufre, bacteria púrpura del azufre, bacteria púrpura no del azufre, extremófilo, levadura, hongo, organismos modificados por ingeniería de las mismas o un organismo sintético. En algunas realizaciones, la célula huésped depende de la luz o fija carbono. En algunas realizaciones, la célula huésped depende de la luz o fija carbono. En algunas realizaciones, la célula huésped tiene actividad autótrofa. En algunas realizaciones, la célula huésped tiene actividad fotoautótrofa, tal como en presencia de luz. En algunas realizaciones, la célula huésped es heterótrofa o mixótrofa en ausencia de luz. En determinadas realizaciones, la célula huésped es una célula de *Arabidopsis thaliana*, *Panicum virgatum*, *Miscanthus giganteus*, *Zea mays*, *Botryococcuse braunii*, *Chlamydomonas reinhardtii*, *Dunaliella salina*, *Synechococcus Sp. PCC 7002*, *Synechococcus Sp. PCC 7942*, *Synechocystis Sp. PCC 6803*, *Thermosynechococcus elongates BP-1*, *Chlorobium tepidum*, *Chloroflexus auranticus*, *Chromatium vinosum*, *Rhodospirillum rubrum*, *Rhodobacter capsulatus*, *Rhodospseudomonas palustris*, *Clostridium ljungdahlii*, *Clostridium thermocellum*, *Penicillium chrysogenum*, *Pichia pastoris*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pseudomonas fluorescens* o *Zymomonas mobilis*.

En otras realizaciones, la célula huésped es una célula CHO, una célula COS, una célula VERO, una célula BHK, una célula HeLa, una célula Cv1, una célula MDCK, una célula 293, una célula 3T3 o una célula PC12.

En aún otras realizaciones, la célula huésped es una célula *E. coli*. En determinadas realizaciones, la célula *E. coli* es un célula *E. coli* de la cepa B, cepa C, cepa K o cepa W.

5 También se da a conocer en el presente documento un método de producción de un aldehído graso, en el que el método incluye poner en contacto un sustrato con (i) un polipéptido de biosíntesis de aldehídos grasos que comprende la secuencia de aminoácidos de SEQ ID NO: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 114, 116, 118, 120 ó 122, o una
10 variante de la misma, o (ii) un polipéptido de biosíntesis de aldehídos grasos codificado por una secuencia de nucleótidos que tiene al menos aproximadamente el 70% de identidad con respecto a la secuencia de nucleótidos de SEQ ID NO: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 113, 115, 117, 119 ó 121, o una variante de la misma. En algunos aspectos, el método incluye además purificar el aldehído graso.

15 En algunas realizaciones, el polipéptido tiene una secuencia de aminoácidos que es idéntica en al menos el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 91%, al menos aproximadamente el 92%, al menos aproximadamente el 93%, al menos aproximadamente el 94%, al menos aproximadamente el 95%, al menos aproximadamente el 96%, al menos aproximadamente el 97%, al menos aproximadamente el 98% o al menos aproximadamente el 99% a la secuencia de aminoácidos de SEQ ID NO: 22. En algunas realizaciones, el polipéptido tiene la secuencia de aminoácidos de SEQ ID NO: 22.

20 También se da a conocer en el presente documento un método de producción de un aldehído graso, en el que el método incluye poner en contacto un sustrato con un polipéptido de biosíntesis de aldehídos grasos que comprende (i) SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 y SEQ ID NO: 10; (ii) SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13 o SEQ ID NO: 14; y/o (iii) SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 10 y SEQ ID NO: 11; en el que el polipéptido tiene actividad ácido carboxílico reductasa. En algunos aspectos, el polipéptido tiene actividad ácido
25 graso reductasa.

En cualquiera de los aspectos de la invención descritos en el presente documento, los métodos pueden producir aldehídos grasos que comprenden un aldehído graso C₆-C₂₆. En algunas realizaciones, el aldehído graso comprende un aldehído graso C₆, C₇, C₈, C₉, C₁₀, C₁₁, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇, C₁₈, C₁₉, C₂₀, C₂₁, C₂₂, C₂₃, C₂₄, C₂₅ o C₂₆. En realizaciones particulares, el aldehído graso es decanal, dodecanal, miristal o hexadecal.

30 En otras realizaciones, el aldehído graso comprende un aldehído graso de cadena lineal. En otras realizaciones, el aldehído graso comprende un aldehído graso de cadena ramificada. En aún otras realizaciones, el aldehído graso comprende un resto cíclico.

35 En algunas realizaciones, el aldehído graso es un aldehído graso insaturado. En otras realizaciones, el aldehído graso es un aldehído graso monoinsaturado. En aún otras realizaciones, el aldehído graso es un aldehído graso saturado.

40 En cualquiera de los aspectos de la invención descritos en el presente documento, un sustrato para un polipéptido de biosíntesis de aldehídos grasos puede ser un ácido graso. En algunas realizaciones, el ácido graso comprende un ácido graso C₆-C₂₆. En algunas realizaciones, el ácido graso comprende un ácido graso C₆, C₇, C₈, C₉, C₁₀, C₁₁, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇, C₁₈, C₁₉, C₂₀, C₂₁, C₂₂, C₂₃, C₂₄, C₂₅ o C₂₆. En realizaciones particulares, el ácido graso es un ácido graso C₆, C₈, C₁₀, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇ o C₁₈.

En otras realizaciones, el ácido graso comprende un ácido graso de cadena lineal. En otras realizaciones, el ácido graso comprende un ácido graso de cadena ramificada. En aún otras realizaciones, el ácido graso comprende un resto cíclico.

45 En algunas realizaciones, el aldehído graso es un aldehído graso insaturado. En otras realizaciones, el aldehído graso es un aldehído graso monoinsaturado. En determinadas realizaciones, el aldehído graso insaturado es un aldehído graso insaturado C₆:1, C₇:1, C₈:1, C₉:1, C₁₀:1, C₁₁:1, C₁₂:1, C₁₃:1, C₁₄:1, C₁₅:1, C₁₆:1, C₁₇:1, C₁₈:1, C₁₉:1, C₂₀:1, C₂₁:1, C₂₂:1, C₂₃:1, C₂₄:1, C₂₅:1 o C₂₆:1. En aún otras realizaciones, el aldehído graso está insaturado en la posición omega-7. En determinadas realizaciones, el aldehído graso insaturado comprende un doble enlace en cis.

50 En otro aspecto, en el presente documento se da a conocer un microorganismo modificado por ingeniería genética que comprende una secuencia de control exógena incorporada de manera estable en el ADN genómico del microorganismo en el sentido de 5' de un polinucleótido que comprende una secuencia de nucleótidos que tiene al menos aproximadamente el 70% de identidad de secuencia con respecto a la secuencia de nucleótidos de SEQ ID NO: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75,
55 77, 79, 81, 83, 85, 87, 89, 91, 113, 115, 117, 119 ó 121, en el que el microorganismo produce un nivel aumentado de un aldehído graso con respecto a un microorganismo silvestre.

En algunos aspectos, la secuencia de nucleótidos tiene al menos aproximadamente el 75%, al menos aproximadamente el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 91%, al menos aproximadamente el 92%, al menos aproximadamente el 93%, al menos aproximadamente el 94%, al menos aproximadamente el 95%, al menos aproximadamente el 96%, al menos aproximadamente el 97%, al menos aproximadamente el 98% o al menos aproximadamente el 99% de identidad de secuencia con respecto a la secuencia de nucleótidos de SEQ ID NO: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 113, 115, 117, 119 ó 121. En algunos aspectos, la secuencia de nucleótidos es SEQ ID NO: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 113, 115, 117, 119 ó 121.

En algunos aspectos, el polinucleótido es endógeno para el microorganismo.

En otros aspectos, el microorganismo se modifica por ingeniería genética para expresar un nivel modificado de un gen que codifica para una ácido graso sintasa en la célula huésped. En determinados aspectos, el microorganismo expresa un gen que codifica para una ácido graso sintasa y/o tiene una expresión o actividad aumentada de una ácido graso sintasa endógena. En aspectos alternativos, el microorganismo tiene expresión atenuada de un gen que codifica para una ácido graso sintasa en la célula huésped y/o tiene una expresión o actividad reducida de una ácido graso sintasa endógena. En algunos aspectos, la ácido graso sintasa es una tioesterasa. En aspectos particulares, la tioesterasa se codifica por *tesA*, *tesA* sin secuencia líder, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA* o *fatA1*.

En otros aspectos, el microorganismo se modifica por ingeniería genética para expresar un nivel atenuado de una enzima de degradación de ácidos grasos con respecto a un microorganismo silvestre. En algunos aspectos, el microorganismo expresa un nivel atenuado de una acil-CoA sintasa con respecto a un microorganismo silvestre. En aspectos particulares, el microorganismo expresa un nivel atenuado de una acil-CoA sintasa codificada por *fadD*, *fadK*, *BH3103*, *yhfL*, *Pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa3p* o el gen que codifica para la proteína ZP_01644857. En determinados aspectos, el microorganismo comprende una desactivación de uno o más genes que codifican para una enzima de degradación de ácidos grasos, tales como los genes de acil-CoA sintasa mencionados anteriormente.

En aún otros aspectos, el microorganismo se modifica por ingeniería genética para expresar un nivel atenuado de una enzima deshidratasa/isomerasa, tal como una enzima codificada por *fabA* o por un gen indicado en la figura 6. En algunos aspectos, el microorganismo comprende una desactivación de *fabA* o un gen indicado en la figura 6. En otros aspectos, el microorganismo se modifica por ingeniería genética para expresar un nivel atenuado de una cetoacil-ACP sintasa, tal como una enzima codificada por *fabB* o por un gen indicado en la figura 7. En otros aspectos, el microorganismo comprende una desactivación de *fabB* o un gen indicado en la figura 7. En aún otros aspectos, el microorganismo se modifica por ingeniería genética para expresar un nivel modificado de un gen que codifica para una enzima desaturasa, tal como *desA*.

En algunos aspectos, el microorganismo es una bacteria. En determinados aspectos, la bacteria es una bacteria Gram-negativa o una Gram-positiva.

En algunos aspectos, el microorganismo es una micobacteria seleccionada del grupo que consiste en *Mycobacterium smegmatis*, *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum* y *Mycobacterium ulcerans*.

En otros aspectos, el microorganismo es *Nocardia* sp. NRRL 5646, *Nocardia farcinica*, *Streptomyces griseus*, *Salinispora arenicola* o *Clavibacter michiganensis*.

En otro aspecto, la invención presenta un aldehído graso producido mediante cualquiera de los métodos o cualquiera de los microorganismos descritos en el presente documento.

En algunas realizaciones, el aldehído graso tiene un $\delta^{13}\text{C}$ de aproximadamente -15,4 o más. En determinadas realizaciones, el aldehído graso tiene un $\delta^{13}\text{C}$ de aproximadamente -15,4 a aproximadamente -10,9, o de aproximadamente -13,92 a aproximadamente -13,84.

En algunas realizaciones, el aldehído graso tiene una $f_{\text{M}}^{14}\text{C}$ de al menos aproximadamente 1,003. En determinadas realizaciones, el aldehído graso tiene una $f_{\text{M}}^{14}\text{C}$ de al menos aproximadamente 1,01 o al menos aproximadamente 1,5. En algunas realizaciones, el aldehído graso tiene una $f_{\text{M}}^{14}\text{C}$ de aproximadamente 1,111 a aproximadamente 1,124.

En cualquiera de los aspectos descritos en el presente documento, se produce un aldehído graso con un rendimiento de aproximadamente 25 mg/l, aproximadamente 50 mg/l, aproximadamente 75 mg/l, aproximadamente 100 mg/l, aproximadamente 125 mg/l, aproximadamente 150 mg/l, aproximadamente 175 mg/l, aproximadamente 200 mg/l, aproximadamente 225 mg/l, aproximadamente 250 mg/l, aproximadamente 275 mg/l, aproximadamente 300 mg/l, aproximadamente 325 mg/l, aproximadamente 350 mg/l, aproximadamente 375 mg/l, aproximadamente 400 mg/l, aproximadamente 425 mg/l, aproximadamente 450 mg/l, aproximadamente 475 mg/l, aproximadamente 500 mg/l, aproximadamente 525 mg/l, aproximadamente 550 mg/l, aproximadamente 575 mg/l, aproximadamente

600 mg/l, aproximadamente 625 mg/l, aproximadamente 650 mg/l, aproximadamente 675 mg/l, aproximadamente 700 mg/l, aproximadamente 725 mg/l, aproximadamente 750 mg/l, aproximadamente 775 mg/l, aproximadamente 800 mg/l, aproximadamente 825 mg/l, aproximadamente 850 mg/l, aproximadamente 875 mg/l, aproximadamente 900 mg/l, aproximadamente 925 mg/l, aproximadamente 950 mg/l, aproximadamente 975 mg/l, aproximadamente 1000 g/L, aproximadamente 1050 mg/l, aproximadamente 1075 mg/l, aproximadamente 1100 mg/l, aproximadamente 1125 mg/l, aproximadamente 1150 mg/l, aproximadamente 1175 mg/l, aproximadamente 1200 mg/l, aproximadamente 1225 mg/l, aproximadamente 1250 mg/l, aproximadamente 1275 mg/l, aproximadamente 1300 mg/l, aproximadamente 1325 mg/l, aproximadamente 1350 mg/l, aproximadamente 1375 mg/l, aproximadamente 1400 mg/l, aproximadamente 1425 mg/l, aproximadamente 1450 mg/l, aproximadamente 1475 mg/l, aproximadamente 1500 mg/l, aproximadamente 1525 mg/l, aproximadamente 1550 mg/l, aproximadamente 1575 mg/l, aproximadamente 1600 mg/l, aproximadamente 1625 mg/l, aproximadamente 1650 mg/l, aproximadamente 1675 mg/l, aproximadamente 1700 mg/l, aproximadamente 1725 mg/l, aproximadamente 1750 mg/l, aproximadamente 1775 mg/l, aproximadamente 1800 mg/l, aproximadamente 1825 mg/l, aproximadamente 1850 mg/l, aproximadamente 1875 mg/l, aproximadamente 1900 mg/l, aproximadamente 1925 mg/l, aproximadamente 1950 mg/l, aproximadamente 1975 mg/l, aproximadamente 2000 mg/l o más.

En cualquiera de los aspectos descritos en el presente documento, se produce un aldehído graso en una célula huésped o un microorganismo descrito en el presente documento a partir de una fuente carbono.

Las siguientes figuras se presentan únicamente con fines de ilustración.

20 Breve descripción de los dibujos

La figura 1 es un esquema de una nueva ruta para la producción de aldehídos grasos.

La figura 2 es una lista de la secuencia de nucleótidos y la secuencia de aminoácidos correspondiente del gen *car* de *Nocardia sp.* NRRL 5646.

La figura 3 es una lista de motivos de secuencia de aminoácidos para homólogos de CAR.

25 La figura 4 es una lista de secuencias de nucleótidos y de aminoácidos de genes homólogos de *car*.

La figura 5 es una tabla que identifica genes a modo de ejemplo que pueden expresarse, sobreexpresarse o atenuarse para aumentar la producción de sustratos particulares.

La figura 6 es una tabla de secuencias de nucleótidos y de aminoácidos para genes relacionados con *fabA*.

La figura 7 es una tabla de secuencias de nucleótidos y de aminoácidos para genes relacionados con *fabB*.

30 Descripción detallada de la invención

A menos que se defina lo contrario, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado que el que entiende comúnmente un experto habitual en la técnica a la que pertenece esta invención. Aunque en la práctica o las pruebas de la presente invención pueden usarse métodos y materiales similares o equivalentes a los descritos en el presente documento, a continuación se describen métodos y materiales adecuados.

Otras características y ventajas de la invención resultarán evidentes a partir de la siguiente descripción detallada, y a partir de las reivindicaciones.

Definiciones

40 A lo largo de toda la memoria descriptiva, puede hacerse referencia usando un nombre de gen o nombre de polipéptido abreviado, pero se entiende que tal nombre de gen o polipéptido abreviado representa el género de genes o polipéptidos. Tales nombres de genes incluyen todos los genes que codifican para el mismo polipéptido y para polipéptidos homólogos que tienen la misma función fisiológica. Los nombres de polipéptidos incluyen todos los polipéptidos que tienen la misma actividad (por ejemplo, que catalizan la misma reacción química fundamental).

45 A menos que se indique lo contrario, los números de registro a los que se hace referencia en el presente documento se derivan de la base de datos de NCBI (National Center for Biotechnology Information) mantenida por el National Institute of Health, EE.UU. A menos que se indique lo contrario, los números de registro son según se proporcionan en la base de datos en octubre de 2008.

50 Los números de EC los establece el Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) (disponible en <http://www.chem.qmul.ac.uk/iubmb/enzyme/>). Los números de EC a los que se hace referencia en el presente documento se derivan de la base de datos KEGG Ligand, mantenida por la Kyoto Encyclopedia of Genes and Genomics, patrocinada en parte por la Universidad de Tokio. A menos que se indique lo contrario, los números de EC son según se proporcionan en la base de datos en octubre de 2008.

Los artículos “un” y “una” se usan en el presente documento para hacer referencia a uno o a más de uno (es decir, a al menos uno) del objeto gramatical del artículo. A modo de ejemplo, “un elemento” significa un elemento o más de un elemento.

5 El término “aproximadamente” se usa en el presente documento para querer decir un valor \pm el 20% de un valor numérico dado. Por tanto, “aproximadamente el 60%” significa un valor de entre $60 \pm$ (el 20% de 60) (es decir, entre 48 y 70).

Tal como se usa en el presente documento, el término “atenuar” significa debilitar, reducir o disminuir. Por ejemplo, un polipéptido puede atenuarse modificando el polipéptido para reducir su actividad (por ejemplo, modificando una secuencia de nucleótidos que codifica para el polipéptido).

10 Tal como se usa en el presente documento, el término “biomasa” se refiere a cualquier material biológico del que se deriva una fuente de carbono. En algunos casos, se procesa una biomasa para dar una fuente de carbono, que es adecuada para la bioconversión. En otros casos, la biomasa puede no requerir un procesamiento adicional para dar una fuente de carbono. La fuente de carbono puede convertirse en un biocombustible. Una fuente de biomasa a modo de ejemplo es materia vegetal o vegetación. Por ejemplo, puede usarse maíz, caña de azúcar o césped de pradera como biomasa. Otro ejemplo no limitativo de biomasa son residuos metabólicos, tales como materia animal, por ejemplo estiércol de vacas. Además, la biomasa puede incluir algas y otras plantas marinas. La biomasa también incluye productos residuales de la industria, agricultura, silvicultura y domésticos. Ejemplos de tales productos residuales que pueden usarse como biomasa son residuos de fermentación, ensilado, paja, madera, aguas residuales, basura, residuos urbanos celulósicos y sobras de comida. La biomasa también incluye fuentes de carbono, tales como hidratos de carbono (por ejemplo, monosacáridos, disacáridos o polisacáridos).

25 Tal como se usa en el presente documento, la frase “fuente de carbono” se refiere a un sustrato o compuesto adecuado para usarse como fuente de carbono para el crecimiento de células procariotas o eucariotas sencillas. Las fuentes de carbono pueden estar en diversas formas, incluyendo, pero sin limitarse a, polímeros, hidratos de carbono, ácidos, alcoholes, aldehídos, cetonas, aminoácidos, péptidos y gases (por ejemplo, CO y CO₂). Estos incluyen, por ejemplo, diversos monosacáridos, tales como glucosa, fructosa, manosa y galactosa; oligosacáridos, tales como fructo-oligosacárido y galacto-oligosacárido; polisacáridos tales como xilosa y arabinosa; disacáridos, tales como sacarosa, maltosa y turanosa; material celulósico, tal como metilcelulosa y carboximetilcelulosa sódica; ésteres de ácidos grasos saturados o insaturados, tales como succinato, lactato y acetato; alcoholes, tales como etanol, metanol y glicerol, o mezclas de los mismos. La fuente de carbono también puede ser un producto de fotosíntesis, incluyendo, pero sin limitarse a, glucosa. Una fuente de carbono preferida es biomasa. Otra fuente de carbono preferida es glucosa.

35 Una secuencia de nucleótidos es “complementaria” a otra secuencia de nucleótidos si cada una de las bases de las dos secuencias se aparean (es decir, pueden formar pares de bases de Watson-Crick). El término “cadena complementaria” se usa en el presente documento de manera intercambiable con el término “complemento”. El complemento de una cadena de ácido nucleico puede ser el complemento de una cadena codificante o el complemento de una cadena no codificante.

40 Tal como se usa en el presente documento, el término “condiciones suficientes para permitir la expresión” significa cualquier condición que permite que una célula huésped produzca un producto deseado, tal como un polipéptido o aldehído graso descrito en el presente documento. Las condiciones adecuadas incluyen, por ejemplo, condiciones de fermentación. Las condiciones de fermentación pueden comprender muchos parámetros, tales como intervalos de temperatura, niveles de aireación y composición de medios. Cada una de estas condiciones, de manera individual y en combinación, permite que la célula huésped crezca. Los medios de cultivo a modo de ejemplo incluyen caldos o geles. Generalmente, el medio incluye una fuente de carbono, tal como glucosa, fructosa, celulosa, o similar, que puede metabolizarse directamente por una célula huésped. Además, pueden usarse enzimas en el medio para facilitar la movilización (por ejemplo, la despolimerización de almidón o celulosa para dar azúcares fermentables) y el posterior metabolismo de la fuente de carbono.

50 Para determinar si las condiciones son suficientes para permitir la expresión, puede cultivarse una célula huésped, por ejemplo, durante aproximadamente 4, 8, 12, 24, 36 ó 48 horas. Durante y/o tras el cultivo, pueden obtenerse muestras y analizarse para determinar si las condiciones permiten la expresión. Por ejemplo, las células huésped en la muestra o el medio en el que se hicieron crecer las células huésped pueden someterse a prueba para determinar la presencia de un producto deseado. Cuando se somete a prueba para determinar la presencia de un producto, pueden usarse ensayos, tales como, pero sin limitarse a, CCF, HPLC, CG/FID, CG/EM, CL/EM, EM.

55 Se entiende que los polipéptidos descritos en el presente documento pueden tener sustituciones de aminoácidos conservativas o no esenciales adicionales, que no tienen un efecto sustancial sobre las funciones del polipéptido. Puede determinarse si se tolerará o no una sustitución particular (es decir, no afectará adversamente a las propiedades biológicas deseadas, tales como actividad ácido carboxílico reductasa) tal como se describe en Bowie *et al.* Science (1990) 247:1306 1310. Una “sustitución de aminoácidos conservativa” es una en la que el residuo de aminoácido se reemplaza por un residuo de aminoácido que tiene una cadena lateral similar. En la técnica se han definido familias de residuos de aminoácido que tienen cadenas laterales similares. Estas familias incluyen

aminoácidos con cadenas laterales básicas (por ejemplo, lisina, arginina, histidina), cadenas laterales ácidas (por ejemplo, ácido aspártico, ácido glutámico), cadenas laterales polares sin carga (por ejemplo, glicina, asparagina, glutamina, serina, treonina, tirosina, cisteína), cadenas laterales no polares (por ejemplo, alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano), cadenas laterales con ramificación beta (por ejemplo, treonina, valina, isoleucina) y cadenas laterales aromáticas (por ejemplo, tirosina, fenilalanina, triptófano, histidina).

Tal como se usa en el presente documento, “elemento de control” significa un elemento de control de la transcripción. Los elementos de control incluyen promotores y potenciadores. El término “elemento promotor”, “promotor” o “secuencia promotora” se refiere a una secuencia de ADN que funciona como interruptor que activa la expresión de un gen. Si el gen está activado, se dice que se transcribe o participa en la transcripción. La transcripción implica la síntesis de ARNm a partir del gen. Por tanto, un promotor sirve como elemento regulador de la transcripción y también proporciona un sitio para el inicio de la transcripción del gen para dar ARNm. Los elementos de control interaccionan específicamente con proteínas celulares implicadas en la transcripción (Maniatis *et al*, Science 236:1237, 1987).

Tal como se usa en el presente documento, el término “ácido graso” significa un ácido carboxílico que tiene la fórmula RCOOH. R representa un grupo alifático, preferiblemente un grupo alquilo. R puede comprender entre aproximadamente 4 y aproximadamente 22 átomos de carbono. Los ácidos grasos pueden estar saturados, monoinsaturados o poliinsaturados. En una realización preferida, el ácido graso se produce a partir de una ruta de biosíntesis de ácidos grasos.

Tal como se usa en el presente documento, el término “ruta de biosíntesis de ácidos grasos” significa una ruta de biosíntesis que produce ácidos grasos. La ruta de biosíntesis de ácidos grasos incluye enzimas de ácidos grasos que pueden modificarse por ingeniería, tal como se describe en el presente documento, para producir ácidos grasos, y en algunas realizaciones pueden expresarse con enzimas adicionales para producir ácidos grasos que tienen características deseadas de la cadena de carbono.

Tal como se usa en el presente documento, el término “enzima de degradación de ácidos grasos” significa una enzima implicada en la descomposición o conversión de un ácido graso o derivado de ácido graso para dar otro producto. Un ejemplo no limitativo de una enzima de degradación de ácidos grasos es una acil-CoA sintasa. En el presente documento se describen ejemplos adicionales de enzimas de degradación de ácidos grasos.

Tal como se usa en el presente documento, el término “derivado de ácido graso” significa productos producidos en parte a partir de la ruta de biosíntesis de ácidos grasos del organismo huésped de producción. “Derivado de ácido graso” también incluye productos producidos en parte a partir de acil-ACP o derivados de acil-ACP. La ruta de biosíntesis de ácidos grasos incluye enzimas ácido graso sintetas que pueden modificarse por ingeniería tal como se describe en el presente documento para producir derivados de ácidos grasos, y en algunos ejemplos pueden expresarse con enzimas adicionales para producir derivados de ácidos grasos que tienen características deseadas de la cadena de carbono. Los derivados de ácidos grasos a modo de ejemplo incluyen, por ejemplo, ácidos grasos, acil-CoA, aldehído graso, alcoholes de cadena corta y larga, hidrocarburos, alcoholes grasos y ésteres (por ejemplo, ceras, ésteres de ácidos grasos o ésteres grasos).

Tal como se usa en el presente documento, el término “enzima de derivados de ácidos grasos” significa cualquier enzima que puede expresarse o sobreexpresarse en la producción de derivados de ácidos grasos. Estas enzimas pueden ser parte de la ruta de biosíntesis de ácidos grasos. Los ejemplos no limitativos de enzimas de derivados de ácidos grasos incluyen ácido graso sintetas, tioesterasas, acil-CoA sintetas, acil-CoA reductasas, alcohol deshidrogenasas, alcohol aciltransferasas, acil-CoA reductasas formadoras de alcoholes grasos, ácido graso (ácido carboxílico) reductasas, aldehído reductasas, acil-ACP reductasas, ácido graso hidroxilasas, acil-CoA desaturasas, acil-ACP desaturasas, acil-CoA oxidasas, acil-CoA deshidrogenasas, éster sintetas y polipéptidos de biosíntesis de alcanos, etc. Las enzimas de derivados de ácidos grasos pueden convertir un sustrato en un derivado de ácido graso. En algunos ejemplos, el sustrato puede ser un derivado de ácido graso que la enzima de derivados de ácidos grasos convierte en un derivado de ácido graso diferente.

Tal como se usa en el presente documento, “enzima de ácidos grasos” significa cualquier enzima implicada en la biosíntesis de ácidos grasos. Las enzimas de ácidos grasos pueden modificarse en células huésped para producir ácidos grasos. Los ejemplos no limitativos de enzimas de ácidos grasos incluyen ácido graso sintetas y tioesterasas. En el presente documento se describen ejemplos adicionales de enzimas de ácidos grasos.

Tal como se usa en el presente documento, “ácido graso sintasa” significa cualquier enzima implicada en la biosíntesis de ácidos grasos. Pueden expresarse o sobreexpresarse ácido graso sintetas en células huésped para producir ácidos grasos. Un ejemplo no limitativo de una ácido graso sintasa es una tioesterasa. En el presente documento se describen ejemplos adicionales de ácido graso sintetas.

Tal como se usa en el presente documento, “aldehído graso” significa un aldehído que tiene la fórmula RCHO caracterizado por un grupo carbonilo insaturado (C=O). En una realización preferida, el aldehído graso es cualquier aldehído producido a partir de un ácido graso o derivado de ácido graso. En una realización, el grupo R tiene al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 ó 20 carbonos de longitud.

R puede ser una cadena lineal o ramificada. Las cadenas ramificadas pueden tener uno o más puntos de ramificación. Además, las cadenas ramificadas pueden incluir ramificaciones cíclicas.

Además, R puede estar saturado o insaturado. Si está insaturado, R puede tener uno o más puntos de insaturación.

En una realización, el aldehído graso se produce de manera biosintética.

5 Los aldehídos grasos tienen muchos usos. Por ejemplo, pueden usarse aldehídos grasos para producir muchos productos químicos especializados. Por ejemplo, se usan aldehídos grasos para producir polímeros, resinas, tintes, saborizantes, plastificantes, perfumes, productos farmacéuticos y otros productos químicos. Algunos se usan como disolventes, conservantes o desinfectantes. Algunos compuestos naturales y sintéticos, tales como vitaminas y hormonas, son aldehídos.

10 Los términos “polipéptido de biosíntesis de aldehídos grasos”, “ácido carboxílico reductasa” y “CAR” se usan de manera intercambiable en el presente documento.

15 Tal como se usa en el presente documento, “fracción de carbono moderno” o “ f_M ” tiene el mismo significado que el definido por los materiales de referencia patrón (SRM) 4990B y 4990C del National Institute of Standards and Technology (NIST), conocidos como patrones de ácidos oxálicos HOxI y HOxII, respectivamente. La definición fundamental se refiere a 0,95 veces la razón de isótopos $^{14}\text{C} / ^{12}\text{C}$ de HOxI (con referencia en AD 1950). Esto es aproximadamente equivalente a la madera previa a la revolución industrial corregida para la descomposición. Para la biosfera viva actual (material vegetal), f_M es de aproximadamente 1,1.

20 “Desactivación génica”, tal como se usa en el presente documento, se refiere a un procedimiento mediante el cual un gen que codifica para una proteína diana se modifica o inactiva para reducir o eliminar la función de la proteína intacta. La inactivación del gen puede realizarse mediante métodos generales tales como mutagénesis por irradiación UV o tratamiento con N-metil-N'-nitro-N-nitrosoguanidina, mutagénesis dirigida al sitio, recombinación homóloga, mutagénesis por inserción-delección o “integración impulsada por Red” (Datsenko *et al.*, Proc. Natl. Acad. Sci. USA, 97:6640-45, 2000). Por ejemplo, en una realización, se introduce un constructo en una célula huésped, de tal manera que es posible seleccionar acontecimientos de recombinación homóloga en la célula huésped. Un experto en la técnica puede diseñar fácilmente un constructo desactivado que incluye genes de selección tanto positiva como negativa para seleccionar eficazmente células transfectadas que experimentan un acontecimiento de recombinación homóloga con el constructo. La alteración en la célula huésped puede obtenerse, por ejemplo, reemplazando mediante una recombinación de cruce simple o doble una secuencia de ADN silvestre por una secuencia de ADN que contiene la alteración. Para una selección conveniente de transformantes, la alteración puede ser, por ejemplo, una secuencia de ADN que codifica para un marcador de resistencia a antibióticos o un gen que complementa una posible auxotrofia de la célula huésped. Las mutaciones incluyen, pero no se limitan a, mutaciones por delección-inserción. Un ejemplo de una alteración de este tipo incluye una alteración génica, es decir, una perturbación de un gen de tal manera que el producto que normalmente se produce a partir de este gen no se produce en una forma funcional. Esto puede deberse a una delección completa, una delección e inserción de un marcador selectivo, una inserción de un marcador selectivo, una mutación del marco de lectura, una delección en marco o una mutación puntual que conduce a la terminación prematura. En algunos casos, todo el ARNm para el gen está ausente. En otras situaciones, la cantidad de ARNm producida varía.

40 Pueden realizarse cálculos de “homología” entre dos secuencias de la siguiente manera. Las secuencias se alinean con fines de comparación óptima (por ejemplo, pueden introducirse huecos en una o ambas de una primera y una segunda secuencia de aminoácidos o de ácido nucleico para una alineación óptima y pueden descartarse secuencias no homólogas con fines de comparación). La longitud de una secuencia de referencia que se alinea con fines de comparación tiene al menos aproximadamente el 30%, preferiblemente al menos aproximadamente el 40%, más preferiblemente al menos aproximadamente el 50%, incluso más preferiblemente al menos aproximadamente el 60% e incluso más preferiblemente al menos aproximadamente el 70%, al menos aproximadamente el 80%, al menos aproximadamente el 90% o aproximadamente el 100% de la longitud de la secuencia de referencia. Entonces se comparan los residuos de aminoácido o nucleótidos en posiciones de aminoácido o posiciones de nucleótido correspondientes. Cuando una posición en la primera secuencia está ocupada por el mismo residuo de aminoácido o nucleótido que la posición correspondiente de la segunda secuencia, entonces las moléculas son idénticas en esa posición (tal como se usa en el presente documento, “identidad” de aminoácido o ácido nucleico es equivalente a “homología” de aminoácido o ácido nucleico). La identidad en porcentaje entre las dos secuencias es una función del número de posiciones idénticas compartidas por las secuencias, teniendo en cuenta el número de huecos y la longitud de cada hueco, que se necesita introducir para obtener una alineación óptima de las dos secuencias.

55 La comparación de secuencias y la determinación de la homología en porcentaje entre dos secuencias pueden lograrse usando un algoritmo matemático. La homología en porcentaje entre dos secuencias de aminoácidos se determina usando el algoritmo de Needleman y Wunsch (1970), J. Mol. Biol. 48:444-453, que se ha incorporado en el programa GAP en el paquete de software GCG, usando o bien una matriz Blossum 62 o bien una matriz PAM250, y un peso de hueco de 16, 14, 12, 10, 8, 6 ó 4 y un peso de longitud de 1, 2, 3, 4, 5 ó 6. La homología en porcentaje entre dos secuencias de nucleótidos se determina usando el programa GAP en el paquete de software GCG, usando una matriz NWSgapdna.CMP y un peso de hueco de 40, 50, 60, 70 ó 80 y un peso de longitud de 1, 2, 3, 4,

5 ó 6. Un conjunto particularmente preferido de parámetros (y el que debe usarse si el técnico no está seguro sobre qué parámetros deben aplicarse para determinar si una molécula está dentro de una limitación de homología de las reivindicaciones) son una matriz de puntuación Blossum 62 con una penalización por hueco de 12, una penalización por extensión de hueco de 4 y una penalización por hueco del marco de lectura de 5.

- 5 Tal como se usa en el presente documento, una “célula huésped” es una célula usada para producir un producto descrito en el presente documento (por ejemplo, un aldehído graso descrito en el presente documento). Una célula huésped puede modificarse para expresar o sobreexpresar genes seleccionados o para tener una expresión atenuada de genes seleccionados. Los ejemplos no limitativos de células huésped incluyen células vegetales, animales, humanas, de bacterias, de levaduras o de hongos filamentosos.
- 10 Tal como se usa en el presente documento, el término “se hibrida en condiciones de rigurosidad baja, rigurosidad media, rigurosidad alta o rigurosidad muy alta” describe condiciones para la hibridación y el lavado. Pueden encontrarse orientaciones para realizar la reacciones de hibridación en *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1 – 6.3.6. En esa referencia se describen métodos acuosos y no acuosos y puede usarse cualquier método. Las condiciones de hibridación específicas a las que se hace referencia en el presente
- 15 documento son las siguientes: 1) condiciones de hibridación de rigurosidad baja en 6X cloruro de sodio/citrato de sodio (SSC) a aproximadamente 45°C, seguido por dos lavados en 0,2X SSC, SDS al 0,1% al menos a 50°C (la temperatura de los lavados puede aumentarse hasta 55°C para condiciones de rigurosidad baja); 2) condiciones de hibridación de rigurosidad media en 6X SSC a aproximadamente 45°C, seguido por uno o más lavados en 0,2X SSC, SDS al 0,1% a 60°C; 3) condiciones de hibridación de rigurosidad alta en 6X SSC a aproximadamente 45°C,
- 20 seguido por uno o más lavados en 0,2X SSC, SDS al 0,1% a 65°C; y preferiblemente 4) las condiciones de hibridación de rigurosidad muy alta son fosfato de sodio 0,5 M, SDS al 7% a 65°C, seguido por uno o más lavados a 0,2X SSC, SDS al 1% a 65°C. Las condiciones de rigurosidad muy alta (4) son las condiciones preferidas a menos que se especifique lo contrario.

- 25 El término “aislado” tal como se usa en el presente documento con respecto a ácidos nucleicos, tales como ADN o ARN, se refiere a moléculas separadas de otros ADN o ARN, respectivamente, que están presentes en la fuente natural del ácido nucleico. Además, por “ácido nucleico aislado” se pretende incluir fragmentos de ácido nucleico, que no se producen de manera natural como fragmentos y no se encontrarían en el estado natural. El término “aislado” también se usa en el presente documento para hacer referencia a polipéptidos, que están aislados de otras
- 30 proteínas celulares y se pretende que abarque tanto polipéptidos purificados como recombinantes. El término “aislado” tal como se usa en el presente documento también se refiere a un ácido nucleico o péptido que está sustancialmente libre de material celular, material viral o medio de cultivo cuando se produce mediante técnicas de ADN recombinante. El término “aislado” tal como se usa en el presente documento también se refiere a un ácido nucleico o péptido que está sustancialmente libre de precursores químicos u otros productos químicos cuando se sintetiza químicamente. El término “aislado”, tal como se usa en el presente documento con respecto a productos,
- 35 tales como aldehídos grasos, se refiere a productos están aislados de componentes celulares, medios de cultivo celular, o precursores químicos o sintéticos.

Tal como se usa en el presente documento, el “nivel de expresión de un gen en una célula” se refiere al nivel de ARNm, transcrito(s) naciente(s) pre-ARNm, productos intermedios de procesamiento de transcritos, ARNm maduro(s) y productos de degradación codificados por el gen en la célula.

- 40 Tal como se usa en el presente documento, el término “microorganismo” significa especie microbiana procarionta y eucariota de los dominios *Archaea*, *Bacteria* y *Eucarya*, incluyendo este último levaduras y hongos filamentosos, protozoos, algas o protistas superiores. Los términos “células microbianas” (es decir, células de microbios) y “microbios” se usan de manera intercambiable y se refieren a células o pequeños organismos que sólo pueden observarse con ayuda del microscopio.
- 45 Tal como se usa en el presente documento, el término “ácido nucleico” se refiere a polinucleótidos, tales como ácido desoxirribonucleico (ADN), y, cuando sea apropiado, ácido ribonucleico (ARN). El término también debe entenderse como que incluye, como equivalentes, análogos de o bien ARN o bien ADN producidos a partir de análogos de nucleótidos, y, según sea aplicable para la realización que esté describiéndose, polinucleótidos monocatenarios (sentido o antisentido) y bicatenarios, EST, cromosomas, ADNc, ARNm y ARNr.
- 50 Tal como se usa en el presente documento, el término “unido operativamente” significa que la secuencia de nucleótidos seleccionada (por ejemplo, que codifica para un polipéptido descrito en el presente documento) está en proximidad a un promotor para permitir que el promotor regule la expresión del ADN seleccionado. Además, el promotor está ubicado en el sentido de 5' de la secuencia de nucleótidos seleccionada en cuanto al sentido de transcripción y traducción. Por “unido operativamente” quiere decirse que una secuencia de nucleótidos y una(s)
- 55 secuencia(s) reguladora(s) están conectadas de tal manera que se permite la expresión génica cuando las moléculas apropiadas (por ejemplo, proteínas activadoras de la transcripción) se unen a la(s) secuencia(s) reguladora(s).

El término “o” se usa en el presente documento para querer decir, y se usa de manera intercambiable con, el término “y/o”, a menos que el contexto indique claramente lo contrario.

Tal como se usa en el presente documento, “sobreexpresar” significa expresar o provocar que se exprese un ácido nucleico, polipéptido o hidrocarburo en una célula a una concentración mayor que la expresada normalmente en una célula silvestre correspondiente. Por ejemplo, un polipéptido puede “sobreexpresarse” en una célula huésped recombinante cuando el polipéptido está presente en una concentración mayor en la célula huésped recombinante en comparación con su concentración en una célula huésped no recombinante de la misma especie.

Tal como se usa en el presente documento, “coeficiente de distribución” o “P”, se define como la concentración en equilibrio de un compuesto en una fase orgánica dividida entre la concentración en equilibrio en una fase acuosa (por ejemplo, caldo de fermentación). En una realización de un sistema bifásico descrito en el presente documento, la fase orgánica se forma por el aldehído graso durante el procedimiento de producción. Sin embargo, en algunos ejemplos, puede proporcionarse una fase orgánica, tal como proporcionando una fase de octano, para facilitar la separación de producto. Cuando se describe un sistema de dos fases, las características de distribución de un compuesto pueden describirse como logP. Por ejemplo, un compuesto con un logP de 1 se distribuirá a 10:1 para la fase orgánica. Un compuesto con un logP de -1 se distribuirá a 1:10 para la fase orgánica. Eligiendo un caldo de fermentación y fase orgánica apropiados, puede separarse un aldehído graso con un alto valor de logP a la fase orgánica incluso a concentraciones muy bajas en el recipiente de fermentación.

Tal como se usa en el presente documento, el término “purificar”, “purificado” o “purificación” significa la retirada o el aislamiento de una molécula de su entorno, por ejemplo, mediante aislamiento o separación. Las moléculas “sustancialmente purificadas” están libres en al menos aproximadamente el 60%, preferiblemente libres en al menos aproximadamente el 75% y más preferiblemente libres en al menos aproximadamente el 90% de otros componentes con los que están asociadas. Tal como se usa en el presente documento, estos términos también se refieren a la retirada de contaminantes de una muestra. Por ejemplo, la retirada de contaminantes puede dar como resultado un aumento en el porcentaje de aldehído graso en una muestra. Por ejemplo, cuando se producen aldehídos grasos en una célula huésped, los aldehídos grasos pueden purificarse mediante la retirada de proteínas de la célula huésped. Tras la purificación, se aumenta el porcentaje de aldehídos grasos en la muestra.

Los términos “purificar”, “purificado” y “purificación” no requieren una pureza absoluta. Son términos relativos. Por tanto, por ejemplo, cuando se producen aldehídos grasos en células huésped, un aldehído graso purificado es uno que se separa sustancialmente de los demás componentes celulares (por ejemplo, ácidos nucleicos, polipéptidos, lípidos, hidratos de carbono u otros hidrocarburos). En otro ejemplo, una preparación de aldehído graso purificado es una en la que el aldehído graso está sustancialmente libre de contaminantes, tales como los que podrían estar presentes tras la fermentación. En algunas realizaciones, un aldehído graso está purificado cuando al menos aproximadamente el 50% en peso de una muestra está compuesto por el aldehído graso. En otras realizaciones, un aldehído graso está purificado cuando al menos aproximadamente el 60%, el 70%, el 80%, el 85%, el 90%, el 92%, el 95%, el 98% o el 99% o más en peso de una muestra está compuesto por el aldehído graso.

Tal como se usa en el presente documento, el término “polipéptido recombinante” se refiere a un polipéptido que se produce mediante técnicas de ADN recombinante, en el que generalmente se inserta ADN que codifica para la proteína expresada o ARN en un vector de expresión adecuado y que a su vez se usa para transformar una célula huésped para producir el polipéptido o ARN.

Tal como se usa en el presente documento, el término “sustancialmente idéntico” (o “sustancialmente homólogo”) se usa para hacer referencia a una primera secuencia de aminoácidos o de nucleótidos que contiene un número suficiente de residuos de aminoácido o nucleótidos idénticos o equivalentes (por ejemplo, con una cadena lateral similar, por ejemplo, sustituciones de aminoácidos conservadas) a una segunda secuencia de aminoácidos o de nucleótidos de tal manera que la primera y la segunda secuencias de aminoácidos o de nucleótidos tienen actividades similares.

Tal como se usa en el presente documento, el término “sintasa” significa una enzima que cataliza un proceso de síntesis. Tal como se usa en el presente documento, el término sintasa incluye sintasas, sintetasas y ligasas.

Tal como se usa en el presente documento, el término “transfección” significa la introducción de un ácido nucleico (por ejemplo, por medio de un vector de expresión) en una célula receptora mediante transferencia génica mediada por ácido nucleico.

Tal como se usa en el presente documento, “transformación” se refiere a un proceso en el que se cambia el genotipo de una célula como resultado de la captación celular de ADN o ARN exógeno. Esto puede dar como resultado que la célula transformada exprese una forma recombinante de un ARN o polipéptido. En el caso de la expresión antisentido a partir del gen transferido, se altera la expresión de una forma que se produce de manera natural del polipéptido.

Tal como se usa en el presente documento, una “proteína de transporte” es un polipéptido que facilita el movimiento de uno o más compuestos hacia dentro y/o hacia fuera de un orgánulo celular y/o una célula.

Tal como se usa en el presente documento, una “variante” de polipéptido X se refiere a un polipéptido que tiene la secuencia de aminoácidos del péptido X en la que se ha alterado uno o más residuos de aminoácido. La variante puede tener cambios conservativos o cambios no conservativos. Pueden encontrarse orientaciones en la

determinación de qué residuos de aminoácido pueden sustituirse, insertarse o delecionarse sin afectar a la actividad biológica usando programas informáticos bien conocidos en la técnica, por ejemplo, el software LASERGENE (DNASTAR).

5 El término “variante”, cuando se usa en el contexto de una secuencia de polinucleótido, puede abarcar una secuencia de polinucleótido relacionada con la de un gen o la secuencia codificante del mismo. Esta definición también puede incluir, por ejemplo, variantes “alélicas”, “de corte y empalme”, “de especie” o “polimórficas”. Una variante de corte y empalme puede tener identidad significativa con respecto a un polinucleótido de referencia, pero generalmente tendrá un número mayor o menor de polinucleótidos debido a corte y empalme alternativo de exones durante el procesamiento del ARNm. El polipéptido correspondiente puede presentar dominios funcionales adicionales o ausencia de dominios. Las variantes de especie son secuencias de polinucleótido que varían de una especie a otra. Los polipéptidos resultantes tendrán generalmente identidad de aminoácido significativa entre sí. Una variante polimórfica es una variación en la secuencia de polinucleótido de un gen particular entre individuos de una especie dada.

15 Tal como se usa en el presente documento, el término “vector” se refiere a una molécula de ácido nucleico que puede transportar otro ácido nucleico al que se ha unido. Un tipo de vector útil es un episoma (es decir, un ácido nucleico que puede realizar replicación extracromosómica). Vectores útiles son aquellos que pueden realizar replicación autónoma y/o expresión de ácidos nucleicos a los que están unidos. Los vectores que pueden dirigir la expresión de genes a los que están unidos operativamente se denominan en el presente documento “vectores de expresión”. En general, los vectores de expresión útiles en técnicas de ADN recombinante están con frecuencia en forma de “plásmidos”, lo que se refiere generalmente a bucles de ADN bicatenario circulares que, en su forma de vector, no están unidos al cromosoma. En la presente memoria descriptiva, “plásmido” y “vector” se usan de manera intercambiable, ya que el plásmido es la forma de vector más comúnmente usada. Sin embargo, también se incluyen aquellas otras formas de vectores de expresión que realizan funciones equivalentes y que lleguen a conocerse en la técnica con posterioridad al presente documento.

25 La invención se basa, al menos en parte, en el descubrimiento de una nueva ruta para la biosíntesis de aldehydos grasos en *E. coli* y la identificación de genes que codifican para polipéptidos de biosíntesis de aldehydos grasos. Los polipéptidos de biosíntesis de aldehydos grasos participan en una ruta de biosíntesis representada en la figura 1. En esta ruta, en primer lugar se activa un ácido graso mediante ATP y después se reduce mediante una enzima de tipo ácido carboxílico reductasa (CAR) para generar un aldehído graso. Por consiguiente, los nucleótidos y polipéptidos de biosíntesis de aldehydos grasos descritos en el presente documento pueden usarse para producir aldehydos grasos.

Genes de biosíntesis de aldehydos grasos y variantes

35 Los métodos y las composiciones descritos en el presente documento incluyen, por ejemplo, genes de biosíntesis de aldehydos grasos, por ejemplo genes de ácido carboxílico reductasa (genes *car*), que tienen una secuencia de nucleótidos indicada en las figuras 2 y 4, así como variantes de polinucleótidos de los mismos. En algunos casos, el gen de biosíntesis de aldehydos grasos codifica para uno o más de los motivos de aminoácido representados en la figura 3. Por ejemplo, el gen puede codificar para un polipéptido que comprende SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 y SEQ ID NO: 10; SEQ ID NO: 11; SEQ ID NO: 12; SEQ ID NO: 13; SEQ ID NO: 14; y/o SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 10 y SEQ ID NO: 11. SEQ ID NO: 7 incluye un dominio de reductasa; SEQ ID NO: 8 y SEQ ID NO: 14 incluyen un dominio de unión a NADP; SEQ ID NO: 9 incluye un sitio de unión a fosfopanteteína; y SEQ ID NO: 10 incluye un dominio de unión a AMP.

45 Las variantes pueden producirse de manera natural o crearse *in vitro*. En particular, tales variantes pueden crearse usando técnicas de ingeniería genética, tales como mutagénesis dirigida al sitio, mutagénesis química al azar, procedimientos de delección con exonucleasa III o técnicas de clonación convencionales. Alternativamente, tales variantes, fragmentos, análogos o derivados pueden crearse usando procedimientos de modificación o síntesis química.

50 En la técnica se conocen bien métodos de producción de variantes. Estos incluyen procedimientos en los que se modifican secuencias de ácido nucleico obtenidas de aislados naturales para generar ácidos nucleicos que codifican para polipéptidos que tienen características que potencian su valor en aplicaciones industriales o de laboratorio. En tales procedimientos, se genera y caracteriza un gran número de secuencias variantes que tienen una o más diferencias de nucleótidos con respecto a la secuencia obtenida a partir del aislado natural. Normalmente, estas diferencias de nucleótidos dan como resultado cambios de aminoácidos con respecto a los polipéptidos codificados por los ácidos nucleicos de los aislados naturales.

55 Por ejemplo, pueden crearse variantes usando PCR propensa a errores (véase, por ejemplo, Leung *et al.*, Technique 1: 11-15, 1989; y Caldwell *et al.*, PCR Methods Applic. 2: 28-33, 1992). En la PCR propensa a errores, se realiza PCR en condiciones en la que la fidelidad de copiado de la ADN polimerasa es baja, de tal manera que se obtiene una alta tasa de mutaciones puntuales a lo largo de toda la longitud del producto de PCR. En resumen, en tales procedimientos, se mezclan ácidos nucleicos que van a mutagenizarse (por ejemplo, una secuencia de polinucleótido de biosíntesis de aldehydos grasos), con cebadores de PCR, tampón de reacción, MgCl₂, MnCl₂, Taq

polimerasa y una concentración apropiada de dNTP para lograr una alta tasa de mutación puntual a lo largo de toda la longitud del producto de PCR. Por ejemplo, la reacción puede realizarse usando 20 fmol de ácido nucleico que va a mutagenizarse (por ejemplo, una secuencia de polinucleótido de biosíntesis de aldehydos grasos), 30 pmol de cada cebador de PCR, un tampón de reacción que comprende KCl 50 mM, Tris HCl 10 mM (pH 8,3), y gelatina al 0,01%, MgCl₂ 7 mM, MnCl₂ 0,5 mM, 5 unidades de Taq polimerasa, dGTP 0,2 mM, dATP 0,2 mM, dCTP 1 mM y dTTP 1 mM. Puede realizarse la PCR durante 30 ciclos de 94°C durante 1 min, 45°C durante 1 min, y 72°C durante 1 min. Sin embargo, se apreciará que estos parámetros pueden variarse según sea apropiado. Entonces se clonan los ácidos nucleicos mutagenizados en un vector apropiado y se evalúan las actividades de los polipéptidos codificados por los ácidos nucleicos mutagenizados.

También pueden crearse variantes usando mutagénesis dirigida por oligonucleótidos para generar mutaciones específicas del sitio en cualquier ADN clonado de interés. La mutagénesis de oligonucleótidos se describe, por ejemplo, en Reidhaar-Olson *et al*, Science 241: 53-57, 1988. En resumen, en tales procedimientos se sintetizan una pluralidad de oligonucleótidos bicatenarios que llevan una o más mutaciones que van a introducirse en el ADN clonado y se insertan en el ADN clonado que va a mutagenizarse (por ejemplo, una secuencia de polinucleótido de biosíntesis de aldehydos grasos). Se recuperan clones que contienen el ADN mutagenizado y se evalúan las actividades de los polipéptidos para los que codifican.

Otro método para generar variantes es PCR de ensamblaje. La PCR de ensamblaje implica el ensamblaje de un producto de PCR a partir de una mezcla de fragmentos de ADN pequeños. Se produce un gran número de reacciones de PCR diferentes en paralelo en el mismo vial, cebando los productos de una reacción a los productos de otra reacción. La PCR de ensamblaje se describe, por ejemplo, en la patente estadounidense n.º 5.965.408.

Todavía otro método de generación de variantes es la mutagénesis por PCR sexual. En la mutagénesis por PCR sexual, se produce recombinación homóloga forzada entre moléculas de ADN con secuencia de ADN diferente, pero altamente relacionada, *in vitro* como resultado de fragmentación al azar de la molécula de ADN basándose en la homología de secuencia. Esto va seguido por fijación del cruce mediante extensión por cebadores en una reacción de PCR. La mutagénesis por PCR sexual se describe, por ejemplo, en Stemmer, PNAS, USA 91: 10747-10751, 1994.

También pueden crearse variantes mediante mutagénesis *in vivo*. En algunas realizaciones, se generan mutaciones al azar en una secuencia de ácido nucleico propagando la secuencia en una cepa bacteriana, tal como una cepa de *E. coli*, que porta mutaciones en una o más de las rutas de reparación del ADN. Tales cepas "mutágenas" tienen una tasa de mutación al azar superior a la de una cepa silvestre. La propagación de una secuencia de ADN (por ejemplo, una secuencia de polinucleótido de biosíntesis de aldehydos grasos) en una de esas cepas generará eventualmente mutaciones al azar dentro del ADN. Se describen cepas mutágenas adecuadas para su uso en la mutagénesis *in vivo*, por ejemplo, en la publicación PCT n.º WO 91/16427.

También pueden generarse variantes usando mutagénesis en casete. En la mutagénesis en casete, se reemplaza una pequeña región de una molécula de ADN bicatenario por un "casete" oligonucleotídico sintético que se diferencia de la secuencia nativa. El oligonucleótido contiene con frecuencia una secuencia nativa completa y/o parcialmente aleatorizada.

También puede usarse mutagénesis de conjunto recursiva para generar variantes. La mutagénesis de conjunto recursiva es un algoritmo para la modificación por ingeniería de proteínas (es decir, mutagénesis de proteínas) desarrollado para producir diversas poblaciones de mutantes fenotípicamente relacionados cuyos miembros se diferencian en cuanto a la secuencia de aminoácidos. Este método usa un mecanismo de retroalimentación para controlar rondas sucesivas de mutagénesis en casete combinatoria. La mutagénesis de conjunto recursiva se describe, por ejemplo, en Arkin *et al*, PNAS, USA 89:7811-7815, 1992.

En algunas realizaciones, se crean variantes usando mutagénesis de conjunto exponencial. La mutagénesis de conjunto exponencial es un procedimiento para generar bibliotecas combinatorias con un alto porcentaje de mutantes únicos y funcionales, en el que pequeños grupos de residuos se aleatorizan en paralelo para identificar, en cada posición alterada, aminoácidos que conducen a proteínas funcionales. La mutagénesis de conjunto exponencial se describe, por ejemplo, en Delegrave *et al*, Biotech. Res. 11: 1548-1552, 1993. Las mutagénesis al azar y dirigida al sitio se describen, por ejemplo, en Arnold, Curr. Opin. Biotech. 4:450-455, 1993.

En algunas realizaciones, se crean variantes usando procedimientos de intercambio en los que partes de una pluralidad de ácidos nucleicos que codifican para distintos polipéptidos se fusionan entre sí para crear secuencias de ácido nucleico quiméricas que codifican para polipéptidos quiméricos tal como se describe, por ejemplo, en las patentes estadounidenses n.ºs 5.965.408 y 5.939.250.

Las variantes de polinucleótido también incluyen análogos de ácido nucleico. Pueden modificarse análogos de ácido nucleico en el resto de base, resto de azúcar o estructura principal de fosfato para mejorar, por ejemplo, la estabilidad, hibridación o solubilidad del ácido nucleico. Las modificaciones en el resto de base incluyen desoxiuridina para desoxitimidina y 5-metil-2'-desoxicidina o 5-bromo-2'-doxicidina para desoxicidina. Las modificaciones del resto de azúcar incluyen modificación del 2'-hidroxilo del azúcar de ribosa para formar azúcares

de 2'-O-metilo o 2'-O-alilo. La estructura principal de fosfato de desoxirribosa puede modificarse para producir ácidos nucleicos de morfolino, en los que cada resto de base está unido a un anillo morfolino de seis miembros, o ácidos nucleicos peptídicos, en los que la estructura principal de desoxifosfato se reemplaza por una estructura principal pseudopeptídica y se conservan las cuatro bases. (Véase, por ejemplo, Summerton *et al.*, *Antisense Nucleic Acid Drug Dev.* (1997) 7:187-195; e Hyrup *et al.*, *Bioorgan. Med. Chem.* (1996) 4: 5-23). Además, la estructura principal de desoxifosfato puede reemplazarse, por ejemplo, por una estructura principal de fosforotioato o fosforoditioato, una estructura principal de fosforoamidito o una de fosfotriéster alquílico.

Puede usarse cualquier secuencia de polinucleótido que codifica para un homólogo indicado en las figuras 2 y 4, o una variante de la misma, como polinucleótido de biosíntesis de aldehydos grasos en los métodos descritos en el presente documento.

Polipéptidos de biosíntesis de aldehydos grasos y variantes

Los métodos y las composiciones descritos en el presente documento también incluyen polipéptidos de biosíntesis de aldehydos grasos que tienen una secuencia de aminoácidos indicada en las figuras 2 y 4, así como variantes de polipéptidos de las mismas. En algunos casos, un polipéptido de biosíntesis de aldehydos grasos es uno que incluye uno o más de los motivos de aminoácido representados en la figura 3. Por ejemplo, el polipéptido puede incluir las secuencias de aminoácidos de SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 y SEQ ID NO: 10. En otras situaciones, el polipéptido incluye una o más de SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13 y SEQ ID NO: 14. En aún otros casos, el polipéptido incluye las secuencias de aminoácidos de SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 10 y SEQ ID NO: 11. SEQ ID NO: 7 incluye un dominio de reductasa; SEQ ID NO: 8 y SEQ ID NO: 14 incluyen un dominio de unión a NADP; SEQ ID NO: 9 incluye un sitio de unión a fosfopanteteína; y SEQ ID NO: 10 incluye un dominio de unión a AMP.

Las variantes de polipéptido de biosíntesis de aldehydos grasos pueden ser variantes en las que uno o más residuos de aminoácido se sustituyen por un residuo de aminoácido conservado o no conservado (preferiblemente un residuo de aminoácido conservado). Tal residuo de aminoácido sustituido puede codificarse o no por el código genético.

Las sustituciones conservativas son aquellas que sustituyen un aminoácido dado en un polipéptido por otro aminoácido de características similares. Las sustituciones conservativas típicas son los siguientes reemplazos: reemplazo de un aminoácido alifático, tal como alanina, valina, leucina e isoleucina, por otro aminoácido alifático; reemplazo de una serina por una treonina o viceversa; reemplazo de un residuo ácido, tal como ácido aspártico y ácido glutámico, por otro residuo ácido; reemplazo de un residuo que porta un grupo amida, tal como asparagina y glutamina, por otro residuo que porta un grupo amida; intercambio de un residuo básico, tal como lisina y arginina, por otro residuo básico; y reemplazo de un residuo aromático, tal como fenilalanina y tirosina, por otro residuo aromático.

Otras variantes de polipéptido son aquellas en las que uno o más residuos de aminoácido incluyen un grupo sustituyente. Todavía otras variantes de polipéptido son aquellas en las que el polipéptido está asociado con otro compuesto, tal como un compuesto para aumentar la semivida del polipéptido (por ejemplo, polietilenglicol).

Variantes de polipéptido adicionales son aquellas en las que se fusionan aminoácidos adicionales con el polipéptido, tal como una secuencia líder, una secuencia secretora, una secuencia proproteína o una secuencia que facilita la purificación, el enriquecimiento o la estabilización del polipéptido.

En algunos casos, las variantes de polipéptido conservan la misma función biológica que un polipéptido que tiene una secuencia de aminoácidos indicada en las figuras 2 y 4 (por ejemplo, conservan actividad de biosíntesis de aldehydos grasos, tal como actividad ácido carboxílico o ácido graso reductasa) y tienen secuencias de aminoácidos sustancialmente idénticas al mismo.

En otros casos, las variantes de polipéptido tienen al menos aproximadamente el 50%, al menos aproximadamente el 55%, al menos aproximadamente el 60%, al menos aproximadamente el 65%, al menos aproximadamente el 70%, al menos aproximadamente el 75%, al menos aproximadamente el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 95% o más de aproximadamente el 95% de homología con una secuencia de aminoácidos indicada en las figuras 2 y 4. En otra realización, las variantes de polipéptido incluyen un fragmento que comprende al menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 ó 150 aminoácidos consecutivos del mismo.

Las variantes de polipéptido o fragmentos de las mismas pueden obtenerse aislando ácidos nucleicos que codifican para las mismas usando técnicas descritas en el presente documento o expresando ácidos nucleicos sintéticos que codifican para las mismas. Alternativamente, pueden obtenerse variantes de polipéptido o fragmentos de las mismas mediante procedimientos de purificación o enriquecimiento bioquímico. La secuencia de variantes de polipéptido o fragmentos puede determinarse mediante digestión proteolítica, electroforesis en gel y/o microsecuenciación. Entonces puede compararse la secuencia de las variantes de polipéptido o fragmentos con una secuencia de aminoácidos indicada en las figuras 2 y 4 usando cualquiera de los programas descritos en el presente documento.

Las variantes de polipéptido y fragmentos de las mismas pueden someterse a ensayo para determinar la actividad

de producción de aldehídos grasos usando métodos rutinarios. Por ejemplo, pueden ponerse en contacto las variantes de polipéptido o fragmento con un sustrato (por ejemplo, un ácido graso, un sustrato derivado de ácido graso u otro sustrato descrito en el presente documento) en condiciones que permiten que funcione la variante de polipéptido. Puede medirse una disminución del nivel del sustrato o un aumento del nivel de un aldehído graso para determinar la actividad de producción de aldehídos grasos.

Anticuerpos anti-polipéptido de biosíntesis de aldehídos grasos

Los polipéptidos de biosíntesis de aldehídos grasos descritos en el presente documento también pueden usarse para producir anticuerpos dirigidos contra polipéptidos de biosíntesis de aldehídos grasos. Tales anticuerpos pueden usarse, por ejemplo, para detectar la expresión de un polipéptido de biosíntesis de aldehídos grasos usando métodos conocidos en la técnica. El anticuerpo puede ser, por ejemplo, un anticuerpo policlonal; un anticuerpo monoclonal o fragmento de unión a antígeno del mismo; un anticuerpo modificado tal como un anticuerpo quimérico, anticuerpo reconformado, anticuerpo humanizado, o fragmento de los mismos (por ejemplo, Fab', Fab, F(ab')₂); o un anticuerpo biosintético, por ejemplo, un anticuerpo de cadena sencilla, anticuerpo de dominio sencillo (DAB), Fv, Fv de cadena sencilla (scFv), o similar.

Se describen métodos de preparación y uso de anticuerpos policlonales y monoclonales, por ejemplo, en Harlow *et al.*, *Using Antibodies: A Laboratory Manual: Portable Protocol I*. Cold Spring Harbor Laboratory (1 de diciembre de 1998). En la técnica se conocen métodos para la producción de anticuerpos modificados y fragmentos de anticuerpo (por ejemplo, anticuerpos quiméricos, anticuerpos reconformados, anticuerpos humanizados, o fragmentos de los mismos, por ejemplo, fragmentos Fab', Fab, F(ab')₂); o anticuerpos biosintéticos (por ejemplo, anticuerpos de cadena sencilla, anticuerpos de dominio sencillo (DAB), Fv, Fv de cadena sencilla (scFv), y similares) y pueden encontrarse, por ejemplo, en Zola, *Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives*, Springer Verlag (15 de diciembre de 2000; 1ª edición).

Sustratos

Las composiciones y los métodos descritos en el presente documento pueden usarse para producir aldehídos grasos a partir de un sustrato apropiado. Sin desear limitarse a la teoría, se cree que los polipéptidos descritos en el presente documento producen aldehídos grasos a partir de sustratos mediante un mecanismo de reducción. En algunos casos, el sustrato es un derivado de ácido graso (por ejemplo, un ácido graso), y puede producirse un aldehído graso que tiene patrones de ramificación y una longitud de cadena de carbono particulares a partir de un derivado de ácido graso que tiene esas características que dará como resultado el aldehído graso deseado.

Por consiguiente, cada etapa dentro de una ruta de biosíntesis que conduce a la producción de un sustrato derivado de ácido graso puede modificarse para producir o sobreproducir el sustrato de interés. Por ejemplo, pueden expresarse, sobreexpresarse o atenuarse genes conocidos implicados en la ruta de biosíntesis de ácidos grasos o la ruta de aldehídos grasos en células huésped para producir un sustrato deseado (véase, por ejemplo, el documento PCT/US08/058788). En la figura 5 se proporcionan genes a modo de ejemplo.

Síntesis de sustratos

Ácido graso sintasa (FAS) es un grupo de polipéptidos que catalizan el inicio y la elongación de cadenas de acilo (Marrakchi *et al.*, *Biochemical Society*, 30: 1050-1055, 2002). La proteína portadora de acilo (ACP) junto con las enzimas en la ruta de FAS controlan la longitud, el grado de saturación y la ramificación de los derivados de ácidos grasos producidos. La ruta de biosíntesis de ácidos grasos implica a los precursores acetil-CoA y malonil-CoA. Las etapas en esta ruta se catalizan mediante enzimas de las familias de genes de biosíntesis de ácidos grasos (*fab*) y acetil-CoA carboxilasa (*acc*) (véase, por ejemplo, Heath *et al.*, *Prog. Lipid Res.* 40(6):467-97 (2001)).

Pueden modificarse por ingeniería células huésped para expresar sustratos derivados de ácidos grasos expresando o sobreexpresando de manera recombinante uno o más genes de ácido graso sintasa, tales como genes de acetil-CoA y/o malonil-CoA sintasa. Por ejemplo, para aumentar la producción de acetil-CoA, puede expresarse uno o más de los siguientes genes en una célula huésped: *pdh* (un complejo multienzimático que comprende *aceEF* (que codifica para el componente de E1p deshidrogenasa, el componente de E2p dihidrolipoamida aciltransferasa de los complejos de piruvato y 2-oxoglutarato deshidrogenasa, y *lpd*), *panK*, *fabH*, *fabB*, *fabD*, *fabG*, *acpP* y *fabF*. Números de registro de GenBank a modo de ejemplo para estos genes son: *pdh* (BAB34380, AAC73227, AAC73226), *panK* (también conocido como CoA, AAC76952), *aceEF* (AAC73227, AAC73226), *fabH* (AAC74175), *fabB* (POA953), *fabD* (AAC74176), *fabG* (AAC74177), *acpP* (AAC74178), *fabF* (AAC74179). Adicionalmente, los niveles de expresión de *fadE*, *gpsA*, *ldhA*, *pflb*, *adhE*, *pta*, *poxB*, *ackA* y/o *ackB* pueden atenuarse o desactivarse en una célula huésped modificada por ingeniería mediante transformación con plásmidos condicionalmente replicantes o no replicantes que contienen mutaciones nulas o por delección de los genes correspondientes o mediante sustitución de secuencias promotoras o potenciadoras. Números de registro de GenBank a modo de ejemplo para estos genes son: *fadE* (AAC73325), *gspA* (AAC76632), *ldhA* (AAC74462), *pflb* (AAC73989), *adhE* (AAC74323), *pta* (AAC75357), *poxB* (AAC73958), *ackA* (AAC75356) y *ackB* (BAB81430). Las células huésped resultantes tendrán niveles de producción de acetil-CoA aumentados cuando se hacen crecer en un entorno apropiado.

La sobreexpresión de malonil-CoA puede verse afectada por la introducción de *accABCD* (por ejemplo, número de

registro AAC73296, EC 6.4.1.2) en una célula huésped. Pueden sobreexpresarse adicionalmente ácidos grasos en células huésped mediante la introducción en la célula huésped de una secuencia de ADN que codifica para una lipasa (por ejemplo, números de registro CAA89087, CAA98876).

- 5 Además, la inhibición de PlsB puede conducir a un aumento en los niveles de acil-ACP de cadena larga, lo que inhibirá las etapas iniciales en la ruta (por ejemplo, *accABCD*, *fabH* y *fabI*). La mutación D311E de *plsB* (por ejemplo, número de registro AAC77011) puede usarse para aumentar la cantidad de ácidos grasos disponibles.

Además, puede modificarse por ingeniería una célula huésped para sobreexpresar un gen *sfa* (supresor de *fabA*, por ejemplo, número de registro AAN79592) para aumentar la producción de ácidos grasos monoinsaturados (Rock et al, J. Bacteriology 178:5382-5387, 1996).

- 10 La longitud de cadena de un sustrato derivado de ácido graso puede seleccionarse modificando la expresión de tioesterasas seleccionadas. La tioesterasa influye sobre la longitud de cadena de ácidos grasos producidos. Por tanto, pueden modificarse por ingeniería células huésped para expresar, sobreexpresar, tener expresión atenuada o no expresar una o más tioesterasas seleccionadas para aumentar la producción de un sustrato derivado de ácido graso preferido. Por ejemplo, pueden producirse ácidos grasos C₁₀ expresando una tioesterasa que tiene preferencia por la producción de ácidos grasos C₁₀ y atenuando tioesterasas que tienen preferencia por la producción de ácidos grasos distintos de ácidos grasos C₁₀ (por ejemplo, una tioesterasa que prefiere producir ácidos grasos C₁₄). Esto dará como resultado una población relativamente homogénea de ácidos grasos que tienen una longitud de cadena de carbono de 10. En otros casos, pueden producirse ácidos grasos C₁₄ atenuando tioesterasas endógenas que producen ácidos grasos distintos de C₁₄ y expresando las tioesterasas que usan C₁₄-ACP. En algunas situaciones, pueden producirse ácidos grasos C₁₂ expresando tioesterasas que usan C₁₂-ACP y atenuando tioesterasas que producen ácidos grasos distintos de C₁₂. La sobreproducción de acetil-CoA, malonil-CoA y ácidos grasos puede verificarse usando métodos conocidos en la técnica, por ejemplo, usando precursores radiactivos, HPLC o CG-EM tras lisis celular. En la tabla 1 se indican ejemplos no limitativos de tioesterasas que pueden usarse en los métodos descritos en el presente documento.

25 Tabla 1: Tioesterasas

Número de registro	Organismo fuente	Gen
AAC73596	<i>E. coli</i>	<i>tesA</i> sin secuencia líder
AAC73555	<i>E. coli</i>	<i>tesB</i>
Q41635, AAA34215	<i>Umbellularia californica</i>	<i>fatB</i>
AAC49269	<i>Cuphea hookeriana</i>	<i>fatB2</i>
Q39513; AAC72881	<i>Cuphea hookeriana</i>	<i>fatB3</i>
Q39473, AAC49151	<i>Cinnamomum camphorum</i>	<i>fatB</i>
CAA85388	<i>Arabidopsis thaliana</i>	<i>fatB</i> [M141T]*
NP_189147; NP_193041	<i>Arabidopsis thaliana</i>	<i>fatA</i>
CAC39106	<i>Bradyrhizobium japonicum</i>	<i>fatA</i>
AAC72883	<i>Cuphea hookeriana</i>	<i>fatA</i>
AAL79361	<i>Helianthus annuus</i>	<i>fatA1</i>

* Mayer et al., BMC Plant Biology 7:1-11, 2007

- 30 En otros casos, un polipéptido de biosíntesis de aldehídos grasos, variante, o un fragmento del mismo, se expresa en una célula huésped que contiene una mutación que se produce de manera natural que da como resultado un aumento del nivel de ácidos grasos en la célula huésped. En algunos casos, la célula huésped se modifica por ingeniería genética para aumentar el nivel de ácidos grasos en la célula huésped con respecto a una célula huésped silvestre correspondiente. Por ejemplo, la célula huésped puede modificarse por ingeniería genética para expresar un nivel reducido de una acil-CoA sintasa con respecto a una célula huésped silvestre correspondiente. En una realización, el nivel de expresión de uno o más genes (por ejemplo, un gen de acil-CoA sintasa) se reduce mediante la modificación por ingeniería genética de una célula huésped "desactivada".

- 35 Cualquier gen de acil-CoA sintasa conocido puede reducirse o desactivarse en una célula huésped. Los ejemplos no limitativos de genes de acil-CoA sintasa incluyen *fadD*, *fadK*, *BH3103*, *yhfL*, *Pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa3p* o el gen que codifica para la proteína ZP_01644857. Los ejemplos específicos de genes de acil-CoA sintasa incluyen *fadDD35* de *M. tuberculosis* H37Rv [NP_217021], *fadDD22* de *M. tuberculosis* H37Rv [NP_217464], *fadD* de *E. coli* [NP_416319], *fadK* de *E. coli* [YP_416216], *fadD* de *Acinetobacter* sp. ADP1 [YP_045024], *fadD* de *Haemophilus influenzae* RdkW20 [NP_438551], *fadD* de *Rhodopseudomonas palustris* Bis B18 [YP_533919], *BH3101* de *Bacillus halodurans* C-125 [NP_243969], *Pfl-4354* de *Pseudomonas fluorescens* Pfo-1 [YP_350082], *EAV15023* de *Comamonas testosterone* KF-1 [ZP_01520072], *yhfL* de *B. subtilis* [NP_388908], *fadD1* de *P. aeruginosa* PAO1 [NP_251989], *fadD1* de *Ralstonia solanacearum* GM1 1000 [NP_520978], *fadD2* de *P. aeruginosa* PAO1 [NP_251990], el gen que codifica para la proteína ZP_01644857 de *Stenotrophomonas maltophilia* R551-3, *faa3p* de *Saccharomyces cerevisiae* [NP_012257], *faa1p* de *Saccharomyces cerevisiae* [NP_014962], *lcfA* de *Bacillus subtilis* [CAA99571], o los descritos en Shockey et al., Plant. Physiol. 129:1710-1722, 2002; Caviglia et al, J. Biol. Chem. 279:1163-1169, 2004; Knoll et al., J. Biol. Chem. 269(23): 16348-

56, 1994; Johnson *et al.*, J. Biol. Chem. 269: 18037-18046, 1994; y Black *et al.*, J. Biol. Chem. 267: 25513-25520, 1992.

Formación de aldehídos grasos ramificados

5 Pueden producirse aldehídos grasos que contienen puntos de ramificación usando derivados de ácidos grasos ramificados como sustratos. Por ejemplo, aunque *E. coli* produce de manera natural ácidos grasos de cadena lineal (sFA), puede modificarse por ingeniería *E. coli* para producir ácidos grasos de cadena ramificada (brFA) introduciendo y expresando o sobreexpresando genes que proporcionan precursores ramificados en *E. coli* (por ejemplo, familias de genes *bkd*, *ilv*, *icm* y *fab*). Adicionalmente, puede modificarse por ingeniería una célula huésped para expresar o sobreexpresar genes que codifican para proteínas para la elongación de brFA (por ejemplo, ACP, FabF, etc.) y/o para delecionar o atenuar los genes correspondientes de la célula huésped que conducen normalmente a sFA.

15 La primera etapa en la formación de brFA es la producción de los α -cetoácidos correspondientes mediante una aminoácido de cadena ramificada aminotransferasa. Las células huésped pueden incluir de manera endógena genes que codifican para tales enzimas o pueden introducirse tales genes de manera recombinante. Por ejemplo, *E. coli* expresa de manera endógena una enzima de este tipo, IlvE (EC 2.6.1.42; registro de GenBank YP_026247). En algunas células huésped, puede no expresarse una aminoácido de cadena ramificada aminotransferasa heteróloga. Sin embargo, si no es endógena, puede introducirse IlvE de *E. coli* o cualquier otra aminoácido de cadena ramificada aminotransferasa (por ejemplo, IlvE de *Lactococcus lactis* (registro de GenBank AAF34406), IlvE de *Pseudomonas putida* (registro de GenBank NP_745648), o IlvE de *Streptomyces coelicolor* (registro de GenBank NP_629657)).

20 En otra realización, la producción de α -cetoácidos puede lograrse usando los métodos descritos en Atsumi *et al.*, Nature 451: 86-89, 2008. Por ejemplo, puede producirse 2-cetoisovalerato sobreexpresando los genes que codifican para IlvI, IlvH, IlvC o IlvD. En otro ejemplo, puede producirse 2-ceto-3-metil-valerato sobreexpresando los genes que codifican para IlvA e IlvI, IlvH (o AlsS de *Bacillus subtilis*), IlvC, IlvD, o sus homólogos correspondientes. En una realización adicional, puede producirse 2-ceto-4-metil-pentanoato sobreexpresando los genes que codifican para IlvI, IlvH, IlvC, IlvD y LeuA, LeuB, LeuC, LeuD, o sus homólogos correspondientes.

30 La segunda etapa es la descarboxilación oxidativa de los α -cetoácidos para dar la acil-CoA de cadena ramificada correspondiente. Esta reacción puede catalizarse mediante un complejo de α -cetoácido de cadena ramificada deshidrogenasa (*bkd*; EC 1.2.4.4.) (Denoya *et al.*, J. Bacteriol. 177:3504, 1995), que consiste en subunidades E1 α / β (descarboxilasa), E2 (dihidrolipoil transacilasa) y E3 (dihidrolipoil deshidrogenasa). Estos complejos de α -cetoácido de cadena ramificada deshidrogenasa son similares a complejos de piruvato y α -cetoglutarato deshidrogenasa. Cualquier microorganismo que presente brFA y/o crezca con aminoácidos de cadena ramificada puede usarse como fuente para aislar genes *bkd* para la expresión en células huésped, por ejemplo, *E. coli*. Además, *E. coli* tiene el componente E3 como parte de su complejo de piruvato deshidrogenasa (*lpd*, EC 1.8.1.4, registro de GenBank NP_414658). Por tanto, puede ser suficiente expresar sólo los genes E1 α/β y E2 *bkd*. La tabla 2 indica ejemplos no limitativos de genes *bkd* de diversos microorganismos que pueden introducirse de manera recombinante y expresarse en una célula huésped para proporcionar precursores de acil-CoA de cadena ramificada.

Tabla 2: Genes *bkd* de microorganismos seleccionados

Organismo	Gen	N.º de registro de GenBank
<i>Streptomyces coelicolor</i>	<i>bkdA1</i> (E1 α) <i>bkdB1</i> (E1 β) <i>bkdC1</i> (E2)	NP_628006 NP_628005 NP_638004
<i>Streptomyces coelicolor</i>	<i>bkdA2</i> (E1 α) <i>bkdB2</i> (E1 β) <i>bkdC2</i> (E2)	NP_733618 NP_628019 NP_628018
<i>Streptomyces avermitilis</i>	<i>bkdA</i> (E1a) <i>bkdB</i> (E1b) <i>bkdC</i> (E2)	BAC72074 BAC72075 BAC72076
<i>Streptomyces avermitilis</i>	<i>bkdF</i> (E1 α) <i>bkdG</i> (E1 β) <i>bkdH</i> (E2)	BAC72088 BAC72089 BAC72090
<i>Bacillus subtilis</i>	<i>bkdAA</i> (E1 α) <i>bkdAB</i> (E1 β) <i>bkdB</i> (E2)	NP_390288 NP_390288 NP_390288
<i>Pseudomonas putida</i>	<i>bkdA1</i> (E1 α) <i>bkdA2</i> (E1 β) <i>bkdC</i> (E2)	AAA65614 AAA65615 AAA65617

En otro ejemplo, puede producirse isobutiril-CoA en una célula huésped, por ejemplo en *E. coli*, mediante la coexpresión de una crotonil-CoA reductasa (Ccr, EC 1.6.5.5, 1.1.1.1) e isobutiril-CoA mutasa (subunidad grande

IcmA, EC 5.4.99.2; subunidad pequeña IcmB, EC 5.4.99.2) (Han y Reynolds, J. Bacteriol. 179:5157, 1997). Crotonil-CoA es un producto intermedio en la biosíntesis de ácidos grasos en *E. coli* y otros microorganismos. En la tabla 3 se indican ejemplos no limitativos de genes *ccr* e *icm* de microorganismos seleccionados.

Tabla 3: Genes *ccr* e *icm* de microorganismos seleccionados

Organismo	Gen	N.º de registro de GenBank
<i>Streptomyces coelicolor</i>	<i>ccr</i> <i>icmA</i> <i>icmB</i>	NP_630556 NP_629554 NP_630904
<i>Streptomyces cinnamonensis</i>	<i>ccr</i> <i>icmA</i> <i>icmB</i>	AAD53915 AAC08713 AJ246005

- 5 Además de la expresión de los genes *bkd*, el inicio de la biosíntesis de brFA usa β -cetoacil-proteína portadora de acilo sintasa III (FabH, EC 2.3.1.41) con especificidad por acil-CoA de cadena ramificada (Li *et al*, J. Bacteriol. 187:3795-3799, 2005). En la tabla 4 se indican ejemplos no limitativos de tales enzimas FabH. Pueden expresarse en una célula huésped genes *fabH* que están implicados en la biosíntesis de ácidos grasos de cualquier microorganismo que contiene brFA. Las enzimas Bkd y FabH de células huésped que no producen de manera natural brFA pueden no soportar la producción de brFA. Por tanto, *bkd* y *fabH* pueden expresarse de manera recombinate. Pueden insertarse vectores que contienen los genes *bkd* y *fabH* en una célula huésped de este tipo. De manera similar, el nivel endógeno de producción de Bkd y FabH puede no ser suficiente para producir brFA. En ese caso, pueden sobreexpresarse. Adicionalmente, pueden expresarse o sobreexpresarse otros componentes de la ruta de biosíntesis de ácidos grasos, tales como proteínas portadoras de acilo (ACP) y β -cetoacil-proteína portadora de acilo sintasa II (*fabF*, EC 2.3.1.41) (en la tabla 4 se indican ejemplos no limitativos de candidatos). Además de expresar estos genes, algunos genes en la ruta de biosíntesis de ácidos grasos endógena pueden atenuarse en la célula huésped (por ejemplo, los genes de *E. coli* *fabH* (n.º de registro de GenBank NP_415609) y/o *fabF* (n.º de registro de GenBank NP_415613)).

Tabla 4: Genes *fabH*, ACP y *fabF* de microorganismos seleccionados con brFA

Organismo	Gen	N.º de registro de GenBank
<i>Streptomyces coelicolor</i>	<i>fabH1</i> <i>acp</i> <i>fabF</i>	NP_626634 NP_626635 NP_626636
<i>Streptomyces avermitilis</i>	<i>fabH3</i> <i>fabC3 (acp)</i> <i>fabF</i>	NP_823466 NP_823467 NP_823468
<i>Bacillus subtilis</i>	<i>fabH_A</i> <i>fabH_B</i> <i>acp</i> <i>fabF</i>	NP_389015 NP_388898 NP_389474 NP_389016
<i>Stenotrophomonas maltophilia</i>	SmalDRAFT_0818 (<i>fabH</i>) SmalDRAFT_0821 (<i>acp</i>) SmalDRAFT_0822 (<i>fabF</i>)	ZP_01643059 ZP_01643063 ZP_01643064
<i>Legionella pneumophila</i>	<i>fabH</i> <i>acp</i> <i>fabF</i>	YP_123672 YP_123675 YP_123676

20 Formación de aldehídos grasos cíclicos

- Pueden producirse aldehídos grasos cíclicos usando derivados de ácidos grasos cíclicos como sustratos. Para producir sustratos derivados de ácidos grasos cíclicos, pueden introducirse genes que proporcionan precursores cíclicos (por ejemplo, las familias de genes *ans*, *chc* y *plm*) en la célula huésped y expresarse para permitir el inicio de la biosíntesis de ácidos grasos a partir de precursores cíclicos. Por ejemplo, para convertir una célula huésped, tal como *E. coli*, en una que pueda sintetizar ácidos grasos ω -cíclicos (cyFA), puede introducirse un gen que proporciona el precursor cíclico ciclohexilcarbonil-CoA (CHC-CoA) (Cropp *et al*, Nature Biotech. 18:980-983, 2000) y expresarse en la célula huésped. Los ejemplos no limitativos de genes que proporcionan CHC-CoA en *E. coli* incluyen: *ansJ*, *ansK*, *ansL*, *chcA* y *ansM* de la agrupación de genes de ansatrienina de *Streptomyces collinus* (Chen *et al*, Eur. J. Biochem. 261: 98-107, 1999) o *plmJ*, *plmK*, *plmL*, *chcA* y *plmM* de la agrupación de genes de foslactomicina B de *Streptomyces* sp. HK803 (Palaniappan *et al.*, J. Biol. Chem. 278:35552-35557, 2003) junto con el gen *chcB* (Patton *et al*, Biochem. 39: 7595-7604, 2000) de *S. collinus*, *S. avermitilis* o *S. coelicolor* (véase la tabla 5). Entonces pueden expresarse los genes indicados en la tabla 4 para permitir el inicio y la elongación de ácidos grasos ω -cíclicos. Alternativamente, pueden aislarse los genes homólogos a partir de microorganismos que producen cyFA y expresarse en una célula huésped (por ejemplo, *E. coli*).

Tabla 5: Genes para la síntesis de CHC-CoA

Organismo	Gen	N.º de registro de GenBank
<i>Streptomyces collinus</i>	<i>ansJK</i> <i>ansL</i> <i>chcA</i> <i>ansM</i> <i>chcB</i>	U72144* AF268489
<i>Streptomyces</i> sp. HK803	<i>pmlJK</i> <i>pmlL</i> <i>chcA</i> <i>pmlM</i>	AAQ84158 AAQ84159 AAQ84160 AAQ84161
<i>Streptomyces coelicolor</i>	<i>chcB/caiD</i>	NP_629292
<i>Streptomyces avermitilis</i>	<i>chcB/caiD</i>	NP_629292

*Sólo se indica *chcA* en la entrada de GenBank U72144, *ansJKLM* son según Chen *et al.* (Eur. J. Biochem. 261:98-107, 1999).

- 5 Los genes indicados en la tabla 4 (*fabH*, *acp* y *fabF*) permiten el inicio y la elongación de ácidos grasos ω -cíclicos porque tienen una amplia especificidad de sustrato. Si la coexpresión de cualquiera de estos genes con los genes indicados en la tabla 5 no proporciona cyFA, entonces pueden aislarse (por ejemplo, usando cebadores de PCR degenerados o sondas de secuencias de ADN heterólogas) homólogos de *fabH*, *acp* y/o *fabF* de microorganismos que producen cyFA (por ejemplo, los indicados en la tabla 6) y coexpresarse.

Tabla 6: Ejemplos no limitativos de microorganismos que contienen ácidos grasos ω -cíclicos

Organismo	Referencia
<i>Curtobacterium pusillum</i>	ATCC19096
<i>Alicyclobacillus acidoterrestris</i>	ATCC49025
<i>Alicyclobacillus acidocaldarius</i>	ATCC27009
<i>Alicyclobacillus cycloheptanicus</i> *	Moore, <i>J. Org. Chem.</i> 62: págs. 2173, 1997

- 10 *Usa cicloheptilcarbonil-CoA y no ciclohexilcarbonil-CoA como precursor para la biosíntesis de cyFA.

Niveles de saturación de aldehídos grasos

- 15 El grado de saturación en ácidos grasos puede controlarse regulando el grado de saturación de productos intermedios de ácidos grasos. Por ejemplo, las familias de genes *sfa*, *gns* y *fab* pueden expresarse, sobreexpresarse, o expresarse a niveles reducidos, para controlar la saturación de ácidos grasos. La figura 5 indica ejemplos no limitativos de genes en estas familias de genes que pueden usarse en los métodos y las células huésped descritos en el presente documento. La figura 6 indica genes relacionados con *fabA* adicionales y la figura 7 indica genes relacionados con *fabB* adicionales.

- 20 Por ejemplo, pueden modificarse por ingeniería células huésped para producir ácidos grasos insaturados modificando por ingeniería el huésped de producción para sobreexpresar *fabB* o haciendo crecer el huésped de producción a bajas temperaturas (por ejemplo, inferiores a 37°C). *FabB* tiene preferencia por *cis*- δ 3decanoil-ACP y da como resultado la producción de ácidos grasos insaturados en *E. coli*. La sobreexpresión de *fabB* da como resultado la producción de un porcentaje significativo de ácidos grasos insaturados (de Mendoza *et al.*, *J. Biol. Chem.* 258:2098-2101, 1983). El gen *fabB* puede insertarse y expresarse en células huésped que no tienen de manera natural el gen. Entonces pueden usarse estos ácidos grasos insaturados como productos intermedios en células huésped que se modifican por ingeniería para producir derivados de ácidos grasos, tales como aldehídos grasos.

- 30 En otros casos, puede delecionarse un represor de la biosíntesis de ácidos grasos, por ejemplo, *fabR* (registro de GenBank NP_418398), lo cual da como resultado un aumento en la producción de ácidos grasos insaturados en *E. coli* (Zhang *et al.*, *J. Biol. Chem.* 277: 15558, 2002). Pueden realizarse delecciones similares en otras células huésped. Puede lograrse un aumento adicional de ácidos grasos insaturados, por ejemplo, sobreexpresando *fabM* (trans-2,*cis*-3-decenoil-ACP isomerasa, registro de GenBank DAA05501) y controlando la expresión de *fabK* (trans-2-enoil-ACP reductasa II, registro de GenBank NP_357969) de *Streptococcus pneumoniae* (Marrakchi *et al.*, *J. Biol. Chem.* 277: 44809, 2002), al tiempo que se deleciona *fabI* de *E. coli* (trans-2-enoil-ACP reductasa, registro de GenBank NP_415804). En algunos ejemplos, puede atenuarse el gen *fabF* endógeno, aumentando por tanto el porcentaje de palmitoleato (C16:1) producido.

- 35 En aún otros ejemplos, pueden modificarse por ingeniería células huésped para producir ácidos grasos saturados reduciendo la expresión de un gen *sfa*, *gns* y/o *fab*.

En algunos casos, puede modificarse por ingeniería una célula huésped para expresar un nivel atenuado de una

deshidratasa/isomerasa y/o una cetoacil-ACP sintasa. Por ejemplo, puede modificarse por ingeniería una célula huésped para expresar un nivel reducido de *fabA*, *fabB*, un gen indicado en la figura 6 y/o un gen indicado en la figura 7. En algunos casos, la célula huésped puede hacerse crecer en presencia de ácidos grasos insaturados. En otros casos, la célula huésped puede modificarse adicionalmente por ingeniería para expresar o sobreexpresar un gen que codifica para una enzima desaturasa. Un ejemplo no limitativo de una desaturasa es DesA de *B. subtilis* (AF037430). En la técnica se conocen otros genes que codifican para enzimas desaturasas y pueden usarse en las células huésped y los métodos descritos en el presente documento, tales como desaturasas que usan acil-ACP, tales como hexadecanoil-ACP u octadecanoil-ACP. Los ácidos grasos saturados pueden usarse para producir derivados de ácidos grasos, tales como aldehídos grasos, tal como se describe en el presente documento.

10 Modificación por ingeniería genética de células huésped para producir aldehídos grasos

Pueden usarse diversas células huésped para producir aldehídos grasos, tal como se describe en el presente documento. Una célula huésped puede ser cualquier célula procariota o eucariota. Por ejemplo, puede expresarse un polipéptido descrito en el presente documento en células bacterianas (tal como *E. coli*), células de insectos, levadura o células de mamíferos (tal como células de ovario de hámster chino (CHO), células COS, células VERO, células BHK, células HeLa, células Cv1, células MDCK, células 293, células 3T3 o células PC12). Otras células huésped a modo de ejemplo incluyen células de los miembros de los géneros *Escherichia*, *Bacillus*, *Lactobacillus*, *Rhodococcus*, *Pseudomonas*, *Aspergillus*, *Trichoderma*, *Neurospora*, *Fusarium*, *Humicola*, *Rhizomucor*, *Kluyveromyces*, *Pichia*, *Mucor*, *Myceliophthora*, *Penicillium*, *Phanerochaete*, *Pleurotus*, *Trametes*, *Chrysosporium*, *Saccharomyces*, *Schizosaccharomyces*, *Yarrowia* o *Streptomyces*. Aún otras células huésped a modo de ejemplo pueden ser una célula *Bacillus lentus*, una célula *Bacillus brevis*, una célula *Bacillus stearothermophilus*, una célula *Bacillus licheniformis*, una célula *Bacillus alkalophilus*, una célula *Bacillus coagulans*, una célula *Bacillus circulans*, una célula *Bacillus pumilis*, una célula *Bacillus thuringiensis*, una célula *Bacillus clausii*, una célula *Bacillus megaterium*, una célula *Bacillus subtilis*, una célula *Bacillus amyloliquefaciens*, una célula *Trichoderma koningii*, una célula *Trichoderma viride*, una célula *Trichoderma reesei*, una célula *Trichoderma longibrachiatum*, una célula *Aspergillus awamori*, una célula *Aspergillus fumigates*, una célula *Aspergillus foetidus*, una célula *Aspergillus nidulans*, una célula *Aspergillus niger*, una célula *Aspergillus oryzae*, una célula *Humicola insolens*, una célula *Humicola lanuginosa*, una célula *Rhizomucor miehei*, una célula *Mucor miehei*, una célula *Streptomyces lividans*, una célula *Streptomyces murinus* o una célula *Actinomycetes*. Otras células huésped son células huésped cianobacterianas.

30 En una realización preferida, la célula huésped es una célula *E. coli*, una célula *Saccharomyces cerevisiae* o una célula *Bacillus subtilis*. En una realización más preferida, la célula huésped es de las cepas B, C, K o W de *E. coli*. Los expertos en la técnica conocen otras células huésped adecuadas.

En los documentos WO2009/111513 y WO2009/111672 se describen células huésped adicionales que pueden usarse en los métodos descritos en el presente documento.

35 Diversos métodos bien conocidos en la técnica pueden usarse para modificar por ingeniería genética células huésped para producir aldehídos grasos. Los métodos pueden incluir el uso de vectores, preferiblemente vectores de expresión, que contienen un ácido nucleico que codifica para un polipéptido de biosíntesis de aldehídos grasos descrito en el presente documento, variante de polipéptido o un fragmento del mismo. Los expertos en la técnica apreciarán que pueden usarse una variedad de vectores virales (por ejemplo, vectores de retrovirus, vectores de lentivirus, vectores de adenovirus y vectores de virus adenoasociados) y vectores no virales en los métodos descritos en el presente documento.

Los vectores de expresión recombinantes descritos en el presente documento incluyen un ácido nucleico descrito en el presente documento en una forma adecuada para la expresión del ácido nucleico en una célula huésped. Los vectores de expresión recombinantes pueden incluir una o más secuencias de control, seleccionadas basándose en la célula huésped que va a usarse para la expresión. La secuencia de control está unida operativamente a la secuencia de ácido nucleico que va a expresarse. Tales secuencias de control se describen, por ejemplo, en Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990). Las secuencias de control incluyen aquellas que dirigen la expresión constitutiva de una secuencia de nucleótidos en muchos tipos de células huésped y aquellas que dirigen la expresión de la secuencia de nucleótidos únicamente en determinadas células huésped (por ejemplo, secuencias reguladoras específicas de tejido). Los expertos en la técnica apreciarán que el diseño del vector de expresión puede depender de factores tales como la elección de la célula huésped que va a transformarse, el nivel de expresión de proteína deseado, etc. La vectores de expresión descritos en el presente documento pueden introducirse en células huésped para producir polipéptidos, incluyendo polipéptidos de fusión, codificados por los ácidos nucleicos tal como se describe en el presente documento.

55 Pueden diseñarse vectores de expresión recombinantes para la expresión de un polipéptido de biosíntesis de aldehídos grasos o variante en células procariotas o eucariotas (por ejemplo, células bacterianas, tales como *E. coli*, células de insectos (por ejemplo, usando vectores de expresión de baculovirus), células de levaduras o células de mamíferos). Se comentan adicionalmente células huésped adecuadas en Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990). Alternativamente, el vector de expresión recombinante puede transcribirse y traducirse *in vitro*, por ejemplo, usando secuencias reguladoras de promotor de

T7 y polimerasa de T7.

La expresión de polipéptidos en procariotas, por ejemplo, *E. coli*, se lleva a cabo con la mayor frecuencia con vectores que contienen promotores constitutivos o inducibles que dirigen la expresión de polipéptidos o bien de fusión o bien no de fusión. Los vectores de fusión añaden varios aminoácidos a un polipéptido codificado en los mismos, habitualmente en el extremo amino-terminal del polipéptido recombinante. Tales vectores de fusión sirven normalmente para tres fines: (1) aumentar la expresión del polipéptido recombinante; (2) aumentar la solubilidad del polipéptido recombinante; y (3) ayudar a la purificación del polipéptido recombinante actuando como ligando en la purificación por afinidad. Con frecuencia, en vectores de expresión de fusión, se introduce un sitio de escisión proteolítica en la unión del resto de fusión y el polipéptido recombinante. Esto permite la separación del polipéptido recombinante del resto de fusión tras la purificación del polipéptido de fusión. Los ejemplos de tales enzimas, y sus secuencias de reconocimiento relacionadas, incluyen factor Xa, trombina y enterocinas. Los vectores de expresión de fusión a modo de ejemplo incluyen pGEX (Pharmacia Biotech Inc; Smith *et al*, Gen (1988) 67:31-40), pMAL (New England Biolabs, Beverly, Mass.) y pRITS (Pharmacia, Piscataway, N.J.), que fusionan glutatión S-transferasa (GST), proteína de unión a maltosa E, o proteína A, respectivamente, con el polipéptido recombinante diana.

Los ejemplos de vectores de expresión de *E. coli* inducibles, no de fusión, incluyen pTrc (Amann *et al*, Gene (1988) 69:301-315) y pET 11d (Studier *et al*, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 60-89). La expresión de genes diana a partir del vector pTrc se basa en la transcripción mediante ARN polimerasa del huésped a partir de un promotor de fusión híbrido trp-lac. La expresión de genes diana a partir del vector pET 11d se basa en la transcripción a partir de un promotor de fusión gn10-lac de T7 mediada por una ARN polimerasa viral coexpresada (gn1 de T7). Esta polimerasa viral se suministra mediante cepas huésped BL21(DE3) o HMS174(DE3) a partir de un profago λ residente que alberga un gen de gn1 de T7 bajo el control transcripcional del promotor lacUV 5.

Una estrategia para maximizar la expresión de polipéptidos recombinantes es expresar el polipéptido en una célula huésped con una capacidad afectada para escindir proteolíticamente el polipéptido recombinante (véase Gottesman, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 119-128). Otra estrategia es alterar la secuencia de ácido nucleico que va a insertarse en un vector de expresión de modo que los codones individuales para cada aminoácido sean aquellos usados de manera preferencial en la célula huésped (Wada *et al*, Nucleic Acids Res. (1992) 20:2111-2118). Tal alteración de secuencias de ácido nucleico puede llevarse a cabo mediante técnicas de síntesis de ADN convencionales.

En otra realización, la célula huésped es una célula de levadura. En esta realización, el vector de expresión es un vector de expresión de levadura. Los ejemplos de vectores para expresión en levadura *S. cerevisiae* incluyen pYepSec1 (Baldari *et al*, EMBO J. (1987) 6:229-234), pMFA (Kurjan *et al*, Cell (1982) 30:933-943), pJRY88 (Schultz *et al*, Gene (1987) 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.) y picZ (Invitrogen Corp, San Diego, Calif.).

Alternativamente, un polipéptido descrito en el presente documento puede expresarse en células de insectos usando vectores de expresión de baculovirus. Los vectores de baculovirus disponibles para la expresión de proteínas en células de insectos en cultivo (por ejemplo, células Sf9) incluyen, por ejemplo, la serie de pAc (Smith *et al*, Mol. Cell Biol. (1983) 3:2156-2165) y la serie de pVL (Lucklow *et al*, Virology (1989) 170:31-39).

En aún otra realización, los ácidos nucleicos descritos en el presente documento pueden expresarse en células de mamíferos usando un vector de expresión de mamíferos. Los ejemplos de vectores de expresión de mamíferos incluyen pCDM8 (Seed, Nature (1987) 329:840) y pMT2PC (Kaufman *et al*, EMBO J. (1987) 6:187-195). Cuando se usa en células de mamíferos, las funciones de control del vector de expresión pueden proporcionarse mediante elementos reguladores virales. Por ejemplo, promotores comúnmente usados se derivan de polioma, adenovirus 2, citomegalovirus y virus del simio 40. Otros sistemas de expresión adecuados para células tanto procariotas como eucariotas se describen en los capítulos 16 y 17 de Sambrook *et al*, eds., Molecular Cloning: A Laboratory Manual. 2ª ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

Pueden introducirse vectores en células procariotas o eucariotas mediante técnicas convencionales de transformación o transfección. Tal como se usan en el presente documento, los términos "transformación" y "transfección" se refieren a una variedad de técnicas reconocidas en la técnica para introducir ácido nucleico foráneo (por ejemplo, ADN) en una célula huésped, incluyendo coprecipitación con fosfato de calcio o cloruro de calcio, transfección mediada por DEAE-dextrano, lipofección o electroporación. Pueden encontrarse métodos adecuados para transformar o transfectar células huésped, por ejemplo, en Sambrook *et al*. (citado anteriormente).

Para la transformación estable de células bacterianas, se conoce que, dependiendo del vector de expresión y la técnica de transformación usados, sólo una pequeña fracción de las células captarán y replicarán el vector de expresión. Con el fin de identificar y seleccionar esos transformantes, puede introducirse un gen que codifica para un marcador seleccionable (por ejemplo, resistencia a antibióticos) en las células huésped junto con el gen de interés. Los marcadores seleccionables incluyen aquellos que confieren resistencia a fármacos, tales como ampicilina, canamicina, cloranfenicol o tetraciclina. Pueden introducirse ácidos nucleicos que codifican para un marcador seleccionable en una célula huésped en el mismo vector que el que codifica para un polipéptido descrito en el

presente documento o pueden introducirse en un vector separado. Pueden identificarse células transfectadas de manera estable con el ácido nucleico introducido mediante selección con fármacos (por ejemplo, las células que han incorporado el gen de marcador seleccionable sobrevivirán, mientras que las demás células morirán).

5 Para la transfección estable de células de mamíferos, se conoce que, dependiendo del vector de expresión y la técnica de transfección usados, sólo una pequeña fracción de células puede integrar el ADN foráneo en su genoma. Con el fin de identificar y seleccionar estos integrantes, puede introducirse un gen que codifica para un marcador seleccionable (por ejemplo, resistencia a antibióticos) en las células huésped junto con el gen de interés. Los marcadores seleccionables preferidos incluyen aquellos que confieren resistencia a fármacos, tales como G418, higromicina y metotrexato. Pueden introducirse ácidos nucleicos que codifican para un marcador seleccionable en
10 una célula huésped en el mismo vector que el que codifica para un polipéptido descrito en el presente documento o pueden introducirse en un vector separado. Pueden identificarse células transfectadas de manera estable con el ácido nucleico introducido mediante selección con fármacos (por ejemplo, las células que han incorporado el gen de marcador seleccionable sobrevivirán, mientras que las demás células morirán).

Proteínas de transporte

15 Las proteínas de transporte pueden exportar polipéptidos y compuestos orgánicos (por ejemplo, aldehydos grasos) fuera de una célula huésped. Muchas proteínas de transporte y eflujo sirven para excretar una amplia variedad de compuestos y pueden modificarse de manera natural para ser selectivas para tipos particulares de hidrocarburos.

Ejemplos no limitativos de proteínas de transporte adecuadas son proteínas de transporte de casete de unión a ATP (ABC), proteínas de eflujo y proteínas transportadoras de ácidos grasos (FATP). Los ejemplos no limitativos
20 adicionales de proteínas de transporte adecuadas incluyen las proteínas de transporte de ABC de organismos tales como *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Alkaligenes eutrophus* y *Rhodococcus erythropolis*. En la figura 5 se indican proteínas de transporte de ABC a modo de ejemplo que pueden usarse (por ejemplo, CER5, AtMRP5, AmiS2 y AtPGP1). También pueden elegirse células huésped por su capacidad endógena para secretar compuestos orgánicos. La eficacia de la producción y secreción de compuestos orgánicos dentro del entorno de la célula huésped (por ejemplo, medio de cultivo, caldo de fermentación) puede expresarse como una razón de producto intracelular con respecto a producto extracelular. En algunos ejemplos, la razón puede ser de aproximadamente 5:1,
25 4:1, 3:1, 2:1, 1:1, 1:2, 1:3, 1:4 ó 1:5.

Fermentación

30 La producción y el aislamiento de aldehydos grasos puede potenciarse empleando técnicas de fermentación beneficiosas. Un método para maximizar la producción al tiempo que se reducen los costes es aumentar el porcentaje de la fuente de carbono que se convierte en productos de hidrocarburos.

Durante ciclos de vida celulares normales, se usa carbono en funciones celulares, tales como producción de lípidos, sacáridos, proteínas, ácidos orgánicos y ácidos nucleicos. La reducción de la cantidad de carbono necesaria para actividades relacionadas con el crecimiento puede aumentar la eficacia de la conversión de la fuente de carbono en
35 producto. Esto puede lograrse, por ejemplo, haciendo crecer en primer lugar células huésped hasta una densidad deseada (por ejemplo, una densidad alcanzada en el pico de la fase de crecimiento logarítmica). En tal punto, pueden emplearse genes de punto de control de la replicación para detener el crecimiento de células. Específicamente, pueden usarse mecanismos de percepción de quórum (revisado en Camilli *et al.*, Science 311:1113, 2006; Venturi FEMS Microbio. Rev. 30:274-291, 2006; y Reading *et al.*, FEMS Microbiol. Lett. 254:1-11, 2006) para activar genes de punto de control, tales como *p53*, *p21* u otros genes de punto de control.

Los genes que pueden activarse para detener la replicación y el crecimiento celular en *E. coli* incluyen genes *umuDC*. La sobreexpresión de genes *umuDC* detiene la progresión desde la fase estacionaria hasta el crecimiento exponencial (Murli *et al.*, J. of Bact. 182:1127, 2000). UmuC es una ADN polimerasa que puede llevar a cabo la síntesis translesión a lo largo de lesiones no codificantes, la base mecánica de la mayor parte de la mutagénesis
45 por UV y química. Los productos del gen *umuDC* están implicados en el proceso de síntesis translesión y también sirven como punto de control de daño de la secuencia de ADN. Los productos del gen *umuDC* incluyen UmuC, UmuD, umuD', UmuD₂C, UmuD'₂ y UmuD₂. Simultáneamente, pueden activarse genes de producción de producto, minimizando así la necesidad de usar rutas de replicación y mantenimiento mientras está produciéndose un aldehydo graso. También pueden modificarse por ingeniería células huésped para expresar *umuC* y *umuD* de *E. coli* en pBAD24 bajo el sistema promotor *prpBCDE* mediante síntesis *de novo* de este gen con los genes de producción de
50 producto final apropiados.

El porcentaje de carbonos introducidos convertidos en aldehydos grasos puede ser un impulsor del coste. Cuanto más eficaz es el procedimiento (es decir, cuanto mayor es el porcentaje de carbonos introducidos convertidos en aldehydos grasos), más económico será el procedimiento. Para fuentes de carbono que contienen oxígeno (por
55 ejemplo, fuentes a base de glucosa y otros hidrato de carbono), el oxígeno debe liberarse en forma de dióxido de carbono. Por cada 2 átomos de oxígeno liberados, también se libera un átomo de carbono conduciendo a una eficacia metabólica teórica máxima de aproximadamente el 34% (p/p) (para productos derivados de ácidos grasos). Sin embargo, esta cifra cambia para otros compuestos orgánicos y fuentes de carbono. Las eficacias típicas en la

bibliografía son aproximadamente inferiores al 5%. Células huésped modificadas por ingeniería para producir aldehídos grasos pueden tener una eficacia de más de aproximadamente el 1, el 3, el 5, el 10, el 15, el 20, el 25 y el 30%. En un ejemplo, las células huésped pueden mostrar una eficacia de aproximadamente el 10% a aproximadamente el 25%. En otros ejemplos, tales células huésped pueden mostrar una eficacia de aproximadamente el 25% a aproximadamente el 30%. En otros ejemplos, las células huésped pueden mostrar una eficacia de más del 30%.

La célula huésped puede modificarse adicionalmente por ingeniería para expresar celulosomas recombinantes, tales como los descritos en la solicitud PCT número PCT/US2007/003736. Estos celulosomas pueden permitir a la célula huésped usar material celulósico como fuente de carbono. Por ejemplo, la célula huésped puede modificarse por ingeniería adicionalmente para expresar invertasas (EC 3.2.1.26) de modo que puede usarse sacarosa como fuente de carbono. De manera similar, la célula huésped puede modificarse por ingeniería usando las enseñanzas descritas en las patentes estadounidenses n.ºs 5.000.000; 5.028.539; 5.424.202; 5.482.846 y 5.602.030; de modo que la célula huésped puede asimilar carbono de manera eficaz y usar materiales celulósicos como fuentes de carbono.

En un ejemplo, la cámara de fermentación puede encerrar una fermentación que está experimentando una reducción continua. En este caso, puede crearse un entorno reductor estable. El equilibrio de electrones puede mantenerse mediante la liberación de dióxido de carbono (en forma gaseosa). Esfuerzos por aumentar el equilibrio de NAD/H y NADP/H también pueden facilitar la estabilización del equilibrio de electrones. También puede potenciarse la disponibilidad de NADPH intracelular modificando por ingeniería la célula huésped para expresar una NADH:NADPH transhidrogenasa. La expresión de una o más NADH:NADPH transhidrogenasas convierte el NADH producido en la glicólisis en NADPH, lo cual puede potenciar la producción de aldehídos grasos.

Para la producción a pequeña escala, pueden hacerse crecer las células huésped modificadas por ingeniería en lotes, por ejemplo, de aproximadamente 100 ml, 500 ml, 1 l, 2 l, 5 l o 10 l; fermentarse; e inducirse para expresar genes de biosíntesis de aldehídos grasos deseados basándose en los genes específicos codificados en los plásmidos apropiados. Para una producción a gran escala, pueden hacerse crecer las células huésped modificadas por ingeniería en lotes de aproximadamente 10 l, 100 l, 1000 l, 10.000 l, 100.000 l, 1.000.000 l o más; fermentarse; e inducirse para expresar genes de biosíntesis de aldehídos grasos deseados basándose en los genes específicos codificados en los plásmidos apropiados o incorporados en el genoma de la célula huésped.

Por ejemplo, un huésped de producción adecuado, tal como células *E. coli*, que alberga plásmidos que contienen los genes de biosíntesis de aldehídos grasos deseados o que tiene los genes de biosíntesis de aldehídos grasos integrados en su cromosoma, puede incubarse en un reactor adecuado, por ejemplo un reactor de 1 l, durante 20 horas a 37°C en medio M9 suplementado con glucosa al 2%, carbenicilina y cloranfenicol. Cuando la DO_{600} del cultivo alcanza 0,9, puede inducirse el huésped de producción con IPTG para activar los sistemas génicos modificados por ingeniería para la producción de aldehídos grasos. Tras la incubación, pueden extraerse los medios gastados y puede examinarse la fase orgánica para determinar la presencia de aldehídos grasos usando CG-EM.

En algunos casos, tras la primera hora de inducción, pueden extraerse cada hora alícuotas de no más de aproximadamente el 10% del volumen celular total y dejarse reposar sin agitación para permitir que los aldehídos grasos suban a la superficie y experimenten separación de fases espontánea o precipitación. Entonces puede recogerse el componente de aldehídos grasos, y devolverse la fase acuosa a la cámara de reacción. La cámara de reacción puede hacerse funcionar de manera continua. Cuando la DO_{600} disminuye por debajo de 0,6, pueden reemplazarse las células por un nuevo lote que se ha hecho crecer a partir de un cultivo de siembra.

Producción de aldehídos grasos usando métodos libres de células

En algunos métodos descritos en el presente documento, puede producirse un aldehído graso usando un polipéptido purificado descrito en el presente documento y un sustrato descrito en el presente documento. Por ejemplo, puede modificarse por ingeniería una célula huésped para expresar un polipéptido de biosíntesis de aldehídos grasos o variante tal como se describe en el presente documento. La célula huésped puede cultivarse en condiciones adecuadas para permitir la expresión del polipéptido. Entonces pueden generarse extractos libres de células usando métodos conocidos. Por ejemplo, pueden someterse las células huésped a lisis usando detergentes o mediante sonicación. Pueden purificarse los polipéptidos expresados usando métodos conocidos. Tras obtener los extractos libres de células, pueden añadirse sustratos descritos en el presente documento a los extractos libres de células y mantenerse en condiciones para permitir la conversión de los sustratos en aldehídos grasos. Entonces pueden separarse los aldehídos grasos y purificarse usando técnicas conocidas.

Procesamiento tras la producción

Los aldehídos grasos producidos durante la fermentación pueden separarse de los medios de fermentación. Puede usarse cualquier técnica conocida para separar aldehídos grasos de medios acuosos. Un procedimiento de separación a modo de ejemplo es un procedimiento de separación de dos fases (bifásico). Este procedimiento implica fermentar las células huésped modificadas por ingeniería genética en condiciones suficientes para producir un aldehído graso, permitiendo que el aldehído graso se acumule en una fase orgánica, y separar la fase orgánica del caldo de fermentación acuoso. Este método puede ponerse en práctica en procedimientos de fermentación tanto

continuos como discontinuos.

La separación bifásica usa la relativa inmiscibilidad de aldehídos grasos para facilitar la separación. Inmiscible se refiere a la relativa incapacidad de un compuesto para disolverse en agua y se define mediante el coeficiente de distribución del compuesto. Un experto habitual en la técnica apreciará que eligiendo un caldo de fermentación y una fase orgánica, de tal manera que el aldehído graso que está produciéndose tenga un alto valor de logP, el aldehído graso puede separarse a la fase orgánica, incluso a concentraciones muy bajas, en el recipiente de fermentación.

Los aldehídos grasos producidos mediante los métodos descritos en el presente documento pueden ser relativamente inmiscibles en el caldo de fermentación, así como en el citoplasma. Por tanto, el aldehído graso puede acumularse en una fase orgánica o bien intracelularmente o bien extracelularmente. La acumulación de los productos en la fase orgánica puede reducir el impacto del aldehído graso sobre la función celular y puede permitir que la célula huésped produzca más producto.

Los métodos descritos en el presente documento pueden dar como resultado la producción de compuestos homogéneos en los que al menos aproximadamente el 60%, el 70%, el 80%, el 90% o el 95% de los aldehídos grasos producidos tendrán longitudes de cadena de carbono que varían en menos de aproximadamente 6 carbonos, menos de aproximadamente 4 carbonos o menos de aproximadamente 2 carbonos. Estos compuestos también pueden producirse con un grado de saturación relativamente uniforme. Estos compuestos pueden usarse directamente como combustibles, aditivos para combustibles, materiales de partida para la producción de otros compuestos químicos (por ejemplo, polímeros, tensioactivos, plásticos, materiales textiles, disolventes, adhesivos, etc.), o aditivos para el cuidado personal. Estos compuestos también pueden usarse como materia prima para reacciones posteriores, por ejemplo, hidrogenación, craqueo catalítico (por ejemplo, mediante hidrogenación, pirólisis, o ambas), para producir otros productos.

En algunas realizaciones, los aldehídos grasos producidos usando métodos descritos en el presente documento pueden contener entre aproximadamente el 50% y aproximadamente el 90% de carbono; o entre aproximadamente 5% y aproximadamente el 25% de hidrógeno. En otras realizaciones, los aldehídos grasos producidos usando métodos descritos en el presente documento pueden contener entre aproximadamente el 65% y aproximadamente el 85% de carbono; o entre aproximadamente el 10% y aproximadamente el 15% de hidrógeno.

Bioproductos

Bioproductos (por ejemplo, aldehídos grasos) que comprenden compuestos orgánicos producidos de manera biológica, particularmente aldehídos grasos producidos de manera biológica usando la ruta de biosíntesis de ácidos grasos, no se han producido a partir de fuentes renovables y, como tales, son nuevas composiciones de materia. Estos nuevos bioproductos pueden distinguirse de compuestos orgánicos derivados de carbono petroquímico basándose en la huella isotópica carbono doble o datación por ^{14}C . Adicionalmente, la fuente específica de carbono de fuente biológica (por ejemplo, glucosa frente a glicerol) puede determinarse mediante la huella isotópica de carbono doble (véase, por ejemplo, la patente estadounidense n.º 7.169.588).

La capacidad para distinguir bioproductos de compuestos orgánicos basados en petróleo es beneficiosa en el rastreo de estos materiales en el comercio. Por ejemplo, los compuestos orgánicos o productos químicos que comprenden perfiles de isótopos de carbono tanto de base biológica como basados en petróleo pueden distinguirse de compuestos orgánicos y productos químicos producidos únicamente a partir de materiales basados en petróleo. Por tanto, puede realizarse un seguimiento de los presentes materiales en el comercio basándose en su perfil de isótopos de carbono único.

Los bioproductos pueden distinguirse de compuestos orgánicos basados en petróleo comparando la razón de isótopos de carbono estable ($^{13}\text{C}/^{12}\text{C}$) en cada combustible. La razón de $^{13}\text{C}/^{12}\text{C}$ en un bioproducto dado es una consecuencia de la razón de $^{13}\text{C}/^{12}\text{C}$ en el dióxido de carbono atmosférico en el momento en el que se fija el dióxido de carbono. También refleja la ruta metabólica precisa. También se producen variaciones regionales. El petróleo, las plantas C_3 (de hoja ancha), las plantas C_4 (las hierbas) y los carbonatos marinos muestran todos ellos diferencias significativas en cuanto a $^{13}\text{C}/^{12}\text{C}$ y los correspondientes valores de $\delta^{13}\text{C}$. Además, la materia lipídica de plantas C_3 y C_4 presenta resultados en análisis diferentes de los materiales derivados de los componentes de hidratos de carbono de las mismas plantas como consecuencia de la ruta metabólica.

Dentro de la precisión de la medición, ^{13}C muestra grandes variaciones debido a efectos de fraccionamiento isotópico, el más significativo de los cuales para los bioproductos es el mecanismo fotosintético. La principal causa de diferencias en la razón de isótopos de carbono en plantas está estrechamente asociada con diferencias en la ruta del metabolismo de carbono fotosintética en las plantas, particularmente la reacción que se produce durante la carboxilación primaria (es decir, la fijación inicial de CO_2 atmosférico). Dos grandes clases de vegetación son aquellas que incorporan el ciclo fotosintético " C_3 " (o de Calvin-Benson) y aquellas que incorporan el ciclo fotosintético " C_4 " (o de Hatch-Slack).

En las plantas C_3 , la reacción de carboxilación o fijación de CO_2 primaria implica la enzima ribulosa-1,5-difosfato carboxilasa, y el primer producto estable es un compuesto de 3-carbonos. Las plantas C_3 , tales como las de madera dura y las coníferas, son dominantes en las zonas de clima templado.

En las plantas C₄, una reacción de carboxilación adicional que implica otra enzima, fosfoenol-piruvato carboxilasa, es la reacción de carboxilación primaria. El primer compuesto de carbono estable es un ácido de 4 carbonos que posteriormente se descarboxila. El CO₂ así liberado vuelve a fijarse mediante el ciclo C₃. Ejemplos de plantas C₄ son hierbas tropicales, maíz y caña de azúcar.

- 5 Tanto las plantas C₄ como las C₃ muestran un intervalo de razones isotópicas ¹³C/¹²C, pero valores típicos son de aproximadamente -7 a aproximadamente -13 por mil para plantas C₄ y de aproximadamente -19 a aproximadamente -27 por mil para plantas C₃ (véase, por ejemplo, Stuiver *et al*, Radiocarbon 19:355, 1977). El carbón y el petróleo se encuentran generalmente en este último intervalo. La escala de medición de ¹³C se definió originalmente mediante un cero establecido por caliza belemnita de Pee Dee (PDB), en la que los valores vienen dados en desviaciones en partes por mil con respecto a este material. Los valores de “δ¹³C” se expresan en partes por cada mil (por mil), abreviado, ‰, y se calculan de la siguiente manera:

$$\delta^{13}\text{C} (\text{‰}) = \left[\left(\frac{^{13}\text{C}/^{12}\text{C}}{\text{muestra}} - \left(\frac{^{13}\text{C}/^{12}\text{C}}{\text{patrón}} \right) \right) / \left(\frac{^{13}\text{C}/^{12}\text{C}}{\text{patrón}} \right) \times 1000 \right]$$

- 15 Dado que el material de referencia (RM) PDB se ha agotado, se ha desarrollado una serie de RM alternativos en colaboración con IAEA, USGS, NIST, y otros laboratorios de isótopos internacionales seleccionados. La notación para las desviaciones por mil con respecto a PDB es δ¹³C. Las mediciones se realizan con CO₂ mediante espectrometría de masas de razón estable de alta precisión (EMRI) con iones moleculares con masas de 44, 45 y 46.

- 20 Las composiciones descritas en el presente documento incluyen bioproductos producidos mediante cualquiera de los métodos descritos en el presente documento. Específicamente, el bioproducto puede tener un δ¹³C de aproximadamente -28 o más, aproximadamente -27 o más, -20 o más, -18 o más, -15 o más, -13 o más, -10 o más, o -8 o más. Por ejemplo, el bioproducto puede tener un δ¹³C de aproximadamente -30 a aproximadamente -15, de aproximadamente -27 a aproximadamente -19, de aproximadamente -25 a aproximadamente -21, de aproximadamente -15 a aproximadamente -5, de aproximadamente -13 a aproximadamente -7, o de aproximadamente -13 a aproximadamente -10. En otros casos, el bioproducto puede tener un δ¹³C de aproximadamente -10, -11, -12 o -12,3.

- 25 También pueden distinguirse bioproductos de compuestos orgánicos basados en petróleo comparando la cantidad de ¹⁴C en cada compuesto. Dado que ¹⁴C tiene una semivida nuclear de 5730 años, los combustibles basados en petróleo que contienen carbono “más antiguo” pueden distinguirse de bioproductos que contienen carbono “más nuevo” (véase, por ejemplo, Currie, “Source Apportionment of Atmospheric Particles”, Characterization of Environmental Particles, J. Buffle y H. P. van Leeuwen, Eds., 1 of Vol. I of the IUPAC Environmental Analytical Chemistry Series (Lewis Publishers, Inc) (1992) 3-74).

- 30 La suposición básica en la datación por radiocarbono es que la constancia de la concentración de ¹⁴C en la atmósfera conduce a la constancia de ¹⁴C en organismos vivos. Sin embargo, debido a las pruebas nucleares atmosféricas desde 1950 y a la quema de combustible fósil desde 1850, ¹⁴C ha adquirido una segunda característica temporal geoquímica. Su concentración en el CO₂ atmosférico, y por tanto en la biosfera viva, aproximadamente se duplicó en el punto máximo de las pruebas nucleares, a mediados de los años 1960. Desde entonces ha estado volviendo gradualmente a la tasa isotópica (¹⁴C / ¹²C) inicial (atmosférica) cosmogénica en estado estacionario de aproximadamente 1,2 x 10⁻¹², con una “semivida” de relajación aproximada de 7-10 años. (Esta última semivida no debe tomarse literalmente; en vez de eso, debe usarse la función de introducción/descomposición nuclear atmosférica detallada para realizar un seguimiento de la variación de ¹⁴C atmosférico y biosférico desde el inicio de la era nuclear).

- 35 Esta última característica temporal de ¹⁴C biosférico es la que conserva la promesa de la datación anual de carbono biosférico reciente. ¹⁴C puede medirse mediante espectrometría de masas con aceleradores (EMA), facilitándose los resultados en unidades de “fracción de carbono moderno” (f_M). f_M se define por los materiales de referencia patrón (SRM) 4990B y 4990C del National Institute of Standards and Technology (NIST). Tal como se usa en el presente documento, la “fracción de carbono moderno” o “f_M” tiene el mismo significado que el definido por los materiales de referencia patrón (SRM) 4990B y 4990C del National Institute of Standards and Technology (NIST), conocidos como patrones de ácidos oxálicos HOxI y HOxII, respectivamente. La definición fundamental se refiere a 0,95 veces la razón de isótopos ¹⁴C / ¹²C de HOxI (con referencia a AD 1950). Esto es aproximadamente equivalente a la madera previa a la revolución industrial corregida para la descomposición. Para la biosfera viva actual (material vegetal), f_M es de aproximadamente 1,1.

- 40 La invención proporciona un bioproducto que puede tener una f_M de ¹⁴C de al menos aproximadamente 1. Por ejemplo, el bioproducto puede tener una f_M de ¹⁴C de al menos aproximadamente 1,01, una f_M de ¹⁴C de aproximadamente 1 a aproximadamente 1,5, una f_M de ¹⁴C de aproximadamente 1,04 a aproximadamente 1,18, o una f_M de ¹⁴C de aproximadamente 1,111 a aproximadamente 1,124.

Otra medición de ¹⁴C se conoce como el porcentaje de carbono moderno, pMC. Para un arqueólogo o geólogo que usa dataciones por ¹⁴C, AD 1950 es igual a “cero años de edad”. Esto también representa 100 pMC. El “carbono bomba” en la atmósfera alcanzó casi el doble del nivel normal en 1963 en el punto máximo de las armas

termonucleares. Su distribución dentro de la atmósfera se ha aproximado desde su aparición, mostrando valores superiores a 100 pMC para plantas y animales vivos desde AD 1950. Ha disminuido gradualmente a lo largo del tiempo siendo el valor actual próximo a 107,5 pMC. Esto significa que un material de biomasa reciente, tal como maíz, dará un valor característico de ¹⁴C próximo a 107,5 pMC. Los compuestos basados en petróleo tendrán un valor de pMC de cero. Combinar carbono fósil con carbono actual dará como resultado una dilución del contenido de pMC actual. Suponiendo que 107,5 pMC representa el contenido en ¹⁴C de los materiales de biomasa actuales y 0 pMC representa el contenido en ¹⁴C de productos basados en petróleo, el valor de pMC medido para ese material reflejará las proporciones de los dos tipos de componentes. Por ejemplo, un material derivado al 100% de semillas de soja actuales dará un valor característico de radiocarbono próximo a 107,5 pMC. Si se diluye ese material al 50% con productos basados en petróleo, dará un valor característico de radiocarbono de aproximadamente 54 pMC.

Un contenido en carbono de base biológica se deriva asignando “el 100%” igual a 107,5 pMC y “el 0%” igual a 0 pMC. Por ejemplo, una muestra que mide 99 pMC dará un contenido en carbono de base biológica equivalente del 93%. Este valor se denomina resultado de carbono de base biológica medio y supone que todos los componentes dentro del material analizado se originan o bien de material biológico actual o bien de material basado en petróleo.

Un bioproducto descrito en el presente documento puede tener un pMC de al menos aproximadamente 50, 60, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99 ó 100. En otros casos, un bioproducto descrito en el presente documento puede tener un pMC de entre aproximadamente 50 y aproximadamente 100; aproximadamente 60 y aproximadamente 100; aproximadamente 70 y aproximadamente 100; aproximadamente 80 y aproximadamente 100; aproximadamente 85 y aproximadamente 100; aproximadamente 87 y aproximadamente 98; o aproximadamente 90 y aproximadamente 95. En aún otros casos, un bioproducto descrito en el presente documento puede tener un pMC de aproximadamente 90, 91, 92, 93, 94 ó 94,2.

La invención se ilustra adicionalmente mediante los siguientes ejemplos. Los ejemplos se proporcionan únicamente con fines ilustrativos. No deben interpretarse como limitativos del alcance o contenido de la invención de ninguna manera.

Ejemplos

EJEMPLO 1

Identificación de homólogos de ácido carboxílico reductasa (CAR)

La ácido carboxílico reductasa (CAR) de la cepa NRRL 5646 de *Nocardia sp.* puede reducir ácidos carboxílicos para dar los aldehídos correspondientes sin enzimas activantes separadas, tales como acil-CoA sintasas (Li *et al.*, J. Bacteriol. 179:3482-3487, 1997; He *et al.*, Appl. Environ. Microbiol. 70:1874-1881, 2004)). Una búsqueda en BLAST usando la secuencia de aminoácidos de CAR de NRRL 5646 (registro de Genpept AAR91681) (SEQ ID NO: 16) como secuencia consulta identificó aproximadamente 20 secuencias homólogas. Tres homólogos, indicados en la tabla 7, se evaluaron para determinar su capacidad para convertir ácidos grasos en aldehídos grasos *in vivo* cuando se expresaban en *E. coli*. A nivel de secuencia de nucleótidos, *carA* (SEQ ID NO: 19), *carB* (SEQ ID NO: 21) y *fadD9* (SEQ ID NO: 17) demostraron una homología del 62,6%, el 49,4% y el 60,5%, respectivamente, con respecto al gen *car* (AY495697) de *Nocardia sp.* NRRL 5646 (SEQ ID NO: 15). A nivel de aminoácidos, CARA (SEQ ID NO: 20), CARB (SEQ ID NO: 22) y FadD9 (SEQ ID NO: 18) demostraron una identidad del 62,4%, el 59,1% y el 60,7%, respectivamente, con respecto a CAR de *Nocardia sp.* NRRL 5646 (SEQ ID NO: 16).

Tabla 7: Proteína de tipo CAR y secuencias codificantes correspondientes.

Registro de Genpept	Etiqueta de locus	Anotación en GenBank	Nombre de gen
NP_217106	Rv 2590	Ácido graso-CoA ligasa probable (FadD9)	<i>fadD9</i>
ABK75684	MSMEG 2956	Proteína de la familia de epimerasa/deshidratasa dependiente de NAD	<i>carA</i>
YP_889972.1	MSMEG 5739	Proteína de la familia de epimerasa/deshidratasa dependiente de NAD	<i>carB</i>

EJEMPLO 2

Expresión de homólogos de CAR en *E. coli*

A. Construcción de plásmidos

Se construyeron tres plásmidos de expresión en *E. coli* para expresar los genes que codificaban para los homólogos de CAR indicados en la tabla 7. En primer lugar, se amplificó *fadD9* a partir de ADN genómico de *Mycobacterium tuberculosis* H37Rv (obtenido de The University of British Columbia, Vancouver, BC Canadá) usando los cebadores *fadD9F* y *FadDR* (véase la tabla 8). En primer lugar se clonó el producto de PCR en PCR-blunt (Invitrogen) y después se liberó como un fragmento de NdeI-AvrII. Entonces se clonó el fragmento de NdeI-AvrII entre los sitios NdeI y AvrII de pACYCDuet-1 (Novogen) para generar pACYCDuet-1-*fadD9*.

Se amplificaron los genes *carA* y *carB* a partir del ADN genómico de *Mycobacterium smegmatis* MC2 155 (obtenido de ATCC (ATCC 23037D-5)) usando los cebadores CARMCaF y CARMCaR o CARMCbF y CARMCbR, respectivamente (véase la tabla 8). En primer lugar se clonó cada producto de PCR en PCR-blunt y después se liberó como fragmento de NdeI-AvrII. Entonces se subclonó cada uno de los dos fragmentos entre los sitios NdeI y AvrII de pACYCDuet-1 (Novogen) para generar pACYCDUET-*carA* y pACYCDUET-*carB*.

Tabla 8. Cebadores usados para amplificar genes que codifican para homólogos de CAR

fadD9F	cat ATGTCGATCAACGATCAGCGACTGAC (SEQ ID NO: 1)
fadD9R	cctagg TCACAGCAGCCCGAGCAGTC (SEQ ID NO: 2)
CARMCaF	cat ATGACGATCGAAACGCG (SEQ ID NO: 3)
CARMCaR	cctagg TTACAGCAATCCGAGCATCT (SEQ ID NO: 4)
CARMCbF	cat ATGACCAGCGATGTTAC (SEQ ID NO: 5)
CARMCbR	cctagg TCAGATCAGACCGAACTCAG (SEQ ID NO: 6)

B. Evaluación de la producción de aldehídos grasos

Se cotransformaron por separado plásmidos que codificaban para los homólogos de CAR (pACYCDUET-*fadD9*, pACYCDUET-*carA* y pACYCDUET-*carB*) en la cepa C41 de *E. coli* (DE3, Δ *fadE*) (descrita en el documento PCT/US08/058788) junto con pETDuet-1-*tesA* (descrito en el documento PCT/US08/058788).

Se hicieron crecer los transformantes de *E. coli* en 3 ml de medio LB suplementado con carbenicilina (100 mg/l) y cloranfenicol (34 mg/l) a 37°C. Tras el crecimiento durante la noche, se transfirieron 15 μ l de cultivo a 2 ml de medio LB reciente suplementado con carbenicilina y cloranfenicol. Tras 3,5 horas de crecimiento, se transfirieron 2 ml de cultivo a un matraz de 125 ml que contenía 20 ml de medio M9 con glucosa al 2% y con carbenicilina y cloranfenicol. Cuando la DO₆₀₀ del cultivo alcanzó 0,9, se añadió 1 mM de IPTG a cada matraz. Tras 20 horas de crecimiento a 37°C, se añadieron 20 ml de acetato de etilo (con el 1% de ácido acético, v/v) a cada matraz para extraer los compuestos orgánicos producidos durante la fermentación. Se analizó directamente el extracto de acetato de etilo bruto con CG/EM tal como se describe a continuación.

La coexpresión de *tesA* sin líder y cualquiera de los tres genes *car* en *E. coli* dio como resultado una producción de aldehídos grasos detectable. En una fermentación, LS9001/pACYCDUET *carB*+ pETDuet-1-*tesA* produjo un promedio de 120 mg/l de aldehídos grasos. Los tiempos de retención fueron de 6,959 minutos para dodecanal, 8,247 minutos para 7-tetradecenal, 8,37 minutos para tetradecanal, 9,433 minutos para 9-hexadecenal, 9,545 minutos para hexadecanal y 10,945 minutos para 11-octadecenal. La presencia de grandes cantidades de aldehídos grasos concuerda con el hecho de que CAR es una ácido graso reductasa generadora de aldehídos (AFAR). Este mecanismo es diferente de las acil graso-CoA reductasas generadoras de alcoholes (FAR), por ejemplo, JjFAR, y acil graso-CoA reductasas, tales como Acr1.

C. Preferencias de sustrato de los homólogos de CAR

Se observaron claras preferencias de sustrato entre los tres homólogos de CAR evaluados. FadD9 mostró una fuerte preferencia por ácidos grasos C₁₂ con respecto a otros ácidos grasos con longitudes de cadena de carbono superiores a 12. Tanto *CarA* como *CarB* demostraron gamas de sustrato más amplias que FadD9.

D. Cuantificación e identificación de aldehídos grasos

Se realizó CG-EM usando un sistema Agilent 5975B MSD equipado con una columna DB-5 de 30 m x 0,25 mm (película de 0,10 μ m). La temperatura de la columna era isotérmica durante 3 min a 100°C. Se programó la columna para subir desde 100°C hasta 320°C a una velocidad de 20°C/min. Cuando se alcanzó la temperatura final, la columna permaneció isotérmica durante 5 minutos a 320°C. El volumen de inyección fue de 1 μ l. El gas portador, helio, se liberó a 1,3 ml/min. El espectrómetro de masas estaba equipado con una fuente de ionización por impacto de electrones. La temperatura de la fuente de ionización se fijó a 300°C.

Antes de la cuantificación, se identificaron diversos aldehídos usando dos métodos. En primer lugar, se comparó el tiempo de retención en CG de cada compuesto con el tiempo de retención de un patrón conocido, tal como laurilaldehído (dodecanal). En segundo lugar, se confirmó la identificación de cada compuesto haciendo coincidir el espectro de masas del compuesto con el espectro de masas de un patrón en la biblioteca de espectros de masas.

REIVINDICACIONES

1. Método de producción de un aldehído graso, comprendiendo el método:
 - (a) proporcionar una célula huésped, estando dicha célula huésped modificada por ingeniería para expresar
 - (i) un gen que codifica para un polipéptido que tiene al menos el 80% de identidad de secuencia con respecto a la secuencia de aminoácidos de SEQ ID NO: 22, en el que el polipéptido tiene actividad ácido carboxílico reductasa, y
 - (ii) un gen que codifica para un polipéptido que tiene actividad tioesterasa (EC 3.1.2.14 o EC 3.1.1.5);
 - (b) cultivar dicha célula huésped modificada por ingeniería en un medio de cultivo que contiene una fuente de carbono, en condiciones eficaces para producir un aldehído graso.
2. Método según la reivindicación 1, que comprende además modificar la expresión de un gen que codifica para un polipéptido que tiene actividad ácido graso sintasa en la célula huésped.
3. Método según la reivindicación 1 ó 2, en el que la célula huésped es una célula huésped modificada por ingeniería genética que expresa un nivel atenuado de una enzima de degradación de ácidos grasos con respecto a una célula huésped silvestre, en el que la enzima de degradación de ácidos grasos tiene actividad acil-CoA sintasa (EC 2.3.1.86).
4. Método según una cualquiera de las reivindicaciones 1 a 3, en el que la célula huésped se selecciona del grupo que consiste en una célula de mamífero, célula vegetal, célula de insecto, célula de hongo y célula bacteriana.
5. Método según una cualquiera de las reivindicaciones 1 a 4, en el que el aldehído graso es decanal, dodecanal, miristal o hexadecal.
6. Método según una cualquiera de las reivindicaciones 1 a 5, en el que dicha fuente de carbono es un monosacárido, preferiblemente glucosa.
7. Método según una cualquiera de las reivindicaciones 1 a 6, en el que el aldehído graso se detecta en el cultivo.
8. Método según una cualquiera de las reivindicaciones 1 a 7, en el que el polipéptido comprende la secuencia de aminoácidos de SEQ ID NO: 22.
9. Método según una cualquiera de las reivindicaciones 1 a 8, en el que la célula huésped es una célula de levadura o una célula de hongo filamentosos.

FIG. 1

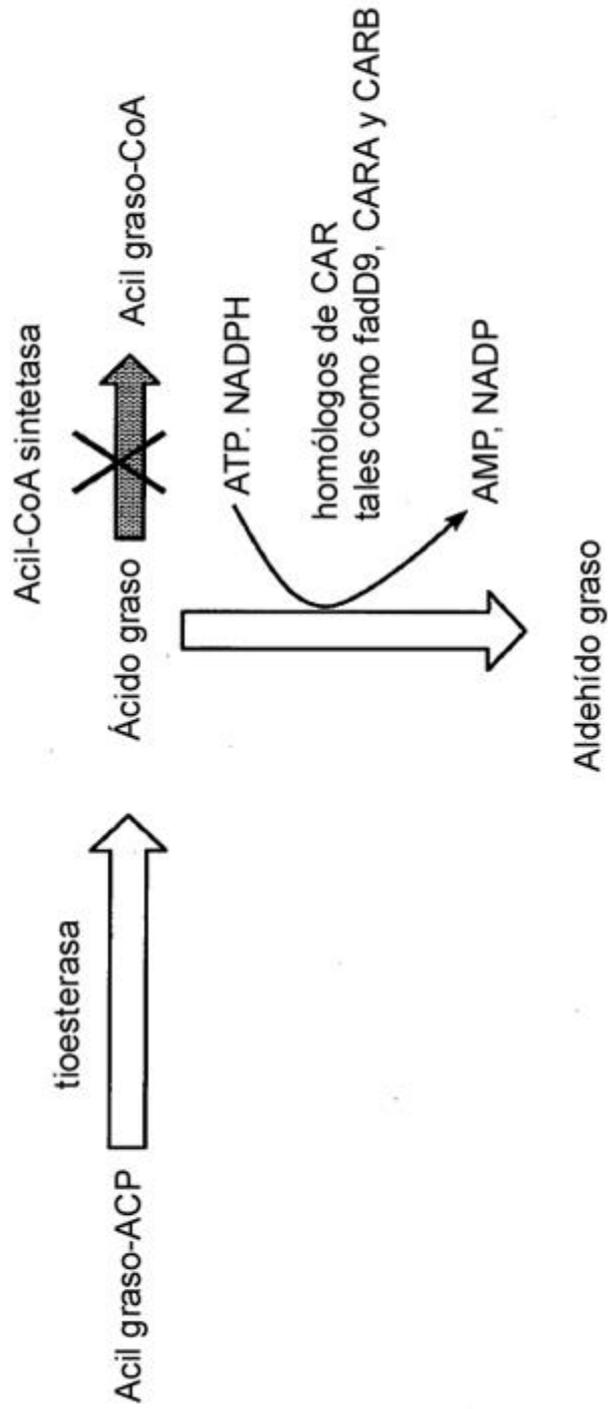


FIG 2

AAR91681.1**Secuencia de nucleótidos (SEQ ID NO:15)**

>gi | 40796034:488-4012 Nocardia sp. NRRL 5646 gen ácido carboxílico reductasa dependiente de ATP/NADPH (car), cds completa

ATGGCAGTGGATTACCCGGATGAGCGGCTACAGCGCCGCATTGCACAGT
 TGTTTGCAGAAGATGAGCAGGTCAAGGCCGCACGTCCGCTCGAAGCGGT
 GAGCGCGGCGGTGAGCGCGCCCGGTATGCGGCTGGCGCAGATCGCCGCC
 ACTGTTATGGCGGGTTACGCCGACCGCCCGGCCGCCGGGCAGCGTGCCT
 TCGAACTGAACACCGACGACGCGACGGGCCGCACCTCGCTGCGGTTACT
 TCCCCGATTGAGACCATCACCTATCGCGAACTGTGGCAGCGAGTCGGC
 GAGGTTGCCGCGGCCTGGCATCATGATCCCGAGAACCCTTGCAGCGCAG
 GTGATTTGTCGCGCCCTGCTCGGCTTACCAGCATCGACTACGCCACCCT
 CGACCTGGCCGATATCCACCTCGGCGCGGTTACCGTGCCGTTGCAGGCC
 AGCGCGGCGGTGTCCAGCTGATCGCTATCCTCACCGAGACTTCGCCCGC
 GGCTGCTCGCCTCGACCCCGGAGCACCTCGATGCGGCGGTTCGAGTGCCT
 ACTCGCGGGCACACACCGGAACGACTGGTGGTCTTCGACTACCACCCC
 GAGGACGACGACCAGCGTGCGGCCTTCGAATCCGCCCGCCGCCGCTTG
 CCGACGCGGGCAGCTTGGTGATCGTCGAAACGCTCGATGCCGTGCGTGC
 CCGGGGCCGCGACTTACCGGCCGCGCCACTGTTCGTTCCCGACACCGAC
 GACGACCCGCTGGCCCTGCTGATCTACACCTCCGGCAGCACCGGAACGC
 CGAAGGGCGCGATGTACACCAATCGGTTGGCCGCCACGATGTGGCAGGG
 GAACTCGATGCTGCAGGGGAACTCGCAACGGGTCGGGATCAATCTCAAC
 TACATGCCGATGAGCCACATCGCCGGTTCGCATATCGCTGTTTCGGCGTGC
 TCGCTCGCGGTGGCACCGCATACTTCGCGGCCAAGAGCGACATGTGCAC
 ACTGTTTCGAAGACATCGGCTTGGTACGTCCCACCGAGATCTTCTTCGTC
 CCGCGCGTGTGCGACATGGTCTTCCAGCGCTATCAGAGCGAGCTGGACC
 GGCGCTCGGTGGCGGGCGCCGACCTGGACACGCTCGATCGGGAAGTGAA
 AGCCGACCTCCGGCAGAACTACCTCGGTGGGCGCTTCCCTGGTGGCGGTC
 GTCGGCAGCGCGCCGCTGGCCGCGGAGATGAAGACGTTTCATGGAGTCCG
 TCCTCGATCTGCCACTGCACGACGGGTACGGGTCGACCGAGGCGGGCGC
 AAGCGTGCTGCTCGACAACCAGATCCAGCGGCCGCCGGTGCTCGATTAC
 AAGCTCGTCGACGTGCCCGAACTGGGTTACTTCCGCACCGACCGGCCGC
 ATCCGCGCGGTGAGCTGTTGTTGAAGGCGGAGACCAGATTCGGGGCTA
 CTACAAGCGGCCCGAGGTCACCGCGGAGATCTTCGACGAGGACGGCTTC
 TACAAGACCGGCGATATCGTGGCCGAGCTCGAGCACGATCGGCTGGTCT
 ATGTCGACCGTTCGCAACAATGTGCTCAAACCTGTCGACAGGGCGAGTTCGT

FIG. 2 (Cont.)

GACCGTCGCCCATCTCGAGGCCGTGTTCCGCCAGCAGCCCGCTGATCCGG
 CAGATCTTCATCTACGGCAGCAGCGAACGTTCCCTATCTGCTCGCGGTGA
 TCGTCCCCACCGACGACGCGCTGCGCGGCCGCGACACCGCCACCTTGAA
 ATCGGCACTGGCCGAATCGATTTCAGCGCATCGCCAAGGACGCGAACCTG
 CAGCCCTACGAGATTCCGCGCGATTTCCCTGATCGAGACCGAGCCGTTCA
 CCATCGCCAACGGACTGCTCTCCGGCATCGCGAAGCTGCTGCGCCCCAA
 TCTGAAGGAACGCTACGGCGCTCAGCTGGAGCAGATGTACACCGATCTC
 GCGACAGGCCAGGCCGATGAGCTGCTCGCCCTGCGCCGCGAAGCCGCCG
 ACCTGCCGGTGCTCGAAACCGTCAGCCGGGCAGCGAAAGCGATGCTCGG
 CGTCGCCCTCCGCCGATATGCGTCCCGACGCGCACTTCACCGACCTGGGC
 GCGGATTCCCTTTCCGCGCTGTCGTTCTCGAACCTGCTGCACGAGATCT
 TCGGGGTTCGAGGTGCCGGTGGGTGTCGTCGTCAGCCCGGCGAACGAGCT
 GCGCGATCTGGCGAATTACATTGAGGCGGAACGCAACTCGGGCGCGAAG
 CGTCCCACCTTCACCTCGGTGCACGGCGGCGGTTCCGAGATCCGCGCCG
 CCGATCTGACCCTCGACAAGTTCATCGATGCCCGCACCTGGCCGCCGC
 CGACAGCATTCCGCACGCGCCGGTGCCAGCGCAGACGGTGCTGCTGACC
 GCGCGAACGGCTACCTCGGCCGGTTCCTGTGCCTGGAATGGCTGGAGC
 GGCTGGACAAGACGGGTGGCACGCTGATCTGCGTCGTGCGCGGTAGTGA
 CGCGGCCCGCGGCCCGTAAACGGCTGGACTCGGCGTTCGACAGCGGCGAT
 CCCGGCCTGCTCGAGCACTACCAGCAACTGGCCGCACGGACCCTGGAAG
 TCCTCGCCGGTGATATCGGCGACCCGAATCTCGGTCTGGACGACGCGAC
 TTGGCAGCGGTTGGCCGAAACCGTCGACCTGATCGTCCATCCCGCCGCG
 TTGGTCAACCACGTCCTTCCCTACACCAGCTGTTTCGGCCCCAATGTCG
 TCGGCACCGCCGAAATCGTCCGGTTGGCGATCACGGCGCGGCGCAAGCC
 GGTACCTACCTGTCGACCGTCCGAGTGGCCGACCAGGTTCGACCCGGCG
 GAGTATCAGGAGGACAGCGACGTCCGCGAGATGAGCGCGGTGCGCGTCG
 TCGCGGAGAGTTACGCCAACGGCTACGGCAACAGCAAGTGGGCGGGGGA
 GGTCTGCTGCGCGAAGCACACGATCTGTGTGGCTTGCCGGTTCGCGGTG
 TTCCGTTTCGGACATGATCCTGGCGCACAGCCGGTACGCGGGTTCAGCTCA
 ACGTCCAGGACGTGTTACCCGGCTGATCCTCAGCCTGGTCGCCACCGG
 CATCGCGCCGTACTCGTTCTACCGAACCGACGCGGACGGCAACCGGCAG
 CGGGCCCACTATGACGGCTTGCCGGCGGACTTCACGGCGGCGGCGATCA
 CCGCGCTCGGCATCCAAGCCACCGAAGGCTTCCGGACCTACGACGTGCT
 CAATCCGTACGACGATGGCATCTCCCTCGATGAATTCGTTCGACTGGCTC
 GTCGAATCCGGCCACCCGATCCAGCGCATCACCGACTACAGCGACTGGT
 TCCACCGTTTCGAGACGGCGATCCGCGCGCTGCCGGAAAAGCAACGCCA
 GGCCTCGGTGCTGCCGTTGCTGGACGCCTACCGCAACCCCTGCCCGGCG
 GTCCGCGGCGCGATACTCCCGGCCAAGGAGTTCGAAGCGGCGGTGCAAA
 CAGCCAAAATCGGTCCGGAACAGGACATCCCGCATTTGTCCGCGCCACT
 GATCGATAAGTACGTCAGCGATCTGGAAGTCTTCAGCTGCTCTGA

FIG 2 (Cont.)

Secuencia de nucleótidos (SEQ ID NO:16)

>gi | 40796035 | gb | AAR91681.1 | ácido carboxílico reductasa dependiente de ATP/NADPH [Nocardia sp. NRRL 5646]

MAVDSPPERLQRRRIAQLFAEDEQVKAARPLEAVSAAVSAPGMRLAQIAA
 TVMAGYADRPAAGQRAFELNTDDATGR TSLRLLPRFETITYRELWQRVG
 EVAAAWHHPENPLRAGDFVALLGFTSIDYATLDLADIHLGAVTVPLQA
 SAAVSQLIAILTETSPRLLASTPEHLDAAVECLLAGTTPERLVVFDYHP
 EDDQRAAFESARRRLADAGSLVIVETLDAVRARGRDLPAAPLFVPD TD
 DDPLALLIY TSGSTGTPKGAMYTNRLAATMWQGN SMLQGN SQRVGINLN
 YMPMSHIAGRISLFGVLARGGTAYFAAKSDMSTLFEDIGLVRPTEIFFV
 PRVCDMV FQRYQSELDRRSVAGADLDTLDREVKADLRQNYLGGRFLVAV
 VGSAPLAAEMKTFMESVLDLPLHDGYGSTEAGASVLLDNQIQRPVLDY
 KLVDVPELGYFRTDRPHPRGELLKKAETTIPGYKRPEVTAEIFDEDGF
 YKTGDIVAELEHDLRVYVDRRNNVLKLSQGEFVTVAHLEAVFASSPLIR
 QIFIYGS SERSYLLAVIVPTDDALRGRDTATLKSALAESIQRIAKDANL
 QPYEIPRDFLIETEPFTIANGLLSGIAKLLRPNLKERYGAQLEQMYTDL
 ATGQADELLALRREAADLPVLETVSRAAKAMLGVASADM RPDAHFTDLG
 GDSLSALSFSNLLHEIFGVEVPVGVVVS PANELRDLANYIEAERN SGAK
 RPTFTSVHGGGSEIRAADLTLDKFIDARTLAAADSI PHAPVPAQTVLLT
 GANGYLGRFLCLEWLERLDKTGGTLICVVRGSDAAAARKRLDSAFDSGD
 PGLLEHYQQLAARTLEVLAGDIGDPNLGLDDATWQRLAETVDLIVHPAA
 LVNHVLPYTQLFGPNVVGTA EIVRLAITARRKPV TYLSTVGVADQVDPA
 EYQEDSDVREMSAVRVVRESYANGYGNSKWAGEVLLREAHDL CGLPVAV
 FRSDMILAH SRYAGQLNVQDV FTRLILSLVATGIAPYSFYRTDADGNRQ
 RAHYDGLPADFTAAAITALGIQATEGFRTYDVLN PYDDGISLDEFVDWL
 VESGHPIQRITDYSDFHRFETAIRALPEKQRQASVLP LLDAYRNPCPA
 VRGAILPAKEFQAAVQTAKIGPEQDIPHL SAPLIDKYVSDLELLQLL

FIG. 3

Motivo 1

-G-Y-X-X-S/A/T-K-W/L (SEQ ID NO:7); y

-G-X-X-G-X-L-G (SEQ ID NO:8); y

-L/V/I-G-G-D-S-X-X-A (SEQ ID NO:9); y

-[LIVMFY]-{E}-{VES}-[STG]-[STAG]-G-[ST]-[STEIA]-[SG]-X-[PASLIVM]-[KR] (SEQ ID NO:10), en el que {X} representa cualquier aminoácido excepto X y [X₁X₂] representa X₁ o X₂

Motivo 2

RTVLLX₁GAX₂GX₃LGRX₄LX₅LX₆WL (SEQ ID NO:11)

en el que

X₁ es S o T;

X₂ es T o N;

X₃ es F o W;

X₄ es F o Y;

X₅ es A o T; y

X₆ es E o Q

Motivo 3

LXXGXXGXLGXXLXLXWLXR (SEQ ID NO:12)

Motivo 4

WAXEVLLR (SEQ ID NO:13), en el que X puede ser cualquier aminoácido; o

LXXGXXGXLGXXLXX₁XX₂LX₃R (SEQ ID NO:14), en el que

X₁ es Leu o Ile;

X₂ es Trp o Leu; y

X₃ varía entre 13 aminoácidos o 14 aminoácidos

FIG. 3 (Cont.)

Motivo 5

-G-Y-X-X-S/A/T-K-W/L (SEQ ID NO:7); y

-L/V/I-G-G-D-S-X-X-A (SEQ ID NO:9); y

-[LIVMFY]- {E} - {VES} - [STG] - [STAG] - G - [ST] - [STEIA] - [SG] - X - [PASLIVM] - [KR] (SEQ ID NO:10), en el que {X} representa cualquier aminoácido excepto X y [X₁X₂] representa X₁ o X₂; y

RTVLLX₁GAX₂GX₃LGRX₄LX₅LX₆WL (SEQ ID NO:11), en el que

- X₁ es S o T;
- X₂ es T o N;
- X₃ es F o W;
- X₄ es F o Y;
- X₅ es A o T; y
- X₆ es E o Q

FIG. 4

NP 217106 (FADD9)**Secuencia de nucleótidos (SEQ ID N0:17)**

>gi | 57116681:2917871-2921377 Mycobacterium tuberculosis H37Rv, genoma completo

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTCGAGGACCTATACGCCAG
 CGACGCCAGTTTCGCCGCCGCCAGTCCCAACGAGGCGATCACCCAGGCGATCG
 ACCAGCCCGGGGTCGCGCTTCCACAGCTCATCCGTATGGTCATGGAGGGCTAC
 GCCGATCGGGCCGGCACTCGGCCAGCGTGCGCTCCGCTTCGTCACCGACCCCGA
 CAGCGGCCGCACCATGGTTCGAGCTACTGCCGCGGTTTCGAGACCATCACCTACC
 GCGAACTGTGGGCCCCGCGCCGGCACATTGGCCACCGCGTTGAGCGCTGAGCCC
 GCGATCCGGCCGGGCGACCGGGTTTTCGTGCTGGGCTTCAACAGCGTCGACTA
 CACAACCATCGACATCGCGCTGATCCGGTTGGGCGCCGTGTCGGTTCCACTGC
 AGACCAGTGCGCCGGTACCGGGTTGCGCCCGATCGTCACCGAGACCGAGCCG
 ACGATGATCGCCACCAGCATCGACAATCTTGGCGACGCCGTCGAAGTGCTGGC
 CGGTCACGCCCCGGCCGGCTGGTTCGTATTTCGATTACCACGGCAAGGTTGACA
 CCCACCGCGAGGCCGTCGAAGCCGCCCGAGCTCGGTTGGCCGGCTCGGTGACC
 ATCGACACACTTGCCGAACTGATCGAACGCGGCAGGGCGCTGCCGGCCACACC
 CATTGCCGACAGCGCCGACGACGCGCTGGCGCTGCTGATTTACACCTCGGGTA
 GTACCGGCGCACCCAAAGGCGCCATGTATCGCGAGAGCCAGGTGATGAGCTTC
 TGGCGCAAGTCGAGTGGCTGGTTCGAGCCGAGCGGTTACCCCTCGATCACGCT
 GAACTTCATGCCGATGAGCCACGTCGGGGGCCGTCAGGTGCTCTACGGGACGC
 TTTCCAACGGCGGTACCGCCTACTTCGTCGCCAAGAGCGACCTGTCGACGCTG
 TTCGAGGACCTCGCCCTGGTGC GGCCACAGAATTGTGCTTCGTGCCGCGCAT
 CTGGGACATGGTGTTCGCAGAGTTCACAGCGAGGTCGACCGCCGCTTGGTGG
 ACGGCGCCGATCGAGCGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAG
 AACGTGCTCGGCGGACGGTTTGTTCATGGCGCTGACCGGTTCCGCGCCGATCTC
 CGCTGAGATGACGGCGTGGGTTCGAGTCCCTGCTGGCCGACGTGCATTTGGTGG
 AGGGTTACGGCTCCACCGAGGCCGGGATGGTCTGAACGACGGCATGGTGGG
 CGCCCCGCGGTGATCGACTACAAGCTGGTTCGACGTGCCCGAGCTGGGCTACTT
 CGGCACCGATCAGCCCTACCCCGGGGCGAGCTGCTGGTCAAGACGCAAACCA
 TGTTCCCCGGCTACTACCAGCGCCCGGATGTCACCGCCGAGGTGTTTCGACCCC
 GACGGCTTCTACCGGACCGGGGACATCATGGCCAAAGTAGGCCCCGACCAGTT
 CGTCTACCTCGACCGCCGCAACAACGTGCTAAAGCTCTCCAGGGCGAGTTCA
 TCGCCGTGTCGAAGCTCGAGGCGGTGTTTCGGCGACAGCCCGCTGGTCCGACAG
 ATCTTCATCTACGGCAACAGTGCCCGGGCCTACCCGCTGGCGGTGGTTGTCCC
 GTCCGGGGACGCGCTTTCTCGCC

FIG. 4 (Cont.)

ATGGCATCGAGAATCTCAAGCCCGTGATCAGCGAGTCCCTGCAGGAGGTAGCG
 AGGGCGGCCGGCCTGCAATCCTACGAGATTCCACGCGACTTCATCATCGAAAC
 CACGCCGTTACCCCTGGAGAACGGCCTGCTCACCGGCATCCGCAAGCTGGCAC
 GCCCGCAGTTGAAGAAGTTCTATGGCGAACGTCTCGAGCGGCTCTATACCGAG
 CTGGCCGATAGCCAATCCAACGAGCTGCGCGAGCTGCGGCAAAGCGGTCCCGA
 TCGGCCGGTGCTTCCGACGCTGTGCCGTGCCGCGGCTGCGTTGCTGGGCTCTA
 CCGCTGCGGATGTGCGGCCGGACGCGCACTTCGCCGACCTGGGTGGTGACTCG
 CTCTCGGCGCTGTCGTTGGCCAACCTGCTGCACGAGATCTTCGGCGTCGACGT
 GCCGGTGGGTGTCATTGTCAGCCCGGCAAGCGACCTGCGGGCCCTGGCCGACC
 ACATCGAAGCAGCGCGCACCCGGCGTCAGGCGACCCAGCTTCGCCTCGATAAC
 GGTGCTCCGCGACGGAAGTGCACGCCAGCGACCTCACGCTGGACAAGTTCAT
 CGACGCTGCCACCCTGGCCGCAGCCCCGAACCTGCCGGCACCGAGCGCCCAAG
 TCGCACCGTACTGCTGACCGGCGCCACCGGCTTTTTGGGTGCTACCTGGCG
 CTGGAATGGCTCGACCGCATGGACCTGGTCAACGGCAAGCTGATCTGCCTGGT
 CCGCGCCAGATCCGACGAGGAAGCACAAAGCCCGGCTGGACGCGACGTTTCGATA
 GCGGCGACCCGTATTTGGTGCGGCACTACCGCGAATTGGGCGCCGGCCGCCTC
 GAGGTGCTCGCCGGCGACAAGGGCGAGGCCGACCTGGGCCTGGACCGGGTCAC
 CTGGCAGCGGCTAGCCGACACGGTGGACCTGATCGTGGACCCCGCGGCCCTGG
 TCAACCACGTGCTGCCGTATAGCCAGCTGTTTCGGCCCAAACGCGGCGGGCACC
 GCCGAGTTGCTTCGGCTGGCGCTGACCGGCAAGCGCAAGCCATACATCTACAC
 CTCGACGATCGCCGTGGGCGAGCAGATCCCGCCGGAGGCGTTCACCGAGGACG
 CCGACATCCGGGCCATCAGCCCGACCCGCAGGATCGACGACAGCTACGCCAAC
 GGCTACGCGAACAGCAAGTGGGCCGGCGAGGTGCTGCTGCGCGAAGCTCACGA
 GCAGTGCGGCCTGCCGGTGACGGTCTTCCGCTGCGACATGATCCTGGCCGACA
 CCAGCTATAACGGTCAGCTCAACCTGCCGGACATGTTACCCGGCTGATGCTG
 AGCCTGGCCGCTACCGGCATCGCACCCGGTTCGTTCTATGAGCTGGATGCGCA
 CGGCAATCGGCAACGCGCCCACATGACGGCTTGCCGGTCGAATTCGTGCGAG
 AAGCCATTTGCACCCTTGGGACACATAGCCCGGACCGTTTTGTACCTACCAC
 GTGATGAACCCCTACGACGACGGCATCGGGCTGGACGAGTTCGTGACTGGCT
 CAACTCCCCAACTAGCGGGTCCGGTTGCACGATCCAGCGGATCGCCGACTACG
 GCGAGTGGCTGCAGCGGTTGAGACTTCGCTGCGTGCTTGCCGGATCGCCAG
 CGCCACGCCTCGCTGCTGCCCTTGCTGCACAACCTACCGAGAGCCTGCAAAGCC
 GATATGCGGGTCAATCGCGCCCACCGACCAAGTTCGCGCTGCCGTCCAAGAAG
 CGAAAATCGGTCCGGACAAAGACATTCCGCACCTCACGGCGGGCGATCATCGCG
 AAGTACATCAGCAACCTGCGACTGCTCGGGCTGCTGTGA

FIG. 4 (Cont.)

Secuencia de aminoácidos (SEQ ID N0:18)

>gi | 15609727 | ref | NP_217106.1 | ácido graso-CoA ligasa
[Mycobacterium tuberculosis H37Rv]

MSINDQRLTRRVEDLYASDAQFAAASPNEAITQ AIDQPGVALPQLIRMVMEGY
ADRPALGQRALRFVTD PDSGRTMVELLPRFETIT YRELWARAGTLATALSAEP
AIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGLRPIVTETEP
TMIATSIDNLGDAVEVLAGHAPARLVVFDYHGKVDTHREAVEAARARLAGSVT
IDTLAELIERGRALPATPIADSADDALALLIYTS GSTGAPKGYRESQVMSF
WRKSSGWFEPSGYP SITLNFMPMSHVGGRRQVLYGTLSNGGTAYFVAKSDLSTL
FEDLALVRPTELCFVPRIWDMVFAEFHSEVDRRLVDGADRAALEAQVKAELRE
NVLGGRFVMALTGSAPI SAEMTAWVESLLADVHLVEGYGSTEAGMVLNDGMVR
RPAVIDYKLVDPPELGYFGTDQPYPRGELLVKTQTMFPGYYQRPDVTAEVDFP
DGFYRTGDIMAKVGPDQFVYLD RRRNNVLKLSQGEFIAVSKLEAVFGDSPLVRQ
IFIYGN SARAYPLAVVVP SGDALSRHG IENLKPVI SE SLQEVARAAGLQSYEI
PRDFI IETTPFTLENGLLTGIRKLARPQLKKFYGERLERLYTELADSQSNELR
ELRQSGPDAPVLP TLCRAAAALLGSTAADVRPDAHFADLGGDSL SALS LANLL
HEIFGVDVPVGVIVSPASDLRALADHIEAARTG VRRPSFASIHGRSATEVHAS
DLTLDKFIDAATLAAAPNLPAPSAQVVRTVLLTGATGFLGRYLAL EWLDRMDLV
NGKLICLVRRARSDEEAQARLDATFDSGDPYLVRHYRELGAGRLEVL AGDKGEA
DLGLDRVTWQRLADTVDLIVDPAALVNHVLPYSQ LFGPNAAGTAE LLRLALTG
KRKPYIYTSTI AVGEQIPPEAFTE DADIRAI SPTRRIDDSYANGYANS KWAGE
VLLREAHEQCGLPVTVFRCDMILADTSYTGQLNLPDMFTRLMLSLAATGIAPG
SFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPDRFV TYHVMNPYDDGIG
LDEFVDWLN SPTS GSGCTIQRIADYGEWLQRFETS LRALPDRQRHASLLPLLH
NYREPAKPICGSIAPTDQFRAAVQEAKIGPDKDIPHLTAAI IAKYISNLRLLG
LL

ABK75684 (CARA)

Secuencia de nucleótidos (SEQ ID N0:19)

>gi | 118168627:3015785-3019291 Mycobacterium smegmatis str. MC2 155, genoma completo

TTACAGCAATCCGAGCATCTGCAGGTTGCTGATGTACTTGACGATCACGTCGG
CCGTGACGTGCGGAATGTCCTTGTCGGGGCCGATCTTCGCGTCCTGCACCGCG
GCACGGAACCGGTCGGTGGGTGCCATGGCACCGCACACGGGGCGGTGAGGGCTG
CTGATAGTTGTGCAGCAGCGGCAGCAGCGAGGCTGACGTTGCCGTTCCGGCA

FIG. 4 (Cont.)

GGGCCCGCAGTGCGGTTTCGAACCGGCTCAGCCAGGTGGCGTAGTCGTCGACG
 CGGTGCACGGGGTAGCCGGCCTCGATCAGCCAGTCCACGTA CT CGTCGAGGCC
 GATGCCGTCGTCGTACGGGTTTCATCACGTGGAACGTCTCGAATCCGTCGGTGA
 CCTGCGAGCCGATGGTGGAGATCGCCTCGGGCGATGAACTCCACGGGCAGCCCG
 TCGTAGTGGGCGCGCTGCCGGTTGCCGTCCGCATCGAGTTCGTAGAACGAACC
 GGGCGCGATGCCGGTCCGACGAGGCTCAGCATCAGGCGGGTGAACATGTCCG
 GCAGGTTACAGTGACCCGAGTAGGTTCGTGTCGGCCAGGATCATGTGCGCAGCGG
 AACACCGAGACCGGCAGACCACACCAGTTCGTGCGCCTCCCGCAGCAGGACCTC
 GCCGGCCACTTGCTGTTGCCGTAGCCGTTGGCGTACGAGTCGTGACCCGGC
 GCGTCGCGCTGATCTCGCGGATGTCGGCGTCCCTCGACGAACGCCTCGGGGGAG
 ATGCCCTGTCCACACCGATCGTCGAGACGTACACGTACGGCTTGATCGTGGT
 GGTCAGCGCGATCCGGATGAGTTCGGCGGTGCCGAGCGCATTGGGTCCGAACA
 TCTGGCTGTACGGCAGGACGTGATTGACCAGGGCGGCCGGATCGACGATCAGA
 TCGACGGTGTCCGGCAGTTCGCTGCCACGTGTCGTGGTTCGAGACCCAGATCGGC
 CTCGCCCTTGTCACCGGCGATCACCTCGAGGTGATCGGCTGCCAGCGCGCGGT
 AGTGCTCGAGCAGTGTGCGCTCCCCGGTGTGCAACGTGGCGTCCAGACGCGCC
 CGGGCCTCGTCGTCGCTGCGGGCGCGCACCAGGCAGATCACCTTGCCGTCCAC
 CAGGTCCATGCGCTCCAGCCATTCCAGCGCCAGATAGCGGCCAGGAACCCGG
 TGGCGCCGGTTCAGCAGCACGGTTCGGGATCTCGGTGCCGAACGCGGCAGACCC
 GCGCGGGCGGACAGGGTCTTGGCGTCGATGAACTTGCCCAGGGCGAGATCACG
 CGCGCGCACCTCGGTGGCGTTCGCGCCCGTGCACCGACGCGTATGTGGGGCGCT
 TGGAGCCGCGCAGTTTCGCCCTCGATGTAGGCCGCGACGCCTGCCAGGTTCGGTG
 GCCGGGCTGACGATGACGCCGACCGGCACGTTCGACATCGAAGATCTCGTGCAA
 CAGGTTCGAGAAGCTCAAGGCCGACAACGAATCTCCACCCAGATCGGTGAAGT
 GCGCATCGGACCGCAGATCCGTGACGGAGGCACCGAGCAGTGCACCGCGGGCG
 CGGCTGACGGTCTCGACCACGGGCCGGTTCGGCTCCGTTGCGGGCGCAACTCGCG
 CAACTCGTTGGCCTGCCCTCGGCCAGGTTCGGTGTAGAGCTGTTTCGAGGCGTT
 CGCCGTAGTGCGCCTTCAGTTTCGGCCGGGCCAGCTTGCCGATACCGGTTCAGC
 AGGCCGTTCTCCAGCGTGAAAGGTGTTGTCTCGACGAGGAAGTCACGCGGGAT
 CTCATACGACTGCAATCCGGCGGGCTCGTGCCGCGTCCCTGCAGTGAGTCGCTGA
 TGCGCGACTTGAGTTCGTACCGTCCCAACGTGACAGTGCCTCTTCGGTTCGGG
 ACCACGACCGCCAGCAGATAGGACCGCGCGCTGTTGCCGTAGACGTAGATCTG
 GCGTACCAGGGGGCTGTCGCCGAACACCGCCTCCAGCTTGAGACCGGTGACGA
 ATTCGCCCTGCGACAGTTTCAGCACGTTGTTGCGGGCGGTTCGAGGTATTCGAGA
 TGGTTCGGGCCCCGAGCTCGGGCAGCATGTCGCCGGTTCGGGTAGTACCCGTCCTC
 GTCGAACATCTCGGCGGTGATCTCCGGACGCTTGTTAGTAGCCGGGGAAACATCT
 GCTCGGACTTGACCAGAAGTTCGCCGCGCGGGTAGGGCCGGTCCGTGGCGAAG
 TAGCCGAGATCGGGCACGTTCGACCAGCTTGTTAGTCGATGACCGGGCGGGCGCTG
 GATCTGCCCGTTCGATGAAACACCGCGCCGGCCTCGGTGGAGCCGTAGCCCTCCA
 GCAGATGCATGTGAGCAGGTTCCTCGACCCAGCTCTTCATCTCCGCCGAGATG
 GGAGCCGATCCGGTTCAGGGCCGAAACGAATCGCCCAGCGAGCAGTTGGGTGCG

FIG. 4 (Cont.)

GACCTCTTCGAGGACTGCGGCTTCGGCTCGGTCTCGGATCCCTCGGCGCGGC
 GGTGTGTCGAGGCGGCTCTGGTACTCCTGGAACAGCATGTCCCAGATGCGAGGA
 ACGAAGTTGAGCTGCGTGGGCCGCACGAGGGCGAGGTCCTCCAGGAAGGTGGA
 CAGGTCGCTGCGTGCGGCGAAGTACGCGGTTCCGCCGCTGGCGAGTGTGCTGC
 ACAGGATGCCGCGCCCCATGACGTGACTCATGGGCATGAAGTTCAGGGTGATC
 GACGGCATCACGCCGAGGGTCTCGTCCCACCGGGCCTTGGACCCGGCCTGCCA
 CATCGTGGCGGTCTTGGACTCGGGGTACATCGCGCCCTTGGGAGTGCCGGTGC
 TGCCGGAGGTGTAGATGAGAAGGGTCAGCGGGTCGGCCTCGTTCGGGCACGTAG
 AGCGGTGCGTTCGGCGAGTGACCGCCCGCGGTCCAGTTCGTCGGTGATCGTCTC
 GACGACGACGCCGGTGCCTGCGAGCTTGCCCTTGGCCGCCTCGAACGCCTCAC
 GCTGATCGTCGACCTCGTGGCTGTAGTCGAACACCACCAGTCGCGACGGCGCG
 GGCCCGGACTCGACGAGAGCGACTGCGTTCGGCGAGGAAGTCGACGCTCGACGC
 GATCACCTTGGGCTCGGTCTCGGCGACGATCGGCTGCAGTTGGGCCACCGGCG
 CACTGGTCTGCAGCGGTACGGACACGGCGCCGAGTTCGAGCAGGGCGATGTCG
 ATCGTTCGTGTAGTCGACACTGGTGAAACCCAGGATGGCCACGCGGTCACCGGC
 ATTCACCGGATGGTTGTGCCAGGCATTGGTCACGGCCTGGATCCGGCCTGCGA
 GCTGACGGTAGGTGATGGTGTGGAAGCGGGGCAGGAGCTTCGCGGTGGTGC GG
 CCTTCTTCGTTCGGTGACGAACTCGACGGCGCGCTTGCCCAGCGCAGGGCGGTC
 CGCATAGCCGGCCAGAATCTGTTTGACCGCGGCAGGAAGGCGCAACTCCGGAT
 CGGCGGCAGCCGCGCTGATCGCCTCGTTCGGGACGGGCGGCGGCGAACTGCGGG
 TCGGTTTCGAACAAGTGGTCAATGCGCCGGTTGAAGCGGTCTTCGCGCGTTTC
 GATCGTCAT

Secuencia de aminoácidos (SEQ ID N0:20)

>gi|11817478|gb|ABK75684.1| Proteína de la familia de epimerasa/deshidratasa dependiente de NAD
 [Mycobacterium smegmatis str. MC2 155]

MTIETREDRFNRRIDHLFETDPQFAAARPDEAISAAAADPELRLPAAVKQILA
 GYADRPALGKRAVEFVTDEEGRTTAKLLPRFDITITYRQLAGRIQAVTNAWHNH
 PVNAGDRVAILGFTSVDYTTIDIALLELGAVSVPLQTSAPVAQLQPIVAETEP
 KVIASSVDFLADAVALVESGPAPSRLLVVFVDSHEVDDQREAFEAAGKLAGTG
 VVETITDALDRGRSLADAPLYVPDEADPLTLIIYTSGSTGTPKGAMYPESKT
 ATMWQAGSKARWDETLGVMPISITLNFMPMSHVMGRGILCSTLASGGTAYFAAR
 SDLSTFLEDLALVRPTQLNFVVPRIWDMLFQEYQSRLDNRRRAEGSEDRAEA AVL
 EEVRTQLLGGRFVSALTGSAPISAEMKSWVEDLLDMHLLLEGYGSTEAGAVFID
 GQIQRPVIDYKLVDPDLGYFATDRPYPRGELLVKSEQMFPGYYKRPEITAE
 MFDEDGYRTGDIVAELGPDHLEYLDRRNNVLKLSQGEFVTVSKLEAVFGDSP
 LVRQIYVYGNARSYLLAVVVPTEEALSRWDGDELKSRI SDSLQDAARAAGLQ
 SYEIPRDFLVETTPFTLENGLLTGIRKLARPKLKAHYGERLEQLYTDLAEGQA

FIG 4 (Cont.)

NELRELRRNGADRPVVE TVSRAAVALLGASVTDLRSDAHFTDLGGDSL SALS F
 SNLLHEIFDVDVPVGVIVSPATDLAGVAAYIEGELRGSKRPTYASVHGRDATE
 VRARDLALGKFIDAKTLSAAPGLPRSGTEIRTVLLTGATGFLGRYLALEWLER
 MDLVDGKVICLVRARSDDEARARLDATFDTGDATLLEHYRALAADHLEVIAGD
 KGEADLGLDHD TWQRLADTVDLIVDPAALVNHVLPYSQMF'GPNALGTAELIRI
 ALTTTIKPYVYVSTIGVGQGISPEAFVEDADIREISATRRVDDSYANGYGNSK
 WAGEVLLREAHDWCGLPVSVFRCDMILADTTYSGQLNLPDMFTRLMLSLVATG
 IAPGSFYELDADGNRQRAHYDGLPVEFIAEAI STIGSQVTDGFETFHVMNPYD
 DGIGLDEYVDWLI EAGYPVHRVDDYATWLSRFETALRALPERQRQASLLPLLH
 NYQQPSPPVCGAMAPTDRFRAAVQDAKIGPDKDIPHVTADVIVKYISNLQMLG
 LL

YP 889972 (CARB)

Secuencia de nucleótidos (SEQ ID N0:21)

>gi | 118467340:5821317-5824838 Mycobacterium smegmatis str. MC2 155, genoma completo

ATGACCAGCGATGTTACGACGCCACAGACGGCGTCACCGAAACCGCACTCGA
 CGACGAGCAGTCGACCCGCCGCATCGCCGAGCTGTACGCCACCGATCCCGAGT
 TCGCCGCCGCCGCACCGTTGCCCGCCGTGGTCGACGCGGCGCACAAACCCGGG
 CTGCGGCTGGCAGAGATCCTGCAGACCCTGTTACCGGCTACGGTGACCGCCC
 GCGCTGGGATAACCGCGCCCGTGA ACTGGCCACCGACGAGGGCGGGCGCACCG
 TGACGCGTCTGCTGCCGCGGTTGACACCCTCACCTACGCCAGGTGTGGTCG
 CGCGTGCAAGCGGTGCGCCGCGGCCCTGCGCCACA ACTTCGCGCAGCCGATCTA
 CCCC GGCGACGCCGTGCGGACGATCGGTTTCGCGAGTCCCGATTACCTGACGC
 TGGATCTCGTATGCGCCTACCTGGGCCTCGTGAGTGTTCCGCTGCAGCACAAC
 GCACCGGTCAGCCGGCTCGCCCCGATCCTGGCCGAGGTGGAACCGCGGATCCT
 CACCGTGAGCGCCGAATACCTCGACCTCGCAGTCGAATCCGTGCGGGACGTCA
 ACTCGGTGTCGCAGCTCGTGGTGTTCGACCATCACCCCGAGGTGACGACCAC
 CGCGACGCACTGGCCCGCGCGCGTGAACA ACTCGCCGGCAAGGGCATCGCCGT
 CACCACCCTGGACGCGATCGCCGACGAGGGCGCCGGGCTGCCGGCCGAACCGA
 TCTACACCGCCGACCATGATCAGCGCCTCGCGATGATCCTGTACACCTCGGGT
 TCCACCGGCGCACCCAAGGGT GCGATGTACACCGAGGCGATGGTGGCGCGGCT
 GTGGACCATGTGTTTCATCACGGGTGACCCACGCCGGTCATCAACGTCAACT
 TCATGCCGCTCAACCACCTGGGCGGGCGCATCCCCATTTCCACCGCCGTGCAG
 AACGGTGGAACCAAGTTACTTCGTACCGGAATCCGACATGTCCACGCTGTTTGA
 GGATCTCGCGCTGGTGCGCCCGACCGAACTCGGCCTGGTTCCGCGCGTCGCCG
 ACATGCTCTACCAGCACCACTCGCCACCGTGCACCGCCTGGTCACGCAGGGC
 GCCGACGAACTGACCGCCGAGAAGCAGGGCCGGTGCCGAACTGCGTGAGCAGGT

FIG 4 (Cont.)

GCTCGGCGGACGCGTGATCACCGGATTCGTCAGCACCGCACCGCTGGCCGCGG
 AGATGAGGGCGTTCCCTCGACATCACCTGGGCGCACACATCGTCGACGGCTAC
 GGGCTCACCGAGACCGGCGCCGTGACACGCGACGGTGTGATCGTGCGGCCACC
 GGTGATCGACTACAAGCTGATCGACGTTCCCGAACTCGGCTACTTCAGCACCG
 ACAAGCCCTACCCGCGTGGCGAACTGCTGGTCAGGTCGCAAACGCTGACTCCC
 GGGTACTACAAGCGCCCCGAGGTACCGCGAGCGTCTTCGACCGGGACGGCTA
 CTACCACACCGGCGACGTCATGGCCGAGACCGCACCCGACCACCTGGTGTACG
 TGGACCGTCGCAACAACGTCCTCAAACCTCGCGCAGGGCGAGTTCGTGGCGGT
 GCCAACCTGGAGGCGGTGTTCTCCGGCGCGGCGCTGGTGCGCCAGATCTTCGT
 GTACGGCAACAGCGAGCGCAGTTTCTTCTGGCCGTGGTGGTCCCGACGCCGG
 AGGCGCTCGAGCAGTACGATCCGGCCGCGCTCAAGGCCGCGCTGGCCGACTCG
 CTGCAGCGCACCGCACGCGACGCCGAACTGCAATCCTACGAGGTGCCGGCCGA
 TTTCATCGTCGAGACCGAGCCGTTTCAGCGCCGCCAACGGGCTGCTGTCGGGTG
 TCGGAAAACCTGCTGCGGCCAACCTCAAAGACCGCTACGGGCAGCGCCTGGAG
 CAGATGTACGCCGATATCGCGGCCACGCAGGCCAACAGTTGCGCGAACTGCG
 GCGCGCGGCCGCCACACAACCGGTGATCGACACCCTCACCCAGGCCGCTGCCA
 CGATCCTCGGCACCGGGAGCGAGGTGGCATCCGACGCCCACTTCACCGACCTG
 GGCGGGGATTCCCTGTGCGCGCTGACACTTTCGAACCTGCTGAGCGATTTCTT
 CGGTTTCGAAGTTCCCGTTCGGCACCATCGTGAACCCGGCCACCAACCTCGCCC
 AACTCGCCCAGCACATCGAGGCGCAGCGCACCGCGGGTGACCGCAGGCCGAGT
 TTCACCACCGTGCACGGCGCGGACGCCACCGAGATCCGGGCGAGTGAGCTGAC
 CCTGGACAAGTTCATCGACGCCGAAACGCTCCGGGCGCACCGGGTCTGCCCA
 AGGTCACCACCGAGCCACGGACGGTGTGCTCTCGGGCGCCAACGGCTGGCTG
 GGCCGGTTCCTCACGTTGCAGTGGCTGGAACGCCTGGCACCTGTGCGGCGGCAC
 CCTCATCACGATCGTGCGGGGCCGCGACGACGCCGCGGCCCGCGCACGGCTGA
 CCCAGGCCTACGACACCGATCCCGAGTTGTCCCGCCGCTTCGCCGAGCTGGCC
 GACCGCCACCTGCGGGTGGTCGCCGGTACATCGGCGACCCGAATCTGGGCCT
 CACACCCGAGATCTGGCACCGGCTCGCCGCCGAGGTCGACCTGGTGGTGCATC
 CGGCAGCGCTGGTCAACCACGTGCTCCCCTACCGGCAGCTGTTTCGGCCCCAAC
 GTCGTGGGCACGGCCGAGGTGATCAAGCTGGCCCTCACCGAACGGATCAAGCC
 CGTCACGTACCTGTCCACCGTGTGCGGTGGCCATGGGGATCCCCGACTTCGAGG
 AGGACGGCGACATCCGGACCGTGAGCCCGGTGCGCCCCGCTCGACGGCGGATAC
 GCCAACGGCTACGGCAACAGCAAGTGGGCGCGGAGGTGCTGCTGCGGGAGGC
 CCACGATCTGTGCGGGCTGCCCGTGGCGACGTTCCGCTCGGACATGATCCTGG
 CGCATCCGCGCTACCGCGGTGAGGTCAACGTGCCAGACATGTTACGCGACTC
 CTGTTGAGCCTCTTGATCACCGGCGTCGCGCCGCGGTTCGTTCTACATCGGAGA
 CGGTGAGCGCCCGCGGGCGCACTACCCCGGCCTGACGGTCGATTTTCGTGGCCG
 AGGCGGTACGACGCTCGGGCGCGCAGCAGCGCGAGGGATACGTGTCCTACGAC
 GTGATGAACCCGCACGACGACGGGATCTCCCTGGATGTGTTTCGTGGACTGGCT
 GATCCGGGCGGGCCATCCGATCGACCGGGTCGACGACTACGACGACTGGGTGC
 GTCGGTTCGAGACCGCGTTGACCGCGCTTCCCGAGAAGCGCCGCGCACAGACC

FIG 4 (Cont.)

GTACTGCCGCTGCTGCACGCGTTCCGCGCTCCGCAGGCACCGTTGCGCGGGCGC
 ACCCGAACCCACGGAGGTGTTCCACGCCGCGGTGCGCACCGCGAAGGTGGGCC
 CGGGAGACATCCCGCACCTCGACGAGGCGCTGATCGACAAGTACATACGCGAT
 CTGCGTGAGTTCGGTCTGATCTGA

Secuencia de aminoácidos (SEQ ID N0:22)

>gi |118469671 |ref |YP_889972.1 | supuesta ácido graso de cadena larga-CoA ligasa [Mycobacterium
 smegmatis str. MC2 155]

MTSDVHDATDGVTE TALDDEQSTRRIAELYATDPEFAAAAAPLPAVVDAAHKPG
 LRLAEILQTLFTGYGDRPALGYRARELATDEGGRTVTRLLPRFDLTLYAQVWS
 RVQAVAAALRHNFAQPIYPGDAVATIGFASPDYLTLDLVCAYLGLVSVPLQHN
 APVSR LAPILAEVEPRILTVSAEYLDLAVESVRDVNSVSQLVVFDHHPVEVDDH
 RDALARAREQLAGKGI AVTTLD AIADEGAGLPAEPIYTADHDQRLAMILYTSG
 STGAPKGAMYTEAMVARLWTMSFITGDPTPVINVNF MPLNHLGGRIPISTAVQ
 NGGTSYFVPESDMSTLFEDLALVRPTELGLVPRVADMLYQHHLATVDRLVTQG
 ADELTAEKQAGAE LREQVLGGRVITGFVSTAPLAAEMRAFLDITLGAHIVDGY
 GLTETGAVTRDGVIVRPPVIDYKLI DVPELGYFSTDKPYPRGELLVRSQTLTP
 GYYKRPEVTASVFD RDGYHTGDVMAETAPDHLVYVDRRNNVLKLAQGEFVAV
 ANLEAVFSGAALVRQIFVYGN SERSFLLAVVVPTPEALEQYDPAALKAALADS
 LQRTARDAELQSYEVPADFIVETEPFSAANGLLSGVGKLLRPNLKDRYGQRLE
 QMYADIAATQANQLREL RRAAATQPV IDTLTQAAATILGTGSEVASDAHF TDL
 GGDSL SALTLSNLLSDFFGFEV PVGTIVNPATNLAQLAQHIEAQR TAGDRRPS
 FTTVHGADATEIRASELTLDKFI DAETLRAAPGLPKVTTEPRTVLLSGANGWL
 GRFLT LQWLERLAPVGGTLITIVRGRDDAAARARLTQAYDTPELSRRFAELA
 DRHLRVVAGDIGDPNLGLTPEIWHRLAAEVDLVVHPAALVNHVLPYRQLFGPN
 VVGTAEVIK LALTERIKPVTYLSTVSVAMGIPDFEEDGDIRTVSPVRPLDGGY
 ANGYGNSKWAGEVLLREAHDL CGLPVATFRSDMILAHPRYRGQVNV PDMFTRL
 LLSLLITGVAPRSFYIGDGERPRAHYPGLTVDFVAEAVTTLGAQQREGYVSYD
 VMNPHDDGISLDVFDWLIRAGHPIDRVDDYDDWVRRFETALTALPEKRR AQT
 VLPLLHAFRAPQAPLRGAPEPTEVFHAAVRTAKVGP GDIPHLDEALIDKYIRD
 LREFGLI

FIG 4 (Cont.)

YP 905678.1**Secuencia de nucleótidos (SEQ ID NO:23)**

>uniprot | A0PPD8 | A0PPD8_MYCUA ácido graso-CoA ligasa FadD9

ATGTCGCCAATCACGCGTGAAGAGCGGCTCGAGCGCCGCATCCAGGACCTCTA
 CGCCAACGACCCGCAGTTCGCCGCCGCAAACCCGTCACGGCGATCACCGCAG
 CAATCGAGCGGCCGGGTCTACCGCTACCCAGATCATCGAGACCGTCATGACC
 GGATACGCCGATCGGCCGGCTCTCGCTCAGCGCTCGGTCTGAATTCGTGACCGA
 TGCCGGCACCGGCCACACCACGCTGCGACTGCTCCCCACTTCGAAACCATCA
 GCTACGGCGAGCTTTGGGACCGCATCAGCGCACTGGCCGACGTGCTCAGCACC
 GAACAGACGGTGAACCCGAGCGACCGGGTCTGCTTGTTGGGCTTCAACAGCGT
 CGACTACGCCACGATCGACATGACTTTGGCGCGGCTGGGCGCGGTGGCTGTAC
 CACTGCAGACCAGCGCGGCGATAACCCAGCTGCAGCCGATCGTCGCCGAGACC
 CAGCCCACCATGATCGCGGCCAGCGTCGACGCACTCGCTGACGCCACCGAATT
 GGCTCTGTCCGGTCAGACCGCTACCCGAGTCCTGGTGTTCGACCACCACCGGC
 AGGTTGACGCACACCGCGCAGCGGTCTGAATCCGCCCGGGAGCGCCTGGCTGGC
 TCGGCGGTCTCGAAACCCTGGCCGAGGCCATCGCGCGCGGCGACGTGCCCCG
 CGGTGCGTCCGCCGGCTCGGCGCCCGGCACCGATGTGTCCGACGACTCGCTCG
 CGCTACTGATCTACACCTCGGGCAGCACCGGTGCGCCCAAGGGCGCGATGTAC
 CCCCAGCGCAACGTTGCGACCTTCTGGCGCAAGCGCACCTGGTTCGAAGGCGG
 CTACGAGCCGTGATCACGCTGAACTTCATGCCAATGAGCCACGTGATGGGCC
 GCCAAATCCTGTACGGCACGCTGTGCAATGGCGGCACCGCCTACTTCGTGGTG
 AAAAGCGATCTCTCCACCTTGTTTGAAGACCTGGCGCTGGTGCGGCCCACCGA
 GCTGACCTTCGTGCCGCGCGTGTGGGACATGGTGTTCGACGAGTTTCAGAGTG
 AGGTCGACCGCCGCTGGTCGACGGCGCCGACCGGGTCGCGCTCGAAGCCCAG
 GTCAAGGCCGAGATACGCAACGACGTGCTCGGTGGACGGTATAACCAGCGCACT
 GACCGGTCCGCCCGATCTCCGACGAGATGAAGGCGTGGGTTCGAGGAGCTGC
 TCGACATGCATCTGGTCGAGGGCTACGGCTCCACCGAGGCCGGGATGATCCTG
 ATCGACGGAGCCATTCGGCGCCCGGCGGTACTCGACTACAAGCTGGTTCGATGT
 TCCCGACCTGGGTTACTTCTGACCGACCGGCCACATCCGCGGGGCGAGTTGC
 TGGTCAAGACCGATAGTTTGTTCGCGGGCTACTACCAGCGAGCCGAAGTCACC
 GCCGACGTGTTTCGATGCTGACGGCTTCTACCGGACCGGCGACATCATGGCCGA
 GGTCGGCCCCGAACAGTTCGTGTACCTCGACCGCCGCAACAACGTGTTGAAGC
 TGTCGCAGGGCGAGTTCGTACCGTCTCCAACTCGAAGCGGTGTTTGGCGAC
 AGCCCACTGGTACGGCAGATCTACATCTACGGCAACAGCGCCCGTGCCTACCT
 GTTGGCGGTGATCGTCCCCACCCAGGAGGCGCTGGACGCCGTGCCTGTCGAGG
 AGCTCAAGGCGCGGCTGGGCGACTCGCTGCAAGAGGTCGCAAAGGCCCGCCGGC
 CTGCAGTCCTACGAGATCCCGCGCGACTTCATCATCGAAACAACACCATGGAC

FIG 4 (Cont.)

GCTGCAGAACGGCCTGCTCACCGGCATCCGCAAGTTGGCCAGGCCGCGAGCTGA
AAAAGCATTACGGCGAGCTTCTCGAGCAGATCTACACGGACCTGGCACACGGC
CAGGCCGACGAACTGCGCTCGCTGCGCCAAAGCGGTGCCGATGCGCCGGTGCT
GGTGACGGTGTGCCGCGCGGCCGCGCTGTTGGGCGGCAGCGCCTCTGACG
TCCAGCCCAGATGCGCACTTCACCGATTTGGGCGGGCGACTCGCTGTCGGCGCTG
TCGTTACCAACCTGCTGCACGAGATCTTCGACATCGATGTGCCGGTGGGCGT
CATCGTCAGCCCCGCCAACGACTTGCAGGCCCTGGCCGACTACGTCGAGGCGG
CTCGCAAACCCGGCTCGTCACGACCGACCTTCGCCTCGGTCCACGGCGCCTCG
AATGAGCAGGTCACCGAGGTGCATGCCGGTGACCTGTCCCTGGACAAATTCAT
CGATGCCGCAACCCTGGCCGAAGCTCCCCGGCTGCCCGCCGAAACACCCAAG
TGCGCACCGTGCTGCTGACCGGCGCCACCGGCTTCCTCGGGCGCTACCTGGCC
CTGGAATGGCTGGAGCGGATGGACCTGGTCGACGGCAAACCTGATCTGCCTGGT
CCGGGCCAAGTCCGACACCGAAGCACGGGCGCGGTGGAAAAGACGTTTCGACA
GCGGGCGCCCCGAACTGCTGGCCCACTACCGCGCACTGGCCGGCGACCACCTC
GAGGTGCTCGCCGGTGACAAGGGCGAAGCCGACCTCGGACTGGACCGGCAGAC
CTGGCAACGCCTGGCCGACACGGTGCACCTGATCGTGGACCCCGCGGCCCTGG
TCAACCACGTACTGCCATACAGCCAGCTGTTCCGGGCCAACGCGCTGGGCACC
GCCGAGCTGCTGCGGCTCGCGCTCACCTCCAAGATCAAGCCCTACAGCTACAC
CTCGACAATCGGTGTCGCCGACCAGATCCCGCCGTCCGGCGTTCACCGAGGACG
CCGACATCCGGGTTCATCAGCGCCACCCGCGCGGTTCGACGACAGCTACGCCAAT
GGCTATTCGAACAGCAAGTGGGCCGGCGAGGTGCTGTTGCGCGAGGCGCATGT
CCTGTGTGGCCTGCCGGTTGCGGTGTTCCGCTGCGACATGATCCTGGCCGACA
CCACATGGGCGGGACAGCTCAACGTGCCGGACATGTTACCCGTATGATCCTG
AGCCTGGCGGCCACCGGTATCGCGCCGGGTTTCGTTCTATGAGCTTGCGGCCGA
CGGCGCCCGGCAACGCGCCCACTATGACGGTCTGCCCGTTCGAGTTCATCGCCG
AGGCGATTTTCGACTTTGGGTGCGCAGAGCCAGGATGGGTTCCACACGTATCAC
GTGATGAACCCTTACGACGACGGCATCGGACTCGACGAGTTCGTCGACTGGCT
CAACGAGTCCGGTTGCCCCATCCAGCGCATCGCTGACTATGGCGACTGGCTGC
AGCGCTTCGAAACCGCACTGCGCGCACTGCCCGATCGGCAGCGGCACAGCTCA
CTGCTGCCGCTGTTGCACAACCTATCGGCAGCCGGAGCGGCCCGTCCGCGGGTC
GATCGCCCCTACCGATCGCTTCCGGGCAGCGGTGCAAGAGGCCAAGATCGGCC
CCGACAAAGACATTCCGCACGTCGGCGCGCCGATCATCGTGAAGTACGTCAGC
GACCTGCGCCTACTCGGCCTGCTCTGA

Secuencia de aminoácidos (SEQ ID N0:24)

>uniprot | A0PPD8 | A0PPD8_MYCUA ácido graso-CoA ligasa FadD9

MSPITREERLERRIQDLYANDPQFAAAKPVTAITAAIERPGLPLPQIIET
VMTGYADRPALAQRSVEFVTDAGTGHTTLRLLPHFETISYGELWDRISAL

FIG 4 (Cont.)

ADVLSTEQTVKPSDRVCLLGFNSVDYATIDMTLARLGAVAVPLQTSAAIT
 QLQPIVAETQPTMIAASVDALADATELÀLSGQTATRVLVFDHHRQVDAHR
 AAVESARERLAGSAVVETLAEAIARGDVPRGASAGSAPGTDVSDDSLALL
 IYTSGSTGAPKGAMYPRRVATFWRKRTWFEAGGYEPSITLNFMPMSHVMG
 RQILYGTLCNGGTAYFVVKSDLSTLFEDLALVRPTELTFVPRVWDMVFDE
 FQSEVDRRLVDGADRVALEAQVKAERNDVLGGRYTSALTGSAPI SDEMK
 AWVEELDMHLVEGYGSTEAGMILIDGAIRPAVLVDYKLVDPDLGYFLT
 DRPHPRGELLVKTDSLFPGYQRAEVTADVFDADGFYRTGDIMAEVGP
 EQFVYLDRRNNVLKLSQGEFVTVSKLEAVFGDSPLVQRQIYIYGNSARAYLLA
 VIVPTQEALDAVPVEELKARLGDSDLQEVAKAAGLQSYEIPRDFI IETTPW
 TLQNGLLTGIRKLARPQLKKHYGELLEQIYTDLAHGQADELRSLRQSGAD
 APVLVTVCRAAAALLGGSASDVQPDAHFTDLGGDSL SALSFTNLLHEIFD
 IDVPVGVIVSPANLQALADYVEAARKPGSSRPTFASVHGASNEQVTEVH
 AGDLSLDKFIDAATLAEAPRLPAANTQVRTVLLTGATGFLGRYLALALEWLE
 RMDLVDGKLICLVRAKSDTEARARLEKTFDSGAPELLAHYRALAGDHLEV
 LAGDKGEADLGLDRQRTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNALG
 TAE LLRLALTSKIKPYSYTSSTIGVADQIPPSAFTEDADIRVISATRAVDD
 SYANGYSNSKWAGEVLLREAHVLCGLPVAVFRCDMILADTTWAGQLNVPD
 MFTRMILSLAATGIAPGSFYELAADGARQRAHYDGLPVEFIAEAI STLGA
 QSQDGFHTYHVMNPYDDGIGLDEFVDWLNESGCPIQRIADYGDWLQRFET
 ÀLRALPDRQRHSSLLPLLHNYRQPERPVRGSIAPTDRFRAAVQEAKIGPD
 KDIPHVGAPI IIVKYVSDLRLLGLL

ZP_04027864

Secuencia de nucleótidos (SEQ ID N0:25)

>gi |227980074:564806-568123 Tsukamurella paurometabola DSM 20162 TpauDRAFT_4083016_Cont3,
 secuenciación genómica completa al azar

ATGTCGATTGAGACGGTGCAGAACGGCGTCCCCGCAGAGGGCTCGGTGCCCCC
 CGCCGACCAGCAGACCGAGCGACTGCCGCAGGTGATCGCCAGGATCTTCGCCC
 AGTTCGCGGATCGTCCGGCCTTCGCGACCCGCGAGGCGGGGCCGGGGACCCCC
 TACGCCACCGTCTCCTATCGGGAGATCTGGCGGCGCGTCACCGCGCTGGTGGC
 CTCCTGGCAGAGCGAAGTGGCTCCGGGAGACTTCGTCGCCATCCTCGGCTTCA
 CCAGCTCGGACTTCGTGACCGTCGACCTCGCGACCACACTGCTCGGCGCCCCG
 AACGTGCCGCTCCAGGCCGGGGCCCCCGCCGCTCGCATCGCGACCATCCTCGA
 TGAGACCCGGCCGAAGATCCTCGCCGTGAGTGCCGATCAGGTCGACCTCGCCC
 AGGAGGCTCTGGCCGAGTCCGCGGCTACCCCGCGGGTGGTCGTCTTCGACGGC
 GAACGCGACGGCTACGÀGGGCATCGAGGCGGACATCCTTTCCGGCTCCGCCCT

FIG 4 (Cont.)

GCCGGCACCGGAGTTCTTCGCGCCCGAGCCCGGCACCGATCCTCTCGTCACGC
 TCATCTACACATCCGGCAGCACCGGTACCCCGAAGGGGGCCATGTACACCGAG
 CAGTTGGTTCGCGATGCCTGGCTCAAGGTGGACAGCATCGTCGACATCGACAT
 GCCGGCCGAGTCGCTCCTGCACTTCTGCCTATGAGCCATATGTACGGGCGCA
 ACTGGCTGATCGCCGGCCTGGCATCGGGCGGGACCGGGTACTTCGCCGGCGCC
 TCCGATATGTCCACCCTGTTCGACGACCTCGCCGCCGCCCGGCCACCGCCAT
 CGGCCTGGTGCCCCGCGTGTGCGAGCTGATACACCAGCGCTATCTGGCCGTCG
 AGGCGGACACTGATGCGGAGACCGCGCGCGTCGAACTGCGTGACCGGGTACTC
 GCGGTTCGGCTGCAGGCCGCGATGTGCGGTAGCGCCGCCCTCTCGTCGGAGCT
 GCAGACCTTCATGGAGTGGTTGCTCGGAATCGATATCCAGATCGGCTACGGAT
 CCACCGAGGCCGGTGGTGTTCATCCGCGACGGAGTGGTCGTTTCGGCCGCCGGTC
 ACGGAGTACAAGCTGATCGATGTCCCCGAACTGGGCTACTTCGTCACCGACTC
 CCCGCATCCACGCGGGCGAACTCCTGGTCAAGTCGACGCAGTTGATTCCCAGGT
 ACTACAACCTCCGACAAGCGGATCCGCGACGACGAAGGCTTCTACCGCACCGGC
 GATGTGATGGCCGAGCTGGGACCCGACCGGCTCGAGTACGTCGACCGGCGGAG
 CAACGTGATCAAGTTGGCACAGGGAGAGTTCGTGCCGATCGCCAACTCGAGG
 CCATCTACGCCGCCGGTCCCGATGTGCACCAGATCTTCCTGTACGGAACCAGC
 GAACGCTCCTACCTGATCGGCGTCGTCGTGCCCGCGCCGGGACCCGACGGCGA
 GACCGATGCGCAGACCCGCACCCGCGTACTCGATGGCCTGGCCGCGATCGCCC
 GTGAGAACGATCTCGCTGCCTACGAGGTGCCGCGCGATGTGCTCATCGAACGT
 GATCCCTTCTCTCAGGAGAACGGGCTGCGGTCGGGGATCGGCAAGCTGGTGCG
 CCCGGCCCTCATCGCCCGCTACGGTGACCGGTTGCACGACCTCTACGCCCAGG
 CCGACACCCGTCAACGCGAGGGCTTTCGCGCTCTCGACGCCTCGGGCCCGATC
 ATCGACACCGTGCTCGGGGCGGCTGCGTTGACGCTCGGCGCGGATATCGCGGA
 CTTTCGACGCCGACACTCGATTTCGGCGACCTCGGTGGCGACTCGTTGTCGGCGC
 TCTCGCTCGCGACGACGCTCGAAGGCCTCTACGACGTGCCCGTCCCCGTGCAG
 ACGATCGTCGGACCGACCGCCACACTCGGCGGGCGTCGCCCGGCACATCGAGAA
 GGCTCGATCGGGTGGCGTCGCGGCACCGACCGCCGACTCGGTGCACGGCGTGG
 GTGCGAGCGTCGCCCGGGCCACCGACCTGACGCTGGAGAAGTTCATCGACCCC
 GAGTCCTTCGCGCTCGCGCCGACGCTTCCC CGGGCGACCGGTGAGCCGAACAC
 CGTGCTGCTCACCGGATCCACCGGCTACCTCGGCCGCTTCTCTGCTGCTGGACT
 GGTGCGACGGGTCGCTCCGCACGGCGGCACCGTGATCGCGCTGGTGC GCGGC
 GCCGACGCCGACGATGCGCGACGCCGCGTCACGGCCGCGATCGGTGACTCGGA
 TCCTGACCTGACACAAGAGTTCACGTCACCTCGCGGAGCATCACCTCCACGTGA
 TCGCCGGTGACTTCGGCAGCCCCGCACTCGGACTCGACGATGCCACCTGGAGC
 GATCTCGCCGGGCGAGTCGATCACGTGGTGC ACTGCGGCGCGCTCGTCAACCA
 CGTGCTGCCCTACGACCAACTGTTTCGGTCCCAATGTGGTGGCCACCGGCCAAG
 TGGTGC GACTCGCACTACCACGCGCCGCAAGTCCGTGGATTACGTCTCCACG
 GTGGCTGTGGTTCCGCAGGATGACGGCCGCGTCTTGGTTCGAGGACGACGATGT
 TCGCGAGCTCGGCGCCGAACGGCGCATCGGGGCCGATGCCTACGCGAACGGCT
 ACGCCGTGAGCAAATGGGCGGGCGAAGTGCTGTTGCATGAGGCAGCCGACCTG

FIG 4 (Cont.)

GCGGACCTGCCGGTGCGGGTGTTCGCTCCGATATGATCTTGGCGCACAGTCG
 ATTCCACGGACAGTTCAACGAGGTTCGACCAGTTCACCCGCCTGCTCCTGAGTA
 TCGCCGAGACCGGACTGGCGCCGGCGTTCGTTCTACACGCCGGATCCGAGTGGA
 CACCGCCC GCACTACGACGGGCTGCCGGTGGACTTCACCGCCGAAGCGATCAC
 CACGCTCAGCGCCGCGGGGCGTTCGGGGTACCGGACCTTCCACGTGCTCAACG
 CCAACGATGACGGCGTGAGCTTGACAGCTTCGTCGACTGGATCGCCGCCTCG
 GGCCGGAGCATCGAACGGATCGACGACTACGACACCTGGTTCGCCCGGTTTCA
 GCAGGCGCTCCAGCAGCTCCCCGATGAGGGCGCGCCAGCGGTCGGTGCTGCCCC
 TGCTGCACGCGGTGCGCGAGCCGGCTCCGGCCGCCGGGACCTCCGCGCTGTCG
 GTGGACCGGTTCCGTGGTGCGGTGCCTGAGACCGGAGTAGGACCGGGGGAÇAT
 CCCGGTGCTCGATCGCGCCCTGATCGAGAAGTACCTGCGCGACTTCGAGACCG
 CGGGCTGGCTCGCGCCCGGTGCGCGCGACTGA

Secuencia de aminoácidos (SEQ ID N0:26)

>gi | 227980601 | ref | ZP_04027864.1 | proteína de tipo tioéster reductasa [Tsukamurella paurometabola DSM 20162]

MSIETVQNGVPAEGSVPPADQQTERLPQVIARIFAQFADRPATREAGPGTP
 YATVSYREIWRRTALVASWQSEVAPGDFVAILGFTSSDFVTVDLATTLLGAP
 NVPLQAGAPAARIATILDETRPKILAVSADQVDLAQEALAESAAATPRVVVFDG
 ERDGYEGIEADILSGSALPAPEFFAPEPGTDPLVTLIYTSGSTGTPKGAMYTE
 QLVRDAWLKVDIVDIDMPAESLLHFLPMSHMYGRNWLIAGLASGGTGYFAGA
 SDMSTLFDLLAAARPTAIGLVPRVCELIHQRYLAVEADTDAETARVELRDRVL
 GGRLQAAMCGSAALSSELQTFMEWLLGIDIQIGYGSTEAGGVIRDGVVVRPPV
 TEYKLI DVPELGYFVTDSPHPRGELLVKSTQLIPGYNSDKRIRDDEGFYRTG
 DVMAELGPDRLEYVDRRSNVIKLAQGEFVPIAQLEAIYAAGPDVHQIFLYGTS
 ERSYLIGVVVPAPGPDGETDAQTRTRVLDGLAAIARENDLAAAYEVRDVLIER
 DPF SQENGLRSGIGKLV RPAL IARYGDRLHDLYAQADTRQREGLRALDASGP I
 IDTVL GAAAL TLGADIADFDADTRFGDLGGDSLSALS LATTLEGLYDVPVPVQ
 TIVGPTATLGGVARHIEKARSGGVAAPTADSVHGVGASVARATDLTLEKFIDP
 ELLALAPTLPAAATGEPNTVLLTGSTGYLGRFLLLDWLRRVAPHGGTVIALVRG
 ADADDARRRVTA AIGDSDPDLTQEFTSLAEHHLHVIAGDFGSPALGLDDATWS
 DLAGRVDHVHCGALVNHVLPYDQLFGPNVVATGEVVR LALTTRRKSVDYVST
 VAVVPQDDGRVLVEDDDVRELGAERRIGADAYANGYAVSKWAGEVLLHEAADL
 ADLPVRVFRSDMILAHSRFHGQFNEVDQFTRL LLSIAETGLAPASFYTPDPSG
 HRPHYDGLPVDF TAEAITTL SAAGRSGYRTFHVLNANDDGVSLDSFVDWIAAS
 GRSIERIDDYDTWFARFEQALQQLPDEARQRSVLP LLHAVREPAPAAGTSALS
 VDRFRGAVRETGVGPGDIPVLDRALIEKYLRDFETAGWLAPGARD

FIG 4 (Cont.)

ZP 05045132**Secuencia de nucleótidos (SEQ ID N0:27)**

>gi |254430111:343253-346687 Cyanobium sp. PCC 7001 scf_1106012173546 estructura principal genómica, secuenciación genómica completa al azar

GTGAATGAGTCTTCCGCGGACCAGAGTTCCGGCAACGTTTCCGAGGGGTGGCC
 TGATGCTTCGGTCACAGCACGGGCCCTGCAGGCTCACCTGCGCTACGAACAGA
 TCATCGATGCCATTCTGAGCGGCTACGCCGAGCGCCCTGCTCTGGCGGAGCGT
 TCCTACCTGGTGC GGCCGGACCCGAGCACAGGTCAAACGGTGC GTGTCCACGA
 GCAGGCCTTCCGCTCGATCAGCTACCGAACCTGCAGGAACGGGTTCATGCCC
 TCACCATGGCCTGGCGCCTTCATCCCGATAGCCCGGTGCAAGCGGGAGCCTTC
 GTGGTGTCTGGTGGGATTTGCCAGCATCGATTACGCCGTTCTTGATCTGGCACT
 GGCTACACCAAGGGCGTGC CGGTGCCCTGT CACCGAACCATTCAGCGAGG
 ACGATGACGCCATCCTCGGCACAGTCCAGCCCGTCACTCTGGCGGTATCGATC
 AGTGAGTTCTCTGGCTGTGTGACCTGATCGCCGATCGACGTGATCCGAAC
 TGTGATCGTCTTTGACCTTGACCCTGCCGTCGACTGCGAGCGCGCCGCACTGG
 AGAGCGGCATCCGGGCACTCAACGAGAAGGGGT CAGACGTTGTGTT CAGACA
 CTGCAGGATCTGATTGACGTTGGGAGAGACGCAGAGTTCAGCTTCCTGCCGAT
 CCAGGCGCAGGATCAAGATGACCTGGCACTTCTGATTCACACATCCGGCAGCA
 CAGGCACACCCAAGGGAGCCTGCATCTCATCCCGTGC ACTGATCAACACCTGG
 CGCCATGTTTCCGGTCCCTATCCAAAAGTGACCGTGGTTCTGGCACCCCTTCCA
 CCACATGATGGGACGAGACTCGATGATCACGGCATTGGGCGCGGGCGGCACCG
 CCTACTTCACGCTCAGGCCTGACCTTTCGACCGTGATTGAAGACATCAGACTG
 GCACGGCCACAGGCCTGGTGTGTTTCCCGCCTCTGCGAAGTGATCGAACA
 CCACCTGACTACTGCGCCGGAGTATTCAGGCAACGAGATCCTCGGAGGCAGAC
 TGCAATCCATTGTGGTGGCCTCGGCTCCGATCACGCCACGCTTGAAGGCATCC
 CTGGAGTGCCCTCTTGGGGTGCCGTGTGAGCGAAGGCTACAGCAGCACGGAAAC
 AGCCAGTGGCGGGCTGGCGATGAATGGACTGCTGAACCGCAACAACATTCTCG
 CGTATCGCCTTCGCGATGTGCCGAGGCAGGGTATTCAGTGAATGATCGGCC
 TTTCCGCGCGGAGAACTCTGCGTGAAGACCCGCTTCGGTATCTCAGGCTATTT
 CAGAAATCCAGAGGCCACTGCAGAGCTGTTTCGACGACGATGGCTTCTATTGCA
 CCGGTGACATCGTTGAAGAGCGGGCCCCCGATCAAATCGCCATCATCGACCGG
 CGAAAGAATGTCATCAAGCTGGCGCAGGGTGAATATGTCGCTGTGGGCAGGCT
 GGAACAGCTTTTCCAGGAAGGTTGTGGTTGCGTGCAGCAGATTCACCTCCACG
 GCGACAGCACAAGGGCCTATCTGCTGGCAGTCGTGGTACCTGATCGCAACACC
 CTTGCACCGCCCGGGTCACGGCAGGCCAGTGAGGCCGAGTTAAAGGCACGGGT
 GCGCGAGGAGATTCTCACCTTGGCAAACCAACGGGAGCTGCGCGGCTTCGAGA

FIG 4 (Cont.)

TCCCTCGAGACCTGATCCTGGCGGAGGAACCCTTCTCCCAGCAGAACGGTCTG
 CTGTCGTCCTTGGGTAAGCCGATCCGCCC GGCCATCCGCGCGCGCTACCGCAG
 CCGGCTGGAGAGCCTGTATGCCAGCCATGAGGCCACCCGAGGCACTGAGCTCG
 AGGCCATCAGAGCGTCAGCTGGCGCGGTGGATGTGGAAACCACCCTGTTGGCG
 CTGCTGAGCAGCACGCTGGGTGTGGTGTGTGGGGCTGCCGATCGGCAGACGAG
 TTTCCGCGAGCTGGGGGGCGACTCCCTGGCCGCTGTGCAGCTGGCGATGGAGA
 TCAAGAAGCAGTTTCGGGGTGGGGCTGGAAGGGAGCCAGATTCTCGGGCCGGGC
 GGCACGGTGAAGCGTGGGCGCGGAGGATCCACACCGCCTCCATCCAGCAGGC
 CCCGCACCAGCGGGTTGGCAGTCCCCTCGCCGCCATTCCGGCCGAGGGGTGGC
 TGAAGCCGGACCACTACAGGCTGGAGAACCTGATCGGGATTCCCATCGGTACA
 CCCTCAGCCGAGGTGGCCAGGCCACAGGCGGGCCCCCTACGGTTCGTCTAC
 CGGTGCCACCGGTTTTCTGGGGGGCGCTTGTGCCTGGAGTGGCTGCAACGGC
 TGGCTGGCCAGGGGGGCGAGGCTGATCTGTCTGGTGCGCCCTCGAACAGCCAT
 TCCGCCTGGGAGCGACTGAGGAACCGCTTCTCCCATCTGGAACCCGAGCAGGT
 GGCACGCTTCCGCGAGCTGGCGGGAAGGCATCTGGAGGTGATTCCGGCGGACA
 TCGGAGAGCCCGCCTGGGGCTTGAACCGGGTTGCCAGGAGCGGCTCGCCACT
 GAGGTGGACGCGATCTGTCACTGCGCAGCGGAAGTGAATCACCGGCTGCCCTA
 TCGCCACCTCTACCGGCCCAATGTGATCGGCACCGCGGAGATCATTCACCTGG
 CGATCACGACGCGGCTGAAATCGGTGGACTTCATCTCCAGCATCGGGGTGGCT
 TCCCTGCCCCGGCGGCCGGGAGGGAGCATCCCGGTGGAGGGCGGCTACGCCCCG
 GGGCTACTTCGCCAGCAAGTGGGCTTGCAGCAACTGCTGCGCTCCACCCATG
 ACTGCACCGGTGTGCCCGTACGGGTGATTTCGGCCCAGCCTCATTCTTCCCGAT
 CGTGTGCTGGCCGGGAGATGAACCCGGACGATCTGCTTTCAAGACTGCTGTA
 CAGCATCCTGGTGACCGGGATCGCCCCGGGTGCTTTGGGGAGGAGTCGCAGA
 ACAGTGGACGATCGGGGTTCTCGGTGCAGGGCCTCCCCGTCGACCAGTTGGCG
 CAGACCATCCTTGCCCTCGGGGAAGCGCGCACGGAGGGATTCATGTGCTCAA
 CCTCAACGCTGACAGTGGCAGCGGTGTTCCCTGGATGCCATCCTCCAGGACA
 TCGCCGCCAAAGGAATCAGGCTGCGACGGGTGGAGGGCTATGACCTCTGGCTC
 GACGCGATCACAACCCGCCTGCGTCGCCTGCCAGCCGAGCAACGGGGCCCGTTC
 CCTGCTGGATGTGGCGGAAGCCTATGCAGGATCAGCAGGCCAGACAACGCAGA
 GCAGCGGTGAAATGCAGGCGGGCAGCAGCTCCTGCCCGGAGGAGATCACCAGC
 CTGCAACCGGACTTCAGTAGGGCCTACAGGCGCAAGATCGTGGATGATCTGGC
 TCGGTGGGGGCTGATCGAGCCTCCAGGACCCGTGGATCAGTGA

Secuencia de aminoácidos (SEQ 10 N0:28)

>gi | 254431429 | ref | ZP_05045132.1 | supuesta ácido graso de cadena larga--CoA ligasa [Cyanobium sp. PCC 7001]

FIG 4 (Cont.)

MNESSADQSSGNVSEGWPDASVTARALQAHLRYEQIIDAILSGYAERPALAER
 SYLVRPDPSTGQTVRVHEQAFRSISYRTLQERVHALTMAWRLHPDSPVQAGAF
 VVLVGFASIDYAVLDLALAYTKGVPVPLSPNHSSSEDDDAILGTVQPVTLAVSI
 SEFSGCVDLIARSTSIRTIVVFDLDPAVDCERAALESGIRALNEKGS DVVVQT
 LQDLIDVGRDAEFSFLPIQAQDQDDLALLIHTSGSTGTPKGACISSRALINTW
 RHVSGPYPKVTVVLAPFHMMGRDSMITALGAGGTAYFTLRPDLSTVIEDIRL
 ARPTGLVLFPRLCIEVIEHHLTTAPEYSGNEILGGRLQSIVVASAPITPRLKAS
 LECLLGVPVSEGYSSSTETASGGLAMNGLLNRRNILAYRLRDVPEAGYSVNDRP
 FPRGELCVKTRFGISGYFRNPEATAELFDDDDGFYCTGDIVEERAPDQIAIIDR
 RKNVTKLAQGEYVAVGRLEQLFQEGCGCVQQIHLHGDSTRAYLLAVVVPDRNT
 LAPPGSRQASEAELKARVREEILTLANQRELRGFEIPRDLILAEEPF SQQNGL
 LSSLGKPIRPAIRARYRSRLESLYASHEATRGTLEAIRASAGAVDVETLLA
 LLSSTLGVVCGAADRQTSFRELGGDSLAAVQLAMEIKKQFGVGLGSGQILGPG
 GTVEAWARRIHTASIQQAPHQRVGSPLAAIPAEGWLKPDHYRLENLIGIPIGT
 PSAEVARPTGGPPTVLLTGATGFLGGRLCLEWLQRLAGQGGRLICLVRPSNSH
 SAWERLRNRF SHLEPEQVARFRELAGRHLVIVIPADIGEPGLGLEPGCQERLAT
 EVDAICHCAAENVHRLPYRHLYRPNVIGTAEI IHLAITTRLKSVDFISSIGVA
 SLPRRPGGSIPVEGGYARGYFASKWACEQLLRSTHDCTGVPVRVIRPSLILPD
 RVLAGE MNPDDLRSRLLYSILVTGIAPGCFGEESQNSGRSGFSVQGLPVDQLA
 QTILALGEARTEGFHVLNLNADSGSGVPLDAILQDIAAKGIRLRRVEGYDLWL
 DAITTRLRRLPAEQRARSLLDVAEAYAGSAGQTTQSSGEMQAGSSSCPEEITS
 LQPDFSRAYRRKIVDDLARWGLIEPPGPVDQ

YP 882653.1

Secuencia de nucleótidos (SEQ ID NO:29)

>uniprot|A0QIB5|A0QIB5_MYCA1 Supuesta acil-CoA deshidrogenasa

ATGTCGACTGCCACCCATGACGAACGACTCGACCGTCGCGTCCACGAACTCAT
 CGCCACCGACCCGCAATTCGCCGCCGCCCAACCCGACCCGGCGATCACCGCCG
 CCCTCGAACAGCCCGGGCTGCGGCTGCCGCAGATCATCCGCACCGTGCTCGAC
 GGCTACGCCGACCGGCCGGCGCTGGGACAGCGCGTGGTGGAGTTCGTACGGGA
 CGCCAAGACCGGGCGCACGTCGGCGCAGCTGCTCCCCGCTTCGAGACCATTA
 CGTACGGCGAAGTGGCGCAGCGTGTTCGGCGCTGGGCCGCGCCCTGTCTGAC
 GACGCGGTGCACCCCGGCGACCGGGTGTGCGTGCTGGGCTTCAACAGCGTCGA
 CTACGCCACCATCGACATGGCGCTGGGCGCCATCGGCGCCGTCTCGGTGCCGC
 TGCAGACCAGCGCGCAATCAGCTCGCTGCAGCCGATCGTGGCCGAGACCGAG
 CCCACCCTGATCGCGTCCAGCGTGAACCAGCTGTCCGACGCGGTGCAGCTGAT

FIG 4 (Cont.)

CACCGGCGCCGAGCAGGCGCCCACCCGGCTGGTGGTGTTCGACTACCACCCGC
 AGGTCGACGACCAGCGCGAGGCCGTCCAGGACGCCGCGGGCGGGCTGTCCGGC
 ACCGGCGTGGCCGTCCAGACGCTGGCCGAGCTGCTGGAGCGCGGCAAGGACCT
 GCCCGCCGTGCGGGAGCCGCCCGCCGACGAGGACTCGCTGGCCCTGCTGATCT
 ACACCTCCGGGTCCACCGGCGCCCCAAGGGCGCGATGTACCCGCAGAGCAAC
 GTCGGCAAGATGTGGCGCCGCGGCAGCAAGAACTGGTTCGGCGAGAGCGCCGC
 GTCGATCACCTGAATTTTCATGCCGATGAGCCACGTGATGGGCCGAAGCATCC
 TCTACGGCACGCTGGGCAACGGCGGCACCCGCCTACTTCGCCGCCCGCAGCGAC
 CTGTCCACCCTGCTCGAGGACCTCGAGCTGGTGC GGCCCACCGAGCTCAACTT
 CGTCCC GCGGATCTGGGAGACGCTGTACGGCGAATTCCAGCGTCAGGTCGAGC
 GCGGGCTCTCCGAGGCCGGGGACGCCGGCGAACGTCGCGCCGTTCGAGGCCGAG
 GTGCTGGCCGAGCAGCGCCAGTACCTGCTGGGCGGGCGGTTACCTTCGCGAT
 GACGGGCTCGGCGCCCATCTCGCCCGAGCTGCGCAACTGGGTTCGAGTCGCTGC
 TCGAAATGCACCTGATGGACGGCTACGGCTCCACGGAGGCCGGAATGGTGTG
 TTCGACGGGGAGATTTCAGCGCCCGCCGGTGGTTCGACTACAAGCTGGTCGACGT
 GCCGGACCTGGGCTACTTCAGCACCGACCGGCCGCATCCGCGCGGGCGAGCTGC
 TGCTGCGCACCGAGAACATGTTCCCGGGCTACTACAAGCGGGCCGAAACCACC
 GCGGGCGTCTTCGACGAGGACGGCTACTACCGCACCGGCCGACGTGTTCCGCCGA
 GATCGCCCCGGACCGGCTGGTCTACGTCGACCGCCGCAACAACGTGCTCAAGC
 TGGCGCAGGGCGAATTCGTCACGCTGGCCAAGCTGGAGGCGGTTGTTCCGGCAAC
 AGCCCGCTGATCCGCCAGATCTACGTCTACGGCAACAGCGCCAGCCCTACCT
 GCTGGCGGTCGTGGTGCCACCGAGGAGGCGCTGGCCTCGGGTGACCCCGAGA
 CGCTCAAGCCCAAGATCGCCGACTCGCTGCAGCAGGTCGCCAAGGAGGCCGGC
 CTGCAGTCTACGAGGTGCCGCGGACTTCATCATCGAGACCACCCCGTTTCAG
 CCTGGAAAACGGTCTGCTGACCGGGATCCGGAAGCTGGCGTGGCCGAAACTGA
 AGCAGCACTACGGGGAACGGCTGGAGCAGATGTACGCCGACCTGGCCGCCGGA
 CAGGCCGACGAGCTGGCCGAGCTGCGCCGCAACGGTGCCAGGGCGCCGGTGT
 GCAGACCGTGAGCCGCGCCGCGGGCGCCATGCTGGGTTCCGGCCGCTCCGACC
 TGTCCCCCGACGCCACTTCACCGATCTGGGCGGAGACTCGTTGTTCGGCGTTG
 ACATTCGGCAACCTGCTGCGCGAGATCTTCGACGTTCGACGTGCCGGTGGGCGT
 GATCGTCAGCCCGGCCAACGACCTGGCGGCCATCGCGAGCTACATCGAGGCCG
 AGCGGCAGGGCAGCAAGCGCCCGACGTTCCGCTCGGTGCACGGCCGGGACGCG
 ACCGTGGTGCGCGCCCGGACCTGACGCTGGACAAGTTCCTCGACGCCGACAC
 GCTGGCCTCCGCGCCGAACCTGCCAAGCCGGCCACCGAGGTGCGCACCGTGC
 TGCTGACCGGCGCCACCGGCTTCCTGGGCCGCTACCTGGCCCTGGAATGGCTG
 GAGCGGATGGACATGGTGGACGGCAAGGTCATCGCCCTGGTCCGGGCCCGCTC
 CGACGAGGAGGCACGCGCCCGGCTGGACAAGACCTTCGACAGCGGCGACCCGA
 AGCTGCTCGCGCACTACCAGCAGCTGGCTGCCGATCACCTGGAGGTCATCGCC
 GGCGACAAGGGCGAGGCCAATCTGGGCCTGCGCCAAGATGTTTGGCAACGGCT
 GGCCGACACGGTCGACGTGATCGTCGACCCCGCCGCGCTGGTCAACCACGTGT
 TGCCGTACAGCGAGCTGTTCCGGGCCAACGCCCTGGGCACCGCGGAGCTGATC

FIG 4 (Cont.)

CGGCTGGCGCTGACGTCCAAGCAGAAGCCGTACACCTACGTGTCCACCATCGG
 CGTGGGCGACCAGATCGAGCCGGGCAAGTTCGTGAGAACGCCGACATCCGGC
 AGATGAGCGCCACCCGGGCGATCAACGACAGCTACGCCAACGGCTACGGCAAC
 AGCAAGTGGGCCGGCGAGGTGCTGCTGCGCGAGGCGCACGACCTGTGCGGGCT
 GCCCGTCGCGGTGTTCCGCTGCGACATGATCCTGGCCGACACCACGTATGCCG
 GGCAGCTCAACCTGCCGGACATGTTACCCGGCTGATGCTGAGCCTGGTGGCC
 ACCGGGATCGCGCCCGGCTCGTTCACGAGCTCGACGCCGACGGCAACCGGCA
 GCGGGCGCACTACGACGGCCTGCCGGTCGAGTTCATCGCCGCGGCGATCTCGA
 CGCTGGGTTCGCAGATCACCGACAGCGACACCGGCTTCCAGACCTACCACGTG
 ATGAACCCCTACGATGACGGCATCGGTCTGGACGAGTACGTGATTGGCTGGT
 GGACGCCGGCTATTCGATCGAGCGGATCGCCGACTACTCCGAATGGCTGCGGC
 GGTTCGAGACCTCGCTGCGGGCCCTGCCGGACCGGCAGCGCCAGTACTCGCTG
 CTGCCGCTGCTGCACAACCTACCGCACGCCCGAGAAGCCGATCAACGGGTCGAT
 AGCTCCCACCGACGTGTTCCGGGCAGCGGTGCAGGAGGCGAAAATCGGCCCCG
 ACAAAGACATTCGCACGTGTCGCCGCGGTCATCGTCAAGTACATCACCGAC
 CTGCAGCTGCTCGGGCTGCTCTGA

Secuencia de aminoácidos (SEQ ID N0:30)

>uniprot|A0QIB5|A0QIB5_MYCA1 Supuesta acil-CoA deshidrogenasa

MSTATHDERLDRRVHELIAIDPQFAAAQDPAITAALEQPGLRLPQIIRTVLD
 GYADRPALGQRVVEFVTDKTRGRTSAQLLPRFETITYGEVAQRVSALGRALSD
 DAVHPGDRVCVLGFNSVDYATIDMALGAIGAVSVPLQTSAAISSLQPIVAETE
 PTLIASSVNQLSDAVQLITGAEQAPTRLVVFVDFHPQVDDQREAVQDAAARLSG
 TGVAVQTLAELLERKDLPAVAEPPADEDSLALLIYTSGSTGAPKGMYPQSN
 VGKMWRRGSKNWFGEASAASITLNFMPMSHVMGRSILYGLNGGTAYFAARSD
 LSTLLEDLELVRPTELNFVPRIWETLYGEFQRQVERRLSEAGDAGERAVEAE
 VLAEQRQYLLGGRFTFAMTGSAPISPELRNWVESLLEMHLMDGYGSTEAGMVL
 FDGEIQRPPVVDYKLVDPDLGYFSTDRPHPRGELLLRTENMFPGYKRAETT
 AGVFDEDGYRTGDVFAEIPDRLVYVDRRNNVLKLAQGEFVTLAKLEAVFGN
 SPLIRQIYVYGNSAQPYLLAVVVPTEEALASGDPETLKPKIADSLQQVAKEAG
 LQSYEVPRDFIIEETPFLENGLLTGIRKLAWPKLKQHYGERLEQMYADLAAG
 QADELAELRRNGAQAPVLQTVSRAAGAMLGSAASDLSPDAHFTDLGGDSLAL
 TFGNLLREIFDVDVPVGVIVSPANDLAAIASYIEAERQGSKRPTFASVHGRDA
 TVVRAADLTLDKFLDADTLASAPNLPKPATEVRTVLLTGATGFLGRYLALEWL
 ERMDMVDGKVIALVRARSDEEARARLDKTFDSGDPKLLAHYQQLAADHLEVIA
 GDKGEANLGLRQDVWQRLADTVDVIVDPAALVNHVLPYSELFGNALGTAELI
 RLALTSKQKPYTYVSTIGVGDQIEPGKFVENADIRQMSATRAINDSYANGYGN
 SKWAGEVLLREAHDLGLPVAVFRCDMILADTTYAGQLNLPDMFTRLMLSLVA

FIG 4 (Cont.)

TGIAPGSFYELDADGNRQRAHYDGLPVEFIAAAI STLGSQITDSDTGFQTYHV
MNPYDDGIGLDEYVDWLVDAGYSIERIADYSEWLRRFETSLRALPDRQRQYSL
LPLLHNYRTPEKPIINGSIAPTDFRAAVQEAKIGPDKDIPHVSPPVIVKYITD
LQLLGLL

YP 887275.1

Secuencia de nucleótidos (SEQ ID N0:31)

>uniprot|A0QWI7|A0QWI7_MYCS2 Proteína de la familia de epimerasa/deshidratasa dependiente de NAD

ATGACGATCGAAACGCGCGAAGACCGCTTCAACCGGCGCATTGACCA
CTTGTTCGAAACCGACCCGCAGTTCGCCGCCGCCCGTCCCGACGAGG
CGATCAGCGCGGCTGCCGCCGATCCGGAGTTGCGCCTTCCTGCCGCG
GTCAAACAGATTCTGGCCGGCTATGCGGACCGCCCTGCGCTGGGCAA
GCGCGCCGTTCGAGTTCGTACCGACGAAGAAGGCCGCACCACCGCGA
AGCTCCTGCCCCGCTTCGACACCATCACCTACCGTCAGCTCGCAGGC
CGGATCCAGGCCGTGACCAATGCCTGGCACAACCATCCGGTGAATGC
CGGTGACCGCGTGGCCATCCTGGGTTTCACCAGTGTCGACTACACGA
CGATCGACATCGCCCTGCTCGAACTCGGCGCCGTGTCCGTACCGCTG
CAGACCAGTGCGCCGGTGGCCAACTGCAGCCGATCGTCGCCGAGAC
CGAGCCCAAGGTGATCGCGTCGAGCGTCGACTTCCTCGCCGACGCAG
TCGCTCTCGTCGAGTCCGGGCCCGCGCCGTTCGCGACTGGTGGTGTTC
GACTACAGCCACGAGGTTCGACGATCAGCGTGAGGCGTTCGAGGCGGC
CAAGGGCAAGCTCGCAGGCACCGGCGTTCGTCGTCGAGACGATCACCG
ACGCACTGGACCGCGGGCGGTCACTCGCCGACGCACCGCTCTACGTG
CCCGACGAGGCCGACCCGCTGACCCTTCTCATCTACACCTCCGGCAG
CACCGGCACTCCCAAGGGCGCGATGTACCCCGAGTCCAAGACCGCCA
CGATGTGGCAGGCCGGGTCCAAGGCCCGGTGGGACGAGACCCCTCGGC
GTGATGCCGTCGATCACCCCTGAACTTCATGCCCATGAGTCACGTCAT
GGGGCGCGGCATCCTGTGCAGCACACTCGCCAGCGGCGGAACCGCGT
ACTTCGCCGCACGCAGCGACCTGTCCACCTTCCTGGAGGACCTCGCC
CTCGTGCGGCCACGCAGCTCAACTTCGTTCCCTCGCATCTGGGACAT
GCTGTTCCAGGAGTACCAGAGCCGCCCTCGACAACCGCCGCGCCGAGG
GATCCGAGGACCGAGCCGAAGCCGCAGTCTCGAAGAGGTCCGCACC
CAACTGCTCGGCGGGCGATTTCGTTTCGGCCCTGACCGGATCGGCTCC

FIG 4 (Cont.)

CATCTCGGCGGAGATGAAGAGCTGGGTTCGAGGACCTGCTCGACATGC
 ATCTGCTGGAGGGCTACGGCTCCACCGAGGCCGGCGCGGTGTTCATC
 GACGGGCAGATCCAGCGCCCGCGGTTCATCGACTACAAGCTGGTCGA
 CGTGCCCGATCTCGGCTACTTCGCCACGGACCGGCCCTACCCGCGCG
 GCGAACTTCTGGTCAAGTCCGAGCAGATGTTCCCCGGCTACTACAAG
 CGTCCGGAGATCACCGCCGAGATGTTTCGACGAGGACGGGTACTACCG
 CACCGGCGACATCGTCGCCGAGCTCGGGCCCGACCATCTCGAATAACC
 TCGACCGCCGCAACAACGTGCTGAAACTGTCGCAGGGCGAATTCGTC
 ACGGTCTCCAAGCTGGAGGCGGTGTTTCGGCGACAGCCCCCTGGTACG
 CCAGATCTACGTCTACGGCAACAGCGCGCGGTCTATCTGCTGGCGG
 TCGTGGTCCCGACCGAAGAGGCACTGTCACGTTGGGACGGTGACGAA
 CTCAAGTCGCGCATCAGCGACTCACTGCAGGACGCGGCACGAGCCGC
 CGGATTGCAGTCGTATGAGATCCCGCGTGACTTCCTCGTCGAGACAA
 CACCTTTCACGCTGGAGAACGGCCTGCTGACCGGTATCCGCAAGCTG
 GCCCGGCCGAAACTGAAGGCGCACTACGGCGAACGCCTCGAACAGCT
 CTACACCGACCTGGCCGAGGGGCAGGCCAACGAGTTGCGCGAGTTGC
 GCCGCAACGGAGCCGACCGGCCCGTGGTCGAGACCGTTCAGCCGCGCC
 GCGGTTCGCACTGCTCGGTGCCTCCGTACGGATCTGCGGTCCGATGC
 GCACTTCACCGATCTGGGTGGAGATTCGTTGTCGGCCTTGAGCTTCT
 CGAACCTGTTGCACGAGATCTTCGATGTCGACGTGCCGGTTCGGCGTC
 ATCGTCAGCCCGGCCACCGACCTGGCAGGCGTTCGCGGCCTACATCGA
 GGGCGAACTGCGCGGCTCCAAGCGCCCCACATACGCGTCGGTGCACG
 GGCGCGACGCCACCGAGGTGCGCGCGCGTGATCTCGCCCTGGGCAAG
 TTCATCGACGCCAAGACCCTGTCCGCCGCGCCGGGTCTGCCGCGTTC
 GGGCACCGAGATCCGCACCGTGCTGCTGACCGGCGCCACCGGGTTCC
 TGGGCCGCTATCTGGCGCTGGAATGGCTGGAGCGCATGGACCTGGTG
 GACGGCAAGGTGATCTGCCTGGTGCGCGCCCGCAGCGACGACGAGGC
 CCGGGCGCGTCTGGACGCCACGTTTCGACACCGGGGACGCGACACTGC
 TCGAGCACTACCGCGCGCTGGCAGCCGATCACCTCGAGGTGATCGCC
 GGTGACAAGGGCGAGGCCGATCTGGGTCTCGACCACGACACGTGGCA
 GCGACTGGCCGACACCGTCGATCTGATCGTCGATCCGGCCGCCCTGG
 TCAATCACGTCCTGCCGTACAGCCAGATGTTTCGGACCCAATGCGCTC
 GGCACCGCCGAACTCATCCGGATCGCGCTGACCACCACGATCAAGCC
 GTACGTGTACGTCTCGACGATCGGTGTGGGACAGGGCATCTCCCCCG
 AGGCGTTCGTCGAGGACGCCGACATCCGCGAGATCAGCGCGACGCGC
 CGGGTTCGACGACTCGTACGCCAACGGCTACGGCAACAGCAAGTGGGC

FIG 4 (Cont.)

CGGCGAGGTCCTGCTGCGGGAGGCGCACGACTGGTGTGGTCTGCCGG
TCTCGGTGTTCCGCTGCGACATGATCCTGGCCGACACGACCTACTCG
GGTCAGCTGAACCTGCCGGACATGTTACCCCGCCTGATGCTGAGCCT
CGTGGCGACCGGCATCGCGCCCGGTTTCGTTCTACGAACTCGATGCGG
ACGGCAACCGGCAGCGCGCCCACTACGACGGGCTGCCCGTGGAGTTC
ATCGCCGAGGCGATCTCCACCATCGGCTCGCAGGTCACCGACGGATT
CGAGACGTTCCACGTGATGAACCCGTACGACGACGGCATCGGCCTCG
ACGAGTACGTGGACTGGCTGATCGAGGCCGGCTACCCCGTGCACCGC
GTCGACGACTACGCCACCTGGCTGAGCCGGTTCGAAACCGCACTGCG
GGCCCTGCCGGAACGGCAACGTCAGGCCTCGCTGCTGCCGCTGCTGC
ACAAC TATCAGCAGCCCTCACCGCCCGTGTGCGGGTGCCATGGCACCC
ACCGACCGGTTCCGTGCCGCGGTGCAGGACGCGAAGATCGGCCCCGA
CAAGGACATTCCGCACGTACGGCCGACGTGATCGTCAAGTACATCA
GCAACCTGCAGATGCTCGGATTGCTGTAA

Secuencia de aminoácidos (SEQ ID N0:32)

>uniprot|A0QWI7|A0QWI7_MYCS2 Proteína de la familia de epimerasa/deshidratasa dependiente de NAD

MTIETREDRFNRRIDHLEFETDPQFAAARPDEAISAAAADPELRLPAAVKQILA
GYADRPALGKRAVEFVTDEEGRTTAKLLPRFDTI TYRQLAGRIQAVTNAWHNH
PVNAGDRVAILGFTSVDYTTIDIALLELGA VSVPLQTSAPVAQLQPIVAETEP
KVIASSVDFLADAVALVESGPAPSRLVVF DYSHEVDDQREAFEAAKGLAGTG
VVVETITDALDRGRSLADAPLYVPDEADPL TLLIYTSGSTGTPKGAMY PESKT
ATMWQAGSKARWDETLGVMP SITLNFMPMSHVMGRGILCSTLASGGTAYFAAR
SDLSTFLEDLALVRPTQLNFVPR IWDMLFQEQSRLDNRRRAEGSEDRAEA AVL
EEVRTQLLGGRFVSAL TGSAPISAEMKSWVEDLLDMHLL EGYGSTEAGAVFID
GQIQRPVIDYKLVDPDLGYFATDRPYPRGELLVKSEQMFP GYYKRPEITAE
MFDEDGYRRTGDIVAE LGPÐHLEYLDRRNNVLKLSQGEFVTVSKLEAVFGDSP
LVRQIYVYGNSARSYLLAVVVPTEEALSRWDGDELKSRI SDSLQDAARAAGLQ
SYEIPRDFLVETTPFTLENGLLTGIRKLARPKLKAHYGERLEQLYTDLAEGQA
NELRELRRNGADRPV VETVSRAAVALLGASVTDLRSDAHFTDLGGDSL SALSF
SNLLHEIFDVDVPVGVIVSPATDLAGVAAYIEGELRGSKRPTYASVHGRDATE
VRARDLALGKFIDAKTLSAAPGLPRSGTEIRTVLLTGATGFLGRYLALEWLER
MDLVDGKVICLVRRARS DDEARARLDATFDTGDATLLEHYRALAADHLEVIAGD
KGEADLGLDHD TWQRLADTVDLIVDPAALVNHVLPYSQMFGPNALGTAE LIRI
ALTTTIKPYVYVSTIGVGQGISPEAFVEDADIREISATRRVDDSYANGYGNSK
WAGEVLLREAHDW CGLPVSVFRCDMILADTTYSGOLNLPDMFTRLMLSLVATG

FIG 4 (Cont.)

IAPGSFYELDADGNRQRAHYDGLPVEFIAEAI STIGSQVTDGFETFHVMNPYD
DGIGLDEYVDWLI EAGYPVHRVDDYATWLSRFETALRALPERQRQASLLPLLH
NYQQPSPPVCGAMAPTDRFRAAVQDAKIGPDKDIPHV TADVIVKYI SNLQMLG
LL

ZP 05224908

Secuencia de nucleótidos (SEQ ID NO:33)

>gi | 163719654:2489-6013 Mycobacterium intracellulare ATCC 13950 NZ_ABIN01000072, secuenciación genómica completa al azar

ATGTCGACTGCCATTCATGACGAACACCTCGACCGTCGCATCGAGGAACTCAT
CGCCAACGACCCCAATTCGCCGCCGCCGACCGGACCCGGCCATCACCGCCG
CCACCGAAGCGCCCGGGCTGCGGCTGCCGCAGATCATCCGGACCGTGCTCGAC
GGCTACGCCGACCGGCCTGCCCTGGCGCAGCGCGTCGTGGAGTTCGTCACCGA
CGCCAAGACCGGGCGGACGACGGCCGAGCTGCTCCCCGTTTCGAGACCATCA
CGTATGGCGAÄCTCGGCGAACGGGTTTCGGCCCTCGGCCGTGCCTGGGCCGGC
GACGCGGTGCGCCCCGGCGACCGCGTCTGCGTGCTCGGCTTCAACAGCGTTGA
CTACGCCACCATCGACATCGCGCTGGGCACCATCGGGGCCGTGTCGGTGCCGC
TGCAGACCAGCGCGGCGATCTCCTCGTTGCAGCCGATCGTCGCCGAGACCGAG
CCCAGCCTGATCGCCTCGAGCGTCAACCAGCTGCCCGACGCGGGTGGAGCTGAT
CCTGGCCGGCGACCACGTGCCCGGCAAGCTCGTCGTGTTCTCGACTACCAGCCCC
AGGTCGACGACCAGCGCGAGGCCGTGGAGGCCGCCGCCGCGCGGTTGGCCGAC
TCCGGCGTCGCGGTCGAGGCTCTCGCCGACGTGCTGCGGCGCGGCAAGGACCT
GCCGGCCGTTCGAGCCGCCGGCGAGCGACGAGGACTCGCTGGCCCTGCTGATCT
ACACCTCCGGCAGCACCGGCGCGCCCAAGGGCGCGATGTACCCGCAGAGCAAC
GTCGGCAAGATGTGGCGGCGCGGGAGCAAGAACTGGTTCGGGGAAAGCGCCGC
GTCGATCACCTCAACTTCATGCCGATGAGCCACGTCATGGGGCGCGGAATCC
TCTACGGCACGCTGGGCAACGGCGGCACCGCGTACTTCGCCGCCCGCAGCGAC
CTGTCCACCCTGCTCGAGGACCTCGAGTTGGTGCGGCCACCGAGATGAACTT
CGTCCCCCGCATCTGGGAGACGCTGTACGGCGAATTCCAGCGCCAGGTCGAGC
GGCGGCTGGCCGACGGCGATGCGGGCCCGGAGGCCCGCGAGACTGTGGCGGCT
GCGGTGTTGGAAGAACAGCGCCAGTACCTGCTGGGCGGGCGGTTTCATCTTCGC
GATGACGGGCTCGGCACCCACCTCGCCGGAGCTCAAGGCGTGGGCCGAGTCGC
TCCTGCAGATGCACCTGATGGACGGCTACGGCTCCACCGAGGCCGGAATGGTG
TTGTTTCGACGGGGAGATTCAGCGTCCGCCGGTTATTGATTACAAGCTGGTTCGA
CGTTCCGGATCTGGGCTATTTACAGCACCGACCGTCCGCATCCGCGCGGTTGAGT
TGTTGCTGCGGACCGAGAACATGTTCCCGGGTTATTACAAGCGGGCCGAGACC
ACCGCGAACGTGTTTCGACGAGGACGGTTATTACCGCACCGGTGACGTGTTTCGC

FIG 4 (Cont.)

CGAGATCGCGCCGGACCGGCTGGTGTATGTCGATCGCCGCAACAACGTGCTCA
 AGTTGGCCCAGGGCGAGTTCGTGACGCTGGCCAAGCTGGAGGCGGTGTTCCGGC
 AACAGCCCCTGATCCGCCAGATCTACGTTTACGGCAACAGCTCCCAGCCCTA
 CCTGCTGGCCGTGGTGGTGCCGACCGAGGAAGCGTTGGCGGACAACGATCTTG
 AGTCGCTCAAGCCGAAGATCGCCGACTCGCTGCAGAAGGTCGCCAAGGAGACC
 GGCTGCAGTCCCTACGAGGTGCCGCGCGACTTCATCATCGAGACCACGCCGTT
 CACCCTGGAAAACGGCCTGCTGACCGGGATCCGCAAGCTGGCGTGGCCCAAGC
 TCAAGGCGCACTACGGGGATCGGCTCGAGCAGATGTATGCCGAGCTGGCCGCG
 GGACAGGCCAACGAGTTGGCCGAACTGCGCCGACGCGGCGGGCGGCCGGT
 GGCCAGACCCTGAGCCGGGCCGCGGCCGCCCTGCTGGGTGCGACGGCCGGGG
 ATCTGTCCGCAGATGCCCACTTCACCGATCTTGGTGGAGACTCGTTGTCGGCG
 TTGACCTTCGGCAACCTGCTGCGCGAGATCTTCGATGTCGACGTGCCGGTGGG
 GGTGATCGTCAGCCCGGCCAACGACCTGGCGGGGATCGCCGCCTACATCGAGG
 CCGAGCGGCAGGGCTCCAAGCGCCCAGCTTCGCCGCCGTGCACGGTCGCGGT
 GCGACCATGGTGCACGCCAGTGACCTCACGCTGGACAAGTTCCTCGACGAGGC
 GACCCTGGCCGCCGCGCCCAGCCTGCCCAAGCCGGCCACCGAGGTGCGCACCG
 TGCTGTTGACCGGCGCGACCGGCTTTTTGGGCCGCTACCTGGCGCTGGACTGG
 CTCGAGCGGATGGACATGGTCGACGGCAAGGTCATCGCCCTGGTGCGGGCCCG
 CACCGATGAGGAGGCGCGCGCCCGGCTGGACAAGACCTTCGACAGCGGCGACC
 CCAAACCTGCTGGCGCACTACCAGCGGCTGGCCGCCGACCACCTCGAGGTCATC
 GCCGGCGACAAGGGTGAGGCCAACCTCGGCCTGGACCCCCAGACCTGGCAGCG
 ACTGGCCGAGGAGGTGCACGTCATCGTCGACCCCGCCGCGCTGGTCAACCACG
 TGCTGCCCTACAGCGAGCTGTTCCGGCCCCAACGCCCTGGGCACCGCGGAGCTG
 ATCCGGATCGCGCTGACCTCCAGGCAAAAGCCCTACACCTACGTGTCGACGAT
 CGGGGTGGGCGATCAGATCCAGCCAGGTGAGTTCGTCGAGAACGCCGACATCC
 GCCAGATCAGCGCCACCCGCGAGATCAACGACGGCTACGCCAACGGCTACGGC
 AACAGCAAGTGGGCCGGCGAGGTGTTGCTGCGCGAGGCCACGACCTGTGCGG
 CCTGCCCGTACGGTGTTCGCTGCGACATGATCCTGGCCGACACCACCTATG
 CCGGGCAGCTCAACCTGCCCGACATGTTACCCCGGCTGATGCTGAGCCTGGTC
 GCCACCGGTATCGCGCCCGGGTCTTCTACGAACTGGACGCCGACGGCAACCG
 CCAGCGGGCACACTACGACGGTTTGCCGGTCGAGTTCATCGCCGCGGCGATCT
 CGACGCTGGGGACCCAAATCACCGACAGCGACACGGGCTTTCAGACCTACCAC
 GTGATGAACCCCTACGACGACGGCATCGGGCTGGATGAGTACATCGATTGGCT
 GATCGAGGCCGGGTATTCGATCGAGCGGATCGCCGATTACTCCGAGTGGCTGC
 GGCGCTTCGAGACCTCGCTGCGGGCCCTGCCCGATCGGCAGCGTCAGTACTCG
 CTGCTGCCGCTGCTGCACAACTACCAGAAGCCGGAAAAGCCGATCAACGGCTC
 GATGGCGCCCACCGACGTGTTCCGTGCCGCGGTGCAGGAAGCGAAAATCGGCC
 CCGACAAAGACATCCCGCACGTCTCGGCGCCGGTGATCGTCAAGTACATCACC
 GACCTGGAGTTGCTCGGACTCCTCTGA

FIG 4 (Cont.)

Secuencia de aminoácidos (SEQ ID N0:34)

>gi | 254819907 | ref | ZP_05224908.1 | FadD9 [Mycobacterium intracellulare ATCC 13950]
MSTAIHDEHLDRRIEELIANDPQFAAARPDPAITAATEAPGLRLPQI IRTVLD
GYADRPALAQRVVEFVTDAKTGRITTAELLPRFETITYGELGERVSALGRAWAG
DAVRPGDRVCVLGFNSVDYATIDIALGTIGAVSVPLQTSAAISSLQP IVAETE
PSLIASSVNQLPDAVELILAGDHVPGKLVVFDYQPQVDDQREAVEAAAARLAD
SGVAVEALADVLRGKDLPAVEPPASDEDSLALLIYTSGSTGAPKGMYPQSN
VGKMWRRGSKNWFGEAASITLNFMPMSHVMGRGILYGTGNGGTAYFAARSD
LSTLLEDLELVRPTEMNFVPRIWETLYGEFQRQVERRLADGDAGPEARETVAA
AVLEEQRQYLLGGRFIFAMTGSAPT SPELKAWAESLLQMHLMDGYGSTEAGMV
LFDGEIQRPPVIDYKLVDPDLGYFSTDRPHPRGELLRTENMFPGYYKRAET
TANVFDEDEGYRTGDVFAEIA PDRLVYVDRRNNVLKLAQGEFVTLAKLEAVFG
NSPLIRQIYVYGNSSQPYLLAVVVPTEEALADNDLES LKPKIADSLQVAKET
GLQSYEVPRDFI IETTPFTLENGLLTGIRKLAWPKLKAHYGDRLEQMYAELAA
GQANELAELRRSGAAAPVAQTVSRAAAALLGATAGDLSADAHFTDLGGDSL SA
LTFGNLLREIFDVDVPGVIVSPANDLAGIAAYIEAERQGSKRPTFAAVHGRG
ATMVHASDLTLDKFLDEATLAAAPSLPKPATEVRTVLLTGATGFLGRYLALDW
LERMDMVDGKVIALVRARTDEEARARLDKTFDSGDPKLLAHYQRLAADHLEVI
AGDKGEANLGLDPQTWQRLAEEVDVIVDPAALVNHVLPYSELFGPNALGTAEL
IRIALTSRQKPYTYVSTIGVGDQIQPGEFVENADIRQISATREINDGYANGYG
NSKWAGEVLLREAHDL CGLPVTVFRCDMILADTTYAGQLNLPDMFTRLMLSLV
ATGIAPGSFYELDADGNRQRAHYDGLPVEFIAAAI STLGTQITDSDTGFQTYH
VMNPYDDGIGLDEYIDWLEAGYSIERIADYSEWLRRFETSLRALPDRQRQYS
LLPLLHNYQKPEKPI NGSMAPTDFRAAVQEAKIGPDKDIPHVSAPVIVKYIT
DLELLGLL

YP 889972.1

Secuencia de nucleótidos (SEQ ID N0:35)

>uniprot | A0R484 | A0R484_MYCS2 Supuesta ácido graso de cadena larga--CoA ligasa
ATGACCAGCGATGTTACACGACGCCACAGACGGCGTCACCGAAACCGCACTCGA
CGACGAGCAGTCGACCCGCCGCATCGCCGAGCTGTACGCCACCGATCCCGAGT
TCGCCGCCGCCGACCGTTGCCCGCCGTGGTCGACGCGGCGCACAAACCCGGG
CTGCGGCTGGCAGAGATCCTGCAGACCCTGTTACCGGCTACGGTGACCGCCC

FIG 4 (Cont.)

GGCGCTGGGATAACCGCGCCCGTGAACCTGGCCACCGACGAGGGCGGGCGCACCCG
TGACGCGTCTGCTGCCGCGGTTTCGACACCCTCACCTACGCCAGGTGTGGTCCG
CGCGTGCAAGCGGTTCGCCGCGGCCCTGCGCCACAACCTTCGCGCAGCCGATCTA
CCCCGGCGACGCCGTCGCGACGATCGGTTTCGCGAGTCCCGATTACCTGACGC
TGGATCTCGTATGCGCCTACCTGGGCCTCGTGAGTGTTCGCTGCAGCACAAC
GCACCGGTCAGCCGGCTCGCCCCGATCCTGGCCGAGGTCGAACCGCGGATCCT
CACCGTGAGCGCCGAATACCTCGACCTCGCAGTCGAATCCGTGCGGGACGTCA
ACTCGGTGTCGCAGCTCGTGGTGTTCGACCATCACCCCGAGGTCGACGACCAC
CGCGACGCACTGGCCCCGCGCGCGTGAACAACCTCGCCGGCAAGGGCATCGCCGT
CACACCCTGGACGCGATCGCCGACGAGGGCGCCGGGCTGCCGGCCGAACCGA
TCTACACCGCCGACCATGATCAGCGCCTCGCGATGATCCTGTACACCTCGGGT
TCCACCGGCGCACCCAAGGGTTCGATGTACACCGAGGCGATGGTGGCGCGGCT
GTGGACCATGTTCGTTTCATCACGGGTGACCCACGCCGGTTCATCAACGTCAACT
TCATGCCGCTCAACCACCTGGGCGGGCGCATCCCCATTTCCACCGCCGTGCAG
AACGGTGAACAGTTACTTCGTACCGGAATCCGACATGTCCACGCTGTTTCGA
GGATCTCGCGCTGGTTCGCCCGACCGAACTCGGCCTGGTTCGCGCGTTCGCCG
ACATGCTCTACCAGCACCACTCGCCACCGTCGACCGCCTGGTTCACGCAGGGC
GCCGACGAACTGACCGCCGAGAAGCAGGCCGGTGCCGAACTGCGTGAGCAGGT
GCTCGGCGGACGCGTGATCACCGGATTCGTCAGCACCGCACCGCTGGCCGCGG
AGATGAGGGCGTTCCTCGACATCACCTGGGCGCACACATCGTCGACGGCTAC
GGGCTCACCGAGACCGGCGCCGTGACACGCGACGGTGTGATCGTGCGGCCACC
GGTGTGACTACAAGCTGATCGACGTTCCCGAACTCGGCTACTTCAGCACCG
ACAAGCCCTACCCGCGTGGCGAACTGCTGGTTCAGGTCGCAAACGCTGACTCCC
GGGTAATAACAAGCGCCCCGAGGTCACCGCGAGCGTCTTCGACCGGGACGGCTA
CTACCACACCGGCGACGTCATGGCCGAGACCGCACCCGACCACCTGGTGTACG
TGGACCGTCGCAACAACGTCCTCAAACCTCGCGCAGGGCGAGTTCGTGGCGGTC
GCCAACCTGGAGGCGGTGTTCTCCGGCGCGGCGCTGGTTCGCCAGATCTTCGT
GTACGGCAACAGCGAGCGCAGTTTCTTCTGGCCGTGGTGGTCCCGACGCCGG
AGGCGCTCGAGCAGTACGATCCGGCCGCGCTCAAGGCCGCGCTGGCCGACTCG
CTGCAGCGCACCGCACGCGACGCCGAACTGCAATCCTACGAGGTGCCGGCCGA
TTTCATCGTCGAGACCGAGCCGTTTCAGCGCCGCCAACGGGCTGCTGTCCGGTG
TCGGAAAACCTGCTGCGGCCAACCTCAAAGACCGCTACGGGCAGCGCCTGGAG
CAGATGTACGCCGATATCGCGGCCACGCAGGCCAACAGTTGCGCGAACTGCG
GCGCGCGGCCGCCACACAACCGGTGATCGACACCCTCACCCAGGCCGCTGCCA
CGATCCTCGGCACCGGGAGCGAGGTGGCATCCGACGCCACTTCACCGACCTG
GGCGGGGATTCCCTGTCCGGCGCTGACACTTTCGAACCTGCTGAGCGATTTCTT
CGGTTTCGAAGTTCCTCGTCCGGCACCATCGTGAACCCGGCCACCAACCTCGCCC
AACTCGCCCAGCACATCGAGGCGCAGCGCACCGCGGGTGACCGCAGGCCGAGT
TTCACCACCGTGCACGGCGCGGACGCCACCGAGATCCGGGCGAGTGAGCTGAC
CCTGGACAAGTTCATCGACGCCGAAACGCTCCGGGCCGACCGGGTCTGCCCA
AGGTCACCACCGAGCCACGGACGGTGTGCTCTCGGGCGCCAACGGCTGGCTG

FIG 4 (Cont.)

GGCCGGTTCCTCACGTTGCAGTGGCTGGAACGCCTGGCACCTGTCGGCGGCAC
 CCTCATCACGATCGTGCGGGGCCGCGACGACGCCGCGGCCCGCGCACGGCTGA
 CCCAGGCCTACGACACCGATCCCGAGTTGTCCCGCCGCTTCGCCGAGCTGGCC
 GACCGCCACCTGCGGGTGGTTCGCCGGTGACATCGGGCGACCCGAATCTGGGCCT
 CACACCCGAGATCTGGCACCGGCTCGCCGCCGAGGTCGACCTGGTGGTGCATC
 CGGCAGCGCTGGTCAACCACGTGCTCCCCTACCGGCAGCTGTTTCGGCCCCAAC
 GTCGTGGGCACGGCCGAGGTGATCAAGCTGGCCCTCACCGAACGGATCAAGCC
 CGTCACGTACCTGTCCACCGTGTCCGGTGGCCATGGGGATCCCCGACTTCGAGG
 AGGACGGCGACATCCGGACCGTGAGCCCGGTGCGCCCGCTCGACGGCGGATAC
 GCCAACGGCTACGGCAACAGCAAGTGGGCGGGCGAGGTGCTGCTGCGGGAGGC
 CCACGATCTGTGCGGGCTGCCCGTGGCGACGTTCCGCTCGGACATGATCCTGG
 CGCATCCGCGCTACCGCGGTTCAGGTCAACGTGCCAGACATGTTTCACGCGACTC
 CTGTTGAGCCTCTTGATCACCGGCGTCGCGCCCGGTCGTTCTACATCGGAGA
 CGGTGAGCGCCCGCGGGCGCACTACCCCGGCCTGACGGTCGATTTCTGTGGCCG
 AGGCGGTCACGACGCTCGGCGCGCAGCAGCGCGAGGGATACGTGTCCTACGAC
 GTGATGAACCCGCACGACGACGGGATCTCCCTGGATGTGTTTCGTGGACTGGCT
 GATCCGGGCGGGCCATCCGATCGACCGGGTTCGACGACTACGACGACTGGGTGC
 GTCGGTTCGAGACCGCGTTGACCGCGCTTCCCGAGAAGCGCCGCGCACAGACC
 GTACTGCCGCTGCTGCACGCGTTCCGCGCTCCGCAGGCACCGTTGCGCGGCGC
 ACCCGAACCCACGGAGGTGTTCCACGCCGCGGTGCGCACCGCGAAGGTGGGCC
 CGGGAGACATCCCGCACCTCGACGAGGCGCTGATCGACAAGTACATACGCGAT
 CTGCGTGAGTTCGGTCTGATCTGA

Secuencia de aminoácidos (SEQ ID N0:36)

>uniprot|A0R484|A0R484_MYCS2 Supuesta ácido graso de cadena larga--CoA ligasa

MTSDVHDATDGVTEETALDDEQSTRRIAELYATDPEFAAAAPLPAVVDAAH
 KPGLRLAEILQTLFTGYGDRPALGYRARELATDEGGRTVTRLLPRFDLTL
 YAQVWSRVQAVAAALRHNFAPPIYPGDAVATIGFASPDYLTLDLVCAYLG
 LVSVPLOHNAPVSRLLAPILAEVEPRIILTVSAEYLDLAVESVRDVNSVSQ
 LVFDHHPEVDDHRDALARAREQLAGKGIAVTTLDIAIADGAGLPAEPIYT
 ADHDQRLAMILYTSGSTGAPKGAMYTEAMVARLWTMSFITGDPTPVINVN
 FMPLNHLGGRIPISTAVQNGGTSYFVPESDMSTLFEDLALVRPTELGLVP
 RVADMLYQHHLATVDRLVTQGADELTAEKQAGAEQVLRGGRVITGFVS
 TAPLAAEMRAFLDITLGAHIVDGYGLTETGAVTRDGVIVRPPVIDYKLID
 VPGLGYFSTDKPYPRGELLVRSQTLTPGYKRPVETASVFDVDRDGYHTGD
 VMAETAPDHLVYVDRRNNVLKLAQGEFVAVANLEAVFSGAALVRQIFVYG
 NSERSFLLAVVVPTPEALEQYDPAALKAALADSLQRTARDAELQSYEVP
 ADFIVETEPFSAANGLLSGVGKLLRPNLKDQRYGQRLEQMYADIAATQANQL

FIG 4 (Cont.)

RELRRAAATQPVIDTLTQAAATILGTGSEVASDAHFTDLGGDSL SALTLS
 NLLSDFFGFEVPVGTIVNPATNLAQLAQHIEAQRTAGDRRPSFTTVHGAD
 ATEIRASELTLDKFIDAETLRAAPGLPKVTTEPRTVLLSGANGWLGRFLT
 LQWLERLAPVGGTLITIVRGRDDAAARARLTQAYDTPELSRRFAELADR
 HLRVVAGDIGDPNLGLTPEIWHRLAAEVDLVVHPAALVNHVLPYRQLFGP
 NVVGTAEVIKLALTERIKPVTYLSTVSVAMGIPDFEEDGDIRTVSPVRPL
 DGGYANGYGNSKWAGEVLLREAHDL CGLPVATFRSDMILAHPRYRGQVNV
 PDMFTRL LLSLLITGVAPRSFYIGDGERPRAHYPGLTVDFVAEAVTTLGA
 QQREGYVSYDVMNPHDDGISLDV FVDWLI RAGHPIDRVDDYDDWVRRFET
 ALTALPEKRRRAQTVLPLLHAFRAPQAPLRGAPEPTEVFHAAVRTAKVGP
 DIPHLDEALIDKYIRD LREFGLI

ZP_04751860

Secuencia de nucleótidos (SEQ 10 N0:37)

>gi|218125542:1370-4894 Mycobacterium kansasii ATCC 12478 NZ_ACBV01000156, secuenciación genómica completa al azar

Atgtcgactaccactcgtgacgagcgcctcgagcgcgcgcacacacctgat
 ccacgacgacgcacagttcgccgcccgaagccggaccocggcaatcgccgccc
 ccctggaaaagcccggcctgagcctgccagagatcatccagaccgcgctgcag
 ggctacgcccgaaccggccggtctcgggcagcgcgcgcctcgagttcgtcácgga
 caccagaccggccgcacctcggtgcccgtgctgacccgcttcgagaccatca
 cctaccgtcagctcgggcaccgggtcggtgcccgtggcgcgcgcacctgaccac
 gactcggtgcacgcggcgaccgggtctgcccgtgctgggcttcaacagcctcga
 ctacaccaccatcgacatggcgcctggcgaaggctcgccgcccgtgtcggtgcccgc
 tgcagaccagtgcggcgggttaccagctgcagccgatcgtggccgagaccgag
 cccaccatgatggcggccagtgtgaatcagctgtcggatgccgtggacgtgct
 gttgagcggccacctgccggccaagctgggtgggtgttcgactaccaccccagag
 tggacgaccagcgcgagggcgtcgacacggcccgggagcgggttggcggacaca
 gcggtgggtgtccagaccctcaaggacgtcctggaccacggcgcacaacgctggc
 ggccggggtcggttgcccgagccgctcgccggcgtcgggcgacaacgactcgctgg
 cgctgctgatctacacctccggcagcaccggcgcacccaaagggtgcgatgtat
 cgacagagcaacgctcggcaagatgtggcgcgggtcgagcaagaactggttcgg
 gccaccgcggcgtcgatcacctcaacttcatgccgatgagccacatcatgg
 gacgtggagtccctacggcacgctcggcaacggcggcaccgcgtacttcgcg
 gccaggagcgacctctcgacgctgctggaagacttgcggctgggtgcggcccac
 cgagttgaacttcgtgccgcccgatctggggagaccttgtacggcgaataaccagc
 gcgcgggtcgaccagcggtcggtcgatcccgggtgaaccgcggcccgcgaagcc

FIG 4 (Cont.)

gt cgaagcccaggtcatggccgagcagcgcaccaggacctgctgggtgggcgcta
 tatcttcgcgatgacgggctcggcgcccatgtcccggagctgcggaactggg
 tggaaagcgtgctggagatcccactgctcgacgggttacggctccaccgaggcc
 gggatggatcatgttcgacgggtgaaattcagcgcgcccggtgatcgactaaa
 gctggctcgacgtgcccgatctgggctacttcagcaccgaccagccgtatccgc
 gggcgagttgctgctcaagaccgagaacatgttcccaggctattacaagcgg
 ccggaggttacggccagcgtgttcgacgcggacggctactaccggaccggaga
 cgtcgtggccgaggtcgtccggaccggctgggtgtacgtggaccgccgcaaca
 acgtgctgaagcttgcccaggcgagttcgtgaccgtcgccaagctggaggcg
 gtgttcggcaacagcccgtgggtgcgccagatctacgtctacggcaacagcgc
 gcatccctacctgttgcccggtgggtgggtgccaacggaagaagcgtcggcgggca
 ctgacatagcggccttgaagccgctgatcgccgactcgtgcagaccgtcgcc
 aaagaggccggcctgcagtcctacgaggtgcccgcgcgacttctcatcgagac
 aacaccattcacgctggagaacgggtctgctcaccgggtatccgcaagctggcgt
 ggccgaagctgaggcagcattacggcgaacggctggagcagctctacacggag
 ctggccgcgagccaggccaacgagttgagcagcgtgcggcgcagcggggcca
 tgcgccgggtgctggaaacgggtgagccgggcccggcggcgcgctgctgggggagg
 cgagcaccgccttgtcaccgcagcgcacttcaccgatctgggtggagactcg
 ttgtcggcgttgacattcggcaacctgctgcgggagatcttcgacgcttgacgt
 accggtgggtgtgatcgtcagcccggccagcgcacctggcggcgatcgccgctt
 acatcgaggggagcggcaggccagcaagcgcaccacgttcgccgtgattcac
 ggtcgggacgcgctagaggtgcatgcgagtgacctcaccctggacaagttcat
 cgacgcacccctggcagccgcaccggctactgcgcctccgagcgcgcgg
 tgcgcaccgtcctgttgaccggcgcgaccggctttttgggcccgtacctggcg
 ctggactggctggagcgcacggacctggctcgacggcaaggtgatcgccctgg
 gcgggcaagtcggacgacgatgcccgggacgcctggacaagacgttcgaca
 gcggagaccccagcgtgctgaccactaccggcggctggcgaccgaccacttg
 gaggtcatcgccggcgacaagggcgaggccaacctcgggctggatcagctgac
 ctggcagcggctggccgacaccgtcgacctgatcgtcgaccggcccgcttg
 tcaaccacgtgctgcctacagcagcgtgttcggccccaacgcgctgggcacc
 gccgaattgatccggatagcgtgaccggcaagctcaagccctacacctacgt
 ctcgaccatcgggggtgggcgaccagatcgagccgggcaagttcaccgaggacg
 ccgacatccggcacatcagcgcgaccgggaagatcaacgacagttacgccaac
 ggctacggcaacagcaagtgggccggcgaggtgctactgcgcgagggcccacga
 cctgtgcgggctgcccgtcgcgggtgttccgctgcgacatgatcctggccgaca
 ccacgtgggcccgggtcagctcaacgtgcccggacatgttcaccggatgatgctg
 agcctgggtggccaccggcattgcaccgggttcgcttctacgagctggacgcga
 cggcaaccggcagcgtgcccactacgacggcttgcccggctcgagttcatcgccg
 aggcgatcgcgacgctgggcgcccgggacgggaagggttccagacctaccac
 gtgatgaaccctacgacgacggcatcgggatggaccgggttcgctcgactggct
 cgtcgacgcggatgcgccatccaccgcacgcgactacggcgactggctgc

FIG 4 (Cont.)

gacgattcgagaccgcgctgcgcgccctgcccgaaaagcagcgtcacgcgtca
 ctactgccgttgctgcacaactaccagaagccggcgcccgctgcgcggggtc
 gatggctccgaccgaccggttccggggcgccgtgcaggacgcgaaagtgggccc
 cggacaaggacatcccgcacatctcgccgcagatcatcgcgaaagtacctcagt
 gatctgcgcttgctcgggctcctctga

Secuencia de aminoácidos (SEQ ID N0:38)

>gi | 240173202 | ref | ZP_04751860.1 | Fad09 [Mycobacterium kansasii ATCC 12478]

MSTTTRDERLERRIDTLIHDDAQFAAAKPDPAIAAALEKPGLSLPEIIQTALQ
 GYADRPALGQRAVEFVTDQTGRTSVRLLTRFETITYRQLGDRVGALARALTH
 DSVHAGDRVCVLGFNSLDYTTIDMALAKVGAVSVPLQTSAAVTQLQPIVAETE
 PTMMAASVNQLSDAVDVLLSGHLPKLVVFDYHPEVDDQREALDTARERLADT
 AVVVQTLKDVLDHGATLAAGSVAEPLAASGDNDLALLIYTSGSTGAPKGYMY
 RQSNVGMWRSSSKNWFPTAASITLNFMPMSHIMGRGVLYGTGNGGTAYFA
 ARSDLSTLLEDLRLVLRPTELNFVPRIWETLYGEYQRAVDQRSVDPGEPAREA
 VEAQVMAEQRQDLLGGRYIFAMTGSAPMSPELRNWVEALLEIPLLDGYGSTE
 GMVMFDGEIQRPVIDYKLVDPDLGYFSTDQPYPRGELLLKTENMFPGYKR
 PEVTASVFDADGYRTGDVVAEVAPDRLVYVDRRNNVLKLAQGEFVTVAKLEA
 VFGNSPLVRQIYVYGNSAHPYLLAVVVPTEEASAGTDIAALKPLIADSLQTV
 KEAGLQSYEVPRDFLIETTPFTLENGLLTGIRKLAWPKLRQHYGERLEQLYTE
 LAASQANSELRRSGAHAPVLETVSRAAGALLGAASTALSPDAHFTDLGGDS
 LSALTFGNLLREIFDVDVPVGVIVSPASDLAAIAAYIEGERQGSKRPTFAVIH
 GRDALEVHASDLTLDKFIDASTLAAAPVLPPPSAAVRTVLLTGATGFLGRYLA
 LDWLERMDLVDGKVIALVRAKSDDARARLDKTFDSGDPELLTHYRRLATDHL
 EVIAGDKGEANLGLDQLTWQRLADTVDLIVDPAALVNHVLPYSELFPGNALGT
 AELIRIALTGKLPYTYVSTIGVGDQIEPGKFTEDADIRHISATRKINDSYAN
 GYGNSKWAGEVLLREAHDLGLPVAVFRCDMILADTTWAGQLNVPDMFTRMML
 SLVATGIAPGSFYELDADGNRQRAHYDGLPVEFIAEAIATLGARDGKGFQTYH
 VMNPYDDGIGMDRFVDWLVDAGCAIHRIDDYGDWLRRFETALRGLPEKQRHAS
 LLPLLHNYQKPAPPLRGSMAPTDRFRAAVQDAKVGPKDIPHISPQIIAKYLS
 DLRLGLL

FIG 4 (Cont.)

YP 978699.1**Secuencia de nucleótidos (SEQ ID NO:39)**

>uniprot|A1KLT8|A1KLT8_MYCBP Probable ácido graso-CoA ligasa fadD9

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTCGAGGACCTATACGCCAG
 CGACGCCCAGTTCGCCGCCGCCAGTCCCAACGAGGGCGATCACCCAGGGCGATCG
 ACCAGCCCCGGGGTCGCGCTTCCACAGCTCATCCGTATGGTCATGGAGGGCTAC
 GCCGATCGGCCGGCACTCGGCCAGCGTGCCTCCGCTTCGTCACCGACCCCGA
 CAGCGGCCGCACCATGGTCGAGCTACTGCCGCGGTTTCGAGACCATCACCTACC
 GCGAACTGTGGGCCCGCGCCGGCACATTGGCCACCGCGTTGAGCGCTGAGCCC
 GCGATCCGGCCGGGCGACCGGGTTTGCCTGCTGGGCTTCAACAGCGTCGACTA
 CACAACCATCGACATCGCGCTGATCCGGTTGGGCGCCGTGTCCGTTCCACTGC
 AGACCAGTGCGCCGGTCCACCGGGTTGCGCCCGATCGTCACCGAGACCGAGCCG
 ACGATGATCGCCACCAGCATCGACAATCTTGGCGACGCCGTGGAAGTGCTGGC
 CGGTCACGCCCCGGCCCGGCTGGTTCGATTACCGGCAAGGTTGACA
 CCCACCGCGAGGCCGTGGAAGCCGCCCGAGCTCGGTTGGCCGGCTCGGTTGACC
 ATCGACACACTTGCCGAACTGATCGAACGCGGCAGGGCGCTGCCGGCCACACC
 CATTGCCGACAGCGCCGACGACGCGCTGGCGCTGCTGATTTACACCTCGGGTA
 GTACCGGCGCACCCAAAGGCGCCATGTATCGCGAGAGCCAGGTGATGAGCTTC
 TGGCGCAAGTCGAGTGGCTGGTTCGAGCCGAGCGGTTACCCCTCGATCACGCT
 GAACTTCATGCCGATGAGCCACGTCGGGGGCCGTCAGGTGCTCTACGGGACGC
 TTTCCAACGGCGGTACCGCCTACTACGTCGCCAAGAGCGACCTGTCGACGCTG
 TTCGAGGACCTCGCCCTGGTGC GGCCACAGAATTGTGCTTCGTGCCGCGCAT
 CTGGGACATGGTGTTCGCAGAGTTCCACAGCGAGGTCGACCGCCGCTTGGTGG
 ACGGCGCCGATCGAGCGGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAG
 AACGTGCTCGGCGGACGGTTTGTTCATGGCGCTGACCGGTTCCGCGCCGATCTC
 CGCTGAGATGACGGCGTGGGTCGAGTCCCTGCTGGCCGACGTGCATTTGGTGG
 AGGGTTACGGCTCCACCGAGGCCGGGATGGTCTGAACGACGGCATGGTGC GG
 CGCCCCGCGGTTGATCGACTACAAGCTGGTTCGACGTGCCCGAGCTGGGCTACTT
 CGGCACCGATCAGCCCTACCCCGGGGCGAGCTGCTGGTCAAGACGCAAACCA
 TGTTCCCCGGCTACTACCAGCGCCCGGATGTCACCGCCGAGGTGTTTCGACCCC
 GACGGCTTCTACCGGACCGGGGACATCATGGCCAAAGTAGGCCCCCGACCAGTT
 CGTCTACCTCGACCGCCGCAACAACGTGCTAAAGCTCTCCAGGGCGAGTTCA
 TCGCCGTGTCGAAGCTCGAGGCGGTGTTTCGGCGACAGCCCGCTGGTCCGACAG
 ATCTTCATCTACGGCAACAGTGCCCGGGCCTACCCGCTGGCGGTGGTTGTCCC
 GTCCGGGGACGCGCTTTCTCGCCATGGCATCGAGAATCTCAAGCCCGTGATCA
 GCGAGTCCCTGCAGGAGGTAGCGAGGGCGGCCGGCCTGCAATCCTACGAGATT
 CCACGCGACTTCATCATCGAAACCACGCCGTTACCCTGGAGAACGGCCTGCT

FIG 4 (Cont.)

CACCGGCATCCGCAAGCTGGCACGCCCGCAGTTGAAGAAGTTCTATGGCGAAC
 GTCTCGAGCGGCTCTATACCGAGCTGGCCGATAGCCAATCCAACGAGCTGCGC
 GAGCTGCGGCAAAGCGGTCCCGATGCGCCGGTGCTTCCGACGCTGTGCCGTGC
 CGCGGCTGCGTTGCTGGGCTCTACCGCTGCGGATGTGCGGGCCGGACGCGCACT
 TCGCCGACCTGGGTGGTGA CTCTCGGCGCTGTCGTTGGCCAACCTGCTG
 CACGAGATCTTCGGCGTCGACGTGCCGGTGGGTGTCATTGTCAGCCCGGCAAG
 CGACCTGCGGGGCCCTGGCCGACCACATCGAAGCAGCGCGCACCGGCGTCAGGC
 GACCCAGCTTCGCCTCGATACCGGTGCTCCGCGACGGAAGTGCACGCCAGC
 GACCTCACGCTGGACAAGTTCATCGACGCTGCCACCCTGGCCGCAGCCCCGAA
 CCTGCCGGCACCGAGCGCCCAAGTGCGCACCGTACTGCTGACCGGCGCCACCG
 GCTTTTTGGGTGCTACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTC
 AACGGCAAGCTGATCTGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAAGC
 CCGGCTGGACGCGACGTTTCGATAGCGGCGACCCGTATTTGGTGCGGCACTACC
 GCGAATTGGGCGCCGGCCGCTCGAGGTGCTCGCCGGCGACAAGGGCGAGGCC
 GACCTGGGCTGGACCGGGTCACCTGGCAGCGGCTAGCCGACACGGTGGACCT
 GATCGTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGT
 TCGGCCCAAACGCGGCGGGCACCGCCGAGTTGCTTCGGCTGGCGCTGACCGGC
 AAGCGCAAGCCATACATCTACACCTCGACGATCGCCGTGGGCGAGCAGATCCC
 GCCGGAGGCGTTCACCGAGGACGCCGACATCCGGGCCATCAGCCCGACCCGCA
 GGATCGACGACAGCTACGCCAACGGCTACGCGAACAGCAAGTGGGCGGGCGAG
 GTGCTGCTGCGCGAAGCTCACGAGCAGTGCGGCCTGCCGGTGACGGTCTTCCG
 CTGCGACATGATCCTGGCCGACACCAGCTATAACCGGTCAGCTCAACCTGCCGG
 ACATGTTACCCGGCTGATGCTGAGCCTGGCCGCTACCGGCATCGCACCCGGT
 TCGTTCTATGAGCTGGATGCGCACGGCAATCGGCAACGCGCCACTATGACGG
 CTTGCCGGTTCGAATTCGTCGCAGAAGCCATTTGCACCCTTGGGACACATAGCC
 CGGACCGTTTTGTACCTACCACGTGATGAACCCCTACGACGACGGCATCGGG
 CTGGACGAGTTCGTCGACTGGCTCAACTCCCCAACTAGCGGGTCCGGTTGCAC
 GATCCAGCGGATCGCCGACTACGGCGAGTGGCTGCAGCGGTTTCGAGACTTCGC
 TCGTGCCTTGCCGGATCGCCAGCGCCACGCCTCGCTGCTGCCCTTGCTGCAC
 AACTACCGAGAGCCTGCAAAGCCGATATGCGGGTCAATCGCGCCCACCGACCA
 GTTCCGCGCTGCCGTCCAAGAAGCGAAAATCGGTCCGGACAAAGACATTCCGC
 ACCTCACGGCGGCGATCATCGCGAAGTACATCAGCAACCTGCGACTGCTCGGG
 CTGCTGTGA

Secuencia de aminoácidos (SEQ ID N0:40)

>uniprot|A1KLT8|A1KLT8_MYCBP Probable ácido graso-CoA ligasa fadD9

MSINDQRLTRRVEDLYASDAQFAAASPNEAITQ AIDQPGVALPQLIRMVM
 EGYADRPALGQRALRFVTDPSGR TMVELLPRFETITYRELWARAGTLAT

FIG 4 (Cont.)

ALSAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGL
 RPIVTETEPTMIATSIDNLGDAVEVLGHPARLVVFDYHGKVDTHREAV
 EAARARLAGSVTIDTLAELIERGRALPATPIADSADDALALLIYTSGSTG
 APKGAMYRESQVMSFWRKSSGWFEPSGYPSITLNFMPMSHVGGRRQVLYGT
 LSNGGTAYYVAKSDLSTLFEDLALVRPTELCFVPRIWDMVFAEFHSEVDR
 RLVDGADRAALEAQVKAELRENVLGGRFVMALTGSAPI SAEMTAWVESLL
 ADVHLVEGYGSTEAGMVLNDGMVRRPAVIDYKLVDVPELGYFGTDQPYPR
 GELLVKTQTMFPGYYQRPDVTAEVFDPDGFYRTGDIMAKVGPDQFVYLDL
 RNNVLKLSQGEFIAVSKLEAVFGDSPLVRQIFIYGN SARAYPLAVVPSG
 DALSRHGIENLKPVI SE SLQEVARAAGLQSYEIPRDFI IETTPFTLENGL
 LTGIRKLARPQLKKFYGERLERLYTE LADSQSNE LREL RQSGPDAPVLPT
 LCRAAAALLGSTAADVRPDAHFADLGGDSL SALS LANLLHEIFGVDVPVG
 VIVSPASDLRALADHIEAARTGVR RPSFASI HGRSATEVHASDLTLDKFI
 DAATLAAAPNLPAPSAQVRTVLLTGATGFLGRYLALEWLD RMDLVNGKLI
 CLVRARSDEEAQARLDATFDSGDPYLVRHYRELGAGRLEVLAGDKGEADL
 GLDRVTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNAAGTAE LLRLALT
 GKRKPYIYTSTIAVGEQIPPEAFTE DADIRAI SPTRRIDDSYANGYANSK
 WAGEVLLREAHEQCGLPVTVFRCDMILADTSYTGQLNLPDMFTRLMLSLA
 ATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPDRFVTYH
 VMNPYDDGIGLDEFVDWLN SPTSGSGCTIQRIADYGEWLQRFETSLRALP
 DRQRHASLLPLLHNYREPAKPICGSIAPTDQFRAAVQEAKIGPDKDIPHL
 TAAIIAKYISNLRLLLGLL

ZP 05227804

Secuencia de nucleótidos (SEQ ID NO:41)

>gi | 163719878:4480-7968 Mycobacterium intracellulare ATCC 13950 NZ_ABIN01000296, secuenciación genómica completa al azar

TTGGCCGCGACCGACGAACAGTTC CGCAACGCCCAACCCGATCTGTGCTGCA
 GCAGGCGGCGCGCCAACCGGGCCTGCGCCTGCCCCAGATTCTGGAGCTGTTCG
 TCGAGGGCTACGCCGATCGGCCCGCGGTGGGCTGGCGGGCCAGGACACTGAGC
 ACCGACCCCGCGACCGGGCGTACCACCACGCGACTGCTCCCCGCTTCGACAC
 CATGACCTACCGCGAGCTGTGGGCCGATGTGCGCGCGATCGCCGCCGCGTGGC
 GGCACGACGCCGCGAACCCCGTGTGCCCCGGGGACTTCGTGCGGACGGTCCGGT
 TTCGCCAGCGCCGAATATCTGACCCTCGACCTGGTCTGCGGCTACCTCGGGCT
 GGTGCGCTTCCGCTGCAACACAATACGACCCCTCACGGCTGCGACCGATCG
 TCGACGAGGTCGAGCCGTCGATACTGGCCGCCGGCGTCCGTTATCTCGACCTC
 CGGTCGAGGCGGCGTCCGGCAGCTCGTCGTTGCGGCGGCTCGTGGTCTTCGA

FIG 4 (Cont.)

TTACCAGCCCGAGGTCGACGAGCAACGCGAGGCGCTGCAGCGCGCCAGGCGA
 CGCTGGCGGCGCCGGCGCGGCGGTGACCATCGAGACCCTCGACGAAATCATC
 GAACGCGGACGCGCCCTGCCGCCGAGCCCATGTACACCGGTGATACCGATCA
 GCGGCTGGCGATGATCATGTACACCTCGGGCAGCACCGGGTTACCCAAGGGCG
 CCATGTACACCGAGCAGATGCTGGCCAAGGTGTGGACCAACGAGCTGATGCC
 GACTTCGCGGACACACCCGTGTTCAACGTCAACTTCATGCCGCTGAATCACCT
 CGGTGGCCGGATAACCGCTGTCGACCGCGTTCCAGGCCGGCGGCACCAGCTATT
 TCGTGCCGGAAAGCGACCTGTCCACGTTGTTTCGACGACTGGAACCTGGTGC
 CCCACCGAGATGGGCTGGTACCCCGGGTGGCGGAAATGCTCTACCAGCGCTA
 CCAGAGCGCCGTCGACCGACTCGTGGCTTCGGGCGCCGACGCCGGCTCCGCCG
 AGGCCCGGGCGCGGGCCGAGCTGCGTGAGCATGTCCTCGGCGGGCGCATCGTG
 ACCGCCCTTTCGCGGGACGGCGCCGCTGGCCGCGGAGATGCGGGCCTTCGT
 CGAACCTGTTTGGACGTCCACGTTCTCGACGGCTACGGGCTGACCGAGGT
 CGGCA TGGTGACCAAGGACGGGCGCATGACCCGTCCCCGGTGCTCGACTACAAGCTC
 ATCGACGTTCCCGAACTCGGCTATTTCCACACCGACAAGCCTTATCCGCGTGG
 CGAATTGCTGGTGAAGTCGCTGACCGCGACGCCGGGCTACTTCAAACGACCGG
 ACGTCACCGCCAACGCGTTCGATCCCGACGGCTACTACCGGACCGGCGATGTG
 ATGGCCGAGCTCGAGCCGGACCGGCTGGCCTACGTCGACCGCCGCAACAACGT
 GTTGAAGTTGGCGCAGGGCGAGTTCGTGCGCCGTCGCCCGCCTGGAGGCCGTCT
 TCGCCAGCGCGCCGCTGATCCGCCAGATCTTCGTATACGGCAACAGCGAACGC
 CCTATCTGCTGGCCGTCGTGTCGCGACGGCCGACGCCGCGGAGCGATTAC
 CGGAGATCCCGAGGGCCTCAAGGCCGCCGTCGCCGAATCCCTGCGCCAGTCGG
 CGCAACTCGCCGAACTGCAGTCTACGAGGTGCCCGTCGACTTCGTGCTCGAG
 ACCGAGCCGTTTCAGCGAGGACAACGGCCTGCTCTCGGGCGTGGGCAAGCTGCT
 GCGGCCGAAGCTCAAGGAGCGCTACGCCGACCGGCTCGAACAGCTCTACGCCG
 AGCTGGCCGAAAACCGCGTGACCGAGTTGCGTGCGCTGCGCGAGGGGGCGGAC
 AAACACCCCGTCTGTTTACCCTCACCCGGGCGCCGAGGCGCTACTGGGTGT
 GGCCGGCGGCCCCGCCGCCCGACGCACTGTTTCATCGAACTCGGCGGCGATT
 CCTGTGCGGCGCTGACCTTCTCCAACCTGCTGCGCGACATCTTCGACGTCGAC
 GTGCCGGTGGGAATGATCACCGGGCCCGCGACCGACCTGGGCCAGCTCGCGGA
 ATACGTTGAATCCGAACGCAAATCGGGATCACGCCGGCCACATTCGCGACGG
 TGCACGGACGCGGGCGCCGCCGAGGTCCGCGCCGCGGAGCTACCCTCGACAAG
 TTCATCGACGCGACGACCTTGGCCGCCGACCGAACCTGCCGCGCGGACCGG
 CACACCCACACGGTCTGCTGACCGGCGCCAACGGCTACCTCGGCCGCTTCC
 TGGCCCTCGAATGGCTCGAGCGCCTCGCCGAGACCGGCGGGAAGCTCGTCTCC
 ATCGTCCGCGCGACGGACACCGCGGCGGCCGTCAAACGGCTGGAGGCCGTTTT
 CGACAGCGGGGATCCGCAGTTGCTGGAGCGGTTCCGGACGCTGGCCGCCGAGC
 ACCTGGAAGTCATCGTCGGCGACATCGGTGAGCCCAATCTCGGCCTGGACCAA
 GCGACTTGGCAGCGCCTGGCCAGAGCGTGGATCTGATCGTCCACCCGGCCGC
 GTTGGTCAACCACGTGCTGCCGTACGACCAACTGTTCCGGTCCGAACGTCGTG
 GCACCGCCGAGTTGATCCGCCTGGCGATCACGACGCGCATCAAGCCCGTCACC

FIG 4 (Cont.)

TATCTGTCGACCGTTCGCCGTGGCGATGACGGTTCGATCCCGGCGAGTTCGCCGA
 AGACGGCGACATCCGCGCGGTTCAGCGCGGTACGCCCGATCGACGACAGCTACG
 CGAACGGGTACCGAACAGCAAGTGGGCCGGTGAGGTGTTGCTGCGTGAGGCG
 CACGACCTGTGCGGGTTGCCGGTTCGCCGTCTTCCGCTCCGACATGATCCTCGC
 GCACAGCCGGTATGCCGGGCAGCTGAACGTGCCAGATGCCTTCACCCGCTTGA
 TGTTACGCCTGCTGACCACCGGCATCGCGCCGACCACGTTCTACCGGACCGAC
 GAACACGGAAACCGAGCCGTGGCCCACTACGACGGGCTGCCCGCCGACTTCGT
 GGCCGAAGCGGTACCACGCTCGGCGAACAGATGGCGGCCGAGGAATCCGGCG
 GGTACCGCTCCTATGACGTGATGAACCCACACGACGACGGCGTCTCCCTGGAC
 GTGTTTCGTCGACTGGCTGATCGCCGCCGGACACGACATCCGGCGCATCGAGGA
 CTATGACGAATGGCTGGGCCGCTTACCACGGCGCTTCGCGCGTTACCGGACA
 AGCAGCGCCAGCATTCCGGTGTGCGCGTGTGGACGCCTACCGGGAACCCGCG
 ACGCCACTGCGGGGAGCGCCGGCCCCACCGACGTCTTCCGCCACGCGGTGCG
 GACGGCCAAAATCGGTGCGGACGAGGACATTCCGCACCTGTCGGCGGGCGTTGA
 TCGACAAGTACGTCGCCGACCTACGCCTGCTGGGCTTGGTGTAG

Secuencia de aminoácidos (SEQ ID N0:42)

>gi |254822803|ref|ZP_05227804.1| supuesta ácido graso de cadena larga--CoA ligasa [Mycobacterium intracellulare ATCC 13950]

MAATDEQFRNAQPDL SLQQAARQPGLRLPQILELFVEGYADRPVAVGWRARTLS
 TDPATGRITRLLPRFDTMTYRELWADVRAIAAWRHDAANPVSPGDFVATVG
 FASAEYLTLDLVCGYLGLVAVPLQHNTTPSRLRPVDEVEPSILAAGVGYLDL
 AVEAASGSSSLRRLVVFQPEVDEQREALQRAQATLAAAGAAVTIETLDEII
 ERGRALPPEPMYTGDTDQRLAMIMYTSGSTGLPKGAMYTEQMLAKVWTNELMP
 DFADTPVFNVNFMPLNHLGGRIPLSTAFQAGGTSYFVPESDLSTLFDDWNLVR
 PTEMGLVPRVAEMLYQRYQSAVDRLVASGADAGSAEARARAE LREHVLGGRIV
 TAFCGTAPLAAEMRAFVETCLDVHVLVDGYGLTEVGMVTKDGRMTRPPVLDYKL
 IDVPELGYFHDKPYPRGELLVKSLTATPGYFKRPDVTANAFDPDGYRTGDV
 MAELEPDRLAYVDRRNNVLKLAQGEFVAVARLEAVFASAPLIRQIFVYGN SER
 PYLLAVVPTADAAERFTGDPEGLKAAVAESLRQSAQLAELQSYEVPVDFVVE
 TEPFSEDNGLLSGVGKLLRPKPKERYADRLEQLYAE LAENRVTELRALREGAD
 KHPVVFTLTRA AEALLGVAGGPPAPDALFIELGGDSL SALTFSNLLRDIFDVD
 VPVGMITGPATDLGQLAEYVESERKSGSRRPTFATVHGRGAAEVRAAELTLDK
 FIDATTLAAAPNLPRATGTPHTVLLTGANGYLGRFLALEWLERLAETGGKLV S
 IVRATDTAAAVKRLEAVFDSDGPQLLERFRTLAAEHLEVIVGDI GEPNLGLDQ
 ATWQRLAQSVDLIVHPAALVNHVLPYDQLFGPNVVGTAELIRLAIITRIKPVT
 YLSTVAVAMTVDPGEFAEDGDIRAVSAVRPIDDSYANGYANSKWAGEVLLREA
 HDLCGLPVAVFRSDMILAH SRYAGQLNVPDAFTRLMFSLTTGIAPTTFYRTD

FIG 4 (Cont.)

EHGNRAVAHYDGLPADFVAEAVTTLGEQMAAEESGGYRSYDVMNPHDDGVSLD
VFVDWLIAAGHDIRRIEDYDEWLGRFTTALRALPKQRQHSVLPPLLDAYREPA
TPLRGAPAPTDVFRHAVRTAKIGADEDIPHLAALIDKYVADLRLGLV

YP 001288539.1

Secuencia de nucleótidos (SEQ ID NO:43)

>uniprot|A1QUM2|A1QUM2_MYCTF Ácido graso-CoA ligasa fadD9

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTCGAGGACCTATACGCCAG
CGACGCCAGTTCGCCGCCGCCAGTCCCAACGAGGCGATCACCCAGGCGATCG
ACCAGCCCGGGGTCGCGCTTCCACAGCTCATCCGTATGGTCATGGAGGGCTAC
GCCGATCGGCCGGCACTCGGCCAGCGTGCGCTCCGCTTCGTCACCGACCCCGA
CAGCGGCCGCACCATGGTCGAGCTACTGCCGCGGTTTCGAGACCATCACCTACC
GCGAACTGTGGGCCCGCGCCGGCACATTGGCCACCGCGTTGAGCGCTGAGCCC
GCGATCCGGCCGGGCGACCGGGTTTGCCTGCTGGGCTTCAACAGCGTCGACTA
CACAACCATCGACATCGCGCTGATCCGGTTGGGCGCCGTGTCGGTTCCTACTGC
AGACCAGTGCGCCGGTCACCGGGTTGCGCCCGATCGTCACCGAGACCGAGCCG
ACGATGATCGCCACCAGCATCGACAATCTTGGCGACGCCGTGGAAGTGCTGGC
CGGTCACGCCCCGGCCCGGCTGGTCGTATTTCGATTACCACGGCAAGGTTGACA
CCCACCGCGAGGCCGTGGAAGCCGCCCGAGCTCGGTTGGCCGGCTCGGTGACC
ATCGACACACTTGCCGAACTGATCGAACGCGGCAGGGCGCTGCCGGCCACACC
CATTGCCGACAGCGCCGACGACGCGCTGGCGCTGCTGATTTACACCTCGGGTA
GTACCGGCGCACCCAAAGGCGCCATGTATCGCGAGAGCCAGGTGATGAGCTTC
TGGCGCAAGTCGAGTGGCTGGTTCGAGCCGAGCGGTTACCCCTCGATCACGCT
GAACTTCATGCCGATGAGCCACGTCGGGGGCGGTCAGGTGCTCTACGGGACGC
TTTCCAACGGCGGTACCGCCTACTTCGTCGCCAAGAGCGACCTGTCGACGCTG
TTCGAGGACCTCGCCCTGGTGC GGCCACAGAATTGTGCTTCGTGCCGCGCAT
CTGGGACATGGTGTTCGCAGAGTTCCACAGCGAGGTCGACCGCCGCTTGGTGG
ACGGCGCCGATCGAGCGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAG
AACGTGCTCGGCGGACGGTTTGTTCATGGCGCTGACCGGTTCCGCGCCGATCTC
CGCTGAGATGACGGCGTGGGTCGAGTCCCTGCTGGCCGACGTGCATTTGGTGG
AGGGTTACGGCTCCACCGAGGCCGGGATGGTCCTGAACGACGGCATGGTGC GG
CGCCCCGCGGTGATCGACTACAAGCTGGTTCGACGTGCCCGAGCTGGGCTACTT
CGGCACCGATCAGCCCTACCCCGGGGCGAGCTGCTGGTCAAGACGCAAACCA
TGTTCCCCGGCTACTACCAGCGCCCGGATGTCACCGCCGAGGTGTTTCGACCCC
GACGGCTTCTACCGGACCGGGGACATCATGGCCAAAGTAGGCCCCGACCGAGTT
CGTCTACCTCGACCGCCGCAACAACGTGCTAAAGCTCTCCAGGGCGAGTTC A

FIG 4 (Cont.)

TCGCCGTGTCTGAAGCTCGAGGCGGTGTTTCGGCGACAGCCCGCTGGTCCGACAG
 ATCTTCATCTACGGCAACAGTGCCCGGGCCTACCCGCTGGCGGTGGTTGTCCC
 GTCCGGGGACGCGCTTTCTCGCCATGGCATCGAGAATCTCAAGCCCGTGATCA
 GCGAGTCCCTGCAGGAGGTAGCGAGGGCGGCCGGCCTGCAATCCTACGAGATT
 CCACGCGACTTCATCATCGAAACCACGCCGTTACCCCTGGAGAACGGCCTGCT
 CACCGGCATCCGCAAGCTGGCACGCCCGCAGTTGAAGAAGTTCTATGGCGAAC
 GTCTCGAGCGGCTCTATACCGAGCTGGCCGATAGCCAATCCAACGAGCTGCGC
 GAGCTGCGGCAAAGCGGTCCCGATGCGCCGGTGCTTCCGACGCTGTGCCGTGC
 CGCGGCTGCGTTGCTGGGCTCTACCGCTGCGGATGTGCGGGCCGGACGCGCACT
 TCGCCGACCTGGGTGGTGA CTGCTCTCGGCGCTGTGCGTTGGCCAACCTGCTG
 CACGAGATCTTCGGCGTCGACGTGCCGGTGGGTGTCATTGTCAGCCCGGCAAG
 CGACCTGCGGGCCCTGGCCGACCACATCGAAGCAGCGCGCACCCGGCGTCAGGC
 GACCCAGCTTCGCCTCGATACACGGTTCGCTCCGCGACGGAAGTGCACGCCAGC
 GACCTCACGCTGGACAAGTTCATCGACGCTGCCACCCTGGCCGCAGCCCCGAA
 CCTGCCGGCACCGAGCGCCCAAGTGCGCACCGTACTGCTGACCGGCGCCACCG
 GCTTTTTGGGTGCTACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTC
 AACGGCAAGCTGATCTGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAGC
 CCGGCTGGACGCGACGTTTCGATAGCGGGCACCCTGATTTGGTGCGGCACTACC
 GCGAATTGGGCGCCGGCCGCTCGAGGTGCTCGCCGGCGACAAGGGCGAGGCC
 GACCTGGGCCTGGACCGGGTACCTGGCAGCGGCTAGCCGACACGGTGGACCT
 GATCGTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGT
 TCGGCCCAAACGCGGGCGGGCACCGCCGAGTTGCTTCGGCTGGCGCTGACCGGC
 AAGCGCAAGCCATACATCTACACCTCGACGATCGCCGTGGGCGAGCAGATCCC
 GCCGGAGGCGTTACCGAGGACGCCGACATCCGGGCCATCAGCCCGACCCGCA
 GGATCGACGACAGCTACGCCAACGGCTACGCGAACAGCAAGTGGGCGGGCGAG
 GTGCTGCTGCGCGAAGCTCACGAGCAGTGCGGCCTGCCGGTGACGGTCTTCCG
 CTGCGACATGATCCTGGCCGACACCAGCTATAACGGTCAAGCTCAACCTGCCGG
 ACATGTTACCCGGCTGATGCTGAGCCTGGCCGCTACCGGCATCGCACCCGGT
 TCGTTCTATGAGCTGGATGCGCACGGCAATCGGCAACGCGCCCACTATGACGG
 CTTGCCGGTCTGAATTCGTGCGAGAAGCCATTTGCACCCTTGGGACACATAGCC
 CGGACCGTTTTTGTACCTACCACGTGATGAACCCCTACGACGACGGCATCGGG
 CTGGACGAGTTCGTGACTGGCTCAACTCCCCAACTAGCGGGTCCGGTTGCAC
 GATCCAGCGGATCGCCGACTACGGCGAGTGGCTGCAGCGGTTTCGAGACTTCGC
 TGCGTGCTTGC CGGATCGCCAGCGCCACGCCTCGCTGCTGCCCTTGCTGCAC
 AACTACCGAGAGCCTGCAAAGCCGATATGCGGGTCAATCGCGCCACCGACCA
 GTTCCGCGCTGCCGTCCAAGAAGCGAAAATCGGTCCGGACAAAGACATTCCGC
 ACCTCACGGCGGGCATCATCGCGAAGTACATCAGCAACCTGCGACTGCTCGGG
 CTGCTGTGA

FIG 4 (Cont.)

Secuencia de aminoácidos (SEQ 10 N0:44)

>uniprot|A1QUM2|A1QUM2_MYCTF Ácido graso-CoA ligasa fadD9

MSINDQRLTRRVEDLYASDAQFAAASPNEAITQAIDQPGVALPQLIRMVM
 EGYADRPALGQRALRFVTDPSGRTMVPELLPRFETITYRELWARAGTLAT
 ALSAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGL
 RPIVTETEPTMIATSIDNLGDAVEVLGHPARLVVFDYHGKVDTHREAV
 EAARARLAGSVTIDTLAELIERGRALPATPIADSADDALALLIYTSGSTG
 APKGAMYRESQVMSFWRKSSGWFEPSGYPSITLNFMPMSHVGGRRQVLYGT
 LSNGGTAYFVAKSDLSTLFEDLALVRPTLFCVVPRIWDMVFAEFHSEVDR
 RLVDGADRAALEAQVKAELRENVLGGRFVMALTGSAPI SAEMTAWVESLL
 ADVHLVEGYGSTEAGMVLNDGMVRRPAVIDYKLVDPVPELGYFGTDQPYPR
 GELLVKTQTMFPGYYQRPDVTAEVDFPDGDFYRTGDIMAKVGPDQFVYLDL
 RNNVLKLSQGEFIAVSKLEAVFGDSPLVRQIF IYGNSARAYPLAVVPSG
 DALSRHGIENLKPVI SE SLQEVARAAGLQSYEIPRDF I IETTPFTLENGL
 LTGIRKLARPQLKKFYGERLERLYTELADSQSNELELRELRSQSGPDAPVLPT
 LCRAAAALLGSTAADVRPDAHFA DLGGDSL SALS LANLLHEIFGVDVPVG
 VIVSPASDLRALADHIEAARTGVRRPSFASIHGRSATEVHASDLTLDKFI
 DAATLAAAPNLPAPSAQVRTVLLTGATGFLGRYLALEWLDLMDLVNGKLI
 CLVRARSDEEAQARLDATFDSGDPYLVRHYRELGAGRLEVLAGDKGEADL
 GLDRVTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNAAGTAELLRLALT
 GKRKPYIYTSTIAVGEQIPPEAFTEDADIRAI SPTRRIDDSYANGYANSK
 WAGEVLLREAHEQCGLPVTVFRCDMILADTSYTGQLNLPDMFTRLMLSLA
 ATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPDRFVTYH
 VMNPYDDGIGLDEFVDWLNSTSGSGCTIQRIADYGEWLQRFETSLRALP
 DRQRHASLLPLLHNYREPAKPICGSIAPTDQFRAAVQEAKIGPDKDIPHL
 TAAIIAKYISNLRLLLGLL

YP 953393.1

Secuencia de nucleótidos (SEQ ID N0:45)

>uniprot|A1T887|A1T887_MYCVP Dominio de tioéster reductasa

ATGCTACTGATACCCGCGAAGACCGCCTTGCCCGCCGCATCGCCGATCTGTA
 CGCCACCGACCCGCAGTTTGCCGCCGCAGCACCCGACGACGCCATCTCCACG
 CCATCGATCAACCAGGGACGCACCTGCCCGTCATCGTGCAGACCGTCCTCGAC

FIG 4 (Cont.)

GGATACGCCGAGCGGCCCGCGCTCGGACAGCGCGCGGTCCGTTTCGTCACCGA
CCCCGCCACGGGCAGAACCACCACCGAGCTGCTCCCCCGCTTCGAGACCATCA
CCTACGCCGAGTTGTCACGCCGCATCCACGCGGTACCCGCCGCGCTGACCGAT
GTGCATCCGGGCGACCGCGTGGCGGTGCTGGGCTTCACCAGCATCGACTACAC
CACCGTCGACATGGCTTTGGCGATGCTGGGCGCGGTGGCGGTACCACTTCAGA
CCAGCGCGCCGGCCACCACCGTCCGGCCGATCGTCGCCGAGACCGAACC GG TG
GTCATCGCATCGTCGGTCGATGCCCTCACCAGCGCCGTCGGCCTGGCTCTCGA
CGCTCCCACCGTGACCCGCTTGTGGTCTTCGATCATCGCGCCGGGGTTCGACG
ATCATCGCGACGCCCTCATCTCCGCGAGCGACCGGTTGCGCGCCGCCAACTCG
CCGATCGAGGTCGAGACCATTACCGACATCGTCGCTCGGGGTTTCGAAACTGCC
TGTACGCGCGCAATTCTCGGCCGACGGTGACGCGCTGAGCCTGCTGATCTACA
CCTCCGGCAGCACCGGCGCACCCAAGGGTGCATGTACCCCAACACCTGGTC
GCCAACTCATGGCGGCGGTTGGCCCGGTCCTTCTGGGGCGACCTGGGGGTCTT
CCCGGCAATCACGCTGAACTTCATGCCGATGAGCCACGTGATGGGCCGCGGAC
TGCTCTACGGCACGCTGGACGCCGGCGGCACCGCGTATTTTCGCGGCCAGGAGC
GATCTGTCGACGTTCTTGGAGGATCTCGCCCTGGTGCGCCCGACGCAGCTGAG
CTTCGTGCCGCGGATCTGGGACACCATCCACGCCGAGGTGTCCAGGAACTCG
AGCGCCGGCCGTCGGATGCGACCGAGGTGATCGCCGATCTGCGGCCGAGCCTG
CTGGGCGGCCGCTACGTCACGGCGATGACGGGCTCCGCGCCGCTGTCACCGGA
GATGCGGGCCTTCGTCGAGAACCTGCTCGACGTGCACCTGATCGACGGGTACG
GCTCGACCGAGGCCGGCGCGGTGTTTCGTCGACGGCCGGGTCCAACGCCCGCCG
GTCATCGATTACAAGCTCGTCGACGTCGCCGACCTCGGCTACTTCTCCACCGA
CCGCCCCCATCCGCGCGGCGAGCTTCTCGTCAAATCCGAGACGCTGTTTCCCG
GCTACTACAAACGCCCCGACGTCACCGCCGAGATGTTTCGACGAAGACGGCTAC
TACCGCACCGGCGACATCGTCGCCGAGACCGGGGCCGACCAGCTGACCTATCT
GGACCGCCGCAACAACGTCCTCAAACGTGTCGCAGGGCGAATTCGTCACCGTCT
CCCGACTGGAGGCAGTGTTTCGGCAACAGCCCGCTGGTCCGCCAGATCTACGTC
TACGGCAACAGCGCCCGCCCTACCTGCTGGCTGTGGTTCGTGCCACCGAAGC
CGCGCTGGCCGGTGCTGACGCCAAAGCCGCTGTGGCCGAGTCACTTCAGGATG
TCGCCAAGGCGACCGGGCTGCAGTCTACGAGATCCCCGCGATTTCTTCTC
GAGACGACGCCGTTACGCTGGAGAACGGCTTGCTGACCGGCATCCGAAA ACT
GGCCCGCCCCAGACTCAGAGAGCGTTACGGCGAACAGCTCGAGGCCCTCTACA
CCATGCTGTCCGAAGAGCAGGCCGACGAGCTGCGGGAGCTGCGCCGCTCCGGC
GGAGAGCGTCCGGCGCTGGAAACCGTCGGACGCGCCGCCGGGGCGCTGCTCGG
CACCACCGCAGGCGAGCTGGAGCCGAGCGCCCACTTCACAGATCTGGGCGGGG
ATTCGCTGTGCGCGTTGACCTTCGCCAACCTGCTGCGCGACATCTTCGACGTC
GACGTCCCCGTCGGTGTGATCGTCAGCCCGGCCACCGATCTGCAGGCCCTTGC
CGACTACGTCGAGTCCGCCCGCCGGCACGGGTCCGGTTCGGGCCCACTTTCGAAT
CGGTGCACGGGCATTCGGGACGACCCGGGACCGAGGTGCATGCCCGCGATCTG
ACGTTGGACGAATTCGTCGACGCCGCGACCCCTGGCGCACGCGCCGACGTTGCC
CGGACCGCGCGCCGAGGTCCGCACCGTCTGCTGACCGGGGCGACCGGCTTCC

FIG 4 (Cont.)

TCGGCCGGTATCTCGCTCTCGAATGGCTTGAGCGGATGGCGCTGGTCGGCGGC
 AAGCTGATCTGCCTGGTCCGCGCCAAAGACGATGCGGCAGCGCGGGTTTCGGCT
 GGACAGCACGTTTCGACAGCGGGGACCCGGAGCTGCTGCGGCACTACCGACGGC
 TGGCAGCCGACCATCTCGAAGTGATCGCCGGCGACAAGGCCGACGCCGATCTC
 GGACTCGACGCGCGGACGTGGCAGCGGCTCGCGGACACCGTCGATCTGATCGT
 CGATCCCGCCGCCCTGGTCAACCACGTCTGCGGTACCGCCAACTGTTTCGCC
 CGAACGTGCTCGGCACCGCCGAACCTGCTCCGCATCGCGCTGACAACGAGGATG
 AAGCCGTTTCGTGTACGTGTCGACGATCGGCGTCGGCGCCGGTATCGAACCCGC
 GAGGTTACCCGAGGACGCCGACATCAGGCAGATCAGCGCGACGCGCAGAATCG
 ACGACAGCTACGCCAACGGCTACGGCAACAGCAAGTGGGCCGGCGAGGTGCTG
 CTGCGCGAGGCGCACGATCTGTGCGGCCTGCCCGTTTTCGGTGTTCGGCTGCGA
 CATGATCCTGGCCGACACCACCTACGCCGGTCAGCTGAACCTGCCGGACATGT
 TCACCCGCCTGATCTTCAGCCTGGTCGCCACCGGCGTCGCGCCCGAGTCGTTT
 TACCACCTCGCCACCGACGGCACCCGGCAACGGGCCCACTATGACGGGCTACC
 GGTGGAGTTCATCGCCGAAGCGATCTCGACCCTCGGCTCTGACGTCGCCTCCG
 GGTTCGGACGTATCACGTGATGAATCCCCACGACGACGGGATAGGCCTCGAC
 GAGTACGTGGACTGGCTCATCGACGCAGGCCATCCGATTCGACGCGTCGGCGA
 CTACCCGACGTGGCTGCAACGGTTCACGGTGGCGATCACCGCGCTGCCGGAAC
 GGCAGCGGCAGGCCTCACTGTTGCCGCTGCTGCACAACCTACCAGCACCCGGAG
 ACGCCGATCCGCGGCTCCATCGCACCGACAGACAGATTCCGTGAAGCGGTCCA
 GGACGCCAAAATCGGTCCGGACAAAGACATTCCACACGTGACACCGCAGATCG
 TCATCAAGTACGTCACCGACCTGCAGCGACTCGGACTGCTCTAA

Secuencia de aminoácidos (SEQ ID N0:46)

>uniprot|A1T887|A1T887_MYCVP Dominio de tioéster reductasa

MSTDTRDLARRIADLYATDPQFAAAAPDDAISHAIDQPGTHLPVIVQT
 VLDGYAERPALGQRAVRFVTD PATGRITTELLPRFETITYAELSRRIHAV
 TAALTDVHPGDRVAVLGFSTIDYTTVDMALAMLGAVAVPLQTSAPATTVR
 PIVAETEPVVIASSVDALTDVGLALDAPTVTRLVVFDHRAGVDDHRDAL
 ISASDRLRAANSPIEVETITDIVARGSKLPVRAQFSADGDALSLLIYTS
 STGAPK GAMYPQH LVANSWRRLARSFWGDLGVFPAILNFMPSHVMGRG
 LLYGTL DAGGTAYFAARSDLSTFLEDLALVRPTQLSFVPRIWDTIHA
 EVS QELERRPSDATEVIADLRRSLLGGRYVTAMTGSAPLSPMRAFVENLLDV
 HLIDGYGSTEAGAVFVDGRVQRPPVIDYKLVDVADLGYFSTDRPHPRGEL
 LVKSETLFPGYKRPDVTAEMFDEDGYRTGDIVAETGADQLTYLDRRNN
 VLKLSQGEFVTVSRLEAVFGNSPLVRQIYVYGNSARPYLLAVVVPTEAAL
 AGADAKAAVAESLQDVAKATGLQSYEIPRDFLLETTPFTLENGLLTGIRK
 LARPRLRERYGEQLEALYTMLSEEQADELRELRRSGGERPALETVGRAAG

FIG 4 (Cont.)

ALLGTTAGELEPSAHFTDLGGDSLSALTFANLLRDIFDVDVPVGVIVSPA
 TDLQALADYVESARRHGSRPTFESVHGHSGRPGTEVHARDLTLDEFVDA
 ATLAHAPTLPGPRAEVRTVLLTGATGFLGRYLALEWLERMALVGGKLI
 VRAKDDAAARVRLDSTFDSDGPELLRHYRRLAADHLEVIAGDKADADLGL
 DARTWQRLADTVDLIVDPAALVNHVLPYRQLFAPNVLGTAE LLRIALTTR
 MKPFVYVSTIGVGAGIEPARFTEDADIRQISATTRIDDSYANGYGNSKWA
 GEVLLREAHDLGLPVSVFRCDMILADTTYAGQLNLPDMFTRLIFSLVAT
 GVAPESFYHLATDGRQRAHYDGLPVEFIAEAI STLGS DVASGFRTYHVM
 NPHDDGIGLDEYVDWLI DAGHP IRRVGDYPTWLQRF TVAIT ALPERQ RQA
 SLLPLLNHYQHPETPIRGS IAPTDRF REAVQDAKIGPDKDIPHVTPQIVI
 KYVTDLQRLGLL

YP 938306.1

Secuencia de nucleótidos (SEQ ID NO:47)

>uniprot|A1UFA8|A1UFA8_MYCSK Dominio de tioéster reductasa

ATGTCCACCGAGACCCGTGAAGCGCGCCTGCAGCAGCGCATCGCGCATCTGTT
 CACCACCGACCCGCGAGTTCGCCGCCGCCGCCGCCGACCCCGGATCAGCGATG
 CCGTCGACCGCGATGACACGCGGCTGACCGCCATCGTCTCCGCGGTCATGTCG
 GGGTACGCCGACCGGCCGCGACTCGGGCAGCGCGCCGCCGAATTCGTCACCGA
 CCCGCAGACCGGCCGACACGATGGAGCTGCTCCCCGCTTCGACACCATCA
 CCTACCGGGAGCTGCTCGACCGCGTCCGGGCGCTACCAACGCGTGGCACGCC
 GACGGGGTGCGCCCCGGCGACCGTGTCGCAATCCTCGGATTCACCGGCATCGA
 CTACACCGTGGTCGACCTCGCCCTGATCCAGCTCGGCGCGGTTCGCGGTGCCGC
 TGCAGACCAGCGCCCGCGTGGAGGCGTTGCGCCCGATCGTGGCCGAGACCGAA
 CCGATGCTCATCGCCACCGGCGTCGATCATGTCGACGCCGCCGCCGAACTCGC
 ACTCACCGGCCACCGTCCGTCCCGGGTGGTGGTCTTCGACCACCGCGAGCAGG
 TCGACGACGAACGCGACGCGGTGCGGGCCGCGACGCCAGGCTGGGAGACGCC
 GTCCCCGTGAGACACTCGCCGAGGTGTTGCGGCGGGGCGCCCATCTGCCCGC
 CGTGGCGCCGCACGTGTTTCGACGAGGCCGATCCACTGCGGCTGCTGATCTACA
 CTTCCGGCAGCGCCGGCGCCCCAAGGGCGCGATGTATCCCGAGAGCAAGGTC
 GCCGGCATGTGGCGCGGTCGGCCAAGGCCGCCTGGAACAACGATCAGACAGC
 GATTCCGTGATCACCTGAACTTCCTGCCGATGAGCCACGTCATGGGTTCGCG
 GCCTGCTGTGCGGCACGCTCAGCACCGGTGGCACC GCGTATTTCCGCCCCCGC
 AGCGATCTGTGACGCTGCTCGAGGACCTGCGCCTGGTACGGCCCACTAACT
 CAGCTTCGTGCCGCGGATCTGGGACATGCTCTTCCAGGAGTTCGTCGGCGAGG
 TCGACCGGCCGGGTGAACGACGGTGCGGACCGCCCCACCGCGGAGGCCGACGTG
 CTGGCCGTACAGCGTCACGAGCTGCTCGGTGGCCGGTTCGTCACCGCGATGAC

FIG 4 (Cont.)

CGGTTTCGGCGCCCATCTCCCTCGAGATGAAGACATGGGTGGAGACCCTGCTCG
 ACATGCACCTGGTTCGAGGGTTACGGCTCGACGGAGGCCGGCGCGGTGTTTCGTC
 GACGGCCACATCCAGCGCCCACCGGTGCTCGACTACAAACTCGTCGACGTCCC
 CGACCTCGGCTACTTCAGCACCGACCGGCCGCACCCGCGCGGTGAGCTGCTGG
 TCCGCTCCACGCAGCTATTCCCCGGTACTACAAACGTCCCAGCTCACCGCC
 GAGGTGTTTCGACGACGACGGCTTCTACCGCACGGGCGACATCGTCGCCGAGGT
 CGGCCCGATCAGGTGCAGTACCTCGACCGCCGCAACAACGTGCTCAAACCTCG
 CCCAGGGTGAGTTCGTCACCATCTCCAAACTCGAGGCGGTCTTCGCCGGCAGC
 GCCCTGGTCCGCCAGATCTACGTGTACGGCAACAGTGC GCGCTCCTACCTGCT
 GGCCGTCGTCGTGCCGACCGACGATGCGGTGGCCCGGCACGACCCGGCATCGC
 TCAAGACCGCGATCAGCGCCTCGCTGCAGCAGGCCGCGAAGACCGCCGGTCTG
 CAGTCCTACGAGCTGCCGCGTGACTTCCTCGTCGAGACTCAACCGTTCACGCT
 GGAGAACGGACTACTGACCGGCATCCGCAAGCTGGCGCGCCCGAAACTCAAGG
 CGCGCTACGGCGATCGGCTCGAGGCGCTCTACGTCGAACTGGTTCGAAGGACAG
 GCAGGCGAATGCGCACCCCTGCGCCGGGACGGCGCGAAGCGTCCGGTGGCCGA
 GACGGTTCGGCCGCGCCGCGGCCGCGCTGCTCGGCGCCGCGCCGCGCCGACGTGC
 GCCCGACGCGCACTTCACCGACCTCGGCGGAGACTCGTTGTCGGCGTTGACC
 TTCGGCAATCTGCTGCAGGAGATCTTCGGCGTCGACGTCCCGGTTCGGGGTGAT
 CGTCAGCCCGGCGGCCGACCTGGCGTCGATCGCGGCGTACATCGAGGCCGAAC
 AGGCCTCGACCGGTAAGCGGCCGACCTACGCGTCGGTGCACGGGCGCGACGCC
 GAACAGGTACACGCGCGCGACCTCACCTGGACAAGTTCATCGACGCCGAAAC
 ACTCTCCGCTGCAACAGAACTGCCCGGCCCGAGCGGTGAGGTGCGCACCGTGC
 TGCTGACCGGGGCCACCGGATTCTCGGCCGCTACCTGGCGCTGGACTGGCTC
 GAACGGATGGCCCTGGTCGACGGCAAGGTCATCTGCCTGGTCCGCGCGAAGGA
 CGATGCGGCCGCCCGCAAACGCCCTCGACGACACCTTCGACAGCGGCGATCCGA
 AGCTGCTGGCGCACTACCGCAAGCTGGCCGCCGACCACCTCGAAGTGCTGGCC
 GGCACAAAGGGTGAGGCGGATCTCGGGCTGCCGCATCCTGTCTGGCAGCGCCT
 GGCCGACACCGTCGACCTCATCGTCGACCCGGCCGCCCTGGTCAACCACGTAC
 TGCCGTACAGCCAGCTGTTTCGGGCCCAACGCGCTGGGCACTGCCGAGCTGATC
 CGGCTTTCGCTCACCACCCGCATCAAACCGTTCACCTACGTGTCGACGATCGG
 CGTCGGCGCCGGTATCGAACC GGCCGTTTCACCGAGGACGACGACATCCGGG
 TGATCAGCCCGACGCGGGCCGTCGACACCGGGTACGCCAACGGCTACGGCAAC
 AGCAAGTGGGCCGGTGAGGTGTTGTTGCGCGAGGCGCACGATCTGTGCGGGCT
 CCCCCTGGCGGTGTTCCGGTGCAGATGATCCTGGCCGACACCACCTACGCCG
 GCCAGCTCAACCTGCCGGACATGTTACCCCGGATGATGCTGAGCCTGGTGACC
 ACGGGTATCGCGCCGAAATCGTTCCACCCACTCGACGCAAAGGGCCACCGGCA
 GAGCGCCCACTACGACGGGCTGCCGGTCGAGTTCGTTCGCCGAATCGATCTCCG
 CGCTGGGAGCGCAGGCGGTTCGACGAGGCGGGAACGGGTTTCGCCACCTACCAC
 GTGATGAACCCCCACGACGACGGGATCGGCCTCGACGAATTCGTTCGACTGGCT
 CGTCGAGGCGGGGTATCGCATCGACCGCATCGACTACTACGCGGCCCTGGCTGC
 AGCGGTTTCGAAACCGCCCTGCGGGCGCTGCCCGAGCGCACTCGGCAGTACTCA

FIG 4 (Cont.)

CTGCTCCCGCTGCTGCACAACACTACCAGCGGCCTGCGCACCCGATCAACGGGGC
GATGGCCCCGACCGACCGGTTCCGCGCTGCGGTGCAGGAGGCAAAGCTCGGCC
CGGACAAGGACATTCCCCACGTCCTCCAGCGGTGATCGTCAAGTACGCCACC
GACCTGGAGCTGCTGGGCCTGATCTAG

Secuencia de aminoácidos (SEQ ID N0:48)

>uniprot|A1UFA8|A1UFA8_MYCSK Dominio de tioéster reductasa

MSTETREARLQQRIAHLFTTDPQFAAARPDPRISDAVDRDDTRLTAIVSAVMS
GYADRPALGQRAAEFVTDPTGRITMELLPRFDITITYRELLDRVRALTNAWHA
DGVPRGDRVAILGFTGIDYTVVDLALIQLGAVAVPLQTSAAVEALRPIVAETE
PMLIATGVDHVDAAELALTGHRPSRVVVFDRHQVDDERDAVRAATARLGDA
VPVETLAEVLRGAHLPAVAPHVFDEADPLRLLIYTSGSAGAPKGMYPESKV
AGMWRASAKAAWNNDQTAIPSIITLNF LPM SHVMGRGLLCGTLSTGGTAYFAAR
SDLSTLLEDLRLVRPTQLSFPRIWDMLFQEFVGEVDRRVNDGADRPTAADV
LAVQRHELLGGRFVTAMTGSAPISLEMKTWVETLLDMHLVEGYGSTEAGAVFV
DGHIQRPVLDYKLVDPDLGYFSTDRPHPRGELLVRSTQLFPGYYKRPDVTA
EVFDDDGFYRTGDIVAEVGPQVQYLDRRNNVLKLAQGEFVTISKLEAVFAGS
ALVRQIYVYGNSARSYLLAVVPTDDAVARHDPASLKTASASLQQAAKTAGL
QSYELPRDFLVETQPFTEENGLLTGIRKLARPKLKARYGDRLEALYVELVEGQ
AGELRTLRRDGAKRPVAETVGRAAAALLGAAAADVDPDAHFTDLGGDSL SALT
FGNLLQEIFGVDVPVGVIVSPAADLASIAAYIEAEQASTGKRPTYASVHGRDA
EQVHARDLTLDFKIDAETLSAATELPGPSGEVRTVLLTGATGFLGRYLALDWL
ERMALVDGKVICLVRAKDDAAARKRLDDTFDSGDPKLLAHYRKLAADHLEVLA
GDKGEADLGLPHPVWQRLADTVDLIVDPAALVNHVLPYSQLFGPNALGTAELI
RLALTRIKPFTYVSTIGVGAGIEPGRFTEDDDIRVISPTRAVDTGYANGYGN
SKWAGEVLLREAHDLGGLPVAVFRCDMILADTTYAGQLNLPDMFTRMMLSLVT
TGIAPKSFHPLDAKGHRQSAHYDGLPVEFVAESI SALGAQAVDEAGTGFATYH
VMNPHDDGIGLDEFVDWLVEAGYRIDRIDYYAAWLQRFETALRALPERTRQYS
LLPLLHNYQRPAHPINGAMAPTDRFRAAVQEA KLGPDKDIPHVTPAVIVKYAT
DLELLGLI

ZP_05217435

Secuencia de nucleótidos (SEQ ID N0:49)

>gi | 222089526:2534-6055 Mycobacterium avium subsp. avium ATCC 25291 NZ_ACFI01000138, secuenciación genómica completa al azar

FIG 4 (Cont.)

ATGTCGACTGCCACCCATGACGAACGACTCGACCGTCGCGTCCACGAACTCAT
 CGCCACCGACCCGCAATTCGCCGCCGCCAACC CGACCCGGCGATCACCGCCG
 CCCTCGAACAGCCCGGGCTGCGGCTGCCGCAGATCATCCGCACCGTGCTCGAC
 GGCTACGCCGACCGGCCGGCGCTGGGACAGCGCGTGGTGGAGTTCGTCACGGA
 CGCCAAGACCGGGCGCACGTCGGCGCAGCTGCTCCCCGCTTCGAGACCATCA
 CGTACGGCGAAGTGGCGCAGCGTGTTCGGCGCTGGGCGCGCCCTGTCCGAC
 GACGCGGTGCACCCCGGGCACC GGGTGTGCGTGCTGGGCTTCAACAGCGTCGA
 CTACGCCACCATCGACATGGCGCTGGGCGCCATCGGCGCCGTCTCGGTGCCGC
 TGCAGACCAGCGCGGCAATCAGCTCGCTGCAGCCGATCGTGGCCGAGACCGAG
 CCCACCCTGATCGCGTCCAGCGTGAACCAGCTGTCCGACGCGGTGCAGCTGAT
 CACCGGCGCCGAGCAGGCGCCCACCCGGCTGGTGGTGTTCGACTACCACCCGC
 AGGTCGACGACCAGCGCGAGGCCGTCCAGGACGCCGCGGGCGGGCTGTCCAGC
 ACCGGCGTGGCCGTCCAGACGCTGGCCGAGCTGCTGGAGCGCGGCAAGGACCT
 GCCCGCCGTCCGGGAGCCGCCCGCCGACGAGGACTCGCTGGCCCTGCTGATCT
 ACACCTCCGGGTCCACCGGCGCCCCAAGGGCGCGATGTACCCGCAGAGCAAC
 GTCGGCAAGATGTGGCGCCGCGGCAGCAAGAACTGGTTCGGCGAGAGCGCCGC
 GTCGATCACCTGAACTTCATGCCGATGAGCCACGTGATGGGCCGAAGCATCC
 TCTACGGCACGCTGGGCAACGGCGGCACCGCCTACTTCGCCGCCCGCAGCGAC
 CTGTCCACCCTGCTCGAGGACCTCGAGCTGGTGC GGCCCACCGAGCTCAACTT
 CGTCCC GCGGATCTGGGAGACGCTGTACGGCGAATTCCAGCGTCAGGTCGAGC
 GCGCGCTCTCCGAGTCCGGGGACGCCGGCGAACGTCGCGCCGTTCGAGGCCGAG
 GTGCTGGCCGAGCAGCGCCAGTACCTGCTGGGCGGGCGGTTACCTTCGCGAT
 GACGGGCTCGGCGCCCATCTCGCCGGAGCTGCGCAACTGGGTTCGAGTCGCTGC
 TCGAAATGCACCTGATGGACGGCTACGGCTCCACCGAGGCCGGAATGGTGTG
 TTCGACGGGGAGATTCAGCGCCCGCCGGTGATCGACTACAAGCTGGTCGACGT
 GCCGGACCTGGGCTACTTCAGCACCGACCGGCCGCATCCGCGCGGGCGAGCTGC
 TGCTGCGCACCGAGAACATGTTCCCGGGCTACTACAAGCGGGCCGAAACCACC
 GCGGGCGTCTTCGACGAGGACGGCTACTACCGCACCGGCGACGTGTTCCCGA
 GATCGCCCCGGACCGGCTGGTCTACGTCGACCGCCGCAACAACGTGCTCAAGC
 TGGCGCAGGGCGAATTTCGTCACGCTGGCCAAGCTGGAGGCGGTGTTCCGGCAAC
 AGCCCGCTGATCCGCCAGATCTACGTCTACGGCAACAGCGCCCAGCCCTACCT
 GCTGGCGGTTCGTGGTGCCACCGAGGAGGCGCTGGCCTCGGGTGACCCCGAGA
 CGCTCAAGCCCAAGATCGCCGACTCGCTGCAGCAGGTCGCCAAGGAGGCCGGC
 CTGCAGTCTTACGAGGTGCCGCGGACTTCATCATCGAGACCACCCGTTTACG
 CCTGGAAAACGGTCTGCTGACCGGGATCCGGAAGCTGGCGTGGCCGAAACTGA
 AGCAGCACTACGGGGAACGGCTGGAGCAGATGTACGCCGACCTGGCCGCCGGA
 CAGGCCAACGAGCTGGCCGAGCTGCGCCGCAACGGTGCCAGGCGCCGGTGCT
 GCAGACCGTGAGCCGCGCCGCGGGCGCCATGCTGGGTTCCGGCCGCTCCGACC
 TGTCCCCGACGCCCACTTCACCGATCTGGGCGGAGACTCGTTGTCGGCGTTG
 ACATTCGGCAACCTGCTGCGCGAGATCTTCGACGTCGACGTGCCGGTGGGCGT
 CATCGTCAGCCCGGCCAACGACCTGGCGGCCATCGCGAGCTACATCGAGGCCG

FIG 4 (Cont.)

AGCGGCAGGGCAGCAAGCGCCCGACGTTTCGCCTCGGTGCACGGCCGGGACGCG
 ACCGTGGTGC GCGCCGCGACCTGACGCTGGACAAGTTCCTCGACGCCGAGAC
 GCTGGCCGCGCGCCGAACCTGCCAAGCCGGCCACCGAGGTGCGCACCGTGC
 TGCTGACCGGCGCCACCGGCTTCCTGGGCGCTACCTGGCCCTGGAATGGCTG
 GAGCGGATGGACATGGTGGACGGCAAGGTCATCGCCCTGGTCCGGGCCCCGCTC
 CGACGAGGAGGCACGCGCCCGGCTGGACAAGACCTTCGACAGCGGCGACCCGA
 AGCTGCTCGCGCACTACCAGCAGCTGGCCGCGATCACCTGGAGGTCATCGCC
 GCGACAAGGGCGAGGCCAATCTGGGCCTGGGCCAAGACGTTTGGCAACGACT
 GGCCGACACGGTTCGACGTGATCGTCGACCCCGCCGCGCTGGTCAACCACGTGT
 TGCCGTACAGCGAGCTGTTCCGGGCCAACGCCCTGGGCACCGCGGAGCTGATC
 CGGCTGGCGCTGACGTCCAAGCAGAAGCCGTACACCTACGTGTCCACCATCGG
 CGTGGGCGACCAGATCGAGCCGGGCAAGTTCGTGAGAACGCCGACATCCGGC
 AGATGAGCGCCACCCGGGCGATCAACGACAGCTACGCCAACGGCTACGGCAAC
 AGCAAGTGGGCGCGGAGGTGCTGCTGCGCGAGGCGCACGACCTGTGCGGGCT
 GCCCGTCGCGGTGTTCCGCTGCGACATGATCCTGGCCGACACCACGTATGCCG
 GGCAGCTCAACCTGCCGGACATGTTACCCGGCTGATGCTGAGCCTGGTGGCC
 ACCGGGATCGCGCCCGGCTCGTTCTACGAGCTCGACGCCGACGGCAACCGGCA
 GCGGGCGCACTACGACGGCCTGCCGGTCGAGTTCATCGCCGCGGGCGATCTCGA
 CGCTGGGTTTCGAGATCACCGACAGCGACACCGGCTTCCAGACCTACCACGTG
 ATGAACCCCTACGATGACGGCATCGGTCTGGACGAGTACGTGATTGGCTGGT
 GGACGCCGGCTATTTCGATCGAGCGGATTGCCGACTACTCCGAATGGCTGCGGC
 GGTTTCGAGACCTCGCTGCGGGCCCTGCCGGACCGGCAGCGCCAGTACTCGCTG
 CTGCCGCTGCTGCACAACTACCGCACGCCGGAGAAGCCGATCAACGGGTCGAT
 AGCTCCCACCGACGTGTTCCGGGCAGCGGTGCAGGAGGCGAAAATCGGCCCCG
 ACAAAGACATTCCGCACGTGTCGCCGCCGGTTCATCGTCAAGTACATCACCGAC
 CTGCAGCTGCTCGGGCTGCTCTGA

Secuencia de aminoácidos (SEQ ID NO:50)

>gi | 254775919 | ref | ZP_05217435.1 | FadD9 [Mycobacterium avium subsp. avium ATCC 25291]

MSTATHDERLDRRVHEL IATDPQFAAAQDPAITAALEQPGLRLPQI IRTVLD
 GYADRPALGQRVVEFVTDKGTGR TSAQLLPRFETIITYGEVAQRVSALGRALSD
 DAVHPGDRVCVLGFNSVDYATIDMALGAIGAVSVPLQTSAAISSLQPIVAETE
 PTLIASSVNQLSDAVQLITGAEQAPTRLVVFVDYHPQVDDQREAVQDAAARLSS
 TGVAVQTLAELLER GKDLPAVGEPPEDESLALLIYTSGSTGAPK GAMYPQSN
 VGKMWRRGSKNWFGE SAASITLNFMPMSHVMGRSILYGT LGNGGTAYFAARSD
 LSTLLEDLELVRPTE LNFPRIWETLYGEFQRQVERRLSESGDAGERRAVEAE
 VLAEQRQYLLGGRFTFAMTGSAPISPELRN WVESLLEMHLMDGYGSTEAGMVL

FIG 4 (Cont.)

FDGEIQRPPVIDYKLVDPDLGYFSTDRPHPRGELLRLTENMFPGYKRAETT
 AGVFDEDEGGYRTGDVFAEIAIPDRLVYVDRRNNVLKLAQGEFVTLAKLEAVFGN
 SPLIRQIYVYGN SAQPYLLAVVVPTEEALASGDPETLKPKIADSLQOVAKEAG
 LQSYEVPRDFI IETTPF SLENGLLTGIRKLAWPKLKQHYGERLEQMYADLAAG
 QANELAELRRNGAQAPVLQTVSRAAGAMLGSAASDLSPDAHFTDLGGDSL SAL
 TFGNLLREIFDVDVPGVIVSPANDLAAIASYIEAERQGSKRPTFASVHGRDA
 TVVRAADLTLDKFLDAETLAAAPNLPKPATEVRTVLLTGATGFLGRYLALAWL
 ERMDMVDGKVI ALVRARSDEEARARLDKTFDSGDPKLLAHYQQLAADHLEVIA
 GDKGEANLGLGQDVWQRLADTVDVIVDPAALVNHVLPYSELF GPNALGTAELI
 RLALTSKQKPYTYVSTIGVGDQIEPGKFVENADIRQMSATRAINDSYANGYGN
 SKWAGEVLLREAHDL CGLPVAVFRCDMILADTTYAGQLNLPDMFTRLMLSLVA
 TGIAPGSFYELDADGNRQRAHYDGLPVEFIAAAI STLGSQITSDTGFQTYHV
 MNPYDDGIGLDEYVDWLVDAGYSIERIADYSEWLRRFETSLRALPDRQRQYSL
 LPLLHNYRTPEKPI NGSIAPTDFRAAVQEA KIGPDKDIPHVSPVIVKYITD
 LQLLGLL

YP 001070587.1

Secuencia de nucleótidos (SEQ ID N0:51)

>uniprot|A3PYW9|A3PYW9_MYCSJ Dominio de tioéster reductasa

ATGTCCACCGAGACCCGTGAAGCGCGCCTGCAGCAGCGCATCGCGCATCTGTT
 CGCCACCGACCCGCAGTTCGCCGCCGCCCGGCCCGACCCCGGATCAGCGATG
 CCGTCGACCGCGATGACGCGCGGCTGACCGCCATCGTCTCCGCCGTCATGTCG
 GGGTACGCCGACCGGCCGCACTCGGGCAGCGCGCCGCGAATTCGCCACCGA
 CCCGCAGACCGGCCGCAACCACGATGGAGCTGCTCCCCGCTTCGACACCATCA
 CCTACCGGGAGCTGCTCGACCGCGTCCGGGCGCTCACCAACGCGTGGCACGCC
 GACGGGGTGCGCCCGGGCAGCGCGTGC AATCCTCGGATTCACCGGCATCGA
 CTACACCGTTCGTCGACCTCGCCCTGATCCAGCTCGGCGCGGTCGCGGTGCCGC
 TGCAGACCAGCGCCCGCTGGAGGCGCTGCGCCCGATCGTGGCCGAGACCGAA
 CCGATGCTCATCGCCACCGGCGTCGATCATGTTCGACGCCGCCGCGGA ACTCGC
 ACTCACCGGCCACCGTCCGTCCCAGGTGGTGGTCTTCGACCACCGCGAGCAGG
 TCGACGACGAACGCGACGCGGTGCGGGCCGCGACGGCCCGGCTGGGAGACGCG
 GTGCCCGTCGAGACACTCGCCGAGGTGTTGCGGCGTGGCGCCCATCTGCCCGC
 CGTGGCACCGCACGTGTTTCGACGAGGCCGATCCACTGCGGCTGCTGATCTACA
 CCTCCGGCAGCACCGGCGCCCCCAAGGGCGCGATGTATCCCAGAGCAAGGTC
 GCCGGCATGTGGCGCGCGTCCGCCAAGGCCGCTGGAACAACGATCAGACGGC
 GATTCCGTCGATCACCTGAACTTCCTGCCGATGAGCCACGTCATGGGTCCGC

FIG 4 (Cont.)

GCCTGCTGTGCGGCACGCTCAGCACCGGTGGCACCGCGTATTTGCGCCGCCCGC
 AGCGATCTGTGACGCTGCTCGAGGACCTGCGCCTGGTACGGCCCACCCAAC
 CAGCTTCGTGCCACGGATCTGGGACATGCTCTTCCAGGAGTTCGTGCGGCGAGG
 TCGACCGGGCGGGTGAACGACGGTGCAGACCGCCCCACCGCGGAGGCCGACGTG
 CTGGCCGAACTGCGCCAGGAGCTGCTCGGTGGCCGGTTCGTACCGCGATGAC
 CGGTTCGGCGCCCATCTCCCCGAGATGAAGACATGGGTGGAGACCCTGCTCG
 ACATGCACCTGGTTCGAGGGTTACGGCTCGACGGAGGCCGGCGCGGTGTTTCGTC
 GACGGCCACATCCAGCGCCCCGCCGGTGTCTGACTACAAACTCGTCGACGTCCC
 CGACCTCGGCTACTTCAGCACCGACCGGCCGCACCCGCGCGGTGAGCTGCTGG
 TCCGCTCCACGCAGCTGTTCCCCGGTACTACAAGCGTCCCGACGTACCCGCC
 GAGGTGTTTCGACGACGACGGCTTCTACCGCACGGGCGACATCGTCGCCGAGCT
 CGGCCCGACCGAGCTGCAGTACCTCGACCGCCGCAACAACGTGCTCAAACCTCG
 CCCAGGGTGAGTTCGTACCATCTCCAAACTCGAGGCGGTCTTCGCCGGCAGC
 GCCCTCGTCCGCCAGATCTTCGTGTACGGCAACAGTGCGCGCTCCTACCTGCT
 GGCCGTCGTGTCGCGACCGACGATGCGGTGGCCCGGCACGACCCGGCATCGC
 TCAAGACCGCGATCAGCGCCTCGCTGCAGCAGGCCGCGAAGACCGCCGGTCTG
 CAGTCTACGAGCTGCCGCGTGACTTCTCGTCGAGACTCAACCGTTCACGCT
 GGAGAACGGACTACTGACCGGCATCCGCAAGCTGGCCCGCCGAAACTCAAGG
 CGCGCTACGGCGATCGGCTCGAGGGCGTCTACGTCGAACCTGGCCGAAGGACAG
 GCAGGCGAACTGCGCACCTGCGCCGGGACGGCGCGAAGCGTCCGGTGGCCGA
 GACGGTTCGGCCGCGCCGCGGCCGCGCTGCTCGGCGCCGCGCCGCGCCGACGTGC
 GCCCGACGCGCACTTCACCGACCTCGGCGGAGACTCGTTGTGCGGCGTTGACC
 TTCGGCAATCTGCTGCAGGAGATCTTCGGCGTCGACGTCCCGGTTCGGGGTGAT
 CGTCAGCCCGGCGGCCGACCTGGCGTCGATCGCGGGGTACATCGAGACCGAAC
 AGGCCTCGACCGGCAAGCGGCCGACCTACGCGTCGGTGCACGGGCGCGACGCC
 GAACAGGTACGCGCGCGGACCTCACCTGGACAAGTTCATCGACGCCGAAAC
 ACTCTCCGCTGCAACAGAACTGCCCGTCCCGATCGGTGAGGTGCGCACCGTGC
 TGCTGACCGGGGCCACCGGATTCTCGGCCGCTACCTGGCGCTGGACTGGCTC
 GAAAGGATGGCCCTGGTTCGACGGCAAGGTCATCTGCCTGGTCCGCGCGAAGGA
 CGATGCGGCCGCCCGCAAACGCCTCGACGACACCTTCGACAGCGGCGATCCGA
 AGCTGCTGGCGCACTACCGCAAGCTGGCCGCGGACACCTCGAAGTGCTGGCC
 GCGACAAGGGTGAGGCGGATCTCGGGCTGCCGCATCAGGTCTGGCAGCGACT
 GGCCGACACCGTCGACCTCATCGTCGACCCGGCCGCCCTGGTCAACCACGTAC
 TGCCGTACAGCCAGCTGTTCCGGGCCAACGCGCTGGGCACTGCCGAGCTGATC
 CGGCTTGCGCTCACCACCCGCATCAAACCGTTCACCTACGTGTCGACGATCGG
 CGTCGGCGCCGGTATCGAACCGGGCCGTTTCACCGAGGACGACGACATCCGGG
 TGATCAGCCCGACGCGGGCCGTCGACACCGGGTACGCCAACGGCTACGGCAAC
 AGCAAGTGGGCCGGTGAGGTGTTGTTGCGCGAGGCGCACGATCTGTGCGGGCT
 CCCCGTGGCGGTGTTCCGGTGCACATGATCCTGGCCGACACCACCTACGCCG
 GCCAGCTCAACCTGCCGGACATGTTACCCGGATGATGGTGAGCCTGGTGACC
 ACGGGTATCGCGCCGAAGTCGTTCCACCCACTCGACGCAAAGGGCCACCGGCA

FIG 4 (Cont.)

GCGCGCCACTACGACGGGCTGCCGGTCGAGTTCGTCGCCGAATCGATCTCCG
 CGCTGGGAGCGCAGGCGGTTCGACGAGGCGGGAACGGGTTTCGCCACCTACCAC
 GTGATGAACCCCCACGACGACGGGATCGGCCTCGACGAATTCGTCGACTGGCT
 CGTCGAGGCGGGGTATCGCATCGACCGCATCGACGACTACGCGGCCTGGCTGC
 AACGGTTCGAAACCGCGCTGCGGGCGCTGCCCGAGCGCACTCGGCAGTACTCA
 CTGCTCCCCTGCTGCACAACCTACCAGCGGCCTGCGCACCCGATCAACGGGGC
 GATGGCCCCGACCGACCGGTTCCGCGCCGCGGTGCAGGAGGCAAAGCTCGGCC
 CGGACAAGGACATTCCCACGTCACCCCGGGGTGATCGTCAAGTACGCCACC
 GACCTGGAAGTCTGGGGCTGATCTAG

Secuencia de aminoácidos (SEQ ID N0:52)

>uniprot|A3PYW9|A3PYW9_MYCSJ Dominio de tioéster reductasa

MSTETREARLQQRIAHLFATDPQFAAARPDPRISDAVDRDDARLTAIVSA
 VMSGYADRPALGQRAAEFATDPQTGRRTMELLPRFDITITYRELLDRVRAL
 TNAWHADGVRPGDRVAIILGFTGIDYTVVDLALIQLGAVAVPLQTSAAVEA
 LRPIVAETEPMLIATGVDHVDAAAELALTGHRPSQVVVFDHREQVDDERD
 AVRAATARLGDAVPVETLAEVLRREGAHLPAVAPHVFDEADPLRLLIYTSG
 STGAPKGAMYPESKVAGMWRASAKAAWNNDQTAIPSIITLNLPM SHVMGR
 GLLCGTLSTGGTAYFAARSDLSTLLEDLRLVRPTQLSFVPRIWDMFLQEF
 VGEVDRRVNDGADRPTAEADVLAELRQELLGGRFVTAMTGSAPISPEMKT
 WVETLLDMHLVEGYGSTEAGAVFVDGHIQRPPVLDYKLVDPDLGYFSTD
 RPHPRGELLVRSTQLFPGYYKRPDVTAEVFDDDDGFYRTGDIVAELGPDQL
 QYLDRRNNVLKLAQGEFVTISKLEAVFAGSALVRQIFVYGN SARSYLLAV
 VVPTDDAVARHDPASLKTAISASLQQAAKTAGLQSYELPRDFLVETQPFT
 LENGLLTGIRKLARPKLKARYGDRLEALYVELAEGQAGELRTLRRDGAKR
 PVAETVGRAAAAALLGAAAADVRPDAHFTDLGGDSLSALTFGNLLQEIFGV
 DVPVGVIVSPAADLASIAAYIETE QASTGKRPTYASVHGRDAEQVRARDL
 TLDKFIDAETLSAATELPVPIGEVRTVLLTGATGFLGRYLALDWLERMAL
 VDGKVICLVRAKDDAAARKRLDDTFDSGDPKLLAHYRKLAADHLEVLGD
 KGEADLGLPHQVWQRLADTVDLIVDPAALVNHVLPYSQLFGPNALGTAEL
 IRLALTTRIKPFTYVSTIGVGAGIEPGRFTEDDDIRVISPTRAVDTGYAN
 GYGNSKWAGEVLLREAHDLCLPVAVFRCMILADTTYAGQLNLPDMFTR
 MMVSLVTTGIAPKSFHPLDAKGHRQRAHYDGLPVEFVAESISALGAQAVD
 EAGTG FATYHVMNPHDDGIGLDEFVDWLVEAGYRIDRIDDYAAWLQRFET
 ALRALPERTRQYSLLPLLHNYQRPAHPINGAMAPTDRFRAAVQEA KLGPD
 KDIPHVTPGVIVKYATDLELLGLI

FIG 4 (Cont.)

CAR70557

Secuencia de nucleótidos (SEQ 10 N0:53)

>gi | 219932450:561817-562104 Mycobacterium leprae Br4923, secuencia de genoma completo

ATGAGCACCGTGGATGGTGTGTGTGATGAAGACATATATTACTATTATGACTG
CATATATTGCGCTGCCGATGTTGCCAGCCCGACTGGGTATTTACCCGGAGGAC
TCTTTAGCTTGCGAAGCTGCTTGTTATACGTGGGGAGAAGCAGTATGTCTTC
GGCCTGAATGACTCGACGAGTGTGCGCCCCGGGCGGGCAACATGCCCTGGCT
'GGACGACCCAGGCCTGTCTGCTGTCATTGCGAGTTCGGTCGCAGCGGCTGAAT
TGGCTGCTGCGCGATCCTGGTAA

Secuencia de aminoácidos (SEQ ID N0:54)

>gi | 219932734|emb| CAR70557.1| Proteína hipotética [Mycobacterium leprae Br4923]

MSTVDGVCEDEIYYYDCIYCAADVASPTGYLPGGLFSLAKLLVIRGEKQYVF
GLNDSTSVGPGPGNMPWLDDPGLSAVIASSVAAAEELAAARSW

YP 001220863.1

Secuencia de nucleótidos (SEQ ID N0:55)

>uniprot| A5CM59| A5CM59_CLAM3 Supuesta acil-CoA sintetasa

ATGAGCACTGAGCAGATGGGCACCGAGCAGATGGGGAGCCAGCACGAGGACAC
GTCGATCGAGGCGATCTTCGCGCAGCACGCCGACCGGACGGCGCTGCGGCAGC
GATCCGGGCGGACATCACGGACATGGGCTTCCGGGAGCTGTGGGACCGGGCG
GGCGCGCTGGCCGCGGCGCTCGGCGAGACCGTGTCCGCCGGAGACCGCATCGC
GGTGCTCGGCACGGCGACCGCGGACGCCGTCACGCTCGACCTCGCCGCGTGGA
TCCTCGGCGCGGTGAGCGTGCCCCTCCAGGCGAGCGCGCCCGTTCGCGGCCCTG
CGCGCGATCGTCGAGGAGACGACGCCGGTGTGGATCGCGGCCACCGCCGACCA
GGCCGCGACGGCCCGGGCGGTGCGCCGAGGCATCGGGCGACGGCATCCGGACGA
TGCGGCTCGACACCGACACCGACGCGGACACCGACACCGACGCCGCCCTGACG
CTCGGCGCGCTCGTCGCCCGCGGGGCGGCCTGCGTCGCCGGAGCCCGTGGCA
CCCCGCACCCGGCGACGACCCGCTCGCGCTCCTGCTCTACACGTCGGGCAGCA
CGGGCACGCCGAAGGGCGCGATGTACACGCGCTCCATGGTTCGAGCGGATGTGG
CACGCCCTCCGGCCCCGACCCCGCCGCGCCCGCCGACGCATCCACGACCGCGGA

FIG 4 (Cont.)

CGACGGAGACGCCGCAGCCATCGTCGGGTACGCGTACCTCCCCATGAGCCACC
 TCACCGGCCGCTCCTCCCTGCTCGCCACGCTCGGCCGCGGGCGGCACGGTTCGCG
 CTCGCGACCTCGACCGACCTGTTCGACGCTCTTCGACGACCTGCGGACCTTCGC
 CCCGACCGAGTTCGTCTTCGTGCCGCGCGTTCGCGGAGCTGGTTCGCGCAGGAGG
 GCGACCGCGAGGAGCAGCGCCGACTCACCGCCGGCAGCACGGACCGGGACGCG
 GTCCGCGCCGAGGTCCAGGCCGACCTGCGCGCCCGCGCGTTTCGGCGGACGGAT
 CCACCGGGCGATCTGCACCAGCGCACCGCTGACGCCGGAGCTCCGCACGTACA
 TCGAGGGATGCCTCGGGCTGACGCTGCACGACCTGTACGGGTTCGACCGAGGCC
 GGCGGCATCCTCCACGACGGGGTGATCCAGCAGCCTCCCGTACACCGAGCACAA
 GCTCGTCGACGTCCCCGAGCTCGGGTACCAGCACCGACCGCCCGCACCCGC
 GGGCGAGCTGCTCGTCAAGAGCGCCTCCGTGATCGCCGGGTACTTCCGCCGG
 CCGGACGTACCGCCGCGGTGTTTCGACGAGGACGGCTTCTACCGGACCGGCGA
 CGTCATGGCGCAGACCGGACCCGGCACCTACGAGTACGTTCGACCGTTCGGAACA
 ACGTGATCAAGCTGTTCGAGGGCGAGTTCGTTCGCGGTGGCATCGCTCGAGGCG
 ACGTACGGCGGGACGCCCGAGGTGCACCAGATCGCGCTGCACGGCGACAGCCG
 GCACGCGTTCCTCGTCGCGGTTCGTTCGCGCGGACCCCGCGGGCGTTCGGAGC
 GCGACATCCTCGCGGCCCTCCAGCGCACCGCCCGCGAGCACGGCCTCGCCCC
 TACGAGGTGCCCGCGGCGTGATCGTTCGAACCGGATCCGTTACGGTTCGACGG
 CGGCATGCTCTCCGACGCCGCAAGCTCCTGCGCCTGCGCCTCACGCAGCGGT
 ACGGCGAGCGCCTCGCCGCCCTTACGACGCGCTCGAGGAGCAGCAGAGCGGC
 ACCCTCGTCGCCGCGCTCCGCGAACCGCGCCGACGACGAGCCGACGGTTCGACAC
 GGTGGTTCGCGCGCCGCCCTCCTGCTCCTCGGGGCGGAGGTGTCGCCCGCGACCG
 CCGCCGCTGCCCGGTTCTCCGACCTCGGCGGGGACTCGCTGTCCGCGCTGACG
 TTCTCCGGGATCCTCGAGGACGTCTTCGGCACCGAGGTGCCCGTTCGGCGTCTCT
 CACCGACCCGACCAACGACCTCGCCGCCGTCGCCGCCTACGTGGAGCGCTCCG
 CCTCCGACGACCGGCCGACCGTGACCCGCGTGCACGGCGCCGGCGCATCCACC
 CTCCGCGTGGGCGACCTCCGGCTCGACCGGATGCTCGGCGGCATCCCGACGCC
 GGTCCCCCGGGCCTCGGCCGCCCGGCCGGGATCCCGCACGGTCTGCTACCG
 GCGCGAACGGTACCTCGGCCGGTTCTGGCCATCGACTGGCTCGAGCGCCTC
 GCTGCGACGGGCGGCACGCTGGTGTGCATCGTTCGCGGGCGCCGACGACGCCGA
 CGCCCGGCGCCGCCCTCGAGGCCGCGTTCGCCGCGGATCCCGCGTTCGCCCGGC
 GCTTCGCCGAGCTGTTCGGGCTCGCTCGAGGTGCTGGCCGGAGACGTCAGCGAG
 CACCGCCTCGGCCTCGACGACGAGCGGTGGATCGACCTGGCCGCGCGGGTTCGA
 CCTCGTCGCGCACGCGGCTGCCCTCGTGAACCACGTCCTCCCGTACTCGGCGC
 TGTTCCGGTCCGAACGTCGTTCGGCACCGCCGAGGCGATCCGCCTGGCGATCGCC
 GCCGGCAGCGTGCCCGTACCTTCGTCTCGAGCGTTCGCGGTTCGCGGGCGGGCGC
 GCGGCCGGGCGCGACCGCCGACGCCGAGCCGTCGGCGCCCGGCGCGCTTCGACG
 AGCACGCCGACATCCGCGCCACGATCCCCGAGTGGGCCGTTCGGCGACGAGTAC
 GCCAACGGGTACGGCGCGAGCAAGTGGGCGAGCGAGGTGCTGCTCCGCGAGGC
 GCACGAGCACACGGCGTCCCGTGGCCGTGTTCCGCTCCGACATGATCCTGG
 CGCACCCCGCTGGCGCGGTTCAGGTGAACCTCCCCGACGTCTTCACCCGGCTG

FIG 4 (Cont.)

ATCTGGAGCGTGCTCACCACCGGCCTCGCCCCGCATCGTTCGTGAGGCGCGG
 CCCCACGCGGCGAGCGGCAGCGGTTCGCACTACGACGGGCTGCCGGCCGACTTCA
 CGGCGGCGGCGATCGACGGGATCGGCGCGGGCGCTCACCGAGGGGCACCGCACC
 TTCAACGTCGTGAACCCCCACGACGACGGCGTCTCGCTCGACACCTTCGTCTGA
 CTGGATCCGCGAGGACGGCCACGACATCGCGCGCGTGGACGACCACGCGGAGT
 GGGTCGACCGGTTCCGCGCGGGCGCTGGGAGCGCTCCCGGACGCGGACCGCGCC
 CGGTCCGTCCTGCCGCTGATGCACGCGTTCGCCTCGCCCCGAGGAGCCGCACGC
 CGGCTCGGCGATCCCGGCGGATGCGTTCGCCGAGGCCGTCCGCGCGGTGCGCC
 CGCTCGGGTCGCCGGACATCCCGTCTCTCGACCACGCGCTCATCGCCAAGGTC
 GCCGACGACCTCGCGTTCCTGGGGCTGCTCGCGCCGGCGCGGGCGGGCGGCTGC
 CTGA

Secuencia de aminoácidos (SEQ ID N0:56)

>uniprot|A5CM59|A5CM59_CLAM3 Supuesta acil-CoA sintetasa

MSTEQMGTEQMGSSQHEDTSIEAIFAQHADRRTALRQRSRGPDIITDMGFRELWDRA
 GALAAALGETVSAGDRIAVLGTATADAVTLDLAAWILGAVSVPLQASAPVAAL
 RAIVEETTPVWIAATADQAATARAVAEASGDGIRTMRLDITDADITDAAALT
 LGALVARGAGLRRRSPWHPAPGDDPLALLLYTSGSTGTPKAMYTRSMVERMW
 HALRPDPAAPADASTTADDGDAAAIVGYAYLPMSHLTGRSSLLATLGRGGTVA
 LATSTDLSLTFDDLRTFAPTEFVVFVPRVAELVRQEGDREEQRRLTAGSTDRDA
 VRAEVQADLRARAFGGRIHRAICTSAPLTPELRTYIEGCLGLTLHDLYGSTE
 GGILHDGVIQQPPVTEHKLVDVPELGYRTTDRPHPRGELLVKSASVIAGYFRR
 PDVTAAVFDEDEGFYRTGDVMAQTGPGTYEYVDRRNNVIKLSQGEFVAVASLEA
 TYGGTPEVHQIALHGDSRHAFVAVVVPADPAASERDILAAALQRTAREHGLAP
 YEVPGRVIVEPDPFTVDGGMLSDAGKLLRLRLTQRYGERLAALYDALEEQQSG
 TLVAALRERADDEPTVDTVVRAALLLLGAEVSPATAAAARFSDLGGDSL SALT
 FSGILEDVFGTEVPVGVLTDPNTDLAAVAAYVERSASDDRPTVTRVHGAGAST
 LRVGDLRLDRMLGGIPTVPRASAARPGSRTVLLTGANGYLGRFLAIDWLERL
 AATGGTLVCIVRGADDADARRRLEAAFAADPAFARRFAELSGSLEVLAGDVSE
 HRLGLDDERWIDLAAARVDLVAHAAALVNHVLPYSALFGPNVVGTAEAIRLAIA
 AGSVPVTFVSSVAVAGGARPGATADAEP SAGALDEHADIRATIPAWAVGDEY
 ANGYGASKWASEVLLREAHEHHGVPVAVFRSDMILAHPRWRGQVNLPDVFTRL
 IWSVLTTGLAPASFVRRGPDGERQRSHYDGLPADFTAAAIDGIGAAALTEGHRT
 FNVVNP HDDGVSLDTFVDWIREDGHDIARVDDHAEWVDRFRAALGALPDADRA
 RSVLPLMHAFASPEEPHAGSAIPADAF AEAVRAVRPLGSPDIPSLDHALIAKV
 ADDLAF LGLLAPARAAAA

FIG 4 (Cont.)

YP 001537947.1

Secuencia de nucleótidos (SEQ ID NO:57)

>uniprot|A8M8D3|A8M8D3_SALAI Dominio de tioéster reductasa

GTGACCACCACGGAGCAGACCCTCACCGAGCGGCTCATCGCCGAGGACGAGCA
 GATCCGGCGGGCCCAGGTCAGCGCCGAAGTCTCCGCCGCGATGCGGGTGCCGG
 GCATGTCGCAGGCCAGATCGTGGCCGCCGATTACCCGGTTACGCCGACCGC
 GCCGCCCTGGGTGAGCGCGCCCGCAAGCCGTCACCGACC^oGGTCACCGGCCG
 CACCACCACCGGCTTCTGCCATGGTTCGACACCATCACCTACGGCGAGGTCC
 GGTCGCGGGTGCTGGCGATCTCCGCCGCTGGTGGCACGACGTGGACGCTCCA
 CTCCGTCCCGGCGCCTTCGTCGTTTTCGGTGGCGTCCCCAGCGCCGACCTCGT
 GACGGTCGAGCTCGCGGTGCTACACACCGGCGCGGTCAGTGTGCCACTGCAGG
 TCAGCTCCACCGCCGAGCAACTGCGCCCGATCCTCGACGAGGCCGCCCGCTC
 ATCGTGGCCACGAGCGTGGACCGGCTCGCTGTGGTGACCGCGGGCGATGTCGGG
 CAACGCGTCGGTGCGCCGGATCATGGTCCTGAACCACGACGCAGCGATCACCG
 CCCACCGGGATGCCGTGGACGCCGCGCATCGGCGCTCGCCGGCACCGCAGTC
 GTCGTGCACACATTGACCGAGGTGTTGGACCGTGGACGGGGCCTGCCCGCCCC
 TGAGCCCTACGCGGGCGCCCACGGGGGAGGATCCCCTGTCGCTGTTGATCTACA
 CCTCGGGCAGTACCGGTACGCCAAGGGCGCAATGTTTTCCGGAGAGCATGACC
 CGCGCCA^oACTGGGTGCGTTTTGACCCCAAGCCGACCGACATGGCGGTCATCCG
 GCTCAACTACCTGCCGCTGAGCCACAACGTCGGCCGCATCGTGCTGTTGAGG
 CGCTCGCGGTGGGCGGCATCGCCTTCTTCACCGCACACAGTGACCTGTCCACG
 CTCCTGGAGGACATGGCCCTCGCCCGGCCACCGACCTGTTCTTATCCCGCG
 GCTGTGCGACATGCTCGCCAGCGCCACGACAGCGAACTGGCCCGCCGCCGCA
 TCACCACCGCGGATCACGAGGGGGTCCGACAGGTCCACACCCATCTGCGCGAG
 GCTGTCTCGGCGGCAGGGTGACCCGCGCGATGTGCTGTCGCGCCGCTGAG
 CCCACAGCTGCGTCCGTTTCGTGGAGTCGTGTCTCGGCTTCGCGGTGCACGATG
 TCTTCGGGTCGACCGAGGCCGGCGGGCTGCTCGTCAACGGCCGGGTGCTGCGC
 CCGCCGGTGCTCGACTATCGCTTGGTCGACGTCCCCGACCTCGGCTACTTCAC
 CACCGACCGTCCGTACCCTCGCGGGGAGCTGCTGGTGCGGACCGCGACGATCA
 TCCCCGGCTACTACCAGCGGCCCGAGCTCAACGCCGAGCTGTTACCGAGGAC
 GGCTACTACCGCACCGGCGACATCATGGCCGAGTACGGCCCCGACCACCTCGG
 CTATGTCGACCGCACACGAGCGTGCTGAAGCTGTCACAGGGCGAGTTCGTGG
 CCGTGTACGGTTGGAGGA^oACTGTTCCCGCCTCCCCGCTGATCCGGCAGATC
 TACCTGTACGGCAACAGCGAGCGGCCGTACCTGCTCGCCGTGGTTCGTGCCAC
 GGAGGAGGCGCACGCCGCCACCCGGGAACCCGCGGCGCTCAAGGCGGTGCTCG
 GCGAGTCGCTGCAACGCATCGCTCAGCAGCACGGCCTGCACCCGTACGAGGTG
 CCGCGCGACCTCCTCATCGAGACCACCCCGTTCAGCACCGCCAACGGTCTGCT

FIG 4 (Cont.)

CTCCGACATCCGTAAGCCCCTGCGTCCGAAGCTCAAGACCCGGTACGCTCCTC
 GACTCGAAGCGCTCTACACCGAGCTCGCCGAGCGCGAGGCCGACCGGATCCGC
 ACGCTGCGCGACGCCGGTTCCGCGCAACCCGTGCTGCCCGCGTTGCGCGAGGC
 TGCCCGGGCGTTTCTCGGCCGCCAGGCGCAGCGCTCGACGTGAACGACCGCT
 TTGTGGACCTCGGGCGGCGACTCCCTGTCCGGCCCTGGCCCTGTGGAACCTGCTG
 AGCGACATCTTCGAGGTCCGCGTCCCGGTCCGGCATCATGATCAGCGCGACCGG
 CACGCTCGGTTCCGTGGCGGCCCTGGATCGAGGCCGAGCGTGCCACCGCCGGAG
 CGGGTATCGGCCGCGCGACGCCACCTCCGTGCACGGTGCGAACCTCACCCAG
 GTACACGCCGATGACCTGACCCTCGGCACGTTTCTCGACGTGACGACCCTCGC
 CGCCGCTGCCTGCCTGCCCCGGGCGCCGCTGTCCGACCCGCGCGTGGTGCTGC
 TGACGGGTGCGACCGGCTATCTGGGCCGGTTTCTGGCCCTCGAGTGGCTGGAC
 CGCCTTTCCCGTAGCGGCGGGACGCTCGTGTGCGTGGTGC GCGCCCGCCGACGA
 TCGGGAAGCCGCGCGCCGCTGGAAAGTGTCTATGGCTCCAGCGACCCCGAGT
 TGCTGGAGCGCTTCCGTTCACTCGCCGGCCACGTGCGCGTGTGGCCGGCGAT
 GTTGCCGAAGCCAGGTTCCGGCCTGCCGGCCGGGGTGTGGCAGGAACTGGCCGA
 AACGGTGGACCTGATCGTGCCTCCGCGGCACTGGTCAACCACGTTCTCCCGT
 ACGAACAGCTGTTCCGGGCCAACGTGGCGGGAACGGCGGAACTGGTGC GCGCTC
 GCCGTACGCGTACGGGTGAAGGGAATTGCCCTTCTCTCCACCGTTGCCGTGAT
 CACCTCGCAGACCACGACACCCGACGAGGACGCGGACATCCGGCAGGCGAGCC
 CGCACCGGGTGCTCGACGACAGCTACGCGAACGGCTACGCGGCCAGCAAGTGG
 GCAGGTGAGGTGCTCCTGCGACGCGCCACGAGGAGTACGGCGTGCCGGTCAG
 CGTGTTCGCTCGGACGTCATCCTGGCCCACTCCCGCTACGCCGGGCAGCTCA
 ACGTCCCGGATATGTTACGCGCCTGCTCCTGAGCATCCTGGCGACCGGTATT
 GCCCCAGCGTCGTTTTATCGCACCCGGCCCGGACGGCGAACGCCAGCCGGCGCA
 CTACGACGGTCTCCCGGTCGACTTACCCGCGGCGGCCGTAGCCGCGGTGGGTG
 TCACCGAGGGACACCGCACCTTCAACGTACTGAATCCACACGAGGACGGCATC
 GGGCTGGATACCTTTGTGGACTGGCTCGTGGCAGCCGGACACCCGGTGCAGCG
 CATCGCGGACCACGACGAGTGGGTGACCCGCTTCGCCACGGCCATGCGTGGGC
 TGCCTGAACGCCAGCGCCGAGCTCGATCCTGCCGCTCCTACACGCCTTTGCC
 GAGCCCGCTCCGCCGACCTTCGGATCCAGACTGCCGACGGACCGGTTTCGCGC
 CGCCGTGAAAGCCGCCAACGTGGTCCCGGCAACGAGATCCCGCACCTCGATG
 CGGCCCTCGTCACCAAGTACGCCGACGACCTCAGGCTGCTCGACCTTCTCTGA

Secuencia de aminoácidos (SEQ ID N0:58)

>uniprot|A8M8D3|A8M8D3_SALAI Dominio de tioéster reductasa

VTTTEQTLTERLIAEDEQIRRAQVSAEVSAAMRVPGMSQAQIVAAGFTGYADR
 AALGERAREAVTDPVTGRTHRLLPWFDTITYGEVRSRVLAI SAAWWHDVDAP
 LRPGAFVVS VGVPSADLVTVELAVLHTGAVSVPLQVSSTAEQLRPILDEAAPL

FIG 4 (Cont.)

IVATSVDRDLAVVTAAMSGNASVRRIMVLNHDAAITAHRDAVDAARSALAGTAV
 VVHTLLEVLDRGRGLPAPEPYAAPPTGEDPLSLLIYTSGSTGTPKGAMFPESMT
 RANWVRFDPKPTDMAVIRLNYLPLSHNVGRIVLFEALAVGGIAFFTAHSDLST
 LLEDMALARPTDLFLIPRLCDMLAQRHDSELARRRITTADHEGVRQVHTHLRE
 AVLGGRVTRAMSLSAPLSPQLRRFVESCLGFVHDVFGSTEAGGLLVNGRVLR
 PPVLDYRLVDVDPDLGYFTTDRPYPRGELLVRTATIIPGYQRPENAEFTED
 GYYRTGDIMAEYGPDLHGYVDRRTSVLKLSQGEFVAVSRLEELFAASPLIRQI
 YLYGNSERPYPYLLAVVPTTEEAHAATREPAALKAVLGESLQRIAQQHGLHPYEV
 PRDLLIETTPFSTANGLLSDIRKPLRPKLKTRYAPRLEALYTELAEREADRIR
 TLRDAGSAQPVLPALREAAAFVLRPGAALDVNDRFVDLGGDSLALSNLL
 SDIFEVRVPGIMISATGTLGSVAAWIEAERATAGAGIGRATPTSVHGANLTQ
 VHADDLTLGTFLDVTTLAAAACLPRAPLSDPRVLLTGATGYLGRFLALEWLD
 RLSRSGGTLVCVVRAADDAEAARRLESVYGSSDPELLERFRSLAGHVRVLAGD
 VAEARFGLPAGVWQELAEVLDLIVHSAALVNHVLPYEQLFGPNVAGTAEVRL
 AVSVRVKGI AFLSTVAVITSQTTTPDEDADIRQASPHRVLDDSYANGYAASKW
 AGEVLLRRAHEEYGVVSVFRSDVILAHSTRYAGQLNVPDMFTRLLLSILATGI
 APASFYRTGPDGERQPAHYDGLPVDFTAAVAAVGVTEGHRTFNVLNPHEDGI
 GLDTFVDWLVAAGHPVQRIADHDEWVTRFATAMRGLPERQRRSSILPLLHAF
 EPAPPTFGSRLPTDRFRAAVKAANVVPVNEIPHLDAALVTKYADDLRLDLL

YP_001703694.1

Secuencia de nucleótidos (SEQ ID NO:59)

>uniprot|B1MCR9|BIMCR9_MYCAB Probable ácido graso-CoA ligasa FadD

ATGACCGTGACCAACGAAACCAACCCACAGCAGGAGCAGCTATCCCGCCGTAT
 TGAAAGTCTGCGCGAAAGCGATCCGCAGTTCCGGGCGGCCAGCCCGACCCGG
 CGGTCGCCGAACAGGTGCTGCGCCCGGGCCTGCATCTTTCTGAAGCCATTGCG
 GCGTTGATGACTGGATACGCTGAGCGCCCGGCGCTCGGTGAGCGCGCACGCGA
 GTTGGTCACCGACCAGGATGGCCGCACCACGCTGCGCCTGTTGCCACGCTTCG
 ACACCACCACATACGGCGAATTATGGTCCCGCACACATCAGTCGCCGCTGCA
 TGGCACCACGACGCCGCCACCCGGTTAAGGCCGGCGATCTGGTGGCCACCCT
 GGGATTCACCAGCATCGACTACCCGTGCTGGATCTGGCGATCATGATCCTCG
 GTGGCGTGGCGGTTCCGCTACAGACCAGCGCCCGGCTTCGCAGTGGACGACC
 ATTCTGGCCGAAGCGGAACCCAACACTCTTGCGGTAAGCATCGAATTGATCGG
 CGCTGCAATGGAATCTGTGCGGGCCACGCCTTCCATCAAGCAGGTCGTCGTGT
 TCGACTACACCCCGAGGTGATGATCAACGGGAGGCATTCGAGGCAGCAAGC
 ACACAACCTCGCCGGCACCAGGCATCGCCATTGAGACCCTCGATGCCGTCATCGC
 CCGCGGCGCCGCACTTCCGGCCGCACCCTCTACGCACCATCGGCCGGCGACG

FIG 4 (Cont.)

ATCCGCTGGCGCTGCTCATCTACACCTCCGGCAGCACCGGGGCTCCAAAGGGC
 GCCATGCACAGCGAAAACATCGTGCGCCGCTGGTGGATTTCGTGAGGACGTCAT
 GGCCGGCACCAGAACCTGCCATGATCGGGCTGAACTTCATGCCGATGAGTC
 ACATCATGGGACGCGGCACCCTCACCTCCACCCTGTCTACCGGTGGAACCGGA
 TACTTCGCGGGCTCCAGTGACATGTCAACGCTCTTCGAGGACATGGAGCTGAT
 CCGCCCGACGGCCCTGGCCTTGGTTCACGCGTGTGCGACATGGTGTTCACGC
 GATTCCAGACCGAGGTGGACCGGCGTCTGGCGAGCAGCGACACCGCCAGTGCC
 GAGGCCGTTGCGGCCGAGGTCAAGGCCGATATCCGTGACAACCTCTTCGGTGG
 CCGCGTATCGGCGGTTCATGGTTCGGTTCTGCTCCGTTGTCCGAGGAGCTGGGTG
 AGTTCATCGAATCCTGCTTCGAGCTGAATCTGACCGATGGCTACGGCTCCACC
 GAAGCCGGCATGGTGTTCGCGACGGCATCGTGCAACGCCCGCCGGTCATTGA
 CTACAAGCTGGTTGACGTGCCCGAACTGGGCTACTTCTCCACCGACAAGCCGC
 ACCCGCGCGGTGAGCTGCTGCTGAAGACCGACGGCATGTTCCCTCGGGTACTAC
 AAACGCCCCGAGGTGACTGCCGGCGTCTTCGACGCGGACGGTTTTTACATGAC
 CGGCGACATCGTCGCCGAGCTGGCCCACGACAACATCGAGATCATCGATCGCC
 GCAACAACGTGCTCAAACCTCTCACAGGGAGAGTTTGTCGCGGTCGCCACCTTG
 GAGGCCGAGTACGCCAATAGCCCTGTGGTGCACCAGATCTACGTCTACGGCAG
 CAGCGAACGGTTCCTACCTGCTAGCAGTCGTGGTGCCGACGCCGGAGGCCGTGG
 CCGCCGCCAAGGGCGACGCGGCGGCACTCAAGACGACCATCGCGGACTCGCTG
 CAGGACATTGCCAAGGAGATCCAGCTGCAGTCTTACGAAGTCCCCCGTGACTT
 CATCATCGAACCGCAGCCATTCACCCAGGGCAACGGCCTGCTGACGGGTATCG
 CCAAGCTGGCGCGTCCGAACCTGAAGGCGCACTATGGACCGCGGCTGGAGCAG
 ATGTACGCCGAAATCGCCGAGCAGCAGGCTGCCGAGCTTCGGGCGTTGCACGG
 AGTGGACCCAGACAAGCCCGCGCTGGAAACGGTCTCAAGGCGGCGCAGGCC
 TGCTCGGCGTCTCGTCGGCCGAACTGGCCGCGGACGCGCATTCACCGATCTA
 GGTGGCGATTTCGCTGTCCGCACTGTCTTCTCGGATCTGCTGCGCGATATCTT
 CGCGGTGCAAGTACCGGTTCGGAGTCATCGTCAGTGCCGCAAACGATCTCAGCG
 GTGTTGCGAAATTTGTTGATGAACAACGCTATTTCGGGCGGGACGCGGCCGACC
 GCGGAGACGGTGCACGGCGCCGGGCATACGGAGATCCGGGCCGCGGACCTGAC
 CCTGGATAAGTTCATCGACGAGGCCACCCTGCATGCGGCACCGTCGCTTCCGA
 AGGCCGTCGGGATCCACACACCGTCTGCTCACCGGGTCCAACGGCTACCTG
 GGCCACTACCTGGCACTGGAATGGCTTGAGCGCCTGGACAAGACAGAAGGCAA
 GCTGATCGCCATCGTCCGCGGTAAGAATGCCGAGGCCGCTACCGCCGCCTCG
 AGGAAGCCTTCGACACCGGCGACACGCAGCTGTTGGCGCACTTCGGGTGCTG
 GCCGACAAGCACCTCGAAGTACTGGCCGGCGATATCGGCGACCCCAACCTTGG
 CCTGGATGCCGACACCTGGCAGCGCCTGGCCGACACCGTCGACGTCATCGTGC
 ACCCCGCCGCCCTGGTCAACCACGTACTGCCCTACAGCCAGCTGTTTCGGACCG
 AATGTCGTCGGCACCGCCGAGATCATCAAGCTGGCCATCACTACCAAGATCAA
 GCCGGTCACTACCTGTCCACGGTCGCGGTCGCGGCATATGTGATCCGACGA
 CATTTCGACGAAGAGTCCGATATCCGGCTCATCAGCGCGGTGCGTCCCGTGGAC
 GAGCTGTACGCGAACGGCTACGGCAACAGCAAGTGGGCCGGCGAGGTACTGCT

FIG 4 (Cont.)

GCGCGAAGCCCACGATCTGTGCGGACTACCCGTCGCGGTCTTCCGCTCCGACA
 TGATCTTGCCCCACAGCCGCTACACCGGACAGCTCAACGTGCCCGACCAGTTC
 ACCCGACTAATCCTCAGCCTCATCGCCACCGGAATCGCACCCGGCTCCTTCTA
 CCAAGCACACGCCACCGGCGAACGCCCACTCGCCACTACGACGGGCTACCCG
 GTGACTTCACCGCCGAGGCGATCACCACGTTGGGCACCCAGGTGGTCGACAGC
 TACGAGACCTACGACTGCGTGAACCCGCATGCAGACGGAGTCTCGCTGGACAA
 CTTCGTCGACTGGCTCATCGAAGCCGGCTACCCCATCGCACGCATCGACAAC
 ACACCGAATGGTTCACCCGCTTCGACACCGCCATCCGAAGCCTCCCCGAAAAA
 CAGAAACAACACTCCCTACTACCACTGCTCCACGCATTTCGAACAGCCGTCCGC
 CGCCGAGAACCACGGCGTCGTCCCGGCAAAGCGTTTCCAGCACGCTGTGCAGG
 CCGCCGGAATCGGTCCGGCCGGGCAAGACGGCACTACCGACATTCCCCACCTG
 TCGCGGCGGCTGATCGTGAAATACGCCAAGGACCTCGAACAGCTCGGACTCCT
 ATGA

Secuencia de aminoácidos (SEQ ID NO:60)

>uniprot|B1MCR91BIMCR9_MYCAB Probable ácido graso-CoA ligasa FadD

MTVTNETNPQQEQLSRRIESLRESDPQFRAAQPDPAVAEQVLRPGLHLSE
 AIAALMTGYAERPALGERARELVTDQDGRITLRLLPREFDTTTYGELWSRT
 TSVAAAWHHDAHPVKAGDLVATLGFTSIDYTVLDLAIMILGGVAVPLQT
 SAPASQWTTILAEAEPNTLAVSIELIGAAMESVRATPSIKQVVVFDYTPE
 VDDQREAFEAASSTQLAGTGIAIETLDAVIARGAALPAAPLYAPSAGDDPL
 ALLIYTSGSTGAPKGAHSENI VRRWWIREDVMAGTENLPMIGLNFMPMS
 HIMGRGTLTSTLSTGGTGYFAASSDMSTLFEDMELIRPTALALVPRVCDM
 VFQRFQTEVDRRLASSDTASAEAVAAEVKADIRDNLFGGRVSAVMVGSAP
 LSEELGEFIESCFELNLTGYYGSTEAGMVFRDGI VQRPPVIDYKLVDPVE
 LGYFSTDKPHPRGELLLKTDGMFLGYYKRPEVTAGVFDADGFYMTGDIVA
 ELAHDNIEI IDRRNNVLKLSQGEFVAVATLEAEYANSPVVHQIYVYGSSE
 RSYLLAVVVPTPEAVAAAKGDAAALKTTIADSLQDIAKEIQLOSIEVPRD
 FIIEPQPFQTQGNGLLTGIAKLARPNLKAHYGPRLEQMYAEIAEQQAELR
 ALHGVDPAKPALETVLKAAQALLGVSSAELAADAHFTDLGGDSLSALSFS
 DLLRDIFAVEVPVGVIVSAANDLSGVAKFVDEQRYSGGTRPTAETVHGAG
 HTEIRAADLTLDKFI DEATLHAAPSLPKAVGIPHTVLLTGSNGYLGHYLA
 LEWLERLDKTEGKLI AIVRGKNAEAAAYRRLEEFDTGDTQLLAHFRSLAD
 KHLEVLAGDIGDPNLGLDADTWQRLADTVDVIVHPAALVNHVLPYSQLFG
 PNVVGTAEI IKLAITTKIKPVTYLSTVAVAAAYVDPTTFDEESDIRLISAV
 RPVDELYANGYGNKSWAGEVLLREAHDLCLPVAVFRSDMILAHSRYTGO
 LNVPDQFTRLILSLIATGIAPGSFYQAHATGERPLAHYDGLPGDFTAETAI
 TTLGTQVVDSYETYDCVNPHADGVSLDNFVDWLI EAGYPIARIDNYTEWF

FIG 4 (Cont.)

TRFDTAIRSLPEKQKQHSLPLLLHAFEQPSAAENHGVVPAKRFQHAVQAA
GIGPAGQDGTDDIPHLRRLIVKYAKDLEQLGLL

YP 001703695.1

Secuencia de nucleótidos (SEQ ID NO:61)

>uniprot|B1MCS0|B1MCS0_MYCAB Probable ácido graso-CoA ligasa FadD

ATGACGATCGACGCCACCGCGGACAACACCAAGGAAGCACGTCGTCAGCGGTT
AGGCGACCGCATCAGGCGCCTATTACCGACGATGAGCAGTTCCGTGCCGCCA
AGCCCGATAACCGCGGTTGATACCGCCGTCGCCAGCCTGGTCTGCGTCTCGCC
CAGGTGGTCGCCACGATCATGAACGGGTACGCGGACCGTCCGGCGCTCGGGCA
CCGAGTCCAGGAGCTCGTCGCCGACGCCGCCGGCCGTTTCGACGCTGCGCCCGT
TGCCCGAGTTCGAGACGGTCACCTACGGAGAGCTGTGGGGCATGGCCCGCGCG
TTGGCCTCCACCTGGTACCACGATCCCGCCGCTCCGGTACGGGCCGGGGACTT
CGTCGCGATGCTCGGCTTACCAGCGTGGACTACACCGCTGTCGACTTGGCAT
GCATCCACCTTGGCGCGGTGGCGGTTCCATTGCAGACCAGCGCATCGGCATCC
AACTGGACCGCGATCCTGGCCGAATCGGAACCTGCCGTCTTGGCGGTAAGCGC
CGAGTTGCTCGATACGGCAATGGAATCGGTGCTCGCCACGCCGTGCTGCGGC
ACATCACCGTCTTTGACTATCATCCCGGTGTCGACGTGCAGCGCGAAAGTCTC
GAATCCGCACAGCACCGGATCGCCGAGGCCGGCCTGCCGATTTCCGGTAGACCC
GATACCCCTGGCGATCGGGCACGGGCGCGCCTTGCCGGATGCGCCGTTGTTC
CCGCAGAGGAGGGTACCGACCCGCTGGCCCTGGTGATCTACACCTCGGGAAGC
ACCGGAACTCCCAAGGGCGCCACCTATAGCGAAAAGATGGTCGCCAAGCCCTG
GCTGCGGGCCGACACGTTGAGCTCTAAGGCCGAAATTCCTTTGATCAACCTGA
ATTTTCATGCCAATGAGCCATGTGATGGGACGCGGTTAGTCTGGTCACTGCCCTG
GCCTGCGGCGGCCTGGCTTACTTTGCCGCGTCCAGCGACATGTCCACGCTGTT
CGAAGACATCACGCTTACGCGCCCCACCGTGGTGACTCGTGCCCCGTGTGT
GCGACATGCTTTTCCAGCGCTACCGCAACGAGGTTGAACGCCGTACCGGGCTT
GATCCGGCGGCCGACCTGGCCACCCTTGATGCCGATGTCAAGACCGATATCCG
CGAAAACCTGTTTCGGCGGGCGTGTCTGACAATCGTGTGCGGCTCTGCCCCAC
TGTCAGAGGAACTGGCCGCCTTCATCGAATCCTGCCTCGATGCCCGTATCACC
GATGGCTACGGCTCCACCGAGGCGGGCGTTCATCGTGCGCAACGGCCGCATTCA
GCGCCCCGCCGTCATCGACTACAAGCTGGTCGACGTGCCTGAGCTCGGTTACT
TCTCCACCGACAAGCCGCACCCGCGCGGCGAGCTGCTCGTGAAAGCCGAATCG
GTGTTTCGGCGGCTACTTCAAACGCCCCGACGTCACCGCCGACGTATTTCGATCC
CGACGGGTACTACAAGACCGGGGACATCGTCGCCGAGCTCGAGCCCGACAAGA
TCCAGATCGTGGACCGGCGCAACAACGTGATCAAGCTGTCCAGGGTGAGTTC
GTGGCGATCGCCAACCTGGAAGCCGAGTTCGCCAATAGTCCACTGGTGCATCA

FIG 4 (Cont.)

GATCTGCGTCTACGGCAGCAGCGAGCGGTTCGTATCTGCTAGCGGTGGTTCGTGC
 CGACCGCTGAGGCATATGAACAAAGCGGTGGAGATGAGGATCTACTCAAACGC
 CTGATCGCGGACTCTCTCGCGCAGGTTGCCCGCGAGGCCAACTGCAGTCCTA
 CGAGGTACCGCGCGACTTCTGCTGGAGACCGAACCCTTACCCTGCGCAACG
 GCCTACTGACCGGCATCGCGAAGCTGGCCCGACCGAAGCTCCATGAGAAGTAC
 GCGGCCCGCCTGGAGCAGCTGTACTCCGATATCGCCGCCGCCAGGCGCTTGA
 GCTGCAAGCACTGCACTCTGCCGGACATGAGGACAAGCCTGTCTTGGATACCG
 TGCAACGCGCGGTACGGCGTTGTTGGGGCTGTCCGGCGGCCGAGGTGAGCCCA
 GACGCGCATTTTCATCGACCTTGGTGGCGATTCACTATCCGCCCTTGCCTTCTC
 GGACCTGCTGCGCGATATCTTCACTGTGGAGGTTCCGGTTGGCGATATCGTCA
 GCGCCGCCAACGATCTGACCGCTATCGCACGCATCGTGGAAAGACACCGGGAA
 GCAGACGGTCATTCGGTAACTCCACCGCCGAATCCGTGCACGGTGCCGGGCA
 CCGCGAGATCCGGGCCGCGGACCTGACGCTGGACAAGTTCATCGACGCGGACA
 CCCTGCGCGCGGCCCGGCACTGTCCACATTCACCGGCACCCCGCACACGGTG
 CTGCTCACCGGCGCCAACGGCTACCTGGGGCGGTTCTTGGCCCTGGAATGGCT
 TGAGCGCCTGGACAAGACAGACGGCAAGCTGATCGCCATCGTCCGCGGTAAGA
 ATGCCGAGGCCGCCTACCGCCGCCTGGAGGAAGCCTTCGACACCGGCGACACG
 CAGCTGTTGGCGCACTTCCGGTCGCTGGCCGACAAGCACCTCGAAGTACTGGC
 CGGCGATATCGGCGACCCCAACCTTGGCCCTGGATGCCGACACCTGGCAGCGCC
 TGGCCGAGACCGTCGACGTCATCGTGCACCCCGCCGCCCTGGTCAACCACGTA
 CTGCCCTACAGCCAGCTGTTCCGACCCAATGTTCGTTGGCACCGCCGAAATCAT
 CAAGCTGGCACTCACCAAGATCAAGCCCATCACCTACCTCTCCACAGTGG
 CCGTGGCAATCTCGGTGGACCCCAAGGTATTCGATGAAGACTCCGACATCCGC
 ACGATCAGCGCGGTACGACCAATCAACGACGGCTACGCCAACGGATACGGCAA
 CGCGAAATGGGCTGGCGAGGTAAGTGTGCGCGAAGCCACGACCTGTGCGGAC
 TACCCGTCGCGGTCTTCCGCTCCGACATGATCTTGGCCCACAGCCGCTACACC
 GGACAGCTCAACGTGCCCGACCAGTTCACCCGACTAATCCTCAGCCTCATCGC
 CACCGGAGTCGCACCCGGCTCCTTCTACCAAGCACACGCCACCGGCGAACGCC
 CACTCGCCCCTACGACGGCCTGCCTGCGGATTTACGGCATCGGCCATCACC
 GCCCTCGGGCCCATCGAGGAGTTCACACCTACGATTCGGTGAACCCGCATGC
 CGATGGGATCTCGCTGGACAACCTCGTTCGACTGGCTCATCGAAGCCGGCTACC
 CCATCGCACGCATCGACAACACCGAATGGTTCACCCGCTTCGACACCGCC
 ATCCGAAGCCTCCCCGAAAAACAGAAACAACACTCCCTACTACCACTACTACA
 CGCGTACAGGCATCCACAACACCCACACAACGGCGCATTCTGCCC GCGATCA
 GGTTCAGTGAAGGCGTCCAGGCCATCTGAACGCCGACATCCCGCACCTCAGG
 CGGGAATCATCGCGAAATACGCGGCCGACCTGAAGCAGCTCGGGTTACTCTA
 G

FIG 4 (Cont.)

Secuencia de aminoácidos (SEQ ID N0:62)

>uniprot|B1MCS0|B1MCS0_MYCAB Probable ácido graso-CoA ligasa FadD

MTIDATADNTKEARRQRLGDRIRRLFTDDEQFRAAKPDTAVDTAVAQPGRLRLA
QVVATIMNGYADRPALGHRVQELVADAAGRSTLRPLPEFETVTYGELWGMARA
LASTWYHDPAAPVRAGDFVAMLGFTSVDYTAVDLACIHLGAVAVPLQTSASAS
NWTAILAESEPAVLAVSAELLDTAMESVLATPSLRHITVFDYHPGVDVQRESL
ESAQHRIAEAGLPISVDPIPLAIGHGRALPDAPLFTAEEGTDPLALVIYTSGS
TGTPKGATYSEKMOVAKPWLADTLSSKAEIPLINLNFMPMSHVMGRGSLVTAL
ACGGLAYFAASSDMSTLFEDITLTRPTVVTLVPRVCDMLFQRYRNEVERRTGL
DPAADLATLDADVKTDIRENLFGGRVLTIVCGSAPLSEELAAFIESCCLDARIT
DGYGSTEAGVIVRNRIQRPPVIDYKLVDPVPELGYFSTDKPHPRGELLVKAES
VFGGYFKRPDVTADVDFDPDGYKYKTDIVAELEPDKIQIVDRRNNVIKLSQGEF
VAIANLEAEFANSPLVHQCIVYGSSERSYLLAVVVPTAEAYEQSGGDEDLLKR
LIADSLAQVAREAQLOQSYEVPRDFLLETETPFTAANGLLTGIAKLARPKLHEKY
GARLEQLYSDIAAAQALELQALHSAGHEDKPVLDTVQRAVTALLGLSAAEVSP
DAHFIDLGGDSLALAFSDLLRDIFTVEVPVGDIVSAANDLTAIARIVERHRE
ADGHSVTPTAESVHGAGHREIRAADLTLDKFIDADTLRAAPALSTFTGTPHTV
LLTGANGYLGRFLALEWLERLDKTDGKLIIVRGKNAEAAAYRRLEEAFTDGT
QLLAHFRSLADKHLEVLGADIGDPNLGLDADTWQRLAETVDVIVHPAALVNHV
LPYSQLFGPNVVGTAEEIKLALTTKIKPITYLSTVAVAI SVDPKVFDEDSDIR
TISAVRPINDGYANGYGNKWAGEVLLREAHDLGCLPVAVFRSDMILAHSRYT
GQLNVPDQFTRLILSLIATGVAPGSFYQAHATGERPLAHYDGLPADFTASAIT
ALGPIIEEFHTYDSVNPHADGISLDNFVDWLIIEAGYPIARIDNYTEWFTFRFTA
IRSLPEKQKQHSLPLHAYRHPQHHPHNGAFLPAIRFSEGVQAHLNADIPHLT
RELIAKYAADLKQLGLL

YP 001704097.1

Secuencia de nucleótidos (SEQ ID N0:63)

>uniprot|B1MDX4|B1MDX4_MYCAB Supuesta ácido graso-CoA ligasa

ATGACGGCTGGTGCGGCGGCTCGCGTTGCCAACTGTTCGAGTCCGATCCCCA
ATTCCGGGCAGCCATGCCGGATCCAGCGGTGATGGACTCGCTGCTGGCGCCCG
GCCTGCGTTTATCCAGGTACTCCACGCGTTGCTCAGCGGTTACGCGGAGCGC
CCGGTGATGGGTTTCCGGTCCCAGGTCGGTGGTCGACACCGCCACCGGCCG
CACGGTCGACCGGCTGCTCCCTGCCTTTGAAACCATCACCTATGGGCAACTCC

FIG 4 (Cont.)

TGGAAGACATCTCGGCCATCCTCGCGGAGTGGCAGCATGGCGACATTCCCATG
 GGCGCCGGCGACTTCATCGCCACCATCGGCTTCTCCAGTCCCGACTACGTCAC
 CCTGGATCTGGCCACCCTCATGAATGGTTTCGGTCTCGATCCCCTGACAGCACA
 ACACATCTGTGGCGCAGCTGCGGATGATGCTGGAGGAGACCAGCCCACGCCTG
 GTGGCGGCGAGCGCGGACTGCCTGGATCTCGCGGTTCGAGGCAGCTGTCGGGCT
 TACCGATCTGCGACGGGTTGTGGTGTTCGATTACCGCGCCGAGACCGACGATC
 ATCGCGAAAAACTGGCCACGGCAAGAGAACGCTTGCACGCGGCCGGTATGGAC
 GTTGTAGTCGAACCGCTCGCAGAGGTGATCGGGAGAGGACGAGACCTACCCGA
 ACCCGTGTGTACACGGCCGGGACGATCAGCGCACGGCCCTGATCATGTACA
 CCTCCGGTAGCACCGGCGCGCCCAAGGGGGCGATGTTACCGAGTGGACGGTG
 ACCCGCTTCTGGTCTTCGGGCGCCGCCCAACCGGGACACCCCGATCATCAA
 CGTGAACCTCCTGCCGCTCAACCACCTTGCGGGCGGGTAGGACTGCTGACGG
 CCTTCATTCGGCGGCACATGCTACTTCGTCCCCGAGAGCGATCTGTCCACC
 CTGTTTCGAGGACTGGCAGCTGGCACGGCCACCCATATGGGTGTGGTTCCCCG
 TGTCGTCGACATGCTCTTCCAGCACTACCAAACGCGAGTGGACGCACTGATGG
 CCGGGGAACCGACGTCGACACCGCCGATCGGCTAGCCAAAACCGAACTGCGC
 GAAGATGTCCTGGGCGGGCGTGTGGTTCGCCGGCATGCTCGCCACCGCGCCGTT
 GTCCCCCGAGATGAAGGCTTTCCTGGAGTCTTCATTGGACTTTCATCTGCTTG
 ATCTGTACGGCCTGACCGAGGTTCGGCGGCGTGTTCGAGACGGCAAGATTTCC
 CGGCCGCCGGTGTCTGACTACAAGCTCGTCGATGTTTCTGAGCTCGGGTACTA
 CACCACCGACAAGCCCCATCCGCGTGGCGAATTGCTGGTCAAGAGTGCCACCG
 CAACGCCCGGCTACTACAAGCGTCCCGACGTCACCGCCGAGGTGTTTCGACGCC
 GATGGCTACTACCGCACGGGCGATGTCATGGCGGAGGTCGCGCCGGACCAATT
 GGTGTACGTGGACAGGCGCAATAACGTCATCAAGCTCGCCAGGGCGAGTTCG
 TCGCGGTTCGCAATTTGGAAACGGTCTATGTGGGTGCGCCGCTGGTTCGCCAG
 ATCTTCGTCTACGGCAACAGCGAACGCGCATACTCCTCGCCGTTGTGGTGCC
 CACCGAGGAAGCCCTGCGGGCACACCCGGACCCCGTCAACTGAAGAATTCGA
 TCCGGGAGTCACTGCAGCGGACCGCCCGCTCCAACCACCTGCATTCTACGAG
 CTGCCC GCCGACTTCATTATCGAAACCACTCCATTACGATCGAGAGTGGGAT
 GCTTGC GGCTGTTCGGTAAGCCGATACGTCCAAGATGATCGAGCACTACGGCG
 ACCGGCTCGAGCAGCTCTACGTGACCTCGCCGAGGCACGCGTCCAGGAACTG
 CGGCAGCTCCGCGATAACGGCGCAACAACGCCCGGTCTCGATAACGTCACCGA
 GGCCGCCAGGCCCTCCTCGGCATGTCTGCGGACGCCGTCCGTCCCGACCACC
 ACTTCATCGACCTCGGCGGAGATTCGCTGTCCGCGTTGACATTCTCCAATCTT
 CTTCGAGACCTCTTCGACGTGAGGTTCCGGTTCGGTGTGATCACCGGCCCGGC
 GGCCGATCTGCGCAAGCTCGCCGCTTACATCCAGCACGAACGGGAGCACAGCA
 CCGCGACCGCTGCCAGCGTGCACGGGCTCGACACCACCGTCATCAGCGCCACC
 GAACTGACACTCGACAAGTTCATCGACGCCGAGACACTCCACAACGCTTCGCA
 ACTCGACGTGCCGGCGGGCGCGGTAGCTACCGTTCTGCTCACCGGCGCCAACG
 GATATCTCGGAAGATTCTCTGCCTGGAGTGGCTGCAACGGCTGTCCAGACA
 GGTGGACAACTGATCTGCCTGGTCCGCGGCGACAACGCCGATCAAGCCCTCGC

FIG 4 (Cont.)

GCGCCTCGTTGCCGCCTACGGCGACACCGATCGCACACTGCTCGAGGAGTTCC
 ACACCCTGGCTCGACGGCACCTGCGCGTGATCGCCGCCGATATCGCTCAGCCG
 CGCTTCGGCGTGGATGACGCCACCTGGGAGCAGCTGGCCCCGCGATGTCGACAA
 GATCGTGCATCCGGCCGCGCTGGTCAACCACGTGCTGCCCTACAACCAGCTGT
 TCGGCCCAATGTGTTTGGCACGGCGGAGGTTATCCGGCTGGCCCTGACCACC
 CGGATCAAGCCGGTGACCTATCTGTGACGATGGCCGTGGCCATGACCGTGCC
 CGATTTGACGAGGACGGGGACATCCGCACGGTGAGTCCCACCCGGCATATCG
 ACCCCGGCTACGCCAACGGGTACGCCAACAGCAAATGGGCCGGCGAGGTGCTG
 CTGCGGGAGGCACACGACATATGCGGCCCTGCCGGTCAGCGTGTTCGGTCCGA
 CATGATCCTGACGCACCGCCGTTACAGCGGACAACCTCAACGTCACCGACGCCT
 TCACCCGCATGCTGCTGAGCCTGGTGTGCTCACCGGCATCGCGCCGCGAAGCTTT
 TACCAAGGCGATGGCAGCGGTGCCCGCCACGCGCTCACTACGAGGGGCTGCC
 GGTCGATTTGTCACCGAAGCCATTACCAGCCTCGGCCTGTCCTCGTCCGAGG
 GATTTGCTCGTACGACGTCATGAATCCTCACGATGACGGCATTCTGTGGAC
 ACCTTTGTCGACTGGCTCATGGAAGATGGGCATTCCATCGACATCATCGACAA
 CTACGACGAATGGCTGTCCCGTTTCGAGACGGCATTGCGAGGTCTGCCCGACG
 AGCAGCGGCGCGCCTCAGTACTTCCGCTCCTCGATGCGTATCGGATAACGGGC
 AACCCGCGCCGTGCTGCCGCCACGCCAATCATGTATTCCGGAAAGCCGTACA
 GGAGAACAACATCGGAGGTGACGGCGCCGATATTCCGCAAATCGATCGTGC
 TGATCGCCAAATACATCGCCGATCTACGAGCACACAGGCTGCTGTGA

Secuencia de aminoácidos (SEQ ID NO:64)

>uniprot|B1MDX4|B1MDX4_MYCAB Supuesta ácido graso-CoA ligasa

MTAGAAARVAKLFE S D P Q F R A A M P D P A V M D S L L A P G L R L S Q V L H A L L S G Y A E R
 P V M G F R S R E S V V D T A T G R T V D R L L P A F E T I T Y G Q L L E D I S A I L A E W Q H G D I P M
 G A G D F I A T I G F S S P D Y V T L D L A T L M N G S V S I P L Q H N T S V A Q L R M M L E E T S P R L
 V A A S A D C L D L A V E A A V G L T D L R R V V V F D Y R A E T D D H R E K L A T A R E R L H A A G M D
 V V V E P L A E V I G R G R D L P E P V L Y T A G D D Q R T A L I M Y T S G S T G A P K G A M F T E W T V
 T R F W S S G A A P N R D T P I I N V N F L P L N H L A G R V G L L T A F I P G G T C Y F V P E S D L S T
 L F E D W Q L A R P T H M G V V P R V D M L F Q H Y Q T R V D A L M A G G T D V D T A D R L A K T E L R
 E D V L G G R V V A G M L A T A P L S P E M K A F L E S S L D F H L L D L Y G L T E V G G V F R D G K I S
 R P P V L D Y K L V D V P E L G Y Y T T D K P H P R G E L L V K S A T A T P G Y Y K R P D V T A E V F D A
 D G Y Y R T G D V M A E V A P D Q L V Y V D R R N N V I K L A Q G E F V A V A N L E T V Y V G A P L V R Q
 I F V Y G N S E R A Y L L A V V V P T E E A L R A H P D P V E L K N S I R E S L Q R T A R S N H L H S Y E
 L P A D F I I E T T P F T I E S G M L A A V G K P I R P K M I E H Y G D R L E Q L Y V D L A E A R V Q E L
 R Q L R D T A Q Q R P V L D T V T E A A Q A L L G M S A D A V R P D H H F I D L G G D S L S A L T F S N L
 L R D L F D V E V P V G V I T G P A A D L R K L A A Y I Q H E R E H S T A T A A S V H G L D T T V I S A T
 E L T L D K F I D A E T L H N A S Q L D V P A G A V A T V L L T G A N G Y L G R F L C L E W L Q R L S Q T

FIG 4 (Cont.)

GGQLICLVRGDNADQALARLVAAAYGDTDRTLLEEFHTLARRHLRVIAADIAQP
 RFGVDDATWEQLARDVDKIVHPAALVNHVLPYNQLFGPNVFGTAEVIRLALTT
 RIKPVTYLSTMAVAMTVPDFDEDDGDIRTVSPTRHIDPGYANGYANSKWAGEVL
 LREAHDICGLPVSVFRSDMILTHRRYSGQLNVTDADFTRMLLSLVLTGIAPRSF
 YQGDGSGARPRAHYEGLPVDFVTEAITSGLSSSEGFRSYDVMNPHDDGISVD
 TFVDWLMEDGHSIDIIDNYDEWLSRFETALRGLPDEQRRASVLPPLLDAYRIPG
 NPRRAATPNHVFRKAVQENNIGGDGADIPQIDRALIAKYIADLRAHRL

YP 001705436.1

Secuencia de nucleótidos (SEQ ID N0:65)

>uniprot|B1MLD7|B1MLD7_MYCAB Probable ácido graso-coa ligasa FadD

ATGACTGAAACGATCTCCACAGCGGCTGTCCCCACTACGGATCTCGA
 AGAGCAGGTGAAGCGACGCATCGAGCAGGTCTGTGTCCAACGATCCGC
 AGCTGGCGGCGCTTCTCCCGGAAGATTCCGGTCACCGAGGCGGTCAAC
 GAGCCCGATCTACCGCTGGTCGAGGTGATCAGGCGACTGCTGGAGGG
 CTACGGTGACCGCCCGGCACTCGGCCAGCGCGCCTTCGAGTTCGTCA
 CCGGGGACGACGGTGCGACCGTGATCGCGCTGAAGCCCGAATACACC
 ACCGTCTCCTACCGCGAGTTGTGGGAACGTGCCGAGGCTATCGCTGC
 CGCGTGGCACGAGCAGGGCATCCGTGACGGCGACTTCGTCTGCTCAGT
 TGGGTTTCACCAGCACGGACTTCGCGTCTGCTCGACGTCGCGGGATTG
 CGTCTGGGCACCGTCTCGGTGCCCTGCAGACGGGCGCGTCTGCTGCA
 GCAGCGCAACGCGATTCTCGAAGAGACCCGGCCCGCAGTCTTTGCCG
 CGAGTATCGAATACCTTGATGCCGCCGTGATTCGGTGCTTGCGACC
 CCCTCGGTGCGACTCCTCTCGGTTTTTCGACTATCACGCGGAGGTCGA
 CAGCCAGCGCGAAGCGCTGGAGGCTGTGCGGGCCCGGCTTGAGAGTG
 CCGGCCGGACGATCGTCGTCGAGGCCCTGGCGGAGGCTCTCGCGCGG
 GGGCGGGACCTGCCCGCCGCGCCGCTGCCAGTGCGAGATCCCGATGC
 CTTGCGTCTGCTCATCTACACCTCCGGCAGCACCGGTACCCCAAGG
 GCGCCATGTATCCGCAATGGCTGGTCGCCAACTTGTGGCAGAAGAAG
 TGGCTCACCGACGATGTGATTCCGTCCATAGGCGTGAACCTTCATGCC
 CATGAGCCACCTGGCGGGTCGCCTCACTCTCATGGGCACCCTTTCCG
 GTGGCGGAACCGCCTACTACATCGCTTCGAGCGATCTTTCGACTTTC
 TTCGAGGACATCGCGCTCATCCGCCCTCCGAAGTGCTCTTCGTGCC

FIG 4 (Cont.)

GCGTGTGGTGGAGATGGTGTTCAGCGTTTTTCAGGCAGAATTGGACC
 GGTCCCTTGCCCCGGGTGAGAGCAACTCCGAGATCGCGGAGCGAATC
 AAGGTCCGCATCCGGGAACAGGACTTCGGCGGGCGTGTGCTCAGTGC
 TGGCTCCGGGTCCGCCCGTTGTCTCCTGAGATGACGGAGTTCATGG
 AGTCGCTGCTGCAGGTGCCGTTGCGCGACGGGTATGGGTCCACCGAG
 GCCGGTGGTGTGTGGCGTGACGGAGTCCTGCAGCGTCCGCCCGTCAC
 CGACTACAAGCTGGTTGACGTTCCGGAACTCGGATACTTCACCACAG
 ATTCGCCGCATCCCCGTGGCGAGCTGCGGTTGAAGTCGGAGACGATG
 TTCCCCGGCTACTACAAGCGCCCGGAGACCACTGCCGATGTCTTCGA
 TGACGAGGGGTACTACAAGACCGGTGACGTGGTCGCCGAGCTCGGGC
 CGGATCACCTCAAGTACCTCGACCGCGTCAAGAACGTCTCAAGCTC
 GCGCAGGGAGAGTTTGTGCGGGTGTCAAAGCTGGAGGCCGCTTACAC
 CGGCAGCCCGCTGGTCCGGCAGATCTTTGTGTACGGGAACAGTGAAC
 GCTCGTTCCTGCTGGCTGTGCTGGTCCCGACACCCGAAGTCCTTGAG
 CGGTACGCAGATTCGCCAGATGCGCTCAAGCCCTTGATCCAGGATTC
 GCTGCAGCAGGTCCCAAGGACGCGGAGCTGCAATCCTATGAGATAC
 CGCGCGACTTCATCGTTGAGACGGTGCCGTTACCGTCGAGTCCGGA
 TTGCTATCGGACGCGCGAAAGCTGCTGCGCCCCAAGCTGAAGGATCA
 CTACGGAGAGAGGCTGGAGGCGCTGTACGCCGAACGGCGGAAAGCC
 AGAATGAGCGGCTGCGCCAGTTGGCCAGGGAGGCAGCCACGCGCCCG
 GTCCTGGAGACGGTGACCGATGCGGCCGCCGCGCTGCTGGGCGCATC
 GTCCTCGGATCTGGCTCCTGATGTGCGATTTCATCGACCTCGGTGGCG
 ACTCACTGTCGGCGCTGTGCTACTCCGAGCTGCTGCGCGACATCTTT
 GAGGTGGACGTTCCGGTGGGCGTCATCAACAGCGTCGCCAACGACCT
 TGCCGCGATCGCCCGGCACATCGAGGCGCAGCGGACCGGCGCCGCTA
 CGCAGCCGACCTTTGCGTCCGGTCCACGGCAAGGACGCGACGGTCATC
 ACCGCCGGTGAACCTCACCTCGACAAGTTCCTTGGACGAGTCACTGTT
 GAAAGCGGCCAAGGACGTTTCAGCCGGCAACGGCCGATGTCAAGACCG
 TTCTAGTGACCGGCGGCAACGGCTGGTTGGGTGCTTGGCTGGTGTCTC
 GATTGGCTGGAGCGGTTGGCACCCAATGGTGGCAAGGTCTACGCCCT
 CATTGCTGGCGCCGATGCCGAAGCAGCCCGGGCACGGTTGGACGCCG
 TGTACGAATCGGGTGATCCCAAGCTGTCCGCGCATTATCGTCACTG
 GCGCAACAGAGTCTGGAAGTTATCGCCGGCGATTTCCGGCGACCAGGA
 TCTCGGTCTATCCCAGGAAGTTTGGCAGAAGCTGGCCAAGGACGTGG
 ACCTGATCGTGCACCTCCGGTGCCCTGGTGAACCACGTGCTGCCGTAC
 AGCCAGTTGTTCGGTCCGAATGTGGCGGGTACCGCCGAGATCATCAA

FIG 4 (Cont.)

GCTGGCAATTTTCGGAGCGGCTCAAGCCGGTCACCTACCTGTCGACGG
 TGGGCATCGCCGACCAGATTCCGGTGACGGAGTTCGAGGAAGACTCC
 GATGTTTCGTGTGATGTTCGGCCGAGCGCCAGATCAATGACGGCTACGC
 GAACGGATACGGCAACTCAAAATGGGCCGGCGAGGTGCTGTTGCGGG
 AGGCTCATGACCTAGCGGGGCTGCCGGTGCGTGTGTTCCGCTCCGAC
 ATGATCCTGGCGCACAGTGACTACCACGGACAGCTCAACGTCACCGA
 CGTGTTACCCGGAGCATCCAGAGTCTGCTGCTCACC GGTTGTCAC
 CGGCCAGCTTCTATGAATTGGATGCCGACGGCAATCGGCAGCGCGCT
 CACTATGACGGTGTGCCCGGCGATTTACCCGCCGCATCGATCACC GC
 CATCGGCGGTGTGAACGTGGTAGACGGTTACCGCAGCTTCGACGTGT
 TCAACCCGCACCATGACGGTGTCTCGATGGATACTTCGTCGACTGG
 CTGATCGACGCAGGCTACAAGATCGCGCGGATCGACGATTACGACCA
 GTGGCTCGCCCGGTTTCGAGCTGGCCCTCAAGGGATTGCCCGAGCAGC
 AGCGGCAACAGTCGGTGTGCCACTTCTCAAGATGTACGAGAAGCCG
 CAACCGGCGATCGACGGAAGTGCACCTCCGACCGCAGAATTCAGTCG
 CGCCGTGCACGAGGCGAAGGTCGGAGACAGCGGTGAGATACCGCACG
 TCACCAAGGAGCTGATCCTCAAGTACGCCAGCGATATTCAGCTGTTG
 GCCTGGTGTAG

Secuencia de aminoácidos (SEQ ID NO:66)

>uniprot|B1MLD7|B1MLD7_MYCAB Probable ácido graso-coa ligasa FadD

MTETI STAAVPTTDLEEQVKRRIEQVVSNDPQLAALLPEDSVTEAVNEPD
 LPLVEVIRRLLEGYGRPALGQRAFEFVTGDDGATVIALKPEYTTVSYRE
 LWERAEAIAAAWHEQGIRDGDFVAQLGFTSTDFASLDVAGLRLGTVSVPL
 QTGASLQQRNAILEETRPVFAASIEYLDAVDSVLATPSVRLLSVFDYH
 AEVDSQREALEAVRARLESAGRTIVVEALAEALARGRDLPAAPLPSADPD
 ALRLLIYTSGSTGTPKGAMYPQWL VANLWQKKWLTDDVIPSIGVNFMPMS
 HLAGRLTLMGTLSSGGTAYYIASSDLSTFFEDIALIRPSEVLFVPRVEM
 VFQRFQAE LDRSLAPGESNSEIAERIKVRIREQDFGGRVLSAGSGSAPLS
 PEMTEFMESLLQVPLRDGYGSTEAGGVWRDGV LQRPPVTDYKLVDVPELG
 YFTTDSPHPRGELRLKSETMFPGYKRPETTADVFDDEGYKYKTGDVVAEL
 GPDHLKYLD R VKNV LKLAQGEFVAVSKLEAAYTGSPLVRQIFVYGN SERS
 FLLAVVPTPEVLERYADSPDALKPLIQDSLQQVAKDAELQSYEIPRDFI
 VETVPFTVESGLLSDARKLLRPKLKDHYGERLEALYAELAESQNERLRQL
 AREAATRPVLETVTDAAAALLGASSDLAPDVRFIDLGGDSL SALS YSEL

FIG 4 (Cont.)

LRDIFEVDVPVGVINSVANDLAAIARHIEAQRRTGAATQPTFASVHGKDAT
 VITAGELTLDKFLDESLLKAAKDVQPATADVKTVLVTGGNGWLGRWLVLVLD
 WLERLAPNGGKVYALIRGADAEAAARARLDAVYESGDPKLSAHYRQLAQQS
 LEVIAGDFGDQDLGLSQEVWQKLAKDVDLIVHSGALVNHVLPYSQLFGPN
 VAGTAEIIKLAISERLKPVTYLSTVGIADQIPVTEFEEDSDVRVMSAERQ
 INDGYANGYGNSKWAGEVLLREAHDLAGLPVRVFRSDMILAHSDYHGQLN
 VTDVFTRSIQSLLLTGVPASFYELDADGNRQRAHYDGVPGDFTAASITA
 IGGVNVVDGYRSFDVFNPHHDGVSMDFVDWLIDAGYKIARIDDDYDQWLA
 RFELALKGLPEQQRQQSVLPLLKMYEKPQPAIDGSALPTAEFSRAVHEAK
 VGDSGEIPHVTKELILKYASDIQLLGLV

YP_001828302.1

Secuencia de nucleótidos (SEQ ID NO:67)

>uniprot|B1VMZ4 | B1VMZ4_STRGG Supuesta ácido carboxílico reductasa

ATGGCCGAACCGCTGGACGCCGCCACCGCGTCGGCGCACGATCCGGGCCAGGG
 GCTCGCCGAGGCCCTGGCCGCCGTCGAACCGGGCCGGGCGCTCGCCGAGGTCA
 TGGCGTCCGTCCTGGAGGGCCACGGGGACCGGCCCGCCCTCGGCGAACGGGCC
 CGGGAGCCGGAGACCGGGCGTCTCCTCCCGCACTTCGACACCATCAGCTACCG
 CGAACTGTGGTCCCGCGTCCGTGCGCTGGCCGGCCGGTGGCACCACGACCCGG
 AATACCCCTGGGCCCCCGGCGACCGGATCTGCACCCTCGGCTTCACCAGCACCC
 GACTACGCGACGCTCGACCTGGCGTGCATCCACCTGGGGGCGGTGCCCGTCCC
 CCTCCCGTCCAACGCCCCGCTGCCCGACTGGCGCCGGTTCGTCGAGGAGTCCG
 GGCCGACGGTACTGGCCGCGAGCGTCGACCGGCTCGACACCGCCATCGACGTC
 GTCCTCGCGTCGAGCACGATCCGCCGCCTCCTCGTCTTCGACGACGGCCCGGG
 GGCCACCCGCCCGGGCGGGGCACTGGCGGGCCGCCCGCCAACGCCTGTCCGGCA
 GCCCCGGTCACCGTCGACACCCTGGCCGGACTCATCGACCGGGGCAGGGACCTG
 CCGCCCCCGCCCTGTACATCCCGGACCCGGGGGAGGACCCGCTCGCCCTGCT
 CATCTACACCTCCGGCAGCACCGGGCGCGCCCAAGGGCGCCATGTACACCCAAC
 GGTCCTGGGCACCGCGTGGTACGGGTTTCAGCTACGGGGCGGCCGACACCCCC
 GCGATCAGCGTCTCTACCTGCCGCAGAGCCACCTCGCGGGCCGCTACGCGGT
 GATGGGATCGCTCGTGAAGGGGGGCACCGGCTACTTCACCGCTGCCGACGACC
 TGTCCACCCTGTTTCGAGGACATCGCCCTGGTCCGCCCCACGGAGCTGACCATG
 GTCCCGCGCCTGTGCGACATGCTCCTCCAGCACTACCGGAGCGAACGGGACCG
 CCGGGCCGACGAACCGGGCGACATCGAGGCGGCGGTACGAAGGCGGTGCGGG
 AGGACTTCCTGGGCGGGCGCGTCCGCAAGGCGTTCGTCGGCACCGCGCCGCTC
 TCCGCCGAACACGCGTTCGTCGAGTCCGTCCTCGGCTTCACCTCTACAC
 CGGCTACGGCTCCACCGAAGCCGGCGGAGTGCTGCTGGACACGGTGGTGCAGC

FIG 4 (Cont.)

GCCCTCCGGTCACCGACTACAAACTGGTTCGACGTCCCCGAACTGGGCTACTAC
 GCGACCGACCTGCCCCATCCGCGCGGGCGAACTGCTGCTGAAGTCCCACACGCT
 CATCCCCGGCTACTACCGGGCGCCCCGACCTCACCGCCGCGATCTTCGACGCGG
 ACGGCTACTACCGCACCGGTGACGTCTTCGCCGAGACCGGACCCGACCGGCTG
 GTCTACGTTCGACCGCACGAAGGACACCCTGAAGCTGTCCCAGGGCGAGTTCGT
 GGCCGTGTCCCGCCTGGAGACCGTCCCTCCTCGACAGCCCTCTCGTCCAGCACC
 TCTACCTGTACGGCAACAGCGAGCGCGCCTACCTCCTCGCGGTGGTGGTGCCC
 ACCCCGGACGCGCTGGCCGGGTGCGGCGGGGACACCGAGGCGCTCAGGCCGCT
 GCTCATGGAATCCCTCCGCAGCGTCCGCCAGGAGGGCCGGGCTCAACGCGTACG
 AGATCCCGCGCGGCATCCTCGTTCGAGCCCGAGCCCTTCAGCCCGGAGAACGGC
 CTCTTCACCGAGAGCCACAAACTGCTGCGCCCCCGCCTCAAGGAGCGCTACGG
 GCCCGCTCTGGAGCTGCTGTACGACCGACTGGCCGACGGGCAGGACCGCCGGC
 TGCGCGAGCTGCGGGCGACCGGTGCGGACCGGCCGGTGCAGGAGACGGTCCCTC
 CGGGCCGCCAGGCCCTGCTGGGATCCCCGGGCTCCGACCTCCGGCCCGGGCGC
 GCACTTCACCGACCTCGGCGGGGACTCCCTCTCCGCGTCTCGTTCTCCGAGC
 TGATGAAGGAGATCTTCCACGTCGACGTCCCGGTCCGGTTCGATCATCGGCCCG
 GCCGCCGACCTGGCGGAGGTGGCGCGGTACATCACGGCGGCCCGTCCGGCCGGC
 CGGGGCACCGCGGCCACGCCCGCCTCCGTGCACGGGGAGCACCGCACCGAAG
 TCCGCGCCGGGGACCTCGCCCCGGAGAAGTTCCCTCGACGCGCCCACGCTCGCC
 GCCGCTCCGGCGCTGCCCCGCCCGACGGCGACGTCCGGACGGTCCCTGCTGAC
 CGGCGCCACCGGCTACCTCGGCCGGTTCCTCTGCCTGGAGTGGCTGGAGCGGC
 TGGCGCCCTCGGGCGGACGGCTGGTCTGCCTCGTCCGCGGCAGCGACGCGACC
 GTCGCGGGCGAGGCGGCTGGAGGCAGCCTTCGACAGCGGCGACACCGCCCTGCT
 CCGGCGCTACCGGAAGGCGGCCGGGAAGACCCTGGACGTGGTTCGCGGGGGACA
 TCGGCGAACCCTGCTGGGCCCTGGCGGAGGAGACCTGGCGGGAACTGGCCGGC
 GCCGTGGACCTGATCGTGCACCCGGCCGCGCTGGTCAACCACCTGCTGCCGTA
 CGGCGAACTGTTCCGCCCAACGTCGTCGGCACCGCCGAAGCGATCCGGCTGG
 CGCTCACCAACCGGCTGAAGCCCGTCAACCACGTCTCGACCGTTCGCGGTCTGC
 CTCGGCACCCCCGCCGAGACGGCCGACGAGAACGCCGACATCCGGGCCGCCGT
 CCCGGTACGGACCACCGGCCAGGGGTACGCCGACGGATACGCGACCAGCAAAT
 GGGCCGGCGAGGTCCCTCCTGCGTGAGGCGCACGAGCGCTATGGCCTCCCCGTC
 GCCGTCTTCCGGTCCGACATGGTCCCTGGCGCACCGCACCTACACCGGACAGGT
 CAACGTCCCGACGTCCCTACCCGGCTGCTGCTCAGCCTGGTCCGCCACCGGCA
 TCGCCCCCGGCTCGTTCACCGCACGGACACCCGTGCCACTACGACGGCCTG
 CCGGTCCGACTTCACCGCGGAGGCCGTCGTCGCGCTGGGCGCACCGATCACCGA
 GGGCCACCGGACCTTCAACGTCCTCAACCCGCACGACGACGGCGTTTCCCTGG
 ACACCTTCGTCGACTGGCTCATCGAGGCCGGCCACCCGATCCGGCGGATCGAC
 GACCACGGTGCCTGGCTCACCCGCTTCACCGCGGCGCTCCGCGCCCTGCCGGA
 GAAGCAGCGGCAGCACTCCCTGCTCCCGCTGATCGGCGCCTGGGCGGAACCCG
 GCGAAGGAGCCCCCGGGCCGCTGCTCCCCGCCGGCGCTTCCACGCCGCCGTC
 CGGGCGGCGGGGGTCCGCCCGAGCGGGACATTCCCCGGGTGTCGCCGGACCT

FIG 4 (Cont.)

CATCCGCAAGTACGTCACCGACCTGCGCGCACTCGGGCTCCTCGCCGGCCCCCT
GA

Secuencia de aminoácidos (SEQ ID N0:68)

>uniprot|B1VMZ4 |B1VMZ4_STRGG Supuesta ácido carboxílico reductasa

MAEPLDAATASAHDPGQGLAEALAAVEPGRALAEVMASVLEGHGDRPALG
ERAREPETGRLLPHFDTISYRELWSRVRALAGRWHHDPEYPLGPGDRICT
LGFTSTDYATLDLACIHLGAVPVPLPSNAPLPR LAPVVEESGPTVLAASV
DRLDTAIDVVLASSTIRLLVFDGPGATRPGGALAAARQRLSGSPVTVD
TLAGLIDRGRDLPPPPLYIPDPGEDPLALLIYTSGSTGAPKGAMYTQRLL
GTAWYGF SYGAADTPAISVLYLPQSHLAGRYAVMGSLVKGGTGYFTAADD
LSTLFEDIALVRPTELTMPRLCDMLLQHYRSEDRRADEPGDIEAAVTK
AVREDFLGGRVAKAFVGTAPLSAELTAFVESVLGFHLYTYGSTEAGGVL
LDTVVQRPPVTDYKLVDPVPELGYATDLPHPRGELLLKSHTLIPGYRRP
DLTAAIFDADGYYRTGDVFAETGPDRLVYVDRTKDTLKLKLSQGEFVAVSRL
ETVLLDSPLVQHLYLYGN SERAYLLAVVPTPDALAGCGGDTEALRPLLM
ESLSVARRAGLNAYEIPRGILVEPEPFPENGLFTESHKLLRPRLKERY
GPALELLYDRLADGQDRRLRELRRTGADRPVQETVLRAAQALLGSPGSDL
RPGAHF TDLGGDSL SAVSFSELMKEIFHVDVPVGAIIGPAADLAEVARYI
TAARRPAGAPRPTPASVHGEHRTEVRAGDLAPEKFLDAPTLAAAPALPRP
DGDVRTVLLTGATGYLGRFLCLEWLERLAPSGGRLVCLVRGSDATVAARR
LEAAFDSGDTALLRRYRKAAGKTL DVVAGDIGEPLLGLAEETWRELAGAV
DLIVHPAALVNHLLPYGELFGPNVVGTAEAIRLALTTRLKPVNHVSTVAV
CLGTPAETADENADIRA AVPVRTTGQGYADGYATSKWAGEVLLREAHERY
GLPVAVFRSDMVL AHRTYTGQVNVPDVLTRLLLSLVATGIAPGSFYRTDT
RAHYDGLPVDF TAEAVVALGAPITEGHRTFNVLNPHDDGVSLDTFVDWLI
EAGHPIRRIDDHGAWLTRFTAALRALPEKQRQHSLPLIGAWAEPGEGAP
GPLLPARRFHA AVRAAGVGP ERDIPRVSPDLIRKYVTDLRALGLLAGP

YP 001851230.1

Secuencia de nucleótidos (SEQ ID N0:69)

>uniprot|B2HE95 |B2HE95_MYCMM ácido graso-CoA ligasa FadD9

TTGTCAATTACCTGTGTGGATACCCGTGCACAGCGGAGCGCCCGTCGCATCGA
GCAGCTTTACTCCACCGATGCGCAATTCGCCGCCGCCCGCCCGCCAGTACGGCGG

FIG 4 (Cont.)

TCGGTATCGCAATCAGCAAGTCCGGGTGGGATTACCACAGATCATTCAAACG
 GTGATGGACGGATACCCGCAACGTCCGGCACTTGGGCAGCGGGCGACGCGCGT
 TGTTACCGATCCGAATACCGGGCGTAGCTCGGGCGCAGCTGTTGGCGGAGTTTCG
 AGACCATCACCTACCGGGAGTTGTGGAACCGCACCAATGCATTGACCAACGCA
 TTCGCCGCCGAGGCACTTGC GGATCGCGGTTCAGCGGGTCTGTGTGCTGGGATT
 CGCGAGCATCGACTACGCCACCATCGACTTGGCGCTGATGTTGCTCGGGCGCGG
 TATCGGTTCCGTTGCCGACGAATGCGGCTCGCGCCAGCTGTGCCATATCGTC
 TCCGAGACCCAGCCCAGCCTGATCGCCTCGAGTACCGAAAACCTGCCCGATGC
 AATCTCTTTGGTGCTGTGCGACCGCGCACACCACCGGGTGGTGGTGTTCGACT
 ACCGCCCCGAACTCGACGCACACCGCGAAGCCCTCGAAGCCGCTCGCGCGCGC
 CTGGCCGCCATCCCGGTGACCGTCGAAACGCTCACCGCCATCATCGCGCGCGG
 TCGAACGGTTCGGGCCGGCCGAGGCCGATTGCGGGCGCCAGTCCGCTGATGCAC
 CGGCGCTTTTGATCTATACCTCCGGAAGCACCGGGGCACCCAAGGGCGTCGTC
 TACACCCGCAACCGGGTGGCGGACTTCTGGCGCACCTCGAAAGCCGAGGTCGA
 AGCGACCGAACAACGAACCGCTCCTTCGATCACCTCAACTTCATGCCGATGA
 GCCACGCGAACGGCCGCCAGGTGCTCTACGGGACGCTGTCCAACGGCGGCACC
 GCGTATTTACGGCCCGCAGCGACCTCTCGACGCTCTTCGATGATCTCGCGTT
 GGTCCGGCCCACCGAATTGGGCTTTCCACCGCGCATTGGGACATGCTGTTGG
 AGAGGTTTGGGCGCGAAGTCGACCGTCGGCTCCGGGACGGCACAGCCGAGGGC
 GCCGACCCGGGCGCGCTGAAGGCTCGCGTGGCGGCCGACCTACGCCAGGTGCT
 GCTCGGCGGACGGTATGCGCTGGCGATGATGGGCTCCGCGCCAATCTCCGAGC
 AGATGAAAGCATCCGTCGAATCCCTGCTCGATCTGGACGTCATGGAGGGCTAT
 GGCTCCACGGAAGCCGGAACGGTCATCATCAACAACGAGGTTGAGCGTCCCCA
 GGTGATCGACTACAAGCTGGTTCGACGTTGCGGAACTGGGCTATTTCCCTTACCG
 ACCGGCCATATCCGCGGGGCGAACTGCTGGTCAAACGCGGACACTGTTTTCC
 GGCTACTACCGGGACCCCAGACGGCGCCCAGGTCTTCGACCCGGACGGCTT
 CTACCGGACCGGCGACATCATGGCCAAAGTCGGCCCCGATCGGCTCGCCTACC
 TCGACCGGCGCAACAACGTGCTGAAGCTGTGCGCAGGGGGAGTTCGTCGCGGTC
 TCGCGACTAGAAGCAATATTTGCCAATAGCCCGTTGGTCCGGCAGATCTTCGT
 CTATGCCAACGGTGTGCTCGCGCCTACCCACTGGCGGTAGTCGTGCCACCCAGG
 ACGCACAGTCGCGCCACGGTCGCGCCGAACTCAAGGCCGAACTCCATACATCG
 CTGCACCGCGTTGCCATGTGCGCCGGTCTGGCACCCCTACGAGATCCCACGCGA
 CTTTCATTGTGAGACAACCCCTTCACGCCGAGAACGGCCTGCTCACCGCAA
 TCCACAAGCTGGCCC GGCCGACCTCACGCAGCGCTATGGCGCACGTCTGGAG
 CTGCTGTACACCGAGCTGGCCGACAGCCAGACCCGCCGGCTGCACCGATTGCG
 CCAAACCGGTGGGCGGCTGCCGGCGCTCGAGACCATCAGGCGTGCCGCCGGGG
 CACTGTTGGGCACGGAGACCACCGAGCCGCGGCCCGAGGCCCACTTCAAAGAT
 CTGGGCGGGGATTCCGGTATCGGCGGTGACGTTCTCCAACTGCTACACGACAT
 CTACGGTTTCGATGTTCCGGTCCGGTGTGATCCTCGGCCCGGCAACCGATTTC
 GGGCGCTGGCCAGCCACGTCGAGAGCCGGCGCGGTGCCGGATGGTCCGGGGCC
 AGCTTCGCGTCGGTGCACGTGCCCGGGCGACCTCGGTACACGCCGGCGACCT

FIG 4 (Cont.)

GAAACTGGCCAAGTTCCTGGACACCAAGACACTCGCAGCTGCCACGAGCCTGC
 CCGCTGCCGATGCCCCGGGCACGGACGGTGCTACTCACCGGGCGCAACCGGATTC
 CTGGGACGCTACCTGGTGCTGGAATGGCTGCGCCGGTTGCGGGCCGTGCGCGG
 CAAGCTGATCTGTCTGGTGCGCGCCGCGTCCGACGAACAAGCCCAGGTTTCGGC
 TGGATACGGCCTTCGATAGCGGCGATCCGCAGCTGCCCCGAGCACTTTCGGCAG
 CTCGCTGTGACCGCCTGGAGGTCCTCGCCGGCGATAAGAGCGAACCAGGTCT
 CGGTCTGGACGGCCCAACCTGGCAGCGACTGGCCGACACGGTCGACCTGATCG
 TCGACCCCGCCACGCTGGTCAACCACGTGCTGTCGTACCGGCAGCTGTTTCGCT
 CCCAACGTGGCGGGCACCGCCGAGTTGCTCCGCCTCGCACTCACCAACAAACG
 CAAGCCCTATGCCTACGTCTCGACCGTCAGCGTGGCCAACCAGATCGAACCCT
 CCGCATTACCGAAGACGCCGACATCCGGGAGATCAGCCGACCCGAACCATC
 GATGACAGCTTTGCCAACGGCTACACCACCAGCAAGTGGGCCAGCGAGGTGCT
 GTTGCCTGAGGCTCACGATCTGTGCGGACTGCCGGTCACGGTCTTTTCGTTGCG
 ACATGATCCTGGCGGACACCAGCTACGCCGGCCAGCTCAACCTCGCCGATAACC
 TTCACCCGGCTGATGCTCAGTGTGGCGGCCACCGGGATCGCGCCCGCCTCGTT
 CTACCGGCTGGGCCCCGACGGCAAACGCCAGCCCGCCCACTTCGACGGATTGC
 CCGTCGAATTCATCGCCGAGGCGGTGGCCACCCTGGGGGCGCGGCCGCCACGAC
 GGGTTCCAGGTCCACCATGTGGCGAATCCGCACCACGACGGCGTTGGGTTGGA
 CGAGTACGTCGACTGGCTAGTCGATGCCGGTTGCCCCATCCGGCGCATTCCCG
 ACTATGACGAGTGGCTGAGTCGATTTCGAGACGGCGCTGCACGCGCTGCCGGAT
 CGCAAGCGTCGTCATTCCTGCTTCCGCTGCTGCAGAACTATCGAGAACCCGC
 CGAGCCGATCCGGGGCGGCATCGCGCCCGCACCACGGTTTCGCGGTGCGGTAC
 GGCAGGCGAAAATCGGCCGCGACAACGACATTCCCCATGTCGGCCCCGGCGATC
 ATCGCCAAGTACGCCAGCGACCTGCAGCTTCTCGGCCTGGCTTGA

Secuencia de aminoácidos (SEQ ID N0:70)

>uniprot|B2HE95 |B2HE95_MYCMM ácido graso-CoA ligasa FadD9

MSITCVDTRAQRSARRIEQLYSTDAQFAAARPSTAVGIAISKSGGLGLPQI
 IQTVMDGYPQRPALGQRATRVVTDPNTRSSAQLLAEFETITYRELWNRT
 NALTNAFAAEALADRGQRVCVLGFASIDYATIDLALMLLGAVSVPLPTNA
 ARAQLCHIVSETQP SLIASSTENLPDAISLVLSHRAPHRVVVFDYRPELD
 AHREALEAARARLAAIPVTVELTAIIARGRTVRPAEADCGAQSADAPAL
 LIYTSGSTGAPKGVVYTRNRVADFWRTSKAEVEATEQRTAP SITLNFMPM
 SHANGRQVLYGTLNNGGTAYFTARS DLSTLFDLALVRPTELGFPPRIWD
 MLLERFGREVDRLRDGTAEGADPGALKARVAADLRQVLLGGRYALAMMG
 SAPISEQMKASVESLLDLDMEGYGSTEAGTVIINNEVQRPQVIDYKLVD
 VAELGYFLTDRPYPRGELLVKTRTLFSGYYRDPEDGAQVFPDGFYRTGD
 IMAQVGPDRLAYLDRRNNVLKLSQGEFVAVSRLEAIFANSPLVRQIFVYA

FIG 4 (Cont.)

NGARAYPLAVVVPTQDAQSRHGRAELKAE LHTSLHRVAMSAGLAPYEIPR
 DFIVETTPFTPQNGLLTAIHKLARPHLTQRYGARLELLYTELADSQTRRL
 HRLRQTGGRLPALETIRRAAGALLGTETTEPRPEAHFKDLGGDSVSAVTF
 SNLLHDIYGFVDPVGVILGPATDLRALASHVESRRGAGWSGSPFASVHVP
 RATSVHAGDLKLAKFLDTKTLAAATSLPAADARARTVLLTGATGFLGRYL
 VLEWLRRLRAVGGKLI CLVRAASDEQARVRLDTAFDSGDPQLPEHFRQLA
 VDRLEVLAGDKSEPGLGLDGPTWQRLADTVDLIVDPATLVNHVLSYRQLF
 APNVAGTAE LLRLALTTKRKPYAYVSTVSVANQIEPSAFTEDADIREISR
 TRTIDDSFANGYTT SKWASEVLLREAHDL CGLPVTVFRCMILADTSYAG
 QNLNADTFTRLMLSVAATGIAPASFYRLGPDGKRQPAHF DGLPVEFIAEA
 VATLGARRHDGFQVHHVANPHHDGVGLDEYVDWLVDAGCPIRRIPDYDEW
 LSRFETALHALPDRKRRLSLPLQLQNYREPAEPIRGGIAPAPRFRGAVRQ
 AKIGRDNDIPHVGPAAI IAKYASDLQLLGLA

YP 001850422.1

Secuencia de nucleótidos (SEQ 10 N0:71)

>uniprot|B2HN69 |B2HN69_MYCMM ácido graso-CoA ligasa FadD9

ATGTCGCCAATCACGCGTGAAGAGCGGCTCGAGCGCCGCATCCAGGACCTCTA
 CGCCAACGACCCGCAGTTCGCCGCCGCAAACCCGCCACGGCGATCACCGCAG
 CAATCGAGCGGCCGGGTCTACCGCTACCCAGATCATCGAGACCGTCATGACC
 GGATACGCCGATCGGCCGGCTCTCGCTCAGCGCTCGGTGCAATTCGTGACCGA
 CGCCGGCACCCGGCCACACCACGCTGCGACTGCTCCCCACTTCGAAACCATCA
 GCTACGGCGAGCTTTGGGACCGCATCAGCGCACTGGCCGACGTGCTCAGCACC
 GAACAGACGGTGAAACCGGGCGACCGGGTCTGCTTGTTGGGCTTCAACAGCGT
 CGACTACGCCACGATCGACATGACTTTGGCGCGGCTGGGCGCGGTGGCCGTAC
 CACTGCAGACCAGCGCGGCGATAACCCAGCTGCAGCCGATCGTCGCCGAGACC
 CAGCCCACCATGATCGCGGCCAGCGTCGACGCACTCGCTGACGCCACCGAATT
 GGCTCTGTCCGGTCAGACCGCTACCCGAGTCTTGGTGTTCGACCACCACCGGC
 AGGTTGACGCACACCGCGCAGCGGTGCAATCCGCCCGGGAGCGCCTGGCCGGC
 TCGGCGGTTCGTCGAAACCCTGGCCGAGGCCATCGCGCGCGGCGACGTGCCCCG
 CGGTGCGTCCGCCGGCTCGGGCGCCCGGCACCGATGTGTCCGACGACTCGCTCG
 CGTACTGATCTACACCTCGGGCAGCACGGGTGCGCCCAAGGGCGCGATGTAC
 CCCCAGCGCAACGTTGCGACCTTCTGGCGCAAGCGCACCTGGTTTCAAGGCGG
 CTACGAGCCGTCGATCACGCTGAACTTCATGCCAATGAGCCACGTCATGGGCC
 GCCAAATCCTGTACGGCACGCTGTGCAATGGCGGCACCGCCTACTTCGTGGCG
 AAAAGCGATCTCTCCACCTTGTTCGAAGACCTGGCGCTGGTTCGGCCACCGA
 GCTGACCTTCGTGCCGCGCGTGTGGGACATGGTGTTCGACGAGTTTCAGAGTG

FIG 4 (Cont.)

AGGTCGACCGCCGCCTGGTTCGACGGCGCCGACCGGGTCGCGCTCGAAGCCCAG
 GTCAAGGCCGAGATACGCAACGACGTGCTCGGTGGACGGTATAACCAGCGCACT
 GACCGGCTCCGCCCTATCTCCGACGAGATGAAGGCGTGGGTTCGAGGAGCTGC
 TCGACATGCATCTGGTTCGAGGGCTACGGCTCCACCGAGGCCGGGATGATCCTG
 ATCGACGGAGCCATTCGGCGCCCGGCGGTACTCGACTACAAGCTGGTCGATGT
 TCCCGACCTGGGTTACTTCTGACCGACCGGCCACATCCGCGGGGCGAGTTGC
 TGGTCAAGACCGATAGTTTGTTCGCGGGTACTACCAGCGAGCCGAAGTCACC
 GCCGACGTGTTTCGATGCTGACGGCTTCTACCGGACCGGCGACATCATGGCCGA
 GGTTCGGCCCCGAACAGTTCGTGTACCTCGACCGCCGCAACAACGTGTTGAAGC
 TGTCGCAGGGCGAGTTCGTACCGTCTCCAACTCGAAGCGGTGTTTGGCGAC
 AGCCCACTGGTACGGCAGATCTACATCTACGGCAACAGCGCCCGTGCCTACCT
 GTTGGCGGTGATCGTCCCACCCAGGAGGCGCTGGACGCCGTGCCTGTCGAGG
 AGCTCAAGGCGCGGCTGGGCGACTCGCTGCAAGAGGTCGCAAAGGCCGCCGCGC
 CTGCAGTCTACGAGATCCCGCGCGACTTCATCATCGAAACAACACCATGGAC
 GCTGGAGAACGGCCTGCTCACCGGCATCCGCAAGTTGGCCAGGCCGCGAGCTGA
 AAAAGCATTACGGCGAGCTTCTCGAGCAGATCTACACGGACCTGGCACACGGC
 CAGGCCGACGAACTGCGCTCGCTGCGCCAAAGCGGTGCCGATGCGCCGGTGCT
 GGTGACGGTGTGCCGTGCGGCGGCCGCGCTGTTGGGCGGCAGCGCCTCTGACG
 TCCAGCCCAGATGCGCACTTCACCGATTTGGGCGGCGACTCGCTGTCGGCGCTG
 TCGTTCACCAACCTGCTGCACGAGATCTTCGACATCGAAGTGCCGGTGGGCGT
 CATCGTCAGCCCCGCCAACGACTTGCAGGCCCTGGCCGACTACGTCGAGGCGG
 CTCGCAAACCCGGCTCGTCACGGCCGACCTTCGCCTCGGTCCACGGCGCCTCG
 AATGGGCAGGTCACCGAGGTGCATGCCGGTGACCTGTCCCTGGACAAATTCAT
 CGATGCCGCAACCCTGGCCGAAGCTCCCCGGCTGCCCGCCGAAACACCCAAG
 TCGCACCGTGCTGCTGACCGGCGCCACCGGCTTCTCGGGCGCTACCTGGCC
 CTGGAATGGCTGGAGCGGATGGACCTGGTTCGACGGCAAACCTGATCTGCCTGGT
 CCGGGCCAAGTCCGACACCGAAGCACGGGCGCGGCTGGACAAGACGTTTCGACA
 GCGGCGACCCCGAACTGCTGGCCACTACCGCGCACTGGCCGGCGACACCTC
 GAGGTGCTCGCCGGTGACAAGGGCGAAGCCGACCTCGGACTGGACCGGCAGAC
 CTGGCAACGCCTGGCCGACACGGTCGACCTGATCGTCGACCCCGCGGCCCTGG
 TCAACCACGTACTGCCATAACAGCCAGCTGTTTCGGGCCCAACGCGCTGGGCACC
 GCCGAGCTGCTGCGGCTGGCGCTCACCTCCAAGATCAAGCCCTACAGCTACAC
 CTCGACAATCGGTGTCGCCGACCAGATCCCGCCGTGCGCGTTCACCGAGGACG
 CCGACATCCGGGTCATCAGCGCCACCCGCGCGGTTCGACGACAGCTACGCCAAT
 GGCTACTCGAACAGCAAGTGGGCGGGCGAGGTGCTGTTGCGCGAGGCGCATGA
 CCTGTGTGGCCTGCCGGTTGCGGTGTTCCGCTGCGACATGATCCTGGCCGACA
 CCACATGGGCGGGACAGCTCAATGTGCCGGACATGTTACCCGGATGATCCTG
 AGCCTGGCGGCCACCGGTATCGCGCCGGGTTTCGTTCTATGAGCTTTCGGCCGA
 CGGCGCCCGGCAACGCGCCCACTATGACGGTCTGCCCGTTCGAGTTCATCGCCG
 AGGCGATTTTCGACTTTGGGTGCGCAGAGCCAGGATGGTTTCCACACGTATCAC
 GTGATGAACCCCTACGACGACGGCATCGGACTCGACGAGTTCGTCGACTGGCT

FIG 4 (Cont.)

CAACGAGTCCGGTTGCCCCATCCAGCGCATCGCTGACTATGGCGACTGGCTGC
 AGCGCTTCGAAACCGCACTGCGCGCACTGCCCGATCGGCAGCGGCACAGCTCA
 CTGCTGCCGCTGTTGCACAACATATCGGCAGCCGGAGCGGCCCGTCCGCGGGTC
 GATCGCCCCTACCGATCGCTTCCGGGCAGCGGTGCAAGAGGCCAAGATCGGCC
 CCGACAAAGACATTCCGCACGTCCGGCGCGCCGATCATCGTGAAGTACGTCAGC
 GACCTGCGCCTACTCGGCCTGCTCTGA

Secuencia de aminoácidos (SEQ ID N0:72)

>uniprot | B2HN69 | B2HN69_MYCMM ácido graso-CoA ligasa FadD9

MSPITREERLERRIQDLYANDPQFAAAKPATAITAAIERPGLPLPQIIET
 VMTGYADRPALAQRSVEFVTDAGTGHTTLRLLPHFETISYGELWDRISAL
 ADVLSTEQTVKPGDRVCLLGFNSVDYATIDMTLARLGAVAVPLQTSAAIT
 QLQPIVAETQPTMIAASVDALADATELALSGQTATRVLVFDHHRQVDAHR
 AAVESARERLAGSAVVETLAEAIARGDVPRGASAGSAPGTDVSDDSLALL
 IYTSGSTGAPKGAMYPRNVATFWRKRTWFEGGYEPSITLNFMPMSHVMG
 RQILYGTLCNGGTAYFVAKSDLSTLFEDLALVRPTELTFVPRVWDMVFDE
 FQSEVDRRLVDGADRVALEAQVKAIEIRNDVLGGRYTSALTGSAPISDEM
 AWVEELDMHLVEGYGSTEAGMILIDGAIRRPVLDYKLVDPDLGYFLT
 DRPHPRGELLVKTDSLFPGYQRAEVTADVFDADGFYRTGDIMAEVGP
 FVYLDRRNNVLKLSQGEFVTVSKLEAVFGDSPLVRQIYIYGNSARAYLLA
 VIVPTQEALDAVPVEELKARLGDSLQEVAKAAGLQSYEIPRDFI IETTPW
 TLENGLLTGIRKLARPQLKKHYGELLEQIYTDLAHGQADELRSRQSGAD
 APVLVTVCAAALLGGSASDVQPDHFDTDLGGDSLALSFTNLLHEIFD
 IEVPVGVIVSPANLQALADYVEAARKPGSSRPTFASVHGASNGQVTEVH
 AGDLSLDKFIDAATLAEAPRLPAANTQVRTVLLTGATGFLGRYLALEWLE
 RMDLVDGKLICLVRAKSDTEARARLDKTFDSGDPELLAHYRALAGDHLEV
 LAGDKGEADLGLDRQWQRLADTVDLIVDPAALVNHVLPYSQLFGPNALG
 TAELLRLALTSKIKPYSYTSITIGVADQIPPSAFTEDADIRVISATRAVDD
 SYANGYSNSKWAGEVLLREAHDLCLPVAVFRCDMILADTTWAGQLNVPD
 MFTRMILSLAATGIAPGSFYELAADGARQRAHYDGLPVEFIAEAI STLGA
 QSQDGFHTYHVMNPYDDGIGLDEFVDWLNESGCPIQRIADYGDWLQRFET
 ALRALPDRQRHSSLLPLLHNYRQPERPVRGSIAPTDRFRAAVQEAKIGPD
 KDIPHVGAPIIVKYVSDLRLLGLL

FIG 4 (Cont.)

O69484**Secuencia de nucleótidos (SEQ ID NO:73)**

>uniprot|069484|069484_MYCLE Supuesta Acil-CoA sintetasa

ATGTCGACTATCACTAAGCAGGAAAAGCAGCTCGCACGCCGCGTTGACGACCT
 CACCGCCAACGACCCGCAGTTCGCCGCCGCCAAACCCGACCCGGCGGTAGCCG
 CCGCCCTTGCCAGCCCGGGCTTCGACTGCCCCAAATCATCCAGACCGCGCTG
 GACGGTTACGCGGAGCGGCCGGCACTGGGCCAGCGCGTCGCCGAGTTCACCAA
 AGACCCTAAGACCGGACGCACCTCGATGGAGCTGCTCCCCAGCTTTGAGACCA
 TCACCTACCGCCAGTTGGGCGACCGTGTTCGGAGCGCTGGCGCGCGCCTGGAGG
 CACGACCTACTGCACGCCGGCTACCGGGTCTGCGTGCTAGGTTTCAACAGTGT
 CGATTACGCCATCATCGACATGGCGCTCGGCGTGATTGGTGCTGTGGCGGTTT
 CACTGCAGACCAGTGCGGCGATCACCCAGCTGCAGTCGATCGTGACCGAGACC
 GAACCCAGTATGATCGCGACGAGCGTAAACCAGCTGCCCGATACTGTTCGAGCT
 GATCCTGTCTGGCCAGGCGCCAGCGAAGCTCGTTGTGTTTGACTACCACCCCG
 AGGTCGACGAGCAGCATGACGCAGTGGCAACCGCCCGGGCGCGGTTGGCGGAC
 AGTAGCGTGGTGGTCGAGAGCCTGACCGAGGTCTTCGGTCGCGGCAAGACGCT
 GCCAGCTACGCCGATCCCCGTGGCCGATGACTCTGCTGACCCGTTGGCGTTGC
 TGATCTACACATCTGGCAGCACCGGCGCACCCAAGGGCGCGATGTATCTGCAA
 AGCAATGTCCGCAAGATGTGGCGCCGGTCAGACGGAAACTGGTTCGGGCCAAC
 CGCCGCGTCAATCACTCTTAACTTCATGCCGATGAGCCACGTCATGGGCCGCG
 GAATCCTCTACGGCACGCTCGGTAACGGCGGCACGGCTTACTTCGCCGCCCGC
 AGCGACCTCTCGACGCTGCTGGAGGATCTCAAGCTGGTGCGGCCGACCGAGTT
 GAACTTTGTACCGCGCATCTGGGAAACCCTCTACGATGAATCCAAACGCGCAG
 TTGACCGTCGGTTAGCCAACAGCGGCTCCGCCGACCGTGCAGCCATCAAAGCC
 GAAGTTATGGATGAACAGCGGCAATCCCTGCTGGGAGGACGGTACATCGCGGC
 TATGACGGGCTCGGCGCCAACCTCCCCGGAGTTGAAACACGGGGTCGAGTCCC
 TACTCGAAATGCATCTGTTGGAAGGCTACGGCTCCACCGAAGCCGGCATGGTC
 TTGTTTGACGGCGAAGTGCAACGTCGCCGGTTATCGATTACAAGCTGGTTCGA
 CGTTCCGGATTTGGGCTACTTCAGCACCGACCAGCCTTATCCGAGAGGTGAAT
 TGCTGCTCAAGACCCAGAACATGTTCCCCGGCTACTACAAGCGTCTTGAGGTT
 ACCGCCACCGTGTTTCGACAGCGACGGTTACTACCAGACCGGAGACATTGTTCG
 CGAAGTCGGTCCCGACCGGCTCGTGTACGTCGATCGCCGCAACAACGTGCTGA
 AACTCGCGCAGGGCCAGTTCGTCACCGTCGCGAAACTCGAGGCAGCGTTCAGC
 AATAGCCCACTGGTCCGGCAGATCTACATCTATGGCAACAGCGCACACCCCTA
 CCTGTTGGCTGTTGTGGTGCCGACCGAGGATGCGTTGGCTACCAATGACATTG
 AGGTGCTCAAACCGCTGATTATCGATTCTTTACAGAAAGTAGCGAAAGAAGCC
 GACCTGCAGTCTACGAGGTGCCGCGCGACTTAATCGTCGAGACTACACCGTT

FIG 4 (Cont.)

CAGCCTGGAGAATGGCCTGCTCACCGGTATTTCGCAAGCTGGCGTGGCCGAAGC
TCAAGCAGCACTACGGCGCGCGACTCGAACAGCTCTACGCCGATCTGGTTGAA
GGTCAGGCAAATGCACTGCACGTGCTAAAACAAAGCGTGGCGAACGCTCCGGT
ACTGCAGACGGTGAGCCGAGCCGTGGGCACCATTCTGGGAGTGGCGACCACCG
ATTTGCCGTCGAATGCGCACTTCACCGACTTAGGAGGAGACTCGTTGTCCGCG
CTGACATTCGGTAGCCTGCTACGCGAACTCTTCGACATCGATGTGCCGGTGGG
CGTCATTGTGAGCCCTGTCAACAACCTGGTGGCGATCGCCGACTACATCGAGC
GCGAGCGGCAGGGCACGAAGCGGCCCACTTTCATTGCCATACACGGTTCGTGAC
GCTGGCAAAGTGCATGCCAGTGACCTCACTCTAGACAAATTCATCGATGTATC
AACGCTGACTGCCGCGCCCGTATTGGCGCAACCCGGCACCGAGGTGCGCACCG
TCCTGTTGACCGGCGCTACCGGCTTCCTGGGGCGCTACTTGGCCCTGAAATGG
CTCGAACGGATGGACCTGGTCGAAGGGAAGGTAATCGCTCTGGTAAGAGCCAA
GTCCAACGAGGACGCTCGGGCCCGGCTCGACAAGACCTTCGATAGCGGAGACC
CCAAACTGCTGGCGCACTACCAGGAACTGGCAACCGACCACCTGGAGGTCATC
GCCGGCGACAAAGGCGAAGTAGATCTGGAATTGGACCGGCAAACGTGGCGACG
ACTGGCCGACACGGTTCGATCTGATCGTCGACCCCGCCGCCCTGGTCAACCACG
TGCTGCCGTACAGCGAGCTATTTCGGCCCAATACGTTAGGCACCGCCGAGCTG
ATTCGGATCGCGCTGACCAGTAAGCAAAAGCCGTACATCTATGTGTCGACAAAT
CGGCGTCGGTAATCAGATTGAGCCAGCAAAATTCACCGAAGACTCCGACATCC
GAGTCATTAGCCCGACGCGCAACATCAACAACAACCTATGCCAACGGCTACGGC
AACAGCAAGTGGGCGGCGAAGTGCTGCTGCGCGAAGCTCACGACCTATGCGG
TCTGCCGGTCACGGTCTTCCGCTGCGACATGATCTTGGCCGACACCAGCTATG
CCGGTCAGCTCAACGTCCCCGACATGTTTACTCGAATGATGCTGAGTCTAGCC
GCCACCGGCATCGCACCCGGCTCGTTCTACGAGCTAGAC
GCCGAGAGCAATCGGCAACGCGCCCACTACGACGGTCTGCCCGTCGAGTTCAT
CGCCGAAGCGATCTCCACCCTGGGAGACCAAAGCCTGCACGATCGAGACGGGT
TCACGACCTATCATGTAATGAACCCGCACGACGACGGCATCGGTATGGACGAG
TTTGTGGACTGGTTAATTGATGCCGGCTGCCCTATAACAACGCATCAACGACTA
CGACGAATGGCTGCGACGGTTTGAGATTTGCTGCGCGCCCTGCCCGAAAGGC
AGCGTCACAGCTCACTGTTGCCGTTGTTGCACAACCTACCAGAAGCCGGAGAAG
CCATTGCACGGGTCGCTGGCACCCACAATCCGGTTCCGTACGGCCGTTCAAAA
CGCGAACATTGGTCAGGACAAAGATATTCCGCATATCTCGCCGGCAATCATCG
CCAAATATGTCAGCGATCTGCAGCTGCTCGGGCTGGTTTGA

Secuencia de aminoácidos (SEQ ID N0:74)

>uniprot|O69484|069484_MYCLE Supuesta Acil-CoA sintetasa

MSTITKQEKQLARRVDDL TANDPQFAAAKPDPAVAAALAQPGLRLPQIIQ
TALDGYAERPALGQRVAEF TKDPKTGR TSMELLP SFETIT YRQLGDRVGA

FIG 4 (Cont.)

LARAWRHDLHAGYRVCVLGFNSVDYAIIDMALGVIGAVAVPLQTSAAIT
 QLQSIVTETEPSMIATSVNQLPDTVELILSGQAPAKLVVFDYHPEVDEQH
 DAVATARARLADSSVVVESLTEVLGRGKTLPATPIPVADDSADPLALLIY
 TSGSTGAPKGAMYLQSNVGKMWRRSDGNWFGPTAASITLNFMPMSHVMGR
 GILYGTLGNGGTAYFAARSDLSTLLEDLKLVRPTELNFVPRIWETLYDES
 KRAVDRRLANSVGSADRAAIKAEVMDEQRQSLGGRYIAAMTGSAPTSP
 KHGVESLLEMHLLEGYGSTAEMVLF DGEVQRPPVIDYKLVDPDLGYFS
 TDQPYPRGELLKTKQNMFPGYKRPVETATVFDSDGYQYQTDIVAEVGP
 RLVYVDRRNNVLKLAQGGQFVTVAKLEAAFSNSPLVRQIYIYGN
 SAHPYLLAVVPTEDALATNDIEVLKPLIIDS LQKVAKEADLQSYEV
 PRDLIVETTFSL ENGLLTGIRKLAWPKLKQHYGARLEQLYADLVEG
 QANALHVLKQSVANAPVLQTVSRAVGTILGVATTDLPSNAHFTDLGG
 DLSLTSALTFGSLLRELFDIDVPVGVIVSPVNNLVAIADYIERERQ
 GTKRPTFIAIHGRDAGKVHASDLTLDKFIDVSTLTAAPVLAQPGTE
 VRTVLLTGATGFLGRYLALKWLERMDLVEGKVIALVRAKSNEDAR
 ARLDKTFDSGDPKLLAHYQELATDHLEVIAGDKGEVDLELDRQT
 WRRLADTVDLIVDPAALVNHVLPYSELF GPNTLGTAE LIRIALTS
 KQKPYIYVSTIGVGNQIEPAKFTEDSDIRVISPTRNINNNYA
 NGYGNSKWAGEVLLREAHDLCLPVTVFRCDMILADTSYAGQLNVP
 DMFTRMMLSLAATGIAPGSFYELDAESNRQRAHYDGLPVEFIAEAI
 STLGDQSLHDRDGFTTYHVMNPHDDGIGMDEFVDWLI DAGCPIQR
 INDYDEWLRRFEISLRALPERQRHSSLLPLLHNYQKPEKPLHGS
 LAPTIRFRTAVQANANIGQDKDIPHISPAAIAKYVSDLQLLGLV

Q10896

Secuencia de nucleótidos (SEQ 10 N0:75)

>uniprot|Q10896|Q10896_MYCTU PROBABLE PÉPTIDO SINTETASA NRP (PÉPTIDO SINTASA)

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTCGAGGACCTATACGCCAG
 CGACGCCAGTTCGCCGCCAGTCCCAACGAGGCGATCACCCAGGCGATCG
 ACCAGCCCGGGGTCGCGCTTCCACAGCTCATCCGTATGGTCATGGAGGGCTAC
 GCCGATCGGCCGGCACTCGGCCAGCGTGCGCTCCGCTTCGTCACCGACCCCGA
 CAGCGGCCGCACCATGGTCGAGCTACTGCCGCGGTTTCGAGACCATCACCTACC
 GCGAACTGTGGGCCCCGCGCCGGCACATTGGCCACCGCGTTGAGCGCTGAGCCC
 GCGATCCGGCCGGGCGACCGGGTTTTCGCTGCTGGGCTTCAACAGCGTCGACTA
 CACAACCATCGACATCGCGCTGATCCGGTTGGGCGCCGTGTCGGTTCCACTGC
 AGACCAGTGCGCCGGTCACCGGGTTGCGCCCGATCGTCACCGAGACCGAGCCG
 ACGATGATCGCCACCAGCATCGACAATCTTGGCGACCGCGTGAAGTGCTGGC
 CGGTCACGCCCCGGCCGGCTGGTCGATTTCGATTACCACGGCAAGGTTGACA

FIG 4 (Cont.)

CCCACCGCGAGGCCGTCTGAAGCCGCCCGAGCTCGGTTGGCCGGCTCGGTGACC
 ATCGACACACTTGCCGAAGTATCGAACGCGGCAGGGCGCTGCCGGCCACACC
 CATTGCCGACAGCGCCGACGACGCGCTGGCGCTGCTGATTTACACCTCGGGTA
 GTACCGGCGCACCCAAAGGCGCCATGTATCGCGAGAGCCAGGTGATGAGCTTC
 TGGCGCAAGTCGAGTGGCTGGTTCGAGCCGAGCGGTTACCCCTCGATCACGCT
 GAACTTCATGCCGATGAGCCACGTCGGGGGCCGTCAGGTGCTCTACGGGACGC
 TTTCCAACGGCGGTACCGCCTACTTCGTCGCCAAGAGCGACCTGTCGACGCTG
 TTCGAGGACCTCGCCCTGGTGCGGCCACAGAATTGTGCTTCGTGCCGCGCAT
 CTGGGACATGGTGTTCGCAGAGTTCACAGCGAGGTGACCGCCGCTTGGTGG
 ACGGCGCCGATCGAGCGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAG
 AACGTGCTCGGCGGACGGTTTGTTCATGGCGCTGACCGGTTCCGCGCCGATCTC
 CGCTGAGATGACGGCGTGGGTCGAGTCCCTGCTGGCCGACGTGCATTTGGTGG
 AGGGTTACGGCTCCACCGAGGCCGGGATGGTCCCTGAACGACGGCATGGTGGCG
 CGCCCCGCGGTGATCGACTACAAGCTGGTTCGACGTGCCCGAGCTGGGCTACTT
 CGGCACCGATCAGCCCTACCCCGGGGCGAGCTGCTGGTCAAGACGCAAACCA
 TGTTCCCGGCTACTACCAGCGCCCGGATGTCACCGCCGAGGTGTTCCGACCCC
 GACGGCTTCTACCGGACCGGGGACATCATGGCCAAAGTAGGCCCCGACCAGTT
 CGTCTACCTCGACCGCCGCAACAACGTGCTAAAGCTCTCCAGGGCGAGTTCA
 TCGCCGTGTCGAAGCTCGAGGCGGTGTTTCGGCGACAGCCCGCTGGTCCGACAG
 ATCTTCATCTACGGCAACAGTGCCCGGGCCTACCCGCTGGCGGTGGTTGTCCC
 GTCCGGGGACGCGCTTTCTCGCCATGGCATCGAGAATCTCAAGCCCGTGATCA
 GCGAGTCCCTGCAGGAGGTAGCGAGGGCGGCCGGCCTGCAATCCTACGAGATT
 CCACGCGACTTCATCATCGAAACCACGCCGTTACCCCTGGAGAACGGCCTGCT
 CACCGGCATCCGCAAGCTGGCACGCCCGCAGTTGAAGAAGTTCTATGGCGAAC
 GTCTCGAGCGGCTCTATAACCGAGCTGGCCGATAGCCAATCCAACGAGCTGCGC
 GAGCTGCGGCAAAGCGGTCCCGATGCGCCGGTGCTTCCGACGCTGTGCCGTGC
 CGCGGCTGCGTTGCTGGGCTCTACCGCTGCGGATGTGCGGGCCGGACGCGCACT
 TCGCCGACCTGGGTGGTGA CTGCTCTCGGCGCTGTCGTTGGCCAACCTGCTG
 CACGAGATCTTCGGCGTCGACGTGCCGGTGGGTGTCATTGTCAGCCCGGCAAG
 CGACCTGCGGGCCCTGGCCGACCACATCGAAGCAGCGCGCACCGGCGTCAGGC
 GACCCAGCTTCGCCTCGATACACGGTCGCTCCGCGACGGAAGTGCACGCCAGC
 GACCTCACGCTGGACAAGTTCATCGACGCTGCCACCCTGGCCGCAGCCCCGAA
 CCTGCCGGCACCGAGCGCCAAAGTGCGCACCGTACTGCTGACCGGCGCCACCG
 GCTTTTTGGGTGCTACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTC
 AACGGCAAGCTGATCTGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAGC
 CCGGCTGGACGCGACGTTTCGATAGCGGCGACCCGTATTTGGTGCGGCACTACC
 GCGAATTGGGCGCCGGCCGCCTCGAGGTGCTCGCCGGCGACAAGGGCGAGGCC
 GACCTGGGCCTGGACCGGGTCACCTGGCAGCGGCTAGCCGACACGGTGGACCT
 GATCGTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGT
 TCGGCCCAAACGCGGCGGGCACCGCCGAGTTGCTTCGGCTGGCGCTGACCGGC
 AAGCGCAAGCCATACATCTACACCTCGACGATCGCCGTGGGCGAGCAGATCCC

FIG 4 (Cont.)

GCCGGAGGCGTTCACCGAGGACGCCGACATCCGGGCCATCAGCCCGACCCGCA
 GGATCGACGACAGCTACGCCAACGGCTACGCGAACAGCAAGTGGGCGCGGAG
 GTGCTGCTGCGCGAAGCTCACGAGCAGTGCGGCCTGCCGGTGACGGTCTTCCG
 CTGCGACATGATCCTGGCCGACACCAGCTATAACGGTCAGCTCAACCTGCCGG
 ACATGTTACCCGGCTGATGCTGAGCCTGGCCGCTACCGGCATCGCACCCGGT
 TCGTTCTATGAGCTGGATGCGCACGGCAATCGGCAACGCGCCCACTATGACGG
 CTTGCCGGTTCGAATTCGTGCGAGAAGCCATTTGCACCCTTGGGACACATAGCC
 CGGACCGTTTTTGTACCTACCACGTGATGAACCCCTACGACGACGGCATCGGG
 CTGGACGAGTTCGTGACTGGCTCAACTCCCCAACTAGCGGGTCCGGTTGCAC
 GATCCAGCGGATCGCCGACTACGGCGAGTGGCTGCAGCGGTTTCGAGACTTCGC
 TCGGTGCCTTGCCGGATCGCCAGCGCCACGCCTCGCTGCTGCCCTTGCTGCAC
 AACTACCGAGAGCCTGCAAAGCCGATATGCGGGTCAATCGCGCCACCGACCA
 GTTCCGCGCTGCCGTCCAAGAAGCGAAAATCGGTCCGGACAAAGACATTCCGC
 ACCTCACGGCGGCGATCATCGCGAAGTACATCAGCAACCTGCGACTGCTCGGG
 CTGCTGTGA

Secuencia de aminoácidos (SEQ ID N0:76)

>uniprot | Q10896 | Q10896_MYCTU PROBABLE PÉPTIDO SINTETASA NRP (PÉPTIDO SINTASA)

MSINDQRLTRVEDLYASDAQFAAASPNEAITQAIDQPGVALPQLIRVMV
 EGYADRPALGQRALRFVTDPSGRTMVELLPRFETITYRELWARAGTLAT
 ALSAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGL
 RPIVTEPETMIATSIDNLGDAVEVLGHPARLVVFDYHGKVDTHREAV
 EAARARLAGSVTIDTLAELIERGRALPATPIADSADDALALLIYTSGSTG
 APKGAMYRESQVMSFWRKSSGWFEPSPGYPISITLNFMPMSHVGGRQVLYGT
 LSNGGTAYFVAKSDLSTLFEDLALVRPTELCFVPRIWDMVFAEFHSEVDR
 RLVDGADRAALEAQVKAELRENVLGGRFVMALTGSAPI SAEMTAWVESLL
 ADVHLVEGYGSTEAGMVLNDGMVRRPAVIDYKLVDVPELGYFGTDQPYP
 R GELLVKQTQTMFPGYYQRPDVTAEVFDPDGFYRTGDIMAKVGPDQFVYLD
 R RNNVLKLSQGEF IAVSKLEAVFGDSPLVRQIF IYGNSARAYPLAVVPSG
 DALSRHG IENLKPVI SE SLQEVARAAGLQSYE IPRDF I IETTPFTLENGL
 LTGIRKLARPQLKKFYGERLERLYTELADSQSNE LREL RQSGPDAPVLPT
 LCRAAAALLGSTAADVRPDAHFADLGGDSL SALSLANLLHEIFGVDVPVG
 VIVSPASDLRALADHIEAARTGVR RPSFASI HGRSATEVHASDLTLDKFI
 DAATLAAAPNLPAPSAQVRTVLLTGATGFLGRYLALEWLD RMDLVNGKLI
 CLVRARSDEEAQARLDATFDSGDPYLVRHYRELGAGRLEVLGDKGEADL
 GLDRVTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNAAGTAELLRLALT
 GKRKPYIYTSTIAVGEQIPPEAFTE DADIRAI SPTRRIDDSYANGYANSK
 WAGEVLLREAHEQCGLPVTVFRCDMILADTSYTGQLNLPDMFTRLMLSLA

FIG 4 (Cont.)

ATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPDRFVTYH
 VMNPYDDGIGLDEFVDWLNSTSGSGCTIQRIADYGEWLQRFETSLRALP
 DRQRHASLLPLLHNYREPAKPICGSIAPTDQFRAAVQEAKIGPDKDIPHL
 TAAIIAKYISNLRLGLL

Q5YY80

Secuencia de nucleótidos (SEQ ID N0:77)

>uniprot| Q5YY80 | QSYY80_NOFA Supuesta ácido carboxílico reductasa

GTGGAATCCACACGAGCGACGCGTCTGCGGCAGCGGATCGCCGCCCTGTACGC
 CGACGACGCGCAGGTGCGCGACGCGCGGCCGACGAGGCGATCAGCACGGCGC
 TGCGGGAACCGGGCCTGCGGCTGCGCGAGCTCGTCGCCACCGTGGTTCGACGGC
 TACCGCGACCGGCCCGCGCTGGCCGCGCGCTCGGTGCAGCCGGCCGTCGACGC
 CGCGACCGGTGCCTGCGTGGCACGGTTGCTGCCCGAGTACACGACGATGAGTT
 ACGGCGAGCTCGGCCTGCGGTTGCGCGCGGTGGCCGCGGCCTGGCAGCACGAC
 GACGAAACCCGCCTTCGCCCCGGCGAATTCGTGCGGACCCCTGGGTTTCACCAG
 CCCCAGCTACGCCGTCGTGACCTGGCCTGCGTGTGGGCGGGCGCGGTGGCGG
 TGCCGCTGCAGGCGAGCGCGTCCGGTGACGCAGCTGACCGCCATCCTGGCCGAG
 ACCGCGCCCGCAATCCTGGCCACCGGCCTGGACACGCTGCCGCACGCGGTGGA
 CTGTGTGCTCGCCGGGGCCACGCCACGGGCACTGCACGTGTTTCGACTTCGACC
 CCGCCATCGACGCGCAGCGCACGGTGTACGAGGCGGCGTGTGCGCGGCTGGCC
 GGTACCGGTGTGCGCGTGCACGCTCGCCGAGGTCGAGGACCGCGGCCGGGC
 GCTGCCGCCTGCCGTGATCGACGACGGCCCCGGCGACGACCGGCTCGCCCTGT
 TGATCTACACCTCCGGCAGTACCGGCACGCCAAGGGGGCGATGTACACCGAG
 CGGCTGGTCGCGCTGATGTGGCTGGGCCAGCCGCGAGGTCGCCGCGCTCACCGT
 CAACTACCTGCCGCTCAGCCACGTCGCCGGGCGGCTGGCGCTGTTTCGGGCTGC
 TCGCGCGCGGGCGGCACCGCCTACTTCACCGCGCGCGCCGACATGTCCACGCTG
 TTCGAGGATCTGGCGCTGGCCAGGCCGACCGAGCTGTTTCGTGGTGGCCGCGCT
 GTGCGAGATGGTGTGCAACGATTCCAGACCGAGCGGCTGCGGCGCCAGGCCG
 ACGACGACCGGGTCAAGGCCGACCTGCGCCTCGAACTGTTTCGGCGACCGGCTG
 CTCTCGGTGGTGTGCGGCAGCGCGCCGCTGGCCCCGGAGCTGAAGGCGTTTCAT
 GGAATCGGTGCTCGACCTGACCCTGCACGACGGCTACGGCTCCACCGAGGCGG
 GCGGCAGCGTGGTTCATCGACACCACCGTGCAGGCGCCGGTGTGACTAC
 CGGCTCGCCGACGTGCCCGAACTGGGCTATTTCCGTACCGACAAGCCGCATCC
 GCGCGGCGAGCTGCTGCTCAAGACCACCACCATGATCCCCGGCTACTACCGGC
 GGCCGGAGCTCAACGCCAGATCTTCGACGAGGACGGCTTCTACCGCACCGGC
 GACGTGGTCGCGGAACTGGCGCCGGACCGGCTCGTGTACGTCGATCGCCGCAA
 CAATGTGCTCAAGCTGGCGCAGGGCGAGTTCGTACCATCGCCGGCTGGAGG
 CGATCTTCGCCAACAGTCCGCTGGTGCAGGATCTTCGTCTACGGCAACAGC

FIG 4 (Cont.)

GAACGCGCCTATCTGCTGGCGGTGATCGTGCCGAGCCGACAGGCGATGGCGGG
 CGATCCGGCCACGCTGAAGACGCGGATCGCGGAGTCGTTGCAGCTCATCGGCC
 GGGACGCCGAGCTGGAGGCCTACGAGATCCCGCGCGACTTCCTGATCGAGACC
 GAGCCGTTACACCACCGAATCCGGGCTGCTCTCGGGCATCGGCAAGATCCTGCG
 TCCCGCCGTCGAGGCGCGCTATCGCGACCGGCTCGAACAGCTCTACGCCGACC
 TGGCCGCGGCCAGCAGGACGAGCTGGCGGCGCTGCGCCGCGAGGCCGGGCAG
 CGTCCGGTGCTCGAGACCGTCACCCGCGCGGCCGCCGCGATCCTCGGCGGCAC
 GGCGAGCGACCTGAGCCC GGCCGCGC ACTTCACCGATCTCGGCGGCGATTGCG
 TGGCGGCGCTGGCGCTGTGAACTGCTGCGTGAGATCTTCGCCGTCGAGGTG
 CCGGTCGGCGTCATCACCGGCCCGCGACCGACCTCCGTGGCCTGGCCGCCA
 CATCGCCGCGGAACGCGAAAACCGCACCGAGACACCGCTGTTGACCGGGTGC
 ATCCCGACCAGATCCTGATCCGGGCCACCGACCTCGCCCTGGAGAAGTTCTTC
 GACGCCGAGGAGTTGGCCGCCGCGGCCACCGCCGCGCCGCCGGTCCGCCGAGCC
 CCGGGTGGTGCTGCTGACCGGTGCCAACGGCTATCTCGGCCGGTTTCTGTGCC
 TGGAATGGCTGGAACGGCTCGACCGCGTCGACGGACGGCTGATCTGCCTGGTG
 CGCGGCGCGGACGAGGCCGCCGCGCTGGCCC GCCTGGAAGCCGCCTTCGACAG
 CGGCGATCCCGAATTGGTGCGCCGCTTCAAGGAATTGGCCCAGCGCAGGCTCA
 CCGTGGTGGCCGGCGACATCGGCGAGCCCGGCCTGGGCCTGGCCACCGCCACG
 TGGCGACGGCTCGCCGCCGAGGTCGAGCACATCGTGCACCCGGCCGCGCTGGT
 CAACCACGTGCTGCCCTACCGGCAGCTGTTCCGGGCCAACGTGGCGGGCACCG
 CGGAGATCCTGCGGCTCGCGCTCACCGAGCGGGCGCAAGCCGATCGACTTCCTG
 TCCACGGTCCCGTTGCCGCGCAGATAACCCGCCGACCGGTTCCGCCGAGGACGG
 CGACATCCGCGTGATCAGCCCGACCCGCACGGTGGACCGCGGCTACGCCAACG
 GCTACGGCAACAGCAAATGGGCCGCCGAGGTGTTGCTGCGTGCGGCGCACGAC
 CGCTTCGATCTCCCGGTGGCGGTGTTCCGCTCGGACATGATCCTGGCCCACGG
 CAGCTTCGCCGGACAGCTCAACATCCCGACGTGTTACCCGGCTGCTGCTCA
 GCCTGCTGGTACCAGGTATCGCGCCCGCCTCGTTCCACGCCGCGACGGTACC
 GGCGAGCGCCCGCGCGCCACTACGACGGGCTGCCCGCGGACTTCACCGCTGC
 CGCGATCACCGCGCTCGGGGCGCGCACCGCGGGATTCCACACCTACGACGTGC
 TCAACCCGCACGACGACGGCATCAGCCTGGACACCTTCGTGGACTGGCTGATC
 GAGGCCGGACATCCCATCGAACGCATCCCGAGCACAGCGAGTGGGTACCCG
 TTTCGAGACGGCGTTGCACGCCCTGCCCGAACGTCAGCGCAAACACTCGCTGC
 TCCCGCTGTTGCACGCCTACCGCAGGCCGGTGCCCGCGCTGCGCGGCTCGGCG
 CTGCCC GCCGCGGAGTTCGGGGCGGCGGTGCGGGCCGCGAGGCATCACCGCCGA
 CGGTGACATCCCGCACCTGACGCGCGCGCTGATCGAGAAGTACGTCCCGGATC
 TCCGCCTGCACGGACTGTTGTAG

FIG 4 (Cont.)

Secuencia de aminoácidos (SEQ ID N0:78)

>uniprot|Q5YY80 | Q5YY80_NOCA Supuesta ácido carboxílico reductasa

VESTRATRLRQRIAALYADDAQVRDARPEAISTALREPGLRLREL VATV
VDGYRDRPALAARSVQPAVDAATGACVARLLPEYTTMSYGELGLRLRAVA
AAWQHDDDETRLRPGFEFVATLGF TSPDYAVVDLACVWAGAVAVPLQASASV
TQLTAILAETAPAILATGLDTLPHAVDCVLAGATPRALHVFDFDPAIDAQ
RTVYEAACARLAGTGVRVRTLAEVEDRGRALPPAVIDDGGPGDDRLALLIY
TSGSTGTPKGAMYTERLVALMWLGGPQVAALTVNYLPLSHVAGRLALFGL
LARGGTAYFTARADMSTLFEDLALARPTELFVVPVRCMVLRQRFQTERLR
RQADDDRKADLRLELFGDRLLSVVCGSAPLAPELKAFMESVLDLTLHDG
YGSTEAGGSVVIDTTVRRPPVLDYRLADVPELGYFRDTPKPHPRGELLKT
TTMIPGYRRPELNAQIFDEDEGFYRTGDVVAELAPDRLVYVDRRNNVLKL
AQGEFVTIARLEAIFANSPLVRQIFVYGNSEYLLAVIVPSRQAMAGDP
ATLKTRIAESLQLIGRDAELEAYEIPRDFLIETEPFTTESGLLSGIGKIL
RPAVEARYRDRLEQLYADLAAAQQDELAALRREAGQRPVLETVTRAAAAI
LGGTASDLSPA AHFTDLGGDSLAAALSNLLREIFAVEVPVGVITGPATD
LRGLAAHIAAERENRTETPLFDRVHPDQILIRATDLALEKFFDAEELAAA
ATAAPPVAEPRVLLTGANGYLGRFLCLEWLERLDRVDGRLICLVRGADE
AAALARLEAAFDSGDPELVRRFKELAQRRLTVVAGDIGEPGLGLATATWR
RLAAEVEHIVHPAALVNHVLPYRQLFGPNVAGTAEILRLALTEERRKPIDF
LSTVAVAAQIPADRFAEDGDIRVISPTRTVDRGYANGYGNKWA AEVLLR
AAHDFRDLPVAVFRSDMILAHGSFAGQLNIPDVFTRLLLSLLVTGIAPAS
FHAATVTGERPRAHYDGLPADFTAAAITALGARTAGFHTYDVLNPHDDGI
SLDTFVDWLI EAGHP IERIPESHSEWVTRFETALHALPERQRKHSLLPLLH
AYRRPVPALRGSALPAAEFRAAVRAAGITADGDIPHLTRALIEKYVADLR
LHGLL

Q6RKB1

Secuencia de nucleótidos (SEQ ID N0:79)

>uniprot|Q6RKB1 | Q6RKB1_9NOCA ácido carboxílico reductasa dependiente de ATP / NADPH

ATGGCAGTGGATTCACCGGATGAGCGGCTACAGCGCCGCATTGCACAGTTGTT
TGCAGAAGATGAGCAGGTCAAGGCCGCACGTCCGCTCGAAGCGGTGAGCGCGG
CGGTGAGCGGCCCGGTATGCGGCTGGCGCAGATCGCCGCCACTGTTATGGCG

FIG 4 (Cont.)

GGTTACGCCGACCGCCCGGCCGCGGGCAGCGTGCGTTCGAACTGAACACCGA
CGACGCGACGGGCCGCACCTCGCTGCGGTTACTTCCCCGATTTCGAGACCATCA
CCTATCGCGAACTGTGGCAGCGAGTCGGCGAGGTTGCCGCGGCCTGGCATCAT
GATCCCGAGAACCCTTGC GCGCAGGTGATTTTCGTCGCCCTGCTCGGCTTAC
CAGCATCGACTACGCCACCCTCGACCTGGCCGATATCCACCTCGGCGCGGTTA
CCGTGCCGTTGCAGGCCAGCGCGGCGGTGTCCCAGCTGATCGCTATCCTCACC
GAGACTTCGCCGCGGCTGCTCGCCTCGACCCCGGAGCACCTCGATGCGGCGGT
CGAGTGCCTACTCGCGGGCACCACACCGGAACGACTGGTGGTCTTCGACTACC
ACCCCGAGGACGACGACCAGCGTGC GGCCTTCGAATCCGCCCGCCGCGCCTT
GCCGACGCGGGCAGCTTGGTGATCGTCGAAACGCTCGATGCCGTGCGTGCCCG
GGGCCGCGACTTACCGGCCGCGCCACTGTTTCGTTCCCGACACCGACGACGACC
CGCTGGCCCTGCTGATCTACACCTCCGGCAGCACCGGAACGCCGAAGGGCGCG
ATGTACACCAATCGGTTGGCCGCCACGATGTGGCAGGGGAACTCGATGCTGCA
GGGAACTCGCAACGGGTCGGGATCAATCTCAACTACATGCCGATGAGCCACA
TCGCCGGTCGCAATATCGCTGTTCCGGCGTGCTCGCTCGCGGTGGCACCCGACAT
TTCGCGGCCAAGAGCGACATGTCGACACTGTTTCGAAGACATCGGCTTGGTACG
TCCCACCGAGATCTTCTTCGTCCCGCGCGTGTGCGACATGGTCTTCCAGCGCT
ATCAGAGCGAGCTGGACCGGCGCTCGGTGGCGGGCGCCGACCTGGACACGCTC
GATCGGGAAAGTAAAAGCCGACCTCCGGCAGAACTACCTCGGTGGGCGCTTCTT
GGTGGCGGTCGTCCGGCAGCGCGCCGCTGGCCGCGGAGATGAAGACGTTTCATGG
AGTCCGTCTCGATCTGCCACTGCACGACGGGTACGGGTCCGACCGAGGCGGGC
GCAAGCGTGCTGCTCGACAACCAGATCCAGCGGCCGCGGTTGCTCGATTACAA
GCTCGTCGACGTGCCCGAACTGGGTTACTTCCGCACCGACCGGCCGCATCCGC
GCGGTGAGCTGTTGTTGAAGGCGGAGACCAGATTCCGGGCTACTACAAGCGG
CCCGAGGTCACCGCGGAGATCTTCGACGAGGACGGCTTCTACAAGACCGGCGA
TATCGTGGCCGAGCTCGAGCACGATCGGCTGGTCTATGTCGACCGTTCGCAACA
ATGTGCTCAAACGTGTCGAGGGCGAGTTCGTGACCGTCGCCCATCTCGAGGCC
GTGTTCCGACGAGCCCGCTGATCCGGCAGATCTTCATCTACGGCAGCAGCGA
ACGTTCTATCTGCTCGCGGTGATCGTCCCACCGACGACGCGCTGCGCGGCC
GCGACACCGCCACCTTGAAATCGGCACTGGCCGAATCGATTCAGCGCATCGCC
AAGGACGCGAACCTGCAGCCCTACGAGATTCCGCGCGATTTCCTGATCGAGAC
CGAGCCGTTACCATCGCCAACGGACTGCTCTCCGGCATCGCGAAGCTGCTGC
GCCCAATCTGAAGGAACGCTACGGCGCTCAGCTGGAGCAGATGTACACCGAT
CTCGCGACAGGCCAGGCCGATGAGCTGCTCGCCCTGCGCCGCGAAGCCGCCGA
CCTGCCGGTGCTCGAAACCGTACGCCGGGCAGCGAAAGCGATGCTCGGCGTCCG
CCTCCGCCGATATGCGTCCCGACGCGCACTTACCGACCTGGGCGGGCGATTCC
CTTCCGCGCTGTCGTTCTCGAACCTGCTGCACGAGATCTTCGGGGTCGAGGT
GCCGGTGGGTGTCGTCGTCAGCCCGGCGAACGAGCTGCGCGATCTGGCGAATT
ACATTGAGGCGGAACGCAACTCGGGCGCGAAGCGTCCACCTTACCTCGGTG
CACGGCGGGCGGTTCCGAGATCCGCGCCGCGATCTGACCCTCGACAAGTTCAT
CGATGCCCGCACCTGGCCGCCGCGACAGCATTCCGCACGCGCCGGTGCCAG

FIG 4 (Cont.)

CGCAGACGGTGCTGCTGACCGGCGCGAACGGCTACCTCGGCCGGTTCCTGTGC
 CTGGAATGGCTGGAGCGGCTGGACAAGACGGGTGGCACGCTGATCTGCGTCGT
 GCGCGGTAGTGACGCGGCCGCGGCCCGTAAACGGCTGGACTCGGCGTTCGACA
 GCGGCGATCCCGGCCTGCTCGAGCACTACCAGCAACTGGCCGCACGGACCCTG
 GAAGTCCTCGCCGGTGATATCGGCGACCCGAATCTCGGTCTGGACGACGCGAC
 TTGGCAGCGGTTGGCCGAAACCGTCGACCTGATCGTCCATCCCGCCGCGTTGG
 TCAACCACGTCCTTCCCTACACCCAGCTGTTTCGGCCCCAATGTCGTCGGCACC
 GCCGAAATCGTCCGGTTGGCGATCACGGCGCGGGCGCAAGCCGGTCACCTACT
 GTCGACCGTCGGAGTGGCCGACCAGGTCGACCCGGCGGAGTATCAGGAGGACA
 GCGACGTCGCGGAGATGAGCGCGGTGCGCGTCGTGCGCGAGAGTTACGCCAAC
 GGCTACGGCAACAGCAAGTGGGCGGGGGAGGTCCTGCTGCGCGAAGCACACGA
 TCTGTGTGGCTTGCCGGTTCGCGGTGTTCCGTTTCGGACATGATCCTGGCGCACA
 GCCGGTACGCGGGTCAGCTCAACGTCCAGGACGTGTTACCCGGCTGATCCTC
 AGCCTGGTCGCCACCGGCATCGCGCCGTAICTCGTTCACCGAACCGACGCGGA
 CGGCAACCGGCAGCGGGCCACTATGACGGCTTGCCGGCGGACTTCACGGCGG
 CGGCGATCACCGCGCTCGGCATCCAAGCCACCGAAGGCTTCCGGACCTACGAC
 GTGCTCAATCCGTACGACGATGGCATCTCCCTCGATGAATTCGTCGACTGGCT
 CGTCGAATCCGGCCACCCGATCCAGCGCATCACCGACTACAGCGACTGGTTCC
 ACCGTTTTCGAGACGGCGATCCGCGCGCTGCCGGAAAAGCAACGCCAGGCCTCG
 GTGCTGCCGTTGCTGGACGCCTACCGCAACCCTGCCCGGCGGTCCGCGGCGC
 GATACTCCCGGCCAAGGAGTTCCAAGCGGCGGTGCAAACAGCCAAAATCGGTC
 CGGAACAGGACATCCCGCATTTGTCCGCGCCACTGATCGATAAGTACGTCAGC
 GATCTGGAACCTGCTTCAGCTGCTCTGA

Secuencia de aminoácidos (SEQ ID NO:80)

>uniprot | Q6RKB1 | Q6RKB1_9NOCA ácido carboxílico reductasa dependiente de ATP/NADPH

MAVDSPPERLQRRIAQLFAEDEQVKAARPLEAVSAAVSAPGMRLAQIAAT
 VMAGYADRPAAGQRAFELNTDDATGRTSLRLLPRFETITYRELWQRVGEV
 AAAWHHPENPLRAGDFVALLGFTSIDYATLDLADIHLGAVTVPLQASAA
 VSQLIAILTETSPRLLASTPEHLDAAVECLLAGTTPERLVVFDYHPEDDD
 QRAAFESARRRLADAGSLVIVETLDAVRARGRDLPAAPLFVPDTHDDPLA
 LLIYTSGSTGTPKGAMYTNRLAATMWQGNMQLQNSQRVGINLNYMPMSH
 IAGRISLFGVLRGGTAYFAAKSDMSTLFEDI GLVVRPTEIFFVPRVCDMV
 FQRYQSELDRRSVAGADLDTLDREVKADLRQNYLGGRFLVAVVGSAPLAA
 EMKTFMESVLDLPLHDGYGSTEAGASVLLDNQIQRPVLDYKLVDPVPELG
 YFRTDRPHPRGELLLKAETTIPGYKRPVTAEIFDEDGFYKTGDIVAEEL
 EHDRLVYVDRRNNVLKLSQGEFVTVAHLEAVFASSPLIRQIFIYSSERS
 YLLAVIVPTDDALRGRDTATLKSALAESIQRIDANLQPYEIPRDFLIE

FIG 4 (Cont.)

TEPFTIANGLLSGIAKLLRPNLKERYGAQLEQMYTDLATGQADELLALRR
 EAADLPVLETVSRAAKAMLGVASADM RPDAHF TD LGGDSL SALSFSNLLH
 EIFGVEVPVGVVVSPANELRDLANYIEAERNSGAKRPTFTSVHGGGSEIR
 AADLTLDKFIDARTLAAADSI PHAPVPAQT VLLTGANGYLGRFLCLEWLE
 RLDKTGGTLICVVRGSDAAAARKRLDSAFD SGDPGLLEHYQQLAARTLEV
 LAGDIGDPNLGLDDATWQRLAETVDLIVHPAALVNHVLPYTQLFGPNVVG
 TAEIVRLAITARRKPVTYLSTVGVADQVDPAEYQEDSDVREMSAVRVVRE
 SYANGYGN SKWAGEVLLREAHDL CGLPVAVFRSDMILAH SRYAGQLNVQD
 VFTRLILSLVATGIAPYSFYRTDADGNRQRAHYDGLPADFTAAAITALGI
 QATEGFRTYDVLN PYDDGISLDEFVDWLVE SGHP IQRITDYSDFH R FET
 AIRALPEKQRQASVLP LLDAYRNPCPAVRGAILPAKEFQA AVQTAKIGPE
 QDIPHLSAPLIDKYVSDLELLQLL

Q741P9**Secuencia de nucleótidos (SEQ ID NO:81)**

>uniprot|Q741P9 | Q741P9_MYCPA FadD9

ATGTCGACTGCCACCCATGACGAACGACTCGACCGTCGCGTCCACGAACTCAT
 CGCCACCGACCCGCAATTCGCCGCCGCCCAACCCGACCCGGCGATCACCGCCG
 CCCTCGAACAGCCCGGGCTGCGGCTGCCGCAGATCATCCGCACCGTGCTCGAC
 GGCTACGCCGACCGGCCGCGCTGGGACAGCGCGTGGTGGAGTTCGTCACGGA
 CGCCAAGACCGGGCGCACGTCGGCGCAGCTGCTCCCCGCTTCGAGACCATCA
 CGTACAGCGAAGTAGCGCAGCGTGTTCGGCGCTGGGCCGCGCCCTGTCCGAC
 GACGCGGTGCACCCCGGCGACCGGGTGTGCGTGCTGGGCTTCAACAGCGTCGA
 CTACGCCACCATCGACATGGCGCTGGGCGCCATCGGCGCCGTCTCGGTGCCGC
 TGCAGACCAGCGCGGCAATCAGCTCGCTGCAGCCGATCGTGGCCGAGACCGAG
 CCCACCCTGATCGCGTCCAGCGTGAACCAGCTGTCCGACGCGGTGCAGCTGAT
 CACCGGCGCCGAGCAGGCGCCCACCCGGCTGGTGGTGTTCGACTACCACCCGC
 AGGTCGACGACCAGCGCGAGGCCGTCAGGACGCCGCGGCGCGGCTGTCCAGC
 ACCGGCGTGGCCGTCCAGACGCTGGCCGAGCTGCTGGAGCGCGGCAAGGACCT
 GCCCGCCGTCGCGGAGCCGCCCGGACGAGGACTCGCTGGCCCTGCTGATCT
 ACACCTCCGGGTCCACCGGCGCCCCCAAGGGCGCGATGTACCCACAGAGCAAC
 GTCGGCAAGATGTGGCGCCGCGGCAGCAAGAACTGGTTCGGCGAGAGCGCCGC
 GTCGATCACCTGAACTTCATGCCGATGAGCCACGTGATGGGCCGAAGCATCC
 TCTACGGCACGCTGGGCAACGGCGGCACCGCCTACTTCGCCGCCCGCAGCGAC
 CTGTCCACCCTGCTTGAGGACCTCGAGCTGGTGC GGCCACCGAGCTCAACTT
 CGTCCCGCGGATCTGGGAGACGCTGTACGGCGAATTCCAGCGTCAGGTCGAGC
 GCGCGCTCTCCGAGGCCGGGGACGCCGGCGAACGTCGCGCCGTCGAGGCCGAG
 GTGCTGGCCGAGCAGCGCCAGTACCTGCTGGGCGGGCGGTTACCTTCGCGAT

FIG 4 (Cont.)

GACGGGCTCGGCGCCCATCTCGCCGGAGCTGCGCAACTGGGTCGAGTCGCTGC
 TCGAAATGCACCTGATGGACGGCTACGGCTCCACCGAGGCCGGAATGGTGTG
 TTCGACGGGGAGATTACGCGCCCGCCGGTATCGACTACAAGCTGGTTCGACGT
 GCCGGACCTGGGCTACTTCAGCACCGACCGGCCGCATCCGCGCGGGCGAGCTGC
 TGCTGCGCACCGAGAACATGTTCCCGGGCTACTACAAGCGGGCCGAAACCACC
 GCGGGCGTCTTCGACGAGGACGGCTACTACCGCACCGGGCGACGTGTTCCCGA
 GATCGCCCCGGACCGGCTGGTCTACGTGACCGCCGCAACAACGTGCTCAAGC
 TGGCGCAGGGCGAATTCGTACGCTGGCCAAGCTGGAGGCGGTGTTCCGGCAAC
 AGCCCGCTGATCCGCCAGATCTACGTCTACGGCAACAGCGCCCAGCCCTACCT
 GCTGGCGGTGCTGGTGCCACCGAGGAGGCGCTGGCCTCGGGTGACCCCGAGA
 CGCTCAAGCCCAAGATCGCCGACTCGCTGCAGCAGGTGCGCAAGGAGGCCGGC
 CTGCAGTCTACGAGGTGCCGCGCGACTTCATCATCGAGACCACCCCGTTCAG
 CCTGGAAAACGGTCTGCTGACCGGGATCCGGAAGCTGGCGTGGCCGAAACTGA
 AGCAGCACTACGGGGAACGGCTGGAGCAGATGTACGCCGACCTGGCCGCCGGA
 CAGGCCAACGAGCTGGCCGAGCTGCGCCGCAACGGTGCCAGGCGCCGGTGT
 GCAGACCGTGAGCCGCGCCGCGGGCGCCATGCTGGGTTCGGCCGCCTCCGACC
 TGTCCCCCGACGCCACTTCACCGATCTGGGCGGAGACTCGTTGTCGGCGTTG
 ACATTCGGCAACCTGCTGCGCGAGATCTTCGACGTGACGTGCCGGTAGGCGT
 GATCGTCAGCCCGGCCAACGACCTGGCGGCCATCGCGAGCTACATCGAGGCCG
 AGCGGCAGGGCAGCAAGCGCCCGACGTTTCGCCTCGGTGCACGGCCGGGACGCG
 ACCGTGGTGCAGCGCCGCGACCTGACGCTGGACAAGTTCCTCGACGCCGAGAC
 GCTGGCCCGCCGCGCCGAACCTGCCCAAGCCGGCCACCGAGGTGCGCACCGTGC
 TGCTGACCGGCGCCACCGGCTTCCTGGGCGCTACCTGGCCCTGGAATGGCTG
 GAGCGGATGGACATGGTGGACGGCAAGGTCATCGCCCTGGTCCGGGCCCCGCTC
 CGACGAGGAGGCACGCGCCCGGCTGGACAAGACCTTCGACAGCGGCGACCCGA
 AACTGCTCGCGCACTACCAGCAGCTGGCCGCCGATCACCTGGAGGTCATCGCC
 GCGACAAGGGCGAGGCCAATCTGGGCCTGGGCCAAGACGTTTGGCAACGACT
 GGCCGACACGGTCGACGTGATCGTGCACCCCGCCGCGCTGGTCAACCACGTGT
 TGCCGTACAGCGAGCTGTTCCGGGCCAACGCCCTGGGCACCGCGGAGCTGATC
 CGGCTGGCGCTGACGTCCAAGCAGAAGCCGTACACCTACGTGTCCACCATCGG
 CGTGGGCGACCAGATCGAGCCGGGCAAGTTCGTGAGAACGCCGACATCCGGC
 AGATGAGCGCCACCCGGGCGATCAACGACAGCTACGCCAACGGCTATGGCAAC
 AGCAAGTGGGCCGGCGAGGTGCTGCTGCGCGAGGCGCACGACCTGTGCGGGCT
 GCCCGTCGCGGTGTTCCGCTGCGACATGATCCTGGCCGACACCACGTATGCCG
 GGCAGCTCAACCTGCCGGACATGTTACCCGGCTGATGCTGAGCCTGGTGGCC
 ACCGGGATCGCGCCCGGCTCGTTCTACGAGCTCGACGCCGACGGCAACCGGCA
 GCGGGCGCACTACGACGGCCTGCCGGTTCGAGTTCATCGCCGCGGGCGATCTCGA
 CGCTGGGTTCGCAGATCACCGACAGCGACACCGGCTTCAGACCTACCACGTG
 ATGAACCCCTACGATGACGGCGTCCGGTCTGGACGAGTACGTGATTGGCTGGT
 GGACGCCGGCTATTCGATCGAGCGGATCGCCGACTACTCCGAATGGCTGCGGC
 GGTTTCGAGACCTCGCTGCGGGCCCTGCCGGACCGGCAGCGCCAGTACTCGCTG

FIG 4 (Cont.)

CTGCCGCTGCTGCACAACCTACCGCACGCCGGAGAAGCCGATCAACGGGTCGAT
AGCTCCCACCGACGTGTTCCGGGCAGCGGTGCAGGAGGCGAAAATCGGCCCGG
ACAAAGACATTCCGCACGTGTCGCCGCCGGTCATCGTCAAGTACATCACCGAC
CTGCAGCTGCTCGGGCTGCTCTGA

Secuencia de aminoácidos (SEQ ID N0:82)

>uniprot|Q741P9 | Q741P9_MYCPA FadD9

MSTATHDERLDRRVHEL IATDPQFAAAQDPAITAAL EQPGLRLPQIIRT
VLDGYADRPALGQRVVEFVTD AKTGRTSAQLLPRFETITYSEVAQRVSAL
GRALSDDAVHPGDRVCV LGFNSVDYATIDMALGAIGAVSVPLQTSAAISS
LQPIVAETEPTLIASSVNQLSDAVQLITGAEQAPTRLVVF DYHPQVDDQR
EAVQDAAARLSSTGVAVQTLAELLER GKDLPAVAEPPADED SLALLIYTS
GSTGAPKGAMYPQSNV GKMWRGSKNWFGE SAASITLNFMPMSHVMGRSI
LYGTLGNGGTAYFAARSDLSTLLEDLELVRPTELN FVPRIWETLYGEFQR
QVERRLSEAGDAGERRAVEAEVLAEQRQYLLGGRFTFAMTGSAPI SPELR
NWVESLLEMHLMDGYGSTEAGMVLFDGEIQRPPVIDYKLVDV PDLGYFST
DRPHPRGELLRLTENMFP GYYKRAETTAGVFDEDGY YRTGDVFAEIAPDR
LVYVDRRNNVLKLAQGEFVTLAKLEAVFGNSPLIRQIYVYGN SAQP YLLA
VVVPTTEALASGDPETLKP KIADSLQOVAKEAGLQSYEVPRDFI IETTPF
SLENGLLTGIRKLAWPKLKQHYGERLEQMYADLAAGQAN ELAELRRNGAQ
APVLQTVSRAAGAMLGSAASDLSPDAHFTDLGGDSL SALTFGNLLREIFD
VDVPVGVIVSPANDLAAIASYIEAERQGSKRPTFASVHGRDATVVRAADL
TLDKFLDAETLAAAPNLPKPATEVRTVLLTGATGFLGRYLALEWLERMDM
VDGKVIALVRARSDEEARARLDKTFDSGDPKLLAHYQQLAADHLEVIAGD
KGEANLGLGQDVWQRLADTVDVIVDPAALVNHVLPYSELFGPNALGT AEL
IRLALTSKQKPYTYVSTIGVGDQIEPGKFVENADIRQMSATRAINDSYAN
GYGNSKWAGEVLLREAHDL CGLPVAVFRCDMILADTTYAGQLNLPDMFTR
LMLS LVATGIAPGSFYELDADGNRQRAHYDGLPVEFIAAAI STLGSQITD
SDTG FQTYHVMNPYDDGVGLDEYVDWLVDAGYSIERIADYSEWLRRFETS
LRALPDRQRQYSLLPLLHNYRTPEKPI NGSIAPTDVFRAAVQEAKIGPDK
DIPHVSPVIVKYITDLQLLGLL

FIG 4 (Cont.)

Q7D6X4**Secuencia de nucleótidos (SEQ ID NO:83)**

>uniprot| Q7D6X4 | Q7D6X4_MYCTU Sustrato--CoA ligasa, supuesta

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTTCGAGGACCTATACGCCAG
 CGACGCCAGTTCGCCGCCGCCAGTCCCAACGAGGCGATCACCCAGGCGATCG
 ACCAGCCCGGGGTCGCGCTTCCACAGCTCATCCGTATGGTCATGGAGGGCTAC
 GCCGATCGGCCGGCACTCGGCCAGCGTTCGCGCTCCGCTTCGTCACCGACCCCGA
 CAGCGGCCGCACCATGGTCGAGCTACTGCCGCGGTTTCGAGACCATCACCTACC
 GCGAACTGTGGGCCCCGCGCCGGCACATTGGCCACCGCGTTGAGCGCTGAGCCC
 GCGATCCGGCCGGGCGACCGGGTTTGCCTGCTGGGCTTCAACAGCGTCGACTA
 CACAACCATCGACATCGCGCTGATCCGGTTGGGCGCCGTGTCGGTTCCACTGC
 AGACCAGTGCGCCGGTTCACCGGGTTGCGCCCGATCGTCACCGAGACCGAGCCG
 ACGATGATCGCCACCAGCATCGACAATCTTGGCGACGCCGTCGAAGTGCTGGC
 CGGTCACGCCCCGGCCCGGCTGGTCGTATTTCGATTACCACGGCAAGGTTGACA
 CCCACCGCGAGGCCGTCGAAGCCGCCGAGCTCGGTTGGCCGGCTCGGTGACC
 ATCGACACACTTGCCGAACCTGATCGAACGCGGCAGGGCGCTGCCGGCCACACC
 CATTGCCGACAGCGCCGACGACGCGCTGGCGCTGCTGATTTACACCTCGGGTA
 GTACCGGGCGCACCCAAAGGCGCCATGTATCGCGAGAGCCAGGTGATGAGCTTC
 TGGCGCAAGTCGAGTGGCTGGTTTCGAGCCGAGCGGTTACCCCTCGATCACGCT
 GAACTTCATGCCGATGAGCCACGTCGGGGGCGCTCAGGTGCTCTACGGGACGC
 TTTCCAACGGCGGTACCGCCTACTTCGTCGCCAAGAGCGACCTGTCGACGCTG
 TTCGAGGACCTCGCCCTGGTTCGGGCCACAGAATTGTGCTTCGTGCCGCGCAT
 CTGGGACATGGTGTTCGCAGAGTTCACACAGCGAGGTCGACCGCCGCTTGGTGG
 ACGGCGCCGATCGAGCGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAG
 AACGTGCTCGGCGGACGGTTTGTTCATGGCGCTGACCGGTTCCGCGCCGATCTC
 CGCTGAGATGACGGCGTGGGTCGAGTCCCTGCTGGCCGACGTGCATTTGGTGG
 AGGGTTACGGCTCCACCGAGGCCGGGATGGTCCTGAACGACGGCATGGTGC GG
 CGCCCCGCGGTGATCGACTACAAGCTGGTCGACGTGCCCGAGCTGGGCTACTT
 CGGCACCGATCAGCCCTACCCCGGGGCGAGCTGCTGGTCAAGACGCAAACCA
 TGTTCCCGGGCTACTACCAGCGCCCGGATGTCACCGCCGAGGTGTTTCGACCCC
 GACGGCTTCTACCGGACCGGGGACATCATGGCCAAAGTAGGCCCCCGACCAGTT
 CGTCTACCTCGACCGCCGCAACAACGTGCTAAAGCTCTCCAGGGCGAGTTCA
 TCGCCGTGTCGAAGCTCGAGGCGGTGTTTCGGCGACAGCCCGCTGGTCCGACAG
 ATCTTCATCTACGGCAACAGTGCCCGGGCCTACCCGCTGGCGGTGGTTGTCCC
 GTCCGGGGACGCGCTTTCTCGCCATGGCATCGAGAATCTCAAGCCCGTGATCA
 GCGAGTCCCTGCAGGAGGTAGCGAGGGGCGGCCGGCCTGCAATCCTACGAGATT

FIG 4 (Cont.)

CCACGCGACTTCATCATCGAAACCACGCCGTTACCCTGGAGAACGGCCTGCT
 CACCGGCATCCGCAAGCTGGCACGCCCGCAGTTGAAGAAGTTCTATGGCGAAC
 GTCTCGAGCGGCTCTATAACCGAGCTGGCCGATAGCCAATCCAACGAGCTGCGC
 GAGCTGCGGCAAAGCGGTCCCGATGCGCCGGTGCTTCCGACGCTGTGCCGTGC
 CGCGGCTGCGTTGCTGGGCTCTACCGCTGCGGATGTGCGGGCCGGACGCGCACT
 TCGCCGACCTGGGTGGTGACTCGCTCTCGGCGCTGTCGTTGGCCAACCTGCTG
 CACGAGATCTTCGGCGTCGACGTGCCGGTGGGTGTCATTGTCAGCCCGGCAAG
 CGACCTGCGGGCCCTGGCCGACCACATCGAAGCAGCGCGCACCGGGCGTCAGGC
 GACCCAGCTTCGCCTCGATACACGGTCGCTCCGCGACGGAAGTGCACGCCAGC
 GACCTCACGCTGGACAAGTTCATCGACGCTGCCACCTTGGCCGCAGCCCCGAA
 CCTGCCGGCACCGAGCGCCCAAGTGCGCACCGTACTGCTGACCGGGCGCCACCG
 GCTTTTTGGGTGCTACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTC
 AACGGCAAGCTGATCTGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAGC
 CCGGCTGGACGCGACGTTTCGATAGCGGCGACCCGTATTTGGTGCGGCACTACC
 GCGAATTGGGCGCCGGCCGCTCGAGGTGCTCGCCGGCGACAAGGGCGAGGCC
 GACCTGGGCCTGGACCGGGTACCTGGCAGCGGCTAGCCGACACGGTGGACCT
 GATCGTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGT
 TCGGCCCAAACGCGGGCGGGCACCGCCGAGTTGCTTCGGCTGGCGCTGACCGGC
 AAGCGCAAGCCATACATCTACACCTCGACGATCGCCGTGGGCGAGCAGATCCG
 CCCGGAGGCGTTCACCGAGGACGCCGACATCCGGGCCATCAGCCCGACCCGCA
 GGATCGACGACAGCTACGCCAACGGCTACGCGAACAGCAAGTGGGCGGGCGAG
 GTGCTGCTGCGCGAAGCTCACGAGCAGTGCGGCCTGCCGGTGACGGTCTTCCG
 CTGCGACATGATCCTGGCCGACACCAGCTATAACCGGTGAGCTCAACCTGCCGG
 ACATGTTACCCCGGCTGATGCTGAGCCTGGCCGCTACCGGCATCGCACCCGGT
 TCGTTCTATGAGCTGGATGCGCACGGCAATCGGCAACGCGCCCACTATGACGG
 CTTGCCGGTCGAATTCGTGCGAGAAGCCATTTGCACCCTTGGGACACATAGCC
 CGGACCGTTTTTGTACCTACCACGTGATGAACCCCTACGACGACGGCATCGGG
 CTGGACGAGTTCGTGACTGGCTCAACTCCCCAACTAGCGGGTCCGGTTGCAC
 GATCCAGCGGATCGCCGACTACGGCGAGTGGCTGCAGCGGTTTCGAGACTTCGC
 TGCGTGCCCTGCCGGATCGCCAGCGCCACGCCTCGCTGCTGCCCTTGCTGCAC
 AACTACCGAGAGCCTGCAAAGCCGATATGCGGGTCAATCGCGCCCACCGACCA
 GTTCCGCGCTGCCGTCCAAGAAGCGAAAATCGGTCCGGACAAAGACATTCCGC
 ACCTCACGGCGGGCGATCATCGCGAAGTACATCAGCAACCTGCGACTGCTCGGG
 CTGCTGTGA

Secuencia de aminoácidos (SEQ ID N0:84)

>uniprot|Q7D6X4 | Q7D6X4_MYCTU Sustrato--CoA ligasa, supuesta

MSINDQRLTRVEDLYASDAQFAASPNEAITQAIQPGVALPQLIRMVM

FIG 4 (Cont.)

EGYADRPALGQRALRFVTDPSGRTMVELLPRFETITYRELWARAGTLAT
 ALSAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGL
 RPIVTETEPTMIATSIDNLGDAVEVLGAPARLVVFDYHGKVDTHREAV
 EAARARLAGSVTIDTLAELIERGRALPATPIADSADDALALLIYTSGSTG
 APKGAMYRESQVMSFWRKSSGWFEPSPGYSITLNFMPMSHVGGQVLYGT
 LSNGGTAYFVAKSDLSTLFEDLALVRPTELCFVPRIWDMVFAEFHSEVDR
 RLVDGADRAALEAQVKAELRENVLGGRFVMALTGSAPI SAEMTAWVESLL
 ADVHLVEGYGSTEAGMVLNDGMVRRPAVIDYKLVDVPELGYFGTDQPYPR
 GELLVKTQTMFPGYYQRPDVTAEVFDPDGFYRTGDIMAKVGPDQFVYLDR
 RNNVLKLSQGEFIAVSKLEAVFGDSPLVRQIFIYGN SARAYPLAVVPSG
 DALSRHGIENLKPVI SESLQEVARAAGLQSYEIPRDFI IETTPFTLENGL
 LTGIRKLARPQLKKFYGERLERLYTELA DSQSNE LREL RQSGPDAPVLP
 LCRAAAALLGSTAADVRPDAHFADLGGDSL SALS LANLLHEIFGVDVPVG
 VIVSPASDLRALADHIEAARTGVR RPSFASI HGRSATEVHASDLTLDKFI
 DAATLAAAPNLPAPSAQVRTVLLTGATGFLGRYLALEWLD RMDLVNGKLI
 CLVRARSDEEAQARLDATFDSGDPYLVRHYRELGAGRLEVLGDKGEADL
 GLDRVTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNAAGTAE LLRLALT
 GKRKPYYITSTIAVGEQIRPEAFTE DADIRAI SPTRRIDDSYANGYANSK
 WAGEVLLREAHEQCGLPVTVFRCDMILADTSYTGQLNLPDMFTRMLSLA
 ATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPDRFV TYH
 VMNPYDDGIGLDEFVDWLN SPTS GSGCTIQRIADYGEWLQRFETSLRALP
 DRQRHASLLPLLHNYREPAK PICGSIAPT DQFRAAVQEAKIGPKDIPHL
 TAAIIAKYISNLRLGLL

Q7TY99

Secuencia de nucleótidos (SEQ ID N0:85)

>uniprot|Q7TY99 | Q7TY99_MYCBO PROBABLE ÁCIDO GRASO-CoA LIGASA FADD9 (ÁCIDO GRASO-COA SINTETASA) (ÁCIDO GRASO- COA SINTASA)

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTCGAGGACCTATACGCCAG
 CGACGCCAGTTCGCCGCCGCGCCAGTCCCAACGAGGCGATCACCCAGGCGATCG
 ACCAGCCCGGGGTCGCGCTTCCACAGCTCATCCGTATGGTCATGGAGGGCTAC
 GCCGATCGGCCGGCACTCGGCCAGCGTGCGCTCCGCTTCGTCACCGACCCCGA
 CAGCGGCCGCACCATGGTCGAGCTACTGCCGCGGTTTCGAGACCATCACCTACC
 GCGAACTGTGGGCCCGCGCCGGCACATTGGCCACCGCGTTGAGCGCTGAGCCC
 GCGATCCGGCCGGGCGACCGGGTTTTCGCTGCTGGGCTTCAACAGCGTCGACTA
 CACAACCATCGACATCGCGCTGATCCGGTTGGGCGCCGTGTCGGTTCCACTGC
 AGACCAGTGCGCCGGTCACCGGGTTGCGCCCGATCGTCACCGAGACCGAGCCG

FIG 4 (Cont.)

ACGATGATCGCCACCAGCATCGACAATCTTGCGGACGCCGTCGAAGTGCTGGC
 CGGTCACGCCCCGGCCCCGGCTGGTCGTATTCGATTACCACGGCAAGGTTGACA
 CCCACCGCGAGGCCGTCGAAGCCGCCGAGCTCGGTTGGCCGGCTCGGTGACC
 ATCGACACACTTGCCGAAC TGATCGAACGCGGCAGGGCGCTGCCGGCCACACC
 CATTGCCGACAGCGCCGACGACGCGCTGGCGCTGCTGATTTACACCTCGGGTA
 GTACCGGCGCACCCAAAGGCGCCATGTATCGCGAGAGCCAGGTGATGAGCTTC
 TGGCGCAAGTCGAGTGGCTGGTTCGAGCCGAGCGGTTACCCCTCGATCACGCT
 GAACTTCATGCCGATGAGCCACGTCGGGGGCCGTCAGGTGCTCTACGGGACGC
 TTTCCAACGGCGGTACCGCCTACTACGTCGCCAAGAGCGACCTGTGACGCTG
 TTCGAGGACCTCGCCCTGGTGC GGCCACAGAATTGTGCTTCGTGCCGCGCAT
 CTGGGACATGGTGTTCGCAGAGTTCACAGCGAGGTCGACCGCCGCTTGGTGG
 ACGGCGCCGATCGAGCGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAG
 AACGTGCTCGGCGGACGGTTTTGTATGGCGCTGACCGGTTCCGCGCCGATCTC
 CGCTGAGATGACGGCGTGGGTGCGAGTCCCTGCTGGCCGACGTGCATTTGGTGG
 AGGGTTACGGCTCCACCGAGGCCGGGATGGTCCTGAACGACGGCATGGTGGCG
 CGCCCCGCGGTGATCGACTACAAGCTGGTTCGACGTGCCCGAGCTGGGCTACTT
 CGGCACCGATCAGCCCTACCCCGGGGCGAGCTGCTGGTCAAGACGCAAACCA
 TGTTCCCGGCTACTACCAGCGCCCGGATGTCACCGCCGAGGTGTTGACCCC
 GACGGCTTCTACCGGACCGGGGACATCATGGCCAAAGTAGGCCCCGACCAGTT
 CGTCTACCTCGACCGCCGAACAACGTGCTAAAGCTCTCCAGGGCGAGTTCA
 TCGCCGTGTGCAAGCTCGAGGCGGTGTTTCGGCGACAGCCCGCTGGTCCGACAG
 ATCTTCATCTACGGCAACAGTGCCCGGGCCTACCCGCTGGCGGTGGTTGTCC
 GTCCGGGGACGCGCTTTCTCGCCATGGCATCGAGAATCTCAAGCCCGTGATCA
 GCGAGTCCCTGCAGGAGGTAGCGAGGGCGGCCGGCCTGCAATCCTACGAGATT
 CCACGCGACTTCATCATCGAAACCACGCCGTTACCCCTGGAGAACGGCCTACT
 CACCGGCATCCGCAAGCTGGCACGCCCGCAGTTGAAGAAGTTCTATGGCGAAC
 GTCTCGAGCGGCTCTATAACCGAGCTGGCCGATAGCCAATCCAACGAGCTGCGC
 GAGCTGCGGCAAAGCGGTCCCGATGCGCCGGTGCTTCCGACGCTGTGCCGTGC
 CGCGGCTGCGTTGCTGGGCTCTACCGCTGCGGATGTGCGGCCGGACGCGCACT
 TCGCCGACCTGGGTGGTGA CTGCTCTCGGCGCTGTCGTTGGCCAACCTGCTG
 CACGAGATCTTCGGCGTCGACGTGCCGGTGGGTGTCATTGTCAGCCCGGCAAG
 CGACCTGCGGGCCCTGGCCGACCACATCGAAGCAGCGCGCACCGGGCGTCAGGC
 GACCCAGCTTCGCCTCGATACACGGTCGCTCCGCGACGGAAGTGCACGCCAGC
 GACCTCACGCTGGACAAGTTCATCGACGCTGCCACCCTGGCCGCAGCCCCGAA
 CCTGCCGGCACCGAGCGCCCAAGTGCGCACCGTACTGCTGACCGGCGCCACCG
 GCTTTTTGGGTGCTACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTC
 AACGGCAAGCTGATCTGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAAGC
 CCGGCTGGACGCGACGTTTCGATAGCGGCGACCCGTATTTGGTGCGGCACTACC
 GCGAATTGGGCGCCGGCCGCCTCGAGGTGCTCGCCGGCGACAAGGGCGAGGCC
 GACCTGGGCCTGGACCGGGTCACCTGGCAGCGGCTAGCCGACACGGTGGACCT
 GATCGTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGT

FIG 4 (Cont.)

TCGGCCCAAACGCGGGCGGGCACCAGCGAGTTGCTTCGGCTGGCGCTGACCGGC
 AAGCGCAAGCCATACATCTACACCTCGACGATCGCCGTGGGCGAGCAGATCCC
 GCCGGAGGCGTTACCGAGGACGCCGACATCCGGGCCATCAGCCCGACCCGCA
 GGATCGACGACAGCTACGCCAACGGCTACGCGAACAGCAAGTGGGCGGGCGAG
 GTGCTGCTGCGCGAAGCTCACGAGCAGTGCAGGCTGCGGGTACGGTCTTCCG
 CTGCGACATGATCCTGGCCGACACCAGCTATACCGGTCAGCTCAACCTGCCGG
 ACATGTTACCCGGCTGATGCTGAGCCTGGCCGCTACCGGCATCGCACCCGGT
 TCGTTCATGAGCTGGATGCGCACGGCAATCGGCAACGCGCCCACTATGACGG
 CTTGCCGGTTCGAATTCGTGCGAGAAGCCATTTGCACCCTTGGGACACATAGCC
 CGGACCGTTTTTGTACCTACCACGTGATGAACCCCTACGACGACGGCATCGGG
 CTGGACGAGTTCGTGACTGGCTCAACTCCCCAACTAGCGGGTCCGGTTGCAC
 GATCCAGCGGATCGCCGACTACGGCGAGTGGCTGCAGCGGTTTCGAGACTTCGC
 TCGTGCCTTGCCGGATCGCCAGCGCCACACCTCGCTGCTGCCCTTGCTGCAC
 AACTACCGAGAGCCTGCAAAGCCGATATGCGGGTCAATCGCGCCCACCGACCA
 GTTCCGCGCTGCCGTCCAAGAAGCGAAAATCGGTCCGGACAAAGACATTCCGC
 ACCTCACGGCGGCGATCATCGCGAAGTACATCAGCAACCTGCGACTGCTCGGG
 CTGCTGTGA

Secuencia de aminoácidos (SEQ ID NO: 86)

>uniprot|Q7TY99 | Q7TY99_MYCBO PROBABLE ÁCIDO GRASO-CoA LIGASA FADD9 (ÁCIDO GRASO-COA SINTETASA) (ÁCIDO GRASO-COA SINTASA)

MSINDQRLTRVEDLYASDAQFAAASPNEAITQAIIDQPGVALPQLIRMVM
 EGYADRPALGQRALRFVTDPSGRITMVELLPRFETITYRELWARAGTLAT
 ALSAEPAIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGL
 RPIVTEPETMIATSIDNLGDAVEVLGHAPARLVVFDYHGKVDTHREAV
 EAARARLAGSVTIDTLAELIERGRALPATPIADSADDALALLIYTSGSTG
 APKGAMYRESQVMSFWRKSSGWFEPSPGYSITLNFMPMSHVGGRQVLYGT
 LSNGGTAYYVAKSDLSTLFEDLALVRPTLFCFVPRIDWVFAEFHSEVDR
 RLVDGADRAALEAQVKAELRENVLGGRFVMALTSAPISAEMTAWVESLL
 ADVHLVEGYGSTEAGMVLNDGMVRRPAVIDYKLVDPPELGYFGTDQPYPR
 GELLVKTQTMFPGYYQRPDVTAEVFDPDGFYRTGDIMAKVGPDPQFVYLDL
 RNNVLKLSQGEFIAVSKLEAVFGDSPLVRQIFIYGN SARAYPLAVVPSG
 DALSRHGIENLKPVI SESLQEVARAAGLQSYEIPRDFI IETTPFTLENGL
 LTGIRKLARPQLKKFYGERLERLYTELADSQSNE LREL RQSGPDAPVLPT
 LCRAAAALLGSTAADVRPDAH FADLGGDSLSALS LANLLHEIFGVDVPVG
 VI VSPASDLRALADHIEAARTG VRRPSFASI HGRSATEVHASDLTLDKFI
 DAATLAAAPNLPAPSAQVRTVLLTGATGFLGRYLALEWLD RMDLVNGKLI
 CLVRARSDEEAQARLDATFDSGDPYLVRRHYRELGAGRLEVL AGDKGEADL

FIG 4 (Cont.)

GLDRVTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNAAGTAELLRLALT
 GKRKPYIYTSTIAVGEQIPPEAFTEADIRAI SPTRRIDDSYANGYANSK
 WAGEVLLREAHEQCGLPVTVFRCDMILADTSYTGQLNLPDMFTRLMLSLA
 ATGIAPGSFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPDRFVTYH
 VMNPYDDGIGLDEFVDWLNSTSGSGCTIQRIADYGEWLQRFETSLRALP
 DRQRHTSLLPLLHNYREPAKPICGSIAPTDQFRAAVQEAKIGPDKDIPHL
 TAAIIAKYISNLRLGLL

Q9CCT4

Secuencia de nucleótidos (SEQ ID NO:87)

>uniprot|Q9CCT4 | Q9CCT4_MYCLE Supuesta acil-CoA sintetasa

GTGTGGCGCCAACAGTCTATTTCTCATCGCAAGGAATCTGTTATGTGCGACTAT
 CACTAAGCAGGAAAAGCAGCTCGCACGCCGCGTTGACGACCTCACCGCCAACG
 ACCCGCAGTTCGCCGCCGCCAAACCCGACCCGGCGGTAGCCGCCGCCCTTGCC
 CAGCCCGGGCTTCGACTGCCCAAATCATCCAGACCCGCGCTGGACGGTTACGC
 GGAGCGGCCGGCACTGGGCCAGCGCGTCGCCGAGTTCACCAAAGACCCTAAGA
 CCGGACGCACCTCGATGGAGCTGCTCCCCAGCTTTGAGACCATCACCTACCGC
 CAGTTGGGCGACCGTGTTCGGAGCGCTGGCGCGCGCCTGGAGGCACGACCTACT
 GCACGCCGGCTACCGGGTCTGCGTGCTAGGTTTCAACAGTGTTCGATTACGCCA
 TCATCGACATGGCGCTCGGCGTGATTGGTGCTGTGGCGGTTCCACTGCAGACC
 AGTGCGGCGATCACCCAGCTGCAGTCGATCGTGACCGAGACCGAACCAGTAT
 GATCGCGACGAGCGTAAACCAGCTGCCCGATACTGTTCGAGCTGATCCTGTCTG
 GCCAGGCGCCAGCGAAGCTCGTTGTGTTTGACTACCACCCCGAGGTCGACGAG
 CAGCATGACGCAGTGGCAACCGCCCGGGCGCGGTTGGCGGACAGTAGCGTGGT
 GGTCGAGAGCCTGACCGAGGTCCTCGGTCGCGGCAAGACGCTGCCAGCTACGC
 CGATCCCCGTGGCCGATGACTCTGCTGACCCGTTGGCGTTGCTGATCTACACA
 TCTGGCAGCACCGGCGCACCCAAGGGCGCGATGTATCTGCAAAGCAATGTCGG
 CAAGATGTGGCGCCGGTCAGACGGAAACTGGTTCGGGCCAACCGCCGCGTCAA
 TCACTCTTAACTTCATGCCGATGAGCCACGTCATGGGCCGCGGAATCCTCTAC
 GGCACGCTCGGTAACGGCGGCACGGCTTACTTCGCCGCCCGCAGCGACCTCTC
 GACGCTGCTGGAGGATCTCAAGCTGGTGCGGCCGACCGAGTTGAACTTTGTAC
 CGCGCATCTGGGAAACCCTCTACGATGAATCCAAACGCGCAGTTGACCGTCGG
 TTAGCCAACAGCGGCTCCGCCGACCGTGCAGCCATCAAAGCCGAAGTTATGGA
 TGAACAGCGGCAATCCCTGCTGGGAGGACGGTACATCGCGGCTATGACGGGCT
 CGGCGCCAACCTCCCCGGAGTTGAAACACGGGGTCGAGTCCCTACTCGAAATG
 CATCTGTTGGAAGGCTACGGCTCCACCGAAGCCGGCATGGTCTTGTGTTGACGG
 CGAAGTGCAACGTCCGCCGGTTATCGATTACAAGCTGGTTCGACGTTCCGGATT

FIG 4 (Cont.)

TGGGCTACTTCAGCACCGACCAGCCTTATCCGAGAGGTGAATTGCTGCTCAAG
 ACCCAGAACATGTTCCCCGGCTACTACAAGCGTCCCTGAGGTTACCGCCACCGT
 GTTCGACAGCGACGGTTACTACCAGACCGGAGACATTGTCGCCGAAGTCGGTC
 CCGACCGGCTCGTGTACGTCGATCGCCGCAACAACGTGCTGAAACTCGCGCAG
 GGCCAGTTCGTCACCGTCGCGAAACTCGAGGCAGCGTTCAGCAATAGCCCCT
 GGTCCGGCAGATCTACATCTATGGCAACAGCGCACACCCCTACCTGTTGGCTG
 TTGTGGTGCCGACCGAGGATGCGTTGGCTACCAATGACATTGAGGTGCTCAA
 CCGCTGATTATCGATTCTTTACAGAAAGTAGCGAAAGAAGCCGACCTGCAGTC
 CTACGAGGTGCCGCGCGACTTAATCGTCGAGACTACACCGTTCAGCCTGGAGA
 ATGGCCTGCTCACCGGTATTTCGCAAGCTGGCGTGGCCGAAGCTCAAGCAGCAC
 TACGGCGCGCGACTCGAACAGCTCTACGCCGATCTGGTTGAAGGTCAGGCAA
 TGCCTGACGTCGCTAAAACAAAGCGTGGCGAACGCTCCGGTACTGCAGACGG
 TGAGCCGAGCCGTGGGCACCATTCTGGGAGTGGCGACCACCGATTTGCCGTCG
 AATGCGCACTTCACCGACTTAGGAGGAGACTCGTTGTCCGCGCTGACATTCGG
 TAGCCTGCTACGCGAACTCTTCGACATCGATGTGCCGGTGGGCGTCATTGTCA
 GCCCTGTCAACAACCTTGGTGGCGATCGCCGACTACATCGAGCGCGAGCGGCAG
 GGCACGAAGCGGCCCACTTTCAATTGCCATACACGGTCGTGACGCTGGCAAAGT
 GCATGCCAGTGACCTCACTCTAGACAAATTCATCGATGTATCAACGCTGACTG
 CCGCGCCCGTATTGGCGCAACCCGGCACCGAGGTGCGCACCGTCCCTGTTGACC
 GGCGCTACCGGCTTCCCTGGGGCGCTACTTGGCCCTGAAATGGCTCGAACGGAT
 GGACCTGGTCGAAGGGAAGGTAATCGCTCTGGTAAGAGCCAAGTCCAACGAGG
 ACGCTCGGGCCCGGCTCGACAAGACCTTCGATAGCGGAGACCCCAAACCTGCTG
 GCGCACTACCAGGAACTGGCAACCGACCACCTGGAGGTCATCGCCGGCGACAA
 AGGCGAAGTAGATCTGGAATTGGACCGGCAAACGTGGCGACGACTGGCCGACA
 CGGTCGATCTGATCGTCGACCCCGCCGCCCTGGTCAACCACGTGCTGCCGTAC
 AGCGAGCTATTTCGGCCCCAATACGTTAGGCACCGCCGAGCTGATTCGGATCGC
 GCTGACCAGTAAGCAAAAGCCGTACATCTATGTGTGACAAATCGGCGTCGGTA
 ATCAGATTGAGCCAGCAAAAATTCACCGAAGACTCCGACATCCGAGTCATTAGC
 CCGACGCGCAACATCAACAACAACCTATGCCAACGGCTACGGCAACAGCAAGTG
 GGCCGGCGAAGTGCTGCTGCGCGAAGCTCACGACCTATGCGGTCTGCCGGTCA
 CGGTCTTCCGCTGCGACATGATCTTGGCCGACACCAGCTATGCCGGTCAGCTC
 AACGTCCCCGACATGTTTACTCGAATGATGCTGAGTCTAGCCGCCACCGGCAT
 CGCACCCGGCTCGTTCTACGAGCTAGACGCCGAGAGCAATCGGCAACGCGCCC
 ACTACGACGGTCTGCCCGTCGAGTTCATCGCCGAAGCGATCTCCACCCTGGGA
 GACCAAAGCCTGCACGATCGAGACGGGTTACGACCTATCATGTAATGAACCC
 GCACGACGACGGCATCGGTATGGACGAGTTTGTGGACTGGTTAATTGATGCCG
 GCTGCCCTATAACAACGCATCAACGACTACGACGAATGGCTGCGACGGTTTGAG
 ATTTGCTGCGCGCCCTGCCCGAAAGGCAGCGTCACAGCTCACTGTTGCCGTT
 GTTGACAACCTACCAGAAGCCGGAGAAGCCATTGCACGGGTGCTGGCACCCA
 CAATCCGGTTCGTACGGCCGTTCAAACGCGAACATTGGTCAGGACAAAGAT

FIG 4 (Cont.)

ATTCCGCATATCTCGCCGGCAATCATCGCCAAATATGTCAGCGATCTGCAGCT
GCTCGGGCTGGTTTGA

Secuencia de aminoácidos (SEQ ID N0:88)

>uniprot|Q9CCT4 | Q9CCT4_MYCLE Supuesta acil-CoA sintetasa

VWRQQSISHRKESVMSTITKQEKQLARRVDDL TANDPQFAAAKPDPAVAA
ALAQPGLRLPQIIQTALDGYAERPALGQRVAEFTKDPKTGR TSMELLPSF
ETITYRQLGDRV GALARAWRHLLHAGYRVCV LGFNSVDYAIIDMALGVI
GAVAVPLQTSAAITQLQSI VTETEPSMIATS VNQLPDTVELILSGQAPAK
LVVFDYHPEVDEQHDAVATARARLADSSVVESL TEVLGRGKTL PATPIP
VADDSADPLALLIY TSGSTGAPKGAMY LQSNVGMWRRSDGNWFGPTAAS
ITLNFMPMSHVMGRGILYGT LGGGTAYFAARSDLSTLLEDLKLVRPT EL
NFVPRIWETLYDESKRAVDRRLANSGSADRAAIKAEVMDEQRQSL LGGRY
IAAMTGSAPTSP ELKHGVESLLEMHLLEGY GSTEAGMVLF DGEVQRPPVI
DYKLVDPDLGYFSTDQPYPRGELL LKTQNMFP GYYKRPEVTATVFDSDG
YYQTGDIVA EVGPDRLVYVDRRNNVLKLA QGQFVTVAKLEAAF SNSPLVR
QIYIYGNSAHPYLLAVVPTEDALATNDIEVLKPLIIDSLQKVAK EADLQ
SYEVPRDLIVETTPF SLENGLLTGIRKLAWPKLKQHYGARLEQLYADLVE
GQANALHVLKQSVANAPVLQTVSRAVG TILGVATTDLPSNAHFTDLGGDS
LSALTFGSLRELFDIDVPVGVIVSPVNNLVAIADYIERERQGTKRPTFI
AIHGRDAGKVHASDLTLDKFI DVSTLTAAPVLAQPGTEVRTVLLTGATGF
LGRYLALKWLERMDLVEGKVIALVRAKSNEDARARLDKTFD SGDPKLLAH
YQELATDHLEVIAGDKGEVDLELDRQTWRRLADTVDLIVDPAALVNHVLP
YSELFPGNTLGTAE LIRIALTSKQKPYIYVSTIGVGNQIEPAKFTEDSDI
RVISPTRNINNNYANGYGNSKWAGEVLLREAHDL CGLPVTVFRCDMILAD
TSYAGQLNVPDMFTRMMLSLAATGIAPGSFYELDAESNRQRAHYDGLPVE
FIAEAI STLGDQSLHDRDGFTTYHVMNPHDDGIGMDEFVDWLIDAGCPIQ
RINDYDEWLRRFEISLRALPERQRHSSLLPLLHNYQKPEKPLHGSLAPTI
RFRTAVQONANIGQDKDIPHISP AIIAKYVSDLQLLGLV

FIG 4 (Cont.)

Q54JKO**Secuencia de nucleótidos (SEQ 10 N0:89)**

>uniprot|Q54JK0 | Q54JKO_DICDI Supuesta proteína no caracterizada

ATGTTAAAACATATTTAAAAATTTTTTAACTAGAAAAGAAGAAAAGAAAAGAAA
 AGAAGTAGAGAAATTTAAAGATGGAGTATCAATAACTGAGGTTAAACAATCGA
 ACCTAGTAGTTTATTCATGCAATGGTTGTGGATCAGAGATATGGCCACCAAAA
 CAAGAGAGATATGCATGTAATGAATGCTCAAATTTTCGATTTTATGTAGTGAGTG
 TTATAGAAAAGAAATGATATTAATAAATGGTACACAAGAAGAGAAAGATAAAT
 TAGTTAGTGGGGAGAGTAATAATGGAATAAAGTACGAGCCAGTGAGACATTAT
 GATCCATCACCATTACCTCATCAATTAACATTAGAGAATGAGACTCAATTTCA
 ATTAGTTTATAGTTTACGTGGTAATTCAACATTTGAAACAATGGAGAAATCAT
 TTAAATACTTTAAAAACAGACCATGTCTTGGTATTAGAGAGAGATTAGGAGAG
 GATAATGTTTTATCAGAAAGATATAAATGGTTAACATATGGTGAAGTGTATGA
 GAAATCTTTAACCTTGGCAAAGGCATTAACATAATTTTCATCGAAAGAAGAGATT
 TCATTTCAATCTATATGGATAATTGCTTAGAATGGTATTTTCACAGATTTTGCA
 TCATTATGGGCCGGATTAATAGTGGTACCATTACATCATGCTTCAAATAGTTT
 TAATCTTTTAGAGATTCTTTGGAATTCCGAATCAAATGTATAGTTTGTCTG
 GTGAATCATTTAAAAATTTAATAGAACTTTATGATCAATTGACTGAGCAAGAT
 AAATTAGAGAAACCAATAGTGTGAAATTGATAGTTCATAAGGAGGATCTATT
 CGATCAGTCATTAGTCGATAGATTGCCAAGTGGCGTAGAATTTAAAACCTTCA
 ATGAGATGATTAAAATTGGGGAATCATTAAGTCAGGCTAAATATGAATTTGTC
 CCAGTTGGTCCAAATGATCTTTCCTCGGTGACTTATAACAAGTGGTAGTACTGG
 TGTACCAAAGGTGTAATGAAGTTAGATTCAATTTTCAATTTACTAATTGTCA
 ATTCCTATGTTCAATTCCCAAATGCAGTTTATAGTTATAATACCCTATCACAT
 AGTCAACGTTTAAGTGATTGGAGATATATTTATATGGGTGGTAGAGTAGCTAT
 CTATTCAGGTGATATGAATCTATTATTTGAAGATTTAGCCTTGGTTAGACCTC
 ATTCATTTTGGGCTGTACCAAGATTTTGGAAATTTATTATTCACCCAGTTCAAG
 AGTGATCTAAAGCAATACATGTTTGAATAATCCACAATTGGATGAAAGAAGTGC
 CACACTCTATTGCTATAAAGGTATTAGAAAGTTATTAGGTGATAGAATTAATA
 ATCTAGTGACTGGTGGTGCTCCGACTGCAAATGAAGTACTCAAATTTATGAGT
 GATTGTTGGAAAGATATAAACATTTCAAATTTCTTATGGTTTAACTGAAGTATC
 AGGTGTTTGTATAGATGGTTATATCTCTGACGAAGTAGAATTCAAAATTGAAC
 CAGTACCTTCTTTTGGATATTACCCAAGTATTACCACATCCTCGTGGTGAA
 TTGGTTGTAAAATCATCAACAATGTCTGCAGGTTATTATAAAAATACTCAATT
 AACTTCAGAATCATTTGAAGGTGGTTGGTTTAAACTGGTGATGTCGTAGAAT
 TAATTGGAGTTAGAAAAGTTAAAATCATTGATAGAATTAACATGCCTTCAA
 TTGGCAAATGGAGAATTCGTTACACCAGAACCATTGGAAAATAATTTCGTTTC

FIG 4 (Cont.)

ACTTTGTATTAATCAAATTTTTATTTATGGTAATTCACCTTAAAACATTTTTAG
 TTGCAATTGTTAAACCATCACAGATTGTTTAAAACAATTAGGACTTCAAGAT
 ATACCAATCGATCAATTAATTGAAAATCCAACTTTAAAATCAAACCTTTTATC
 AGAGATTAATAAAATTTCAAAGAAAAAAACTAGCAAATTATGAAATTCCAA
 AAATTATTACAATAGATTTCACTGAATGGACAATTGATAATAAATTAATCACT
 GGTTCTGGTAAATTTAATAGAGGTGAATTATATAAATTTTATAAAATTAATA
 TAATAATATGTTTGATATAATTGATAAAATTCACAAGGTTTAAAGAAATAATA
 ATAATAATAATAATAATGATAATATTAATAATAATGATAATAATAATAATAAT
 GAATCAAATAAAGATAATTTTGAAAATTATATAAAATCAATTTTAAATTTAGA
 CGGTAGAATTGAAGATAATAATTTAATTTAGAGAATTTATCATTATTCAA
 TTGGTGGTGATTCGTTAGGTGCTGTTAAATTATCATCACTTTTAAAAGAAAA
 GAAAATATTGATATTTACCTTCAACAATTTTAAATCAAATTTTAAATTTATC
 TTCATTATCGAAATTAATAAATGAAAAGAATCAAATCAATCAATTGTTGAAG
 ATTTTAAAGAAAATTTTAAAATCAATTGGAATGAAGAGATGATTTTAGATGAA
 GATATTAATAAATCAATTGACCAAATTAATAAATGCACAACCATCATCAACTCC
 CTCTTCATCAAATCAACACCATCACAAATCATCATCACCACCACCATCAT
 TAAATTCAAATAATATTGGTCAAATGCTTTTCATATGAAATCAATATTTATT
 ACAGGTGTTACAGGTTATTTAGGTACATTTTATTATTTAATTTATTAGAGGA
 TAAATCAATTGGTATTGAGAGAATTTATTGTTTAGTTAGAAATGTAAAGAATG
 AAGAAGAAGGTTTTAAATTAATTGAAAGAATATTTGAAAATCTTGTATCAAT
 GGTATGAATGAAAAGATTAGAGAAAAGGTAATTCAGTTTGTGGTGACTTATC
 AAAACCATTTTTCGGTGTTTCTACTGAAACCTTCAAATGTTATCTTTAGCGG
 TCGATATGGTAATTCACAATGGTGCCATTGTTAATATGGCCTATCCATATGCG
 AATATGAAATCAACAAATGTTACATCAACTCGTGATATCCTAAGATTATGCAC
 TACCGGAAGAGCCTCTTTTAAAAGTTGGTCTACGTTTCAACAGTTGGTGTAT
 TCTTTGAAATGGTGATGAAAAGATAGATGAATCAACAGCACCATCAACTTTC
 TTTTAGATCATGGTAATGGTTATAACCAAACAAAACCTAATATCAGATATACT
 TGTTAGAGAAGCGGCTTCATATGGTTTACCAACAATGATTTTCAGACCAGGTA
 CAATCTTTAGTCATAACCAATCTGGTTTCAACAATCAAATGATTCAATCGGT
 TTAATAATTAAGGTATCCTAGGTTTCGAGTTCTATCCAACCTAAAAAGATTA
 CTCTAGTGGCGATTTAAATCTTTCACCAGTTGATTGGGTATCATCTTCAATGG
 TTTCTTTAATTAAGCATCTTCCATTTTGGTGTAATAATACAAAATTTTATCAT
 ATGGTAAATGATAATCGTTTATCTTTAGATTTGCTATGTCAATATATAAATAA
 AGAAAAACAATTAGAAGAAATTAATTATTTGATTGGATCGATGCTCAACTTA
 ATTCTTCAAATAATCCATTGTATTCTATTAACATTTATTTAAAAGAATGAT
 CGTTTCCCAATTGGTTCTCAGTCAATTAATAAATCCAAAACCTATTAAGATTT
 AGAATCAATTGGTGAACCTCAATGTCAACCAATATCTGATTCCACAGTAATCA
 ATTATGTAAAATATTTAATCTCAAATAATTTAATTCAAACAATTAATAAATAA

FIG 4 (Cont.)

Secuencia de aminoácidos (SEQ ID N0:90)

>uniprot| Q54JK0 | Q54JK0_DICDI Supuesta proteína no caracterizada

MLKHIKNFLTRKEERKEKEVEKLDGVSITEVKQSNLVVYSCNGCGSEIWPPK
 QERYACNECSNFDLCSECYRKEMILINGTQEEKDKLVSGESNNGIKYEPVRHY
 DPSPLPHQLTLENETQFQLVYSLRGNSTFETMEKSFKYFKNRPCLGIRERLGE
 DNVLSERYKWLTYGEVYEKSLTLAKALTNFIERRDFISIYMDNCLEWYFTDFA
 SLWAGLIVVPLHHSNSFNLEILWNSESKCIVCSGESFKNLIELYDQLTEQD
 KLEKPIVLKLIHVHKEDLFDQSLVDRLPSGVEFKTFNEMIKIGESLSQAKYEFV
 PVGPNDLSSVITYTSGSTGVPKGVMLKDSIFNLLIVNSYVQFPNAVYSYNTLSH
 SQRLSDWRYIYMGGRVAIYSGDMNLLFEDLALVRPHSFWAVPRFWNLLFTQFK
 SDLKQYMFENPQLDERTATLYCYKGIRKLLGDRINNLVTGGAPTANEVLKFS
 DCWKDINISNSYGLTEVSGVCIDGYISDEVEFKIEPVPSFGYYPTDLPHPRGE
 LVVKSSTMSAGYYKNTQLTSESFEGGWFKTGDVVELIGVRKVKIIDRIKHAFK
 LANGEFVTPEPLENNFVSLCINQIFIYGNLSLKTFLVAIVKPSQDCLKQLGLQD
 IPIDQLIENPTLTKSKLLSEINKISKEKLANYEIPKIITIDFTEWTIDNKLIT
 GSGKFNRGELYKFYKIKINMFDIIDKIQQGLRNNNNNNNDNINNNNDNNNNN
 ESNKDNFENYIKSILNLDGRIEDNNFNLENLSFIQIGGDSLGAVKLSSLLKEK
 ENIDISPSTILNQNFNLSLSSKLINEKESNQSIVEDFKENFKINWNEEMILDE
 DIKKSIDQIKNAQPSSTPSSSKSTPSQSSSSPPPSLNSNIGQNAFHMKSIFI
 TGVTGYLGTFLFNLLEDKSIGIERIYCLVRNVKNEEEGFKLIERIFEKSCIN
 GMNEKIREKVIPVCGDLSKPFVGVSTETFKMLSLAVDMVIHNGAIVNMAYPYA
 NMKSTNVTSTRDILRLCTTGRASFKKLVYVSTVGVVFFGNGDEKIDESTAPSTF
 FLDHGNNGYNQTKLISDILVREAASYGLPTMIFRPGTIFSHSQSGFNNQNDSIG
 LIIKGI LGSSSYPTKKDYSSGDLNLSVDWVSSSMVSLIKHLPFWCNNTKIYH
 MVNDNRLSLDLLCQYINKEKQLEEINYFDWIDAQLNSSNNPLYSIKHLFKKND
 RFPIGSQSIKNPKTIKDLÉSIGELQCQPI SDSTVINYVKYLISNNLIQTINK

Q2MFQ3

Secuencia de nucleótidos (SEQ ID N0:91)

>uniprot| Q2MFQ3 | Q2MFQ3_STRRY Supuesta péptido sintetasa no ribosomal

ATGACCGACACGTACGTTTCTTCACGCCCGTTATCCAAGCGGCCCCAGGTCCC
 GGGTGCCCGCACGCCGGCGCCCGGCTACCCACGGGACAGCCGCATCCCGGAGC
 TGTTTCGAGGCACAGGCCGCGGCGCTGCCGCAGGCCCGGCGGCCCGGCACGGG
 GACCGCACCTGACCTACGGCCAACCTCGACGCCACCGCGACGCGTTGGCGGA

FIG 4 (Cont.)

CCGGCTGGCTGCCGGCGGGGTCCGGCCGGGTGACCTGATCGGCGTGTGCGGCA
 GCCGTTCCCTGGAGGCGCTGGTGGCGCTGCTGGGCATCCTCAAGGCCGGCTGC
 GCGTACGTACCGCTCGACGAGGAACTGCCGCCGGCCGGCTGCGGGCCATGGC
 CGAGGACGCGGGCATCAGCGCCGCGGTACCCTGCCGGGCAGCACGCGCCGGG
 TGCGGGGTCTACGCGTGTGAGTCGAGGTCGGCTCCCTCGGCCGGCCCGCCCC
 GAGCGCGGAGCGGCCCGCCCCGACCGGGCCACCGGCTCCGCCGCCGACTG
 CGCCTACGTGCCTTCACTTCCGGCACGACCGGCCGGCCCAAGCCCGTAGCGC
 TGTCCCACCGGGGCGTGGTCCGCCTCGTGCTGTCCGACCCCGGCCTCACGCCA
 CCCGGACCGGGCGACGGAGTGCTGCACGCCTACAGCCTGTCCCTCCGACGCCTC
 GACCATCGAGATCTGGGGCGCGCTGCTGACCGGCGCCTGCCTGGTCGTGCGCG
 ACCGCGAGGAACTGCTCTCGCCACCGCCCTGGAACGGCTGCTCCGCGCGGGC
 GCGTACCGTGGCGTACCTGACGACGAGCGTCTTCCACCTCGTCGCCCGGAC
 CCGCCCCGAGGCGCTGGCCGGCCTGCGGTTTCGTCTCCGCGGGGCGGGGAGGCGA
 TGGACCCGCGCCTGGCGAACGCCGTCCTCGCGGCCTGCCCCCGCACCCACGGTG
 GTCAACTTCTACGGCCCGACCGAGAACGCCGTGGTCTCCACCGCCCATGTGCT
 CACCCCCCTCCCCGAGGACGCCGCACACGTCCCCCTGGGACGCCCTTCGGCG
 CTTCCACCTGCCACGTCTGCGGGCCGACGGCTCGCCCGCGCGGCCGGGCGAG
 GAAGGGGAGCTGTACGTGCGCGGGGACGGGCTGGCGCTGGGCTACCTCGGCGA
 CCCGCAGCTGACCGCCGAGCGGTTTCGTGACGCTGCCCGCGGTCGAGCCGGACG
 GACCGCTGTACCGGACCGGGCGACCGGGCCGTACGGCACGCCGACGGGCTGCTG
 GAGTACCGCGGACGGCTCGACCGCCAGGTCAAGCTGCGCGGGCGCCCGCATCGA
 GCTGGACGAGGTGGAGACCCGCTTACGGGCCACCCCGAGGTCGGCGAAGCGG
 CCGTCGAGGTCGACGGGCACTCCCTGACCGCCTACGTACGGCCACCGTCCCC
 GGCCGCCCGCTGCCGCTGGCCGACCTGCGCGCGTACTGCGCCAAGTGGCTGCC
 CCCGCAGGCCGTCCCGGCCCTGATACCCCTGGACCGCTTCCCGGTACCAGCG
 GCGGCAAGATCGACCGCAGCCGTCTGAAGCCGACCGCGCCACCGCCCGGCCCC
 GAAGACACCGCGGAGGCCGCGCGGCCCGGACGAGCCGGAGGCCACCGACGG
 CCTGTCCGGTCTCCTCTCGGAAGTGTGGCACCCAGGTGCTGCGTGTCCGGCCCA
 CGCCCCGGGACGACTTCTTCTCCTCGGGCGGCGACTCCCTGCTCGCCTCGGAG
 ACCGTCACCCGCACCCTCGCCGTACTCGGCCTCGACGCGGCCCTGGGCTCCAC
 CCTCATCAGGGCGCTGCTGGCCGCGCCACCCTCGAAAGCTTACCAGCCGCCG
 TACGCGGAGTCCGCGGCGGCACCGGCGGACCGGCCGGCGGCCAGGAACCGGCC
 GTCGACTTCGCCGCGGAGACCGGACTCGGCTTCGCCCTCCCGCCCGCCGAAGG
 CCCGGCGCCGAACCCGCACGACCCCGAGGACGTCCTGCTCACGGGCGCTTCCG
 GCTTCGTGCGCGGATTCTGCTGCACCGTCTGCTGCACGCGACGGCCGCCCGC
 GTCCACTGCCCCGTACGGGCGACGAGCCCCGCCACGCCCGGCAGCGGGTCCG
 CACCGCCCTCACCCGCTACGGGCTGCACCTCGACGAGGCCGACTGGCAGCGCG
 TGGAGTGCTTCCCGGGGACCTGACCCAGCCGCGCCTGGGGCTCGACCACGAG
 CGCGCCGACGCACTGGCCAGCGCCTGGACCTGATCGTGACAACGGTGCCCG
 GGTCAACTTCTCTACCCCTACCAGCAGTTGCGCCCGGCGAACGTCGACGGAA
 CCCGCGAGGTCGTCCGGATCGCCGCGCGCCGCCGGGTGCCGGTGCACCTTCGTG

FIG 4 (Cont.)

TCCACCGTCGCAGTCGTCGCGGGCTTCGGCACCGCCGGGGTGC GCGAGGTGGA
 CGAAGATCTGCCGCCGGCCACGCCGACGGGCTGACCATGGGGTACGCGGAGA
 GCAAGTGGGTGCGCCGAAGGGGTGCTGCGGCAGGCGGCCGCGCAGGGCCTGCCG
 GTGGCCGTGTACCGGCCGTACGAGGTCACGGGCGACCGGACGCACGGCGCGTG
 CAACACCGAGACGGCCATCTGCTCGCTGTTCAAGATGATCGCCGACACGGGAG
 TGGCCCCGACATCAAGCTGCCGATGGACTTCGTACCCGTCGACCACCTCGCC
 GAGTCCCTGGTGCACATCGCCACGCACCGGCCGGCCGACGGCCGGGTCTACCA
 CCTGACCAACCCGCGCCCGGCGATGCTGTCGGACGTCCTCGACCGGATGCGCG
 CGGCGGGCTTACCCTGCGCACCTGCCGTACGACGCGTGGGTGCGCGAGCTC
 GTCCGGCACGTCGCCGAGAACCCGACGAGCGCCACGGCTCCGTTGCTGCCCT
 GTGCGTGGACCGCAGCCGCACCGCCGACATGTCCGTCAAGGAGATGTACCTCA
 AGGGCACCTTCCCGTCTGGGGCGGGCGCAACGCCGAGGAGGCGCTGGCCGGC
 AGCGGGCTGCACTGCCC GCCGGT C GACTCCGCTCTGCTGGACCGCTACCTGGA
 G TACTTCTT CACCTCCGGCTACCTCACGCGCCCGGGCGGCCCGGGCCCCGGCCCCG
 AATCCGAGGCCGAGCGAATACCGGAGGACGAGCCGGTGTCCGGGACCGAACC
 ATATCCGGGACCGAACC GATTTCTGCCGCCGAGCCGATATCCGGGACCGAACC
 GATTTCTGCCGCCGGGCCGATATCCGGGACCGAGCCGATACCCGCCGCCGAGC
 CGATATCCGGGACCGCAGCCGCAGCCCGCACGGAGCGCAGCCGATGA

Secuencia de aminoácidos (SEQ ID N0:92)

>uniprot| Q2MFQ31Q2MFQ3_STRRY Supuesta péptido sintetasa no ribosomal

MTDTYVSSRPLSKRPQVPGARTPAPGYPRDSRIPELFEAQAAALPQAPAARHG
 DRTLTYGQLDAHADALADRLAAGGVRPGDLIGVCGSRSLEALVALLGILKAGC
 AYVPLDEELPPARLRAMAEDAGISA AVTLPGSTRVRGLRVSVEVGS LGRPAP
 ERASGPAPDRATGSAADCAYVAFTSGTTGRPKPVALSHRGVVRVLSDPGLTP
 PPGDGV LHAYSLSSDASTIEIWGALLTGACLVVADREELLSPTALERLLRAG
 GVTVAYLTTSVFHLVARTRPEALAGLRFVSAGGEAMPRLANAVLAACPRTTV
 VNFYGP TENAVVSTAHVLTPLPEDAAHVPLGRPFGASTCHVLRADGSPARPGE
 EGELYVGGDGLALGYLGDPQLTAERFVTLPAVEPDGPLYRTGDRAVRHADGLL
 EYRGRLDRQVKLRGARIELDEVETRLRAHPEVGEAAVEVDGHSLTAYVTATVP
 GRPLPLADLRAYCAKWLPPQAVPALIPLDRFPVTS GGKIDRSRLKPTAPPPGP
 EDTAEAARRPDEPEATDGLSGLLSEVWHQVLRVRPTPRDDFFLLGGDSLLASE
 TVTRTLAVLGLDAALGSTLIRALLAAPTLESFTA AVRGVRGGTGGPAGGQEP
 VDFAAETGLGFALPPAEGPAPNPHDPEDVLLTGASGFVGGFLLHRL LHATAAR
 VHC PVRATSPA HARQVR TALTRYGLHLDEADWQRVECFPGDLTQPR LGLDHE
 RADALAQRLDLIVHNGARVNF LYPYQQLR PANVDGTREVVRI AARRRVPVHFV
 STVAVVAGFGTAGVREVEDLPPAHADGLTMGYAESKWAEGVLRQAAAQGLP
 VAVYRPYEVTGDRTHGACNTETAICSLFKMIADTGVAPDIKLPMD FVPVDHLA

FIG 4 (Cont.)

ESLVHIATHRPADGRVYHLTNPRPAMLSVLDLDRMRAAGFTLRTLPHYDAWVDEL
 VRHVAENPTSATAPFVSLCVDRSRTADMSVKEMYLKGTFFVLGRRNAEEALAG
 SGLHCPPVDSALLDRYLEYFFFTSGYLTRPAAGPGPESEAERIPEDPVSSTEP
 ISGTEPISAAEPISGTEPISAAGPISGTEPIPAAEPISGTAAAARTERSR

ZP_03429464

Secuencia de nucleótidos (SEQ ID N0:113)

>gi | 189214978:54695 - 58201 Mycobacterium tuberculosis EAS054 NZ_ABOV0100087, secuenciación genómica completa al azar

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTTCGAGGACCTATACGCCAG
 CGACGCCAGTTTCGCCGCCGCGAGTCCCAACGAGGCGATCACCCAGGCGATCG
 ACCAGCCCGGGGTCGCGCTTCCACAGCTCATCCGTATGGTCATGGAGGGCTAC
 GCCGATCGGCCGGCACTCGGCCAGCGTGCCTCCGCTTCGTCACCGACCCCGA
 CAGCGGCCGCACCATGGTCGAGCTACTGCCGCGGTTTCGAGACCATCACCTACC
 GCGAACTGTGGGCCCCGCGCCGGCACATTGGCCACCGCGTTGAGCGCTGAGCCC
 GCGATCCGGCCGGCGACCGGGTTTTCGCTGCTGGGCTTCAACAGCGTCGACTA
 CACAACCATCGACATCGCGCTGATCCGGTTGGGCGCCGTGTCGGTTCCACTGC
 AGACCAGTGCGCCGGTACCGGGTTGCGCCCGATCGTCACCGAGACCGAGCCG
 ACGATGATCGCCACCAGCATCGACAATCTTGGCGACGCCGTGGAAGTGCTGGC
 CGGTCACGCCCCGGCCGGCTGGTTCGATTTCGATTACCACGGCAAGGTTGACA
 CCCACCGCGAGGCCGTCGAAGCCGCCGAGCTCGGTTGGCCGGCTCGGTTGACC
 ATCGACACACTTGCCGAACTGATCGAACGCGGCAGGGCGCTGCCGGCCACACC
 CATTGCCGACAGCGCCGACGACGCGCTGGCGCTGCTGACTTACACCTCGGGTA
 GTACCGGCGCACCCAAAGGCGCCATGTATCGCGAGAGCCAGGTGATGAGCTTC
 TGGCGCAAGTCGAGTGGCTGGTTCGAGCCGAGCGGTTACCCCTCGATCACGCT
 GAACTTCATGCCGATGAGCCACGTCGGGGGCGTCAGGTGCTCTACGGGACGC
 TTTCCAACGGCGGTACCGCCTACTTCGTCGCCAAGAGCGACCTGTCGACGCTG
 TTCGAGGACCTCGCCCTGGTGC GGCCACAGAATTGTGCTTCGTGCCGCGCAT
 CTGGGACATGGTGTTCGCAGAGTTCCACAGCGAGGTCGACCGCCGCTTGGTGG
 ACGGCGCCGATCGAGCGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAG
 AACGTGCTCGGCGGACGGTTTGTTCATGGCGCTGACCGGTTCCGCGCCGATCTC
 CGCTGAGATGACGGCGTGGTTCGAGTCCCTGCTGGCCGACGTGCATTTGGTGG
 AGGGTTACGGCTCCACCGAGGCGGGATGGTTCCTGAACGACGGCATGGTGC GG
 CGCCCCGCGGTGATCGACTACAAGCTGGTTCGACGTGCCCGAGCTGGGCTACTT
 CGGCACCGATCAGCCCTACCCCGGGGCGAGCTGCTGGTCAAGACGCAAACCA
 TGTTCCCCGGCTACTACCAGCGCCCGGATGTCACCGCCGAGGTGTTTCGACCCC
 GACGGCTTCTACCGGACCGGGGACATCATGGCCAAAGTAGGCCCGACCGATT
 CGTCTACCTCGACCGCCGCAACAACGTGCTAAAGCTCTCCAGGGCGAGTTCA

FIG 4 (Cont.)

TCGCCGTGTCTGAAGCTCGAGGCGGTGTTTCGGCGACAGCCCGCTGGTCCGACAG
 ATCTTCATCTACGGCAACAGTGCCCGGGCCTACCCGCTGGCGGTGGTTGTCCC
 GTCCGGGGACGCGCTTTCTCGCCATGGCATCGAGAATCTCAAGCCCGTGATCA
 GCGAGTCCCTGCAGGAGGTAGCGAGGGCGGCCGGCCTGCAATCCTACGAGATT
 CCACGCGACTTCATCATCGAAACCACGCCGTTACCCCTGGAGAACGGCCTGCT
 CACCGGCATCCGCAAGCTGGCACGCCCGCAGTTGAAGAAGTTCTATGGCGAAC
 GTCTCGAGCGGCTCTATACCGAGCTGGCCGATAGCCAATCCAACGAGCTGCGC
 GAGCTGCGGCAAAGCGGTCCCGATGCGCCGGTGCTTCCGACGCTGTGCCGTGC
 CGCGGCTGCGTTGCTGGGCTCTACCGCTGCGGATGTGCGGCCGGACGCGCACT
 TCGCCGACCTGGGTGGTGA CTGCTCTCGGCGCTGTCGTTGGCCAACCTGCTG
 CACGAGATCTTCGGCGTCGACGTGCCGGTGGGTGTCATTGTCAGCCCGGCAAG
 CGACCTGCGGGCCCTGGCCGACCACATCGAAGCAGCGCGCACCGGGCGTCAGGC
 GACCCAGCTTCGCCTCGATACACGGTCGCTCCGCGACGGAAGTGCACGCCAGC
 GACCTCACGCTGGACAAGTTCATCGACGCTGCCACCCTGGCCGCAGCCCCGAA
 CCTGCCGGCACCGAGCGCCCAAGTGCGCACCGTACTGCTGACCGGGCGCCACCG
 GCTTTTTGGGTCGCTACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTC
 AACGGCAAGCTGATCTGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAGC
 CCGGCTGGACGCGACGTTTCGATAGCGGCGACCCGTATTTGGTGCGGCACTACC
 GCGAATTGGGCGCCGGCCGCCTCGAGGTGCTCGCCGGCGACAAGGGCGAGGCC
 GACCTGGGCCTGGACCGGGTCACCTGGCAGCGGCTAGCCGACACGGTGGACCT
 GATCGTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGT
 TCGGCCCAAACGCGGGCGGGCACCGCCGAGTTGCTTCGGCTGGCGCTGACCGGC
 AAGCGCAAGCCATACATCTACACCTCGACGATCGCCGTGGGCGAGCAGATCCC
 GCCGGAGGCGTTCACCGAGGACGCCGACATCCGGGCCATCAGCCCGACCCGCA
 GGATCGACGACAGCTACGCCAACGGCTACGCGAACAGCAAGTGGGCGGGCGAG
 GTGCTGCTGCGCGAAGCTCACGAGCAGTGCGGCCTGCCGGTGACGGTCTTCCG
 CTGCGACATGATCCTGGCCGACACCAGCTATAACCGGTCAGCTCAACCTGCCGG
 ACATGTTACCCGGCTGATGCTGAGCCTGGCCGCTACCGGCATCGCACCCGGT
 TCGTTCTATGAGCTGGATGCGCACGGCAATCGGCAACGCGCCCACTATGACGG
 CTTGCCGGTCTGAATTCGTGCGAGAAGCCATTTGCACCCTTGGGACACATAGCC
 CGGACCGTTTTGTACCTACCACGTGATGAACCCCTACGACGACGGCATCGGG
 CTGGACGAGTTCGTGACTGGCTCAACTCCCCAACTAGCGGGTCCGGTTGCAC
 GATCCAGCGGATCGCCGACTACGGCGAGTGGCTGCAGCGGTTTCGAGACTTCGC
 TCGGTGCCTTGCCGGATCGCCAGCGCCACGCCTCGCTGCTGCCCTTGCTGCAC
 AACTACCGAGAGCCTGCAAAGCCGATATGCGGGTCAATCGCGCCCACCGACCA
 GTTCCGCGCTGCCGTCCAAGAAGCGAAAATCGGTCCGGACAAAGACATTCCGC
 ACCTCACGGCGGCGATCATCGCGAAGTACATCAGCAACCTGCGACTGCTCGGG
 CTGCTGTGA

FIG 4 (Cont.)

Secuencia de aminoácidos (SEQ ID N0:114)

>gi | 215431545 | ref | ZP_03429464.1 | Ácido graso-CoA ligasa fadD9 [Mycobacterium tuberculosis EAS054]

MSINDQRLTRRVEDLYASDAQFAAASPNEAITQAI DQPGVALPQLIRMVMEGY
 ADRPALGQRALRFVTDPSGRMTVELLPRFETITYRELWARAGTLATALSAEP
 AIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGLRPIVTETEP
 TMIATSIDNLGDAVEVLGHAHAPARLVVFDYHGKVDTHREAVEAARARLAGSVT
 IDTLAELIERGRALPATPIADSADDALALLTYTSGSTGAPKGYRESQVMSEF
 WRKSSGWFEPSPGYSITLNFMPMSHVGGQVLYGTL SNGGTAYFVAKSDLSTL
 FEDLALVRPTLFCFVPRIWDMVFAEFHSEVDRRLVDGADRAALEAQVKAELRE
 NVLGGRFVMALTSAPISAEMTAWVESLLADVHLVEGYGSTEAGMVLNDGMVR
 RPAVIDYKLVDPVPELGYFGTDQPYPRGELLVKTQTMFPGYYQRPDVTAEVDFP
 DGFYRTGDIMAKVGPDQFVYLD RRNNVLKLSQGEFIAVSKLEAVFGDSPLVRQ
 IFIYGNSARAYPLAVVPSGDALSRHG IENLKPVI SE SLQEVARAAGLQSYEI
 PRDFIIEETPF TLENGLLTGIRKLARPQLK KFYGERLERLYTELADSQSNELR
 ELRQSGPDAPVLP TLCRAAAALLGSTAADV RPD AHFADLGGDSLSALS LANLL
 HEIFGVDVPVGVIVSPASDLRALADH IEAARTG VRRPSFASI HGRSATEVHAS
 DLTLDKFIDAATLAAAPNLPAPSAQV RTVLLTGATGFLGRYLALEWLD RMDLV
 NGKLICLVRRARSDEEAQARLDATFDSGDPYLVRHYRELGAGRLEVLAGDKGEA
 DLGLDRVTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNAAGTAE LLRLALTG
 KRKPYIYTSTIAVGEQIPPEAFTE DADIRAI SPTRRIDDSYANGYANSKWAGE
 VLLREAHEQCGLPVTVFRCDMILADTSYTGQLNLPDMFTRLMLSLAATGIAPG
 SFYELDAHGNRQRAHYDGLPVEFVAEAICTLGTHSPDRFV TYHVMNPYDDGIG
 LDEFVDWLN SPTS GSGCTIQRIADYGEWLQRFETSLRALPDRQRHASLLPLLH
 NYREPAKPICGSIAPTDQFRAAVQEAKIGPDKDIPHLTAAII AKYISNLRLLG
 LL

ZP_03533123

Secuencia de nucleótidos (SEQ ID N0:115)

>gi | 192384451:15291-18572 Mycobacterium tuberculosis GM 1503 NZ_ABQG01000169, secuenciación genómica completa al azar

ATGGTCGAGCTACTGCCGCGGTTTCGAGACCATCACCTACCGCGAACTGTGGGC
 CCGCGCCGGCACATTGGCCACCGCGTTGAGCGCTGAGCCC GCGATCCGGCCGG
 GCGACCGGGTTTGCGTGCTGGGCTTCAACAGCGTCGACTACACAACCATCGAC
 ATCGCGCTGATCCGGTTGGGCGCCGTGTCGGTTCCTACTGCAGACCAGTGCGCC

FIG 4 (Cont.)

GGTCACCGGGTTGCGCCCGATCGTACCCGAGACCGAGCCGACGATGATCGCCA
CCAGCATCGACAATCTTGGCGACGCCGTCGAAGTGCTGGCCGGTCACGCCCCG
GCCCGGCTGGTCGTATTTCGATTACCACGGCAAGGTTGACACCCACCGCGAGGC
CGTCGAAGCCGCCCGAGCTCGGTTGGCCGGCTCGGTGACCATCGACACACTTG
CCGAAGTGAATCGAACGCGGCAGGGCGCTGCCGGCCACACCCATTGCCGACAGC
GCCGACGACGCGCTGGCGCTGCTGATTTACACCTCGGGTAGTACCGGCGCACC
CAAAGGCGCCATGTATCGCGAGAGCCAGGTGATGAGCTTCTGGCGCAAGTCGA
GTGGCTGGTTCGAGCCGAGCGGTTACCCCTCGATCACGCTGAACTTCATGCCG
ATGAGCCACGTCGGGGGCCGTCAGGTGCTCTACGGGACGCTTTC AACGGCGG
TACCGCCTACTTCGTCGCCAAGAGCGACCTGTCGACGCTGTTTCGAGGACCTCG
CCCTGGTGC GGCCACAGAATTGTGCTTCGTGCCGCGCATCTGGGACATGGTG
TTCGCAGAGTTCACAGCGAGGTCGACCCGCGCTTGGTGGACGGCGCCGATCG
AGCGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCCGGGAGAACGTGCTCGGCG
GACGGTTTGTTCATGGCGCTGACCGGTTCCGCGCCGATCTCCGCTGAGATGACG
GCGTGGGTCGAGTCCCTGCTGGCCGACGTGCATTTGGTGGAGGGTTACGGCTC
CACCGAGGCCGGGATGGTCCTGAACGACGGCATGGTGC GGCGCCCCGCGGTGA
TCGACTACAAGCTGGTCGACGTGCCCGAGCTGGGCTACTTCGGCACCGATCAG
CCCTACCCCCGGGGCGAGCTGCTGGTCAAGACGCAAACCATGTTCCCCGGCTA
CTACCAGCGCCCGGATGTCACCCGCCGAGGTGTTTCGACCCCGACGGCTTCTACC
GGACCGGGGACATCATGGCCAAAGTAGGCCCCGACCAGTTCGTCTACCTCGAC
CGCCGCAACAACGTGCTAAAGCTCTCCAGGGCGAGTTCATCGCCGTGTCGAA
GCTCGAGGGCGGTGTTTCGGCGACAGCCCGCTGGTCCGACAGATCTTCATCTACG
GCAACAGTGCCCGGGCCTACCCGCTGGCGGTGGTTGTCCCGTCCGGGGACGCG
CTTTCTCGCCATGGCATCGAGAATCTCAAGCCCGTGATCAGCGAGTCCCTGCA
GGAGGTAGCGAGGGCGGCCGGCCTGCAATCCTACGAGATTCCACGCGACTTCA
TCATCGAAACCACGCCGTTACCCCTGGAGAACGGCCTGCTCACCGGCATCCGC
AAGCTGGCACGCCCGCAGTTGAAGAAGTTCATATGGCGAACGTCTCGAGCGGCT
CTATACCGAGCTGGCCGATAGCCAATCCAACGAGCTGCGCGAGCTGCGGGCAA
GCGGTCCCGATGCGCCGGTGCTTCCGACGCTGTGCCGTGCCGCGGCTGCGTTG
CTGGGCTCTACCGCTGCGGATGTGCGGCCGGACGCGCACTTCGCCGACCTGGG
TGGTGAATCGCTCTCGGCGCTGTCGTTGGCCAACCTGCTGCACGAGATCTTCG
GCGTCGACGTGCCGGTGGGTGTCATTTGTCAGCCCGGCAAGCGACCTGCGGGCC
CTGGCCGACCACATCGAAGCAGCGCGCACCCGGCGTCAGGCGACCCAGCTTCGC
CTCGATACACGGTCGCTCCGCGACGGAAGTGCACGCCAGCGACCTCACGCTGG
ACAAGTTCATCGACGCTGCCACCCTGGCCGACGCCCCGAACCTGCCGGCACCG
AGCGCCCAAGTGCGCACCGTACTGCTGACCGGGCGCCACCGGCTTTTTGGGTGCG
CTACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTCAACGGCAAGCTGA
TCTGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAAGCCCGGCTGGACGCG
ACGTTTCGATAGCGGCGACCCGTATTTGGTGCGGCACTACCGCGAATTGGGCGC
CGGCCGCTCGAGGTGCTCGCCGGCGACAAGGGCGAGGCCGACCTGGGCCTGG
ACCGGGTCACCTGGCAGCGGCTAGCCGACACGGTGGACCTGATCGTGGACCCC

FIG 4 (Cont.)

GCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGTTTCGGCCCAAACGC
 GCGGGCACC GCCGAGTTGCTTCGGCTGGCGCTGACCGGCAAGCGCAAGCCAT
 ACATCTACACCTCGACGATCGCCGTGGGCGAGCAGATCCCGCCGGAGGCGTTC
 ACCGAGGACGCCGACATCCGGGCCATCAGCCCGACCCGCAGGATCGACGACAG
 CTACGCCAACGGCTACGCGAACAGCAAGTGGGCCGGCGAGGTGCTGCTGCGCG
 AAGCTCACGAGCAGTGCGGCCCTGCCGGTGACGGTCTTCCGCTGCGACATGATC
 CTGGCCGACACCAGCTATAACGGTCAGCTCAACCTGCCGGACATGTTACCCCG
 GCTGATGCTGAGCCTGGCCGCTACCGGCATCGCACCCGGTTCGTTCTATGAGC
 TGGATGCGCACGGCAATCGGCAACGCGCCCACTATGACGGCTTGCCGGTCGAA
 TTCGTCGCAGAAGCCATTTGCACCCTTGGGACACATAGCCCGGACCGTTTTGT
 CACCTACCACGTGATGAACCCCTACGACGACGGCATCGGGCTGGACGAGTTCG
 TCGACTGGCTCAACTCCCCAACTAGCGGGTCCGGTTGCACGATCCAGCGGATC
 GCCGACTACGGCGAGTGGCTGCAGCGGTTTCGAGACTTCGCTGCGTGCCTTGCC
 GGATCGCCAGCGCCACGCCTCGCTGCTGCCCTTGCTGCACAACTACCGAGAGC
 CTGCAAAGCCGATATGCGGGTCAATCGCGCCCACCGACCAGTTCGCGCTGCC
 GTCCAAGAAGCGAAAATCGGTCCGGACAAAGACATTCCGCACCTCACGGCGGC
 GATCATCGCGAAGTACATCAGCAACCTGCGACTGCTCGGGCTGCTGTGA

Secuencia de aminoácidos (SEQ ID NO:116)

>gi |218754327|ref | ZP_03533123.1 | Ácido graso-CoA ligasa fadD9 [Mycobacterium tuberculosis GM 1503]

MVELLPRFETITYRELWARAGTLATALSAEPAIRPGDRVCVLGFNSVDYTTID
 IALIRLGAVSVPLQTSAPVTGLRPIVTEPETMIATSIDNLGDAVEVLGHAP
 ARLVVF DYHGKVDTHREAVEAARARLAGSVTIDTLAELIERGRALPATPIADS
 ADDALALLIYTSGSTGAPKGAMYRESQVMSFWRKSSGWFEPSPYSITLNFMP
 MSHVGGRQVLYGTLNNGGTAYFVAKSDLSTLFEDLALVRPTELCFVPRIWDMV
 FAEFHSEVDRRLVDGADRAALEAQVKAELRENVLGGRFVMALTSAPI SAEMT
 AWVESLLADVHLVEGYGSTEAGMVLNDGMVRRPAVIDYKLVDVPELGYFGTDQ
 PYPRGELLVKTQTMFPGYYQRPDVTAEVFDPDGFYRTGDIMAKVGPDQFVYLD
 RRNNVLKLSQGEFIAVSKLEAVFGDSPLVRQIFIYGN SARAYPLAVVPSGDA
 LSRHGIENLKPVI SESLQEVARAAGLQSYEIPRDFI IETTPFTLENGLLTGIR
 KLARPQLKKFYGERLERLYTELADSQSNELREL RQSGPDAPVLPTLCRAAAAL
 LGSTAADV RPD AHFADLGGDSL SALS LANLLHEIFGVDVPVGVIVSPASDLRA
 LADHIEAARTGVR RPSFASI HGRSATEVHASDLTLDKF IDAATLAAAPNLPAP
 SAQVRTVLLTGATGFLGRYLAL EWLD RMDLVNGKLI CLVRARSDEEAQARLDA
 TFDSDPYLV RHYRELGAGRLEVL AGDKGEADLGLDRVTWQRLADTVDLIVDP
 AALVNHVLPYSQLFGPNAAGTAE LLRLALTGKRKPYIYTSTIAVGEQIPPEAF
 TEDADIRAI SPTRRIDDSYANGYANSKWAGEVLLREAH EQGLPVTVFRCDMI

FIG 4 (Cont.)

LADTSYTGQLNLPDMFTRLMLSLAATGIAPGSFYELDAHGNRQRAHYDGLPVE
FVAEAICTLGTHSPDRFVITYHVMNPYDDGIGLDEFVDWLNSPTSGSGCTIQRI
ADYGEWLQRFETSLRALPDRQRHASLLPLLHNYREPAKPICGSIAPTDQFRAA
VQEAKIGPDKDIPHLTAAIIAKYISNLRLLLGLL

ZP_03433592

Secuencia de nucleótidos (SEQ ID NO:117)

>gi | 189214744:1-3126 Mycobacterium tuberculosis T85 NZ_ABOW01000178, secuenciación genómica completa al azar

GACATCGCGCTGATCCGGTTGGGCGCCGTGTCGGTTCCTACTGCAGACCAGTGC
GCCGGTCACCGGGTTGCGCCCGATCGTCACCGAGACCGAGCCGACGATGATCG
CCACCAGCATCGACAATCTTGCGGACGCCGTCGAAGTGCTGGCCGGTCACGCC
CCGGCCCGGCTGGTCGTATTCGATTACCACGGCAAGGTTGACACCCACCGCGA
GGCCGTCGAAGCCGCCGAGCTCGGTTGGCCGGCTCGGTTGACCATCGACACAC
TTGCCGAAC TGATCGAACGCGGCAGGGCGCTGCCGGCCACACCCATTGCCGAC
AGCGCCGACGACGCGCTGGCGCTGCTGATTTACACCTCGGGTAGTACCGGCGC
ACCCAAAGGCGCCATGTATCGCGAGAGCCAGGTGATGAGCTTCTGGCGCAAGT
CGAGTGGCTGGTTCGAGCCGAGCGGTTACCCCTCGATCACGCTGAACTTCATG
CCGATGAGCCACGTCGGGGGCCGTCAGGTGCTCTACGGGACGCTTTCCAACGG
CGGTACCGCTACTTCGTCGCCAAGAGCGACCTGTCGACGCTGTTTCGAGGACC
TCGCCCTGGTTCGGGCCACAGAATTGTGCTTCGTGCCGCGCATCTGGGACATG
GTGTTTCGCAGAGTTCACAGCGAGGTCGACCGCCGCTTGGTGGACGGCGCCGA
TCGAGCGGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAGAACGTGCTCG
GCGGACGGTTTGTTCATGGCGCTGACCGGTTCCGCGCCGATCTCCGCTGAGATG
ACGGCGTGGGTCGAGTCCCTGCTGGCCGACGTGCATTTGGTGGAGGGTTACGG
CTCCACCGAGGCCGGGATGGTCTGAACGACGGCATGGTTCGGCGCCCCGCGG
TGATCGACTACAAGCTGGTTCGACGTGCCCGAGCTGGGCTACTTCGGCACCGAT
CAGCCCTACCCCCGGGGCGAGCTGCTGGTCAAGACGCAAACCATGTTCCCCGG
CTACTACCAGCGCCCGGATGTCACCGCCGAGGTGTTTCGACCCCGACGGCTTCT
ACCGGACCGGGGACATCATGGCCAAAGTAGGCCCGACAGTTCGTCTACCTC
GACCGCCGCAACAACGTGCTAAAGCTCTCCAGGGCGAGTTCATCGCCGTGTC
GAAGCTCGAGGCGGTGTTTCGGCGACAGCCCGCTGGTCCGACAGATCTTCATCT
ACGGCAACAGTGCCCGGGCCTACCCGCTGGCGGTGGTTGTCCCGTCCGGGGAC
GCGCTTTCTCGCCATGGCATCGAGAATCTCAAGCCCGTGATCAGCGAGTCCCT
GCAGGAGGTAGCGAGGGCGGCCGCTGCAATCCTACGAGATTCCACGCGACT
TCATCATCGAAACCACGCCGTTACCCCTGGAGAACGGCCTGCTCACCGGCATC
CGCAAGCTGGCACGCCCGCAGTTGAAGAAGTTCTATGGCGAACGTCTCGAGCG

FIG 4 (Cont.)

GCTCTATACCGAGCTGGCCGATAGCCAATCCAACGAGCTGCGCGAGCTGCGGC
AAAGCGGTCCCGATGCGCCGGTGCTTCCGACGCTGTGCCGTGCCGCGGCTGCG
TTGCTGGGCTCTACCGCTGCGGATGTGCGGCCGGACGCGCACTTCGCCGACCT
GGGTGGTGACTCGCTCTCGGCGCTGTTCGTTGGCCAACCTGCTGCACGAGATCT
TCGGCGTCGACGTGCCGGTGGGTGTCATTGTCAGCCCGGCAAGCGACCTGCGG
GCCCTGGCCGACCACATCGAAGCAGCGCGCACCCGGCGTCAGGGCAGCCAGCTT
CGCCTCGATACACGGTCGCTCCGCGACGGAAGTGCACGCCAGCGACCTCACGC
TGGACAAGTTCATCGACGCTGCCACCCTGGCCGCAGCCCCGAACCTGCCGGCA
CCGAGCGCCCAAGTGCGCACCGTACTGCTGACCGGCGCCACCCGGCTTTTTGGG
TCGCTACCTGGCGCTGGAATGGCTCGACCCGCATGGACCTGGTCAACGGCAAGC
TGATCTGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAGCCCGGCTGGAC
GCGACGTTTCGATAGCGGCGACCCGTATTTGGTTCGCGGCACTACCGCGAATTGGG
CGCCGGCCGCCTCGAGGTGCTCGCCGGCGACAAGGGCGAGGCCGACCTGGGCC
TGGACCGGGTACCTGGCAGCGGCTAGCCGACACGGTGGACCTGATCGTGGAC
CCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGTTTCGGCCAAA
CGCGGCGGGCACCGCCGAGTTGCTTCGGCTGGCGCTGACCGGCAAGCGCAAGC
CATACTACACCTCGACGATCGCCGTGGGCGAGCAGATCCCGCCGGAGGCG
TTCACCGAGGACGCCGACATCCGGGCCATCAGCCCGACCCGCAGGATCGACGA
CAGCTACGCCAACGGCTACGCGAACAGCAAGTGGGCCGGCGAGGTGCTGCTGC
GCGAAGCTCACGAGCAGTGGCCCTGCCGGTACGGTCTTCCGCTGCGACATG
ATCCTGGCCGACACCAGCTATAACCGGTCAGCTCAACCTGCCGGACATGTTTAC
CCGGCTGATGCTGAGCCTGGCCGCTACCGGCATCGCACCCGGTTCGTTCTATG
AGCTGGATGCGCACGGCAATCGGCAACGCGCCCACTATGACGGCTTGCCGGTC
GAATTCGTCGCGAGAAGCCATTTGCACCCTTGGGACACATAGCCCGGACCGTTT
TGTCACCTACCACGTGATGAACCCCTACGACGACGGCATCGGGCTGGACGAGT
TCGTCGACTGGCTCAACTCCCCAACTAGCGGGTCCGGTTGCACGATCCAGCGG
ATCGCCGACTACGGCGAGTGGCTGCAGCGGTTTCGAGACTTCGCTGCGTGCCTT
GCCGGATCGCCAGCGCCACGCCTCGCTGCTGCCCTTGCTGCACAACCTACCGAG
AGCCTGCAAAGCCGATATGCGGGTCAATCGCGCCCACCGACCAGTTCGCGCT
GCCGTCCAAGAAGCGAAAATCGGTCCGGACAAAGACATTCCGCACCTCACGGC
GGCGATCATCGCGAAGTACATCAGCAACCTGCGACTGCTCGGGCTGCTGTGA

Secuencia de aminoácidos (SEQ ID NO:118)

>gi | 215446840 | ref | ZP_03433592.1 | ácido graso-CoA ligasa fadD9 [Mycobacterium tuberculosis T85]

DIALIRLGAVSVPLQTSAPVTGLRPIVTEPETMIATSIDNLGDAVEVLGHA
PARLVVFDYHGKVDTHREAVEAARARLAGSVTIDTLAELIERGRALPATPIAD
SADDALALLIYTSGSTGAPKGYMSQVMSFWRKSSGWFEPGYPISITLNF
PMSHVGGRQVLYGTLSNGGTAYFVAKSDLSTLFEDLALVRPTLFCFVPRIDM

FIG 4 (Cont.)

VFAEFHSEVDRRLVDGADRAALEAQVKAELRENVLGGRFVMALTGSAPI SAEM
 TAWVESLLADVHLVEGYGSTEAGMVLNDGMVRRPAVIDYKLVDPVPELGYFGTD
 QPYPRGELLVKQTMTMFPGYQRPDVTAEVFDPDGFYRTGDIMAKVGPDQFVYL
 DRRNNVLKLSQGEF IAVSKLEAVFGDSPLVRQIF IYGNSARAYPLAVVPSGD
 ALSRHGIENLKPVI SE SLQEVARAAGLQSYEIPRDF I IETTPFTLENGLLTGI
 RKLARPQLKKFYGERLERLYTELADSQSNELELRQSGPDAPVLP TLCRAAAA
 LLGSTAADV RPDAHFADLGGDSL SALS LANLLHEIFGVDPVPGVIVSPASDLR
 ALADHIEAARTG VRRP SFASIHGRSATEVHASDLTLDKFI DAATLAAAPNLPA
 PSAQVRTVLLTGATGFLGRYLALEWLD RMDLVNGKLI CLVRARSDEEAQARLD
 ATFDSGDPYLVRHYRELGAGRLEVL AGDKGEADLGLDRVTWQRLADTVDLIVD
 PAALVNHVLPYSQLFGPNAAGTAE LLRLAL TGKRKPYIYTSTIAVGEQIPPEA
 FTEDADIRAI SPTRRIDDSYANGYANS KWAGEVLLREAHEQCGLPVTVFRCDM
 ILADTSYTGQLNLPDMFTRLMLS LAATGIAPGSFYELDAHGNRQRAHYDGLPV
 EFVAEAICTLGTHSPDRFV TYHVMNPYDDGIGLDEFVDWLN SPTSGSGCTIQR
 IADYGEWLQRFETSLRALPDRQR HASLLPLLHNYREPAKPI CGSIAPT DQFRA
 AVQEAKIGPDKDIPHLTAAI IAKYISNLRLLLGLL

ZP 03537669

Secuencia de nucleótidos (SEQ ID NO:119)

>gi |192384126:19547-22546 Mycobacterium tuberculosis- T17 NZ_ABQH01000288, secuenciación genómica completa al azar

ATGTCGATCAACGATCAGCGACTGACACGCCGCGTCGAGGACCTATACGCCAG
 CGACGCCCAGTTCGCCGCCGCCAGTCCCAACGAGGCGATCACCCAGGCGATCG
 ACCAGCCCCGGGGTCGCGCTTCCACAGCTCATCCGTATGGTCATGGAGGGCTAC
 GCCGATCGGCCGGCACTCGGCCAGCGTGCGCTCCGCTTCGTCACCGACCCCGA
 CAGCGGCCCGCACCATGGTCGAGCTACTGCCGCGGTTTCGAGACCATCACCTACC
 GCGAACTGTGGGCCCGCGCCGGCACATTGGCCACCGCGTTGAGCGCTGAGCCC
 GCGATCCGGCCGGGCGACCGGGTTTGCCTGCTGGGCTTCAACAGCGTCGACTA
 CACAACCATCGACATCGCGCTGATCCGGTTGGGCGCCGTGTCGGTTCCACTGC
 AGACCAGTGCGCCGGTCAACGGGTTGCGCCCGATCGTCACCGAGACCGAGCCG
 ACGATGATCGCCACCAGCATCGACAATCTTGCGACGCCGTGCAAGTGCTGGC
 CGGTCACGCCCGGCCCGGCTGGTCGTATTCGATTACCACGGCAAGGTTGACA
 CCCACCGCGAGGCCGTCGAAGCCGCCCGAGCTCGGTTGGCCGGCTCGGTGACC
 ATCGACACACTTGCCGAAGTATCGAACGCGGCAGGGCGCTGCCGGCCACACC
 CATTGCCGACAGCGCCGACGACGCGCTGGCGCTGCTGATTTACACCTCGGGTA
 GTACCGGCGCACCCAAAGGCGCCATGTATCGCGAGAGCCAGGTGATGAGCTTC
 TGGCGCAAGTCGAGTGGCTGGTTTCGAGCCGAGCGGTTACCCCTCGATCACGCT
 GAACTTCATGCCGATGAGCCACGTCCGGGGCCGTCAGGTGCTCTACGGGACGC

FIG 4 (Cont.)

TTCCAACGGCGGTACCGCCTACTTCGTGCGCAAGAGCGACCTGTGCGACGCTG
 TTCGAGGACCTCGCCCTGGTGC GGCCACAGAATTGTGCTTCGTGCCGCGCAT
 CTGGGACATGGTGTTCGCAGAGTTCACAGCGAGGTCGACCGCCGCTTGGTGG
 ACGGCGCCGATCGAGCGGCGCTGGAAGCGCAGGTGAAGGCCGAGCTGCGGGAG
 AACGTGCTCGGCGGACGGTTTGTTCATGGCGCTGACCGGTTCCGCGCCGATCTC
 CGCTGAGATGACGGCGTGGGTGCGAGTCCCTGCTGGCCGACGTGCATTTGGTGG
 AGGGTTACGGCTCCACCGAGGCCGGGATGGTCTGAACGACGGCATGGTGC GG
 CGCCCCGCGGTGATCGACTACAAGCTGGTGCAGCTGCCCGAGCTGGGCTACTT
 CGGCACCGATCAGCCCTACCCCGGGGCGAGCTGCTGGTCAAGACGCAAACCA
 TGTTCGCCGGCTACTACCAGCGCCCGGATGTCACCGCCGAGGTGTTGACCCCC
 GACGGCTTCTACCGGACCGGGGACATCATGGCCAAAGTAGGCCCGACAGTT
 CGTCTACCTCGACCGCCGCAACAACGTGCTAAAGCTCTCCAGGGCGAGTTCA
 TCGCCGTGTGGAAGCTCGAGGCGGTGTTGCGCGACAGCCCGCTGGTCCGACAG
 ATCTTCATCTACGGCAACAGTGCCCGGGCCTACCCGCTGGCGGTGGTTGTCC
 GTCCGGGGACGCGCTTTCTCGCCATGGCATCGAGAATCTCAAGCCCGTGATCA
 GCGAGTCCCTGCAGGAGGTAGCGAGGGCGGCCGGCCTGCAATCCTACGAGATT
 CCACGCGACTTCATCATCGAAACCACGCCGTTACCCCTGGAGAACGGCCTGCT
 CACCGGCATCCGCAAGCTGGCACGCCCGCAGTTGAAGAAGTTCTATGGCGAAC
 GTCTCGAGCGGCTCTATACCGAGCTGGCCGATAGCCAATCCAACGAGCTGCGC
 GAGCTGCGGCAAAGCGGTCCCGATGCGCCGGTGCTTCCGACGCTGTGCCGTGC
 CGCGGCTGCGTTGCTGGGCTCTACCGCTGCGGATGTGCGGCCGGACGCGCACT
 TCGCCGACCTGGGTGGTGACTCGCTCTCGGCGCTGTCGTTGGCCAACCTGCTG
 CACGAGATCTTCGGCGTCGACGTGCCGGTGGGTGTCATTGTCAGCCCGGCAAG
 CGACCTGCGGGCCCTGGCCGACCACATCGAAGCAGCGCGCACCGGGCGTCAGGC
 GACCCAGCTTCGCCTCGATACACGGTCGCTCCGCGACGGAAGTGCACGCCAGC
 GACCTCACGCTGGACAAGTTCATCGACGCTGCCACCCTGGCCGACGCCCCGAA
 CCTGCCGGCACCGAGCGCCCAAGTGCGCACCGTACTGCTGACCGGGCGCCACCG
 GCTTTTTGGGTGCTACCTGGCGCTGGAATGGCTCGACCGCATGGACCTGGTC
 AACGGCAAGCTGATCTGCCTGGTCCGCGCCAGATCCGACGAGGAAGCACAAAGC
 CCGGCTGGACGCGACGTTGATAGCGGCGACCCGTATTTGGTGCGGCACTACC
 GCGAATTGGGCGCCGGCCGCCTCGAGGTGCTCGCCGGCGACAAGGGCGAGGCC
 GACCTGGGCCTGGACCGGGTACCTGGCAGCGGCTAGCCGACACGGTGGACCT
 GATCGTGGACCCCGCGGCCCTGGTCAACCACGTGCTGCCGTATAGCCAGCTGT
 TCGGCCCAAACGCGGCGGGCACCGCCGAGTTGCTTCCGGCTGGCGCTGACCGGC
 AAGCGCAAGCCATACATCTACACCTCGACGATCGCCGTGGGCGAGCAGATCCC
 GCCGGAGGCGTTACCGAGGACGCCGACATCCGGGCCATCAGCCCGACCCGCA
 GGATCGACGACAGCTACGCCAACGGCTACGCGAACAGCAAGTGGGCGGGCGAG
 GTGCTGCTGCGCGAAGCTCACGAGCAGTGCGGCCTGCCGGTGACGGTCTTCCG
 CTGCGACATGATCCTGGGCGGACACCAGCTATACCGGTGAGCTCAACCTGCCG
 GACATGTCACCCGGGCTGATGCTGAGCCTGGC

FIG 4 (Cont.)

Secuencia de aminoácidos (SEQ ID NO:120)

>gi | 219558593 | ref | ZP_03537669.11 ácido graso-CoA ligasa fadD9 [Mycobacterium tuberculosis T17]

MSINDQRLTRRVEDLYASDAQFAAASPNEAITQAIQPGVALPQLIRMVMEGY
 ADRPALGQRALRFVTDPSGRTMVELLPRFETITYRELWARAGTLATALSAEP
 AIRPGDRVCVLGFNSVDYTTIDIALIRLGAVSVPLQTSAPVTGLRPIVTETEP
 TMIATSIDNLGDAVEVLGHAPARLVVFDYHGKVDTHREAVEAARARLAGSVT
 IDTLAELIERGRALPATPIADSADDALALLIYTSGSTGAPKGAMYRESQVMSF
 WRKSSGWFEPSGYPISITLNFMPMSHVGGRRQVLYGTLNNGGTAYFVAKSDLSTL
 FEDLALVRPTLFCFVPRIWDMVFAEFHSEVDRRLVDGADRAALEAQVKAELRE
 NVLGGRFVMALTGSAPI SAEMTAWVESLLADVHLVEGYGSTEAGMVLNDGMVR
 RPAVIDYKLVDPPELGYFGTDQPYPRGELLVKTQTMFPGYYQRPDVTAEVFDP
 DGFYRTGDIMAKVGPDQFVYLDRRNNVLKLSQGEFIAVSKLEAVFGDSPVLRQ
 IFIYGNSARAYPLAVVPSGDALSRHGIENLKPVI SE SLQEVARAAGLQSYEI
 PRDFI IETTPFTLENGLLTGIRKLARPQLKKFYGERLERLYTELADSQSNELR
 ELRQSGPDAPVLP TLCRAAAALLGSTAADVRPDAHFADLGGDSLSALSANLL
 HEIFGVDVPVGVIVSPASDLRALADHIEAARTGVRPSPFASIHGRSATEVHAS
 DLTLDKFIDAATLAAAPNLPAPSAQVRTVLLTGATGFLGRYLALWLDLMDLV
 NGKLI CLVRARSDEEAQARLDATFDSGDPYLVVRHYRELGAGRLEVLAGDKGEA
 DLGLDRVTWQRLADTVDLIVDPAALVNHVLPYSQLFGPNAAGTAE LLRLALTG
 KRKPYIYTSTIAVGEQIPPEAFTEADIRAI SPTRRIDDSYANGYANSKWAGE
 VLLREAHEQCGLPVTVFRCDMILGRHQLYRSAQPAGHVTRADAEPG

ZP_05224908

Secuencia de nucleótidos (SEQ ID NO:121)

>gi | 163719654:2489-6013 Mycobacterium intracellulare ATCC 13950 NZ_ABIN01000072, secuenciación genómica completa al azar

ATGTCGACTGCCATTCATGACGAACACCTCGACCGTCGCATCGAGGAACTCAT
 CGCCAACGACCCCAATTCGCCGCGCCCGACCGGACCCGGCCATCACCGCCG
 CCACCGAAGCGCCCGGGCTGCGGCTGCCGCAGATCATCCGGACCGTGCTCGAC
 GGCTACGCCGACCGGCTGCCCTGGCGCAGCGCGTCGTGGAGTTCGTCACCGA
 CGCCAAGACCGGGCGGACGACGGCCGAGCTGCTCCCCGTTTCGAGACCATCA
 CGTATGGCGAACTCGGCGAACGGGTTTCGGCCCTCGGCCGTGCCTGGGCCGGC

FIG 4 (Cont.)

GACGCGGTGCGCCCCGGCGACCGCGTCTGCGTGCTCGGCTTCAACAGCGTTGA
 CTACGCCACCATCGACATCGCGCTGGGCACCATCGGGGCCGTGTGCGGTGCCGC
 TGCAGACCAGCGCGGCGATCTCCTCGTTGCAGCCGATCGTCGCCGAGACCGAG
 CCCAGCCTGATCGCCTCGAGCGTCAACCAGCTGCCCGACGCGGTGGAGCTGAT
 CCTGGCCGGCGACCACGTGCCCGCAAGCTCGTCGTGTTGACTACCAGCCCC
 AGGTCGACGACCAGCGCGAGGCCGTGGAGGCCGCCGCCGCGCGGTTGGCCGAC
 TCCGGCGTCGCGGTCGAGGCTCTCGCCGACGTGCTGCGGCGCGGCAAGGACCT
 GCCGGCCGTGAGCCGCCGGCGAGCGACGAGGACTCGCTGGCCCTGCTGATCT
 ACACCTCCGGCAGCACCGGCGCGCCCAAGGGCGCGATGTACCCGCAGAGCAAC
 GTCGGCAAGATGTGGCGGCGCGGGAGCAAGAACTGGTTCGGGGAAAGCGCCGC
 GTCGATCACCTCAACTTCATGCCGATGAGCCACGTCATGGGGCGCGGAATCC
 TCTACGGCACGCTGGGCAACGGCGGCACCGCGTACTTCGCCGCCCGCAGCGAC
 CTGTCCACCCTGCTCGAGGACCTCGAGTTGGTGCGGCCACCAGATGAACTT
 CGTCCCCCGCATCTGGGAGACGCTGTACGGCGAATTCAGCGCCAGGTCGAGC
 GCGGGCTGGCCGACGGCGATGCGGGCCCGGAGGCCCGCGAGACTGTGGCGGCT
 GCGGTGTTGGAAGAACAGCGCCAGTACCTGCTGGGCGGGCGGTTTCATCTTCGC
 GATGACGGGCTCGGCACCCACCTCGCCGGAGCTCAAGGCGTGGGCCGAGTCGC
 TCCTGCAGATGCACCTGATGGACGGCTACGGCTCCACCGAGGCCGGAATGGTG
 TTGTTGACGGGGAGATTACAGCGTCCGCCGGTTATTGATTACAAGCTGGTTCGA
 CGTTCGGATCTGGGCTATTTACGACCCGACCGTCCGCATCCGCGCGGTGAGT
 TGTTGCTGCGGACCGAGAACATGTTCCCGGGTTATTACAAGCGGGCCGAGACC
 ACCGCGAACGTGTTTCGACGAGGACGGTTATTACCGCACCGGTGACGTGTTTCGC
 CGAGATCGCGCCGGACCGGCTGGTGTATGTCGATCGCCGCAACAACGTGCTCA
 AGTTGGCCCAGGGCGAGTTCGTGACGCTGGCCAAGCTGGAGGCGGTGTTTCGGC
 AACAGCCCGCTGATCCGCCAGATCTACGTTTACGGCAACAGCTCCCAGCCCTA
 CCTGCTGGCCGTGGTGGTGCCGACCGAGGAAGCGTTGGCGGACAACGATCTTG
 AGTCGCTCAAGCCGAAGATCGCCGACTCGCTGCAGAAGGTCGCCAAGGAGACC
 GGCCTGCAGTCTACGAGGTGCCGCGCGACTTCATCATCGAGACCACGCCGTT
 CACCCTGGAAAACGGCCTGCTGACCGGGATCCGCAAGCTGGCGTGGCCCAAGC
 TCAAGGCGCACTACGGGGATCGGCTCGAGCAGATGTATGCCGAGCTGGCCGCG
 GGACAGGCCAACGAGTTGGCCGAAGTGCGCCGACGCGGCGCGGCGGCCGGT
 GGCCAGACCGTGAGCCGGGCCCGCGGCCCGCCCTGCTGGGTGCGACGGCCGGGG
 ATCTGTCCGCAGATGCCCACTTCACCGATCTTGGTGGAGACTCGTTGTGCGCG
 TTGACCTTCGGCAACCTGCTGCGCGAGATCTTCGATGTCGACGTGCCGGTGGG
 GGTGATCGTCAGCCCGGCCAACGACCTGGCGGGGATCGCCGCCTACATCGAGG
 CCGAGCGGCAGGGCTCCAAGCGCCCGACGTTCCGCCCGCTGCACGGTTCGCGGT
 GCGACCATGGTGCACGCCAGTGACCTCACGCTGGACAAGTTCCTCGACGAGGC
 GACCCTGGCCGCCGCGCCAGCCTGCCCAAGCCGGCCACCAGGTGCGCACCG
 TGCTGTTGACCGGCGCGACCGGCTTTTTGGGCCGCTACCTGGCGCTGGACTGG
 CTCGAGCGGATGGACATGGTCGACGGCAAGGTCATCGCCCTGGTGCGGGCCCG
 CACCGATGAGGAGGCGCGCGCCCGGCTGGACAAGACCTTCGACAGCGGCGACC

FIG 4 (Cont.)

CCAAACTGCTGGCGCACTACCAGCGGCTGGCCGCGACCACCTCGAGGTCATC
 GCCGGCGACAAGGGTGAGGCCAACCTCGGCCTGGACCCCAGACCTGGCAGCG
 ACTGGCCGAGGAGGTCGACGTCATCGTCGACCCCGCCGCGCTGGTCAACCACG
 TGCTGCCCTACAGCGAGCTGTTCCGGCCCCAACGCCCTGGGCACCGCGGAGCTG
 ATCCGGATCGCGCTGACCTCCAGGCCAAAAGCCCTACACCTACGTGTGACGAT
 CGGGGTGGGCGATCAGATCCAGCCAGGTGAGTTCGTGCGAGAACGCCGACATCC
 GCCAGATCAGCGCCACCCGCGAGATCAACGACGGCTACGCCAACGGCTACGGC
 AACAGCAAGTGGGCCGCGAGGTGTTGCTGCGCGAGGCCACGACCTGTGCGG
 CCTGCCCCGTCACGGTGTTCGCTGCGACATGATCCTGGCCGACACCACCTATG
 CCGGGCAGCTCAACCTGCCCGACATGTTACCCGGCTGATGCTGAGCCTGGTC
 GCCACCGGTATCGCGCCCGGGTCGTTCTACGAACTGGACGCCGACGGCAACCG
 CCAGCGGGCACACTACGACGGTTTGCCGGTCGAGTTCATCGCCGCGGCGATCT
 CGACGCTGGGGACCCAAATCACCGACAGCGACACGGGCTTTCAGACCTACCAC
 GTGATGAACCCCTACGACGACGGCATCGGGCTGGATGAGTACATCGATTGGCT
 GATCGAGGCCGGGTATTTCGATCGAGCGGATCGCCGATTACTCCGAGTGGCTGC
 GCGCTTCGAGACCTCGCTGCGGGCCCTGCCCGATCGGCAGCGTCAGTACTCG
 CTGCTGCCGCTGCTGCACAACCTACCAGAAGCCGGAAAAGCCGATCAACGGCTC
 GATGGCGCCCACCGACGTGTTCCGTGCCGCGGTGCAGGAAGCGAAAATCGGCC
 CCGACAAAGACATCCCGCACGTCTCGGGCGCCGGTGCATCGTCAAGTACATCACC
 GACCTGGAGTTGCTCGGACTCCTCTGA

Secuencia de aminoácidos (SEQ ID NO:122)

>gi | 254819907 | ref | ZP_05224908.1 | FadD9 [Mycobacterium intracellulare ATCC 13950]

MSTAIHDEHLDRRIEELIANDPQFAAARPDPAITAATEAPGLRRLPQIIRTVLD
 GYADRPALAQRVVEFVTDAKTGRTTAELLPRFETITYGELGERVVSALGRAWAG
 DAVRPGDRVCVLGFNSVDYATIDIALGTIGAVSVPLQTSAAISSLQPIVAETE
 PSLIASSVNQLPDAVELILAGDHVPGKLVVFDYQPQVDDQREAVEAAAARLAD
 SGVAVEALADVLRGKDLPAVEPPASDEDSLALLIYTSGSTGAPKGMYPQSN
 VGKMWRRGSKNWFGEAASITLNFMPMSHVMGRGILYGTGNGGTAYFAARSD
 LSTLLEDLELVRPTEMNFVPRIWETLYGEFQRQVERRLADGDAGPEARETVAA
 AVLEEQRQYLLGGRFIFAMTGSAPTSPKAWAESLLQMHLMMDGYGSTEAGMV
 LFDGEIQRPVIDYKLVDPDLGYFSTDRPHPRGELLLRTENMFPGYKRAET
 TANVFDEDGYYRTGDVFAEIAPDRLVYVDRRNNVLKLAQGEFVTLAKLEAVFG
 NSPLIRQIYVYGNSSQPYLLAVVVPTEEALADNDLES LKPKIADSLQVAKET
 GLQSYEVPRDFI IETTPFTLENGLLTGIRKLAWPKLKAHYGDRLEQMYAELAA
 GQANELAELRRSGAAAPVAQTVSRAAAALLGATAGDLSADAHFTDLGGDSL
 LTFGNLLREIFDVDVPVGVIVSPANDLAGIAAYIEAERQGSKRPTFAAVHGRG
 ATMVHASDLTLDFKFLDEATLAAAPSLPKPATEVRTVLLTGATGFLGRYLALDW

FIG 4 (Cont.)

LERMDMVDGKVIALVRARTDEEARARLDKTFDSGDPKLLAHYQRLAADHLEVI
AGDKGEANLGLDPQTWQRLAEEVDVIVDPAALVNHVLPYSELFPGNALGTAEI
IRIALTSRQKPYTYVSTIGVGDQIQPGEFVENADIRQISATREINDGYANGYG
NSKWAGEVLLREAHDLGCLPVTVFRCDMILADTTYAGQLNLPDMFTRLMLSLV
ATGIAPGSFYELDADGNRQRAHYDGLPVEFIAAAISTLGTQITDSDTGFQTYH
VMNPYDDGIGLDEYIDWLI EAGYSIERIADYSEWLRRFETSLRALPDRQRQYS
LLPLLHNYQKPEKPINGSMAPTDVFRAAVQEAKIGPDKDIPHVSAPVIVKYIT
DLELLGLL

FIG 5

Los números de registro son del NCBI, GenBank, versión 159.0 de marzo de 2008
 Los números de EC son de KEGG, publicación 42.0 de abril de 2007 (más actualizaciones diarias hasta marzo de 2008)

CATEGORIA	GEN	NOMBRE	REGISTRO	NUMERO DE EC	MODIFICACION	USO	MICROORGANISMO
1. Aumenta la producción de ácidos grasos / aumenta la producción de producto							
<i>aumenta acil-Coa</i>							
reduce el catabolismo de derivados y productos intermedios							
reduce la inhibición de retroalimentación							
atenúa otras rutas que consumen ácidos grasos							
	accA	Acetil-CoA carboxilasa, subunidad A (carboxiltransferasa alfa)	AAG73296, NP_414727	6.4.12	Sobreexpresada	aumenta la producción de malonil-CoA	<i>Escherichia coli</i> , <i>Lactococci</i>
	accB	Acetil-CoA carboxilasa, subunidad B (BGGP: proteína portadora de biotina-carboxilo)	NP-417721	6.4.12	Sobreexpresada	aumenta la producción de malonil-CoA	
	accC	Acetil-CoA carboxilasa, subunidad C (biotina carboxilasa)	NP-417721	6.4.12 6.3.4.14	Sobreexpresada	aumenta la producción de malonil-CoA	<i>Escherichia coli</i> , <i>Lactococci</i>

FIG 5 (cont.)

accD	Acetil-CoA carboxilasa, subunidad D (carboxiltransferasa beta)	NP_416819	6.4.12	Sobreexpresada	aumenta la producción de malonil-CoA	<i>Escherichia coli</i> , <i>Lactococci</i>				
accE	piruvato deshidrogenasa, subunidad E1	NP_414656 AAC73226	1.2.4.1	Sobreexpresada	aumenta la producción de acetil-CoA	<i>Escherichia coli</i>				
accF	piruvato deshidrogenasa, subunidad E2	NP_414657	2.3.1.12	Sobreexpresada	aumenta la producción de acetil-CoA	<i>Escherichia coli</i>				
ackA	acetato cinasa	AAC75356, NP_416799	2.7.2.1	Deleción o reduce	aumenta la producción de acetil-CoA	<i>Escherichia coli</i>				
ackB	acetato cinasa ackB	BAB81430	2.7.2.1	Deleción o reduce	aumenta la producción de acetil-CoA	<i>Escherichia coli</i>				
acpP	proteína portadora de acilo	AAC74178	NINGUNO	Sobreexpresada	aumenta la producción de acetil-CoA	<i>Escherichia coli</i>				
facD	acil-CoA sintasa	AP002424	2.3.1.86, 6.2.1.3	Sobreexpresada	aumenta la producción de acetil-CoA	<i>Escherichia coli</i> W3110				
adhE	alcohol deshidrogenasa	CAA47743	1.1.1.1, 1.2.1.10	Deleción o reduce	aumenta la producción de acetil-CoA	<i>Escherichia coli</i> W3111				

FIG 5 (cont.)

acr1	Acil-Coa reductasa	YP_047869, AAC45217	1.2.1.42	Sobreexpresión	para producción de alcoholes grasos	<i>Acinetobacter sp., es decir, calcoaceticus</i>
GST, gshB	Glutación sintasa sn-glicerol biosintético 3-fosfato deshidrogenasa	P04425	6.3.2.3	Deleción o reduce	incrementa la producción de acetil-CoA	<i>E. coli</i> K12
gpsA	lactato deshidrogenasa	AAC76632, NP_418065	EC: 1.1.1.94	Deleción o reduce	incrementa la producción de acetil-CoA	<i>E. coli</i> K12
ldhA	lactato deshidrogenasa	AAC74462, NP_415898	EC: 1.1.1.27, 1.1.1.28	Deleción o reduce	incrementa la producción de acetil-CoA	<i>E. coli</i> K12
Lipasa	Triglicérido lipasa	CAA89087, CAA98876	3.1.1.3	expresión	incrementa la producción de ácidos grasos	<i>Saccharomyces cerevisiae</i>
panD	Malonil-CoA descarboxilasa	AAA26500	4.1.1.9, 4.1.1.41	Sobreexpresión		<i>Saccharopolyspora erytraea</i>
PanK a.k.a. coA	aspartato 1- descarboxilasa	BAB96708	4.1.1.11	Sobreexpresión	incrementa la producción de acetil-CoA	<i>Escherichia coli</i> W3110
panK a.k.a. coA, R106K	pantotenato cinasa	AAC76952	2.7.1.33	Sobreexpresión	incrementa la producción de acetil-CoA	<i>E. coli</i>
pdh	piruvato deshidrogenasa	AAC76952 BAB34380, AAC73226, NP_415392	2.7.1.33	Expresión, sobreexpresión, mutación R106K	incrementa la producción de acetil-CoA	<i>E. coli</i>
			1.2.4.1	Sobreexpresión	incrementa la producción de acetil-CoA	

FIG 5 (cont.)

pfiB	formiato acetil transferasa (piruvato formiato liasa)	AAC73989, P09373	EC: 2.3.1.54	Deleciona o reduce	aumenta la producción de acetil-CoA reduce límites sobre reserva de acil-CoA	<i>E. coli</i> K12
plsB	acetiltransferasa	AAC77011	2.3.1.15	Mutación D311E	aumenta la producción de acetil-CoA	
poxB	acetiltransferasa	AAC73958, NP_415392	1.2.2.2	Deleciona o reduce	aumenta la producción de acetil-CoA	
pta	fosfotransacetilasa	AAC75357, NP_416800	2.3.1.8	Deleciona o reduce	aumenta la producción de acetil-CoA	
udhA	piridina nucleótido transhidrogenasa 3-hidroxiacetil-CoA epimerasa fusionada/delta(3)-cis-delta(2)-trans-enoil-Coa isomerasa/enoil-CoA hidratasa y 3-hidroxiacetil-CoA deshidrogenasa	CAA46822	1.6.1.1	Sobreexpresada	Conversión de NADH en NADPH o viceversa	
fadB	y 3-hidroxiacetil-CoA deshidrogenasa; K01692 enoil-CoA hidratasa; K01782 3-hidroxiacetil-CoA epimerasa	AP 003956	4.2.1.17, 5.1.2.3, 5.3.3.8, 1.1.1.35	Deleciona o reduce	Bloquea la degradación de ácidos grasos	<i>E. coli</i>
fadJ		AAC75401	1.1.1.35, 4.2.1.17, 5.1.2.3	Deleciona o reduce	Bloquea la degradación de ácidos grasos	<i>E. coli</i>

FIG. 5 (Cont.)

expresa FabH de <i>B. subtilis</i> y desactiva FabH endógena									
bdk – E3 – subunidad de dihidrolipouil deshidrogenasa			EC 1.2.4.4						
bkd – E2 – subunidad alfa/beta			EC 1.2.4.4						
bkd – E2 – subunidad dihidrolipouil transacilasa			EC 1.2.4.4						
bkdA1			EC 1.2.4.4	NP_628006	subunidad a de α -cetoácido de cadena ramificada deshidrogenasa (E1a)	Expresa sobreexpresa	o	producir precursores de acil-CoA de cadena ramificada	<i>Streptomyces coelicolor</i>
bkdB1			EC 1.2.4.4	NP_628005	subunidad a de α -cetoácido de cadena ramificada deshidrogenasa (E1b)	Expresa sobreexpresa	o	producir precursores de acil-CoA de cadena ramificada	<i>Streptomyces coelicolor</i>
bkdC1			EC 2.3.1.168	NP_628004	dihidrolipouil transacilasa (E2)	Expresa sobreexpresa	o	producir precursores de acil-CoA de cadena ramificada	<i>Streptomyces coelicolor</i>

FIG. 5 (Cont.)

bkdA2	subunidad a de a-cetoácido de cadena ramificada deshidrogenasa (E1a)	NP_733618	EC 1.2.4.4	Expresa sobreexpresada	o producir precursores de acil-CoA de cadena ramificada	<i>Streptomyces coelicolor</i>
bkdB2	subunidad b de a-cetoácido de cadena ramificada deshidrogenasa (E1b)	NP_628019	EC 1.2.4.4	Expresa sobreexpresada	o producir precursores de acil-CoA de cadena ramificada	<i>Streptomyces coelicolor</i>
bkdC2	dihidrolipoil transacetilasa (E2)	NP_628018	EC 2.3.1.168	Expresa sobreexpresada	o producir precursores de acil-CoA de cadena ramificada	<i>Streptomyces coelicolor</i>
bkdA	subunidad a de a-cetoácido de cadena ramificada deshidrogenasa (E1a)	BAC72074	EC 1.2.4.4	Expresa sobreexpresada	o producir precursores de acil-CoA de cadena ramificada	<i>Streptomyces avermitilis</i>
bkdB	subunidad b de a-cetoácido de cadena ramificada deshidrogenasa (E1b)	BAC72075	EC 1.2.4.4	Expresa sobreexpresada	o producir precursores de acil-CoA de cadena ramificada	<i>Streptomyces avermitilis</i>
bkdC	dihidrolipoil transacetilasa (E2)	BAC72076	EC 2.3.1.168	Expresa sobreexpresada	o producir precursores de acil-CoA de cadena ramificada	<i>Streptomyces avermitilis</i>
bkdF	subunidad a de a-cetoácido de cadena ramificada deshidrogenasa (E1a)	BAC72088	EC 1.2.4.4	Expresa sobreexpresada	o producir precursores de acil-CoA de cadena ramificada	<i>Streptomyces avermitilis</i>

FIG. 5 (Cont.)

bkdG	subunidad b de a-cetoácido de cadena ramificada deshidrogenasa (E1b)	BAC72089	EC 1.2.4.4	Expresa sobreexpresada	o	producir precursores de acil-CoA de cadena ramificada	<i>Streptomyces avermitilis</i>
bkdH	dihidrolipoil transacetilasa (E2)	BAC72090	EC 2.3.1.168	Expresa sobreexpresada	o	producir precursores de acil-CoA de cadena ramificada	<i>Streptomyces avermitilis</i>
bkdAA	subunidad a de a-cetoácido de cadena ramificada deshidrogenasa (E1a)	NP_390285	EC 1.2.4.4	Expresa sobreexpresada	o	producir precursores de acil-CoA de cadena ramificada	<i>Bacillus subtilis</i>
bkdAB	subunidad b de a-cetoácido de cadena ramificada deshidrogenasa (E1b)	NP_390284	EC 1.2.4.4	Expresa sobreexpresada	o	producir precursores de acil-CoA de cadena ramificada	<i>Bacillus subtilis</i>
bkdB	dihidrolipoil transacetilasa (E2)	NP_390283	EC 2.3.1.168	Expresa sobreexpresada	o	producir precursores de acil-CoA de cadena ramificada	<i>Bacillus subtilis</i>
bkdA1	subunidad a de a-cetoácido de cadena ramificada deshidrogenasa (E1a)	AAA65614	EC 1.2.4.4	Expresa sobreexpresada	o	producir precursores de acil-CoA de cadena ramificada	<i>Pseudomonas putida</i>
bkdA2	subunidad b de a-cetoácido de cadena ramificada deshidrogenasa (E1b)	AAA65615	EC 1.2.4.4	Expresa sobreexpresada	o	producir precursores de acil-CoA de cadena ramificada	<i>Pseudomonas putida</i>

FIG. 5 (Cont.)

bkdC	dihidrolipoil transacetilasa (E2)	AAA65617	EC 2.3.1.168	Expresa sobreexpresada	o	producir precursores de acil-CoA de cadena ramificada	<i>Pseudomonas putida</i>
lpd	dihidrolipoamida deshidrogenasa (E3)	NP_414658	1.8.1.4	Expresa sobreexpresada	o	producir precursores de acil-CoA de cadena ramificada	<i>Escherichia coli</i>
llvE	aminoácido de cadena ramificada aminotransferasa	YP_026247	2.6.1.42	Expresa sobreexpresada	o	producir a-cetoácidos ramificados	<i>Escherichia coli</i>
llvE	aminoácido de cadena ramificada aminotransferasa	AAF34406	2.6.1.42	Expresa sobreexpresada	o	producir a-cetoácidos ramificados	<i>Lactococcus lactis</i>
llvE	aminoácido de cadena ramificada aminotransferasa	NP_745648	2.6.1.42	Expresa sobreexpresada	o	producir a-cetoácidos ramificados	<i>Pseudomonas putida</i>
llvE	aminoácido de cadena ramificada aminotransferasa	NP_629657	2.6.1.42	Expresa sobreexpresada	o	producir a-cetoácidos ramificados	<i>Streptomyces coelicolor</i>
ccr	crotonil-CoA reductasa	NP_630556	1.6.5.5, 1.1.1.1	Expresa sobreexpresada	o	convertir crotonil-CoA en butiril-CoA	<i>Streptomyces coelicolor</i>
ccr	crotonil-CoA reductasa	AAD53915	1.6.5.5, 1.1.1.1	Expresa sobreexpresada	o	convertir crotonil-CoA en butiril-CoA	<i>Streptomyces cinnamomensis</i>
lcmA, isobutilil-CoA mutasa	isobutilil-CoA mutasa, subunidad A	NP_629554	5.4.99.2	Expresa sobreexpresada	o	convertir butiril-CoA en isobutilil-CoA	<i>Streptomyces coelicolor</i>
lcmA, isobutilil-CoA mutasa	isobutilil-CoA mutasa, subunidad A	AAC08713	5.4.99.2	Expresa sobreexpresada	o	convertir butiril-CoA en isobutilil-CoA	<i>Streptomyces cinnamomensis</i>

FIG. 5 (Cont.)

IcmB, isobutiril-CoA mutasa	isobutiril-CoA mutasa, subunidad B	NP_630904	5.4.99.2	Expresa sobreexpresa	convertir butiril-CoA en isobutiril-CoA	<i>Streptomyces coelicolor</i>
IcmB, isobutiril-CoA mutasa	isobutiril-CoA mutasa, subunidad B	CAB59633	5.4.99.2	Expresa sobreexpresa	convertir butiril-CoA en isobutiril-CoA	<i>Streptomyces cinnamomensis</i>
Genes de FabH, ACP y fabF con especificidad por acil-CoA de cadena ramificada						
IlvE	aminoácido de cadena ramificada aminotransferasa	CAC12788	EC2.6.1.42	Sobreexpresa	aminoácido de cadena ramificada aminotransferasa	<i>Staphylococcus carnosus</i>
FabH1	beta-cetoacil-ACP sintasa III	NP_626634	2.3.1.180	Expresa sobreexpresa	inicio de la biosíntesis de ácidos grasos de cadena ramificada	<i>Streptomyces coelicolor</i>
ACP	proteína portadora de acilo	NP_626635	NINGUNO	Expresa sobreexpresa	inicio y elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Streptomyces coelicolor</i>

FIG. 5 (Cont.)

FabF	beta-cetoacil-ACP sintasa II	NP_626636	2.3.1.179	Expresa sobreexpresada	o elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Streptomyces coelicolor</i>
FabH3	beta-cetoacil-ACP sintasa III	NP_823466	2.3.1.180	Expresa sobreexpresada	o inicio de la biosíntesis de ácidos grasos de cadena ramificada	<i>Streptomyces avermitilis</i>
FabC3 (ACP)	proteína portadora de acilo	NP_823467	NINGUNO	Expresa sobreexpresada	o inicio y elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Streptomyces avermitilis</i>
FabF	beta-cetoacil-ACP sintasa II	NP_823468	2.3.1.179	Expresa sobreexpresada	o elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Streptomyces avermitilis</i>
FabH_A	beta-cetoacil-ACP sintasa III	NP_389015	2.3.1.180	Expresa sobreexpresada	o inicio de la biosíntesis de ácidos grasos de cadena ramificada	<i>Bacillus subtilis</i>
FabH_B	beta-cetoacil-ACP sintasa III	NP_388898	2.3.1.180	Expresa sobreexpresada	o inicio de la biosíntesis de ácidos grasos de cadena ramificada	<i>Bacillus subtilis</i>

FIG. 5 (Cont.)

ACP		proteína portadora de acilo	NP_389474	NINGUNO	Expresa sobreexpresa	o inicio y elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Bacillus subtilis</i>
FabF		beta-cetoacil-ACP sintasa II	NP_389016	2.3.1.179	Expresa sobreexpresa	o elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Bacillus subtilis</i>
SmaIDRAFT_08 18		beta-cetoacil-ACP sintasa III	ZP_01643059	2.3.1.180	Expresa sobreexpresa	o inicio de la biosíntesis de ácidos grasos de cadena ramificada	<i>Stenotrophomonas maltophilia</i>
SmaIDRAFT_08 21		proteína portadora de acilo	ZP_01643063	NINGUNO	Expresa sobreexpresa	o inicio y elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Stenotrophomonas maltophilia</i>
SmaIDRAFT_08 22		beta-cetoacil-ACP sintasa II	ZP_01643064	2.3.1.179	Expresa sobreexpresa	o elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Stenotrophomonas maltophilia</i>
FabH		beta-cetoacil-ACP sintasa III	YP_123672	2.3.1.180	Expresa sobreexpresa	o inicio de la biosíntesis de ácidos grasos de cadena ramificada	<i>Legionella pneumophila</i>

FIG. 5 (Cont.)

ACP	proteína portadora de acilo	YP_123675	NINGUNO	Expresa sobreexpresa	o	inicio y elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Legionella pneumophila</i>
FabF	beta-cetoacil-ACP sintasa II	YP_123676	2.3.1.179	Expresa sobreexpresa	o	elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Legionella pneumophila</i>
FabH	beta-cetoacil-ACP sintasa III	NP_415609	2.3.1.180	Deleciona reduce	o	inicio de la biosíntesis de ácidos grasos de cadena ramificada	<i>Escherichia coli</i>
FabF	beta-cetoacil-ACP sintasa II	NP_415613	2.3.1.179	Deleciona reduce	o	elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Escherichia coli</i>
<i>Para producir ácidos grasos cíclicos</i>							
AnSJ	deshidratasa (supuesta)	no disponible	no disponible	Expresa sobreexpresa	o	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces collinus</i>
AnSK	CoA ligasa (supuesta)	no disponible	no disponible	Expresa sobreexpresa	o	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces collinus</i>
AnSL	deshidrogenasa (supuesta)	no disponible	no disponible	Expresa sobreexpresa	o	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces collinus</i>

FIG. 5 (Cont.)

ChcA		enoil-CoA reductasa (supuesta)	U72144	1.3.1.34	Expresa sobreexpresa	o	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces collinus</i>
AnzM		oxidoreductasa (supuesta)	no disponible	no disponible	Expresa sobreexpresa	o	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces collinus</i>
PlmJ		deshidratasa (supuesta)	AAQ84158	no disponible	Expresa sobreexpresa	o	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces sp. HK803</i>
PlmK		CoA ligasa (supuesta)	AAQ84158	no disponible	Expresa sobreexpresa	o	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces sp. HK803</i>
PlmL		deshidrogenasa (supuesta)	AAQ84159	no disponible	Expresa sobreexpresa	o	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces sp. HK803</i>
ChcA		enoil-CoA reductasa	AAQ84160	EC 1.1.1.34	Expresa sobreexpresa	o	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces sp. HK803</i>
PlmM		oxidoreductasa (supuesta)	AAQ84161	no disponible	Expresa sobreexpresa	o	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces sp. HK803</i>
ChcB		enoil-CoA isomerasa	AF268489	no disponible	Expresa sobreexpresa	o	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces collinus</i>
ChcB/CaiD		enoil-CoA isomerasa	NP_629292	4.2.1.-	Expresa sobreexpresa	o	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces coelicolor</i>

FIG. 5 (Cont.)

ChcB/CaiD	enoil-CoA isomerasa	NP_824296	4.2.1.-	Expresa sobreexpresa	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces avermitilis</i>
<u>2C. Control del nivel de saturación</u>						
Sta	Supresor de FabA	AAN79592, AAC44390	NINGUNO	Sobreexpresa	aumentar ácidos grasos monoinsaturados	<i>E. coli</i>
véase también FabA en la sec. 1				Expresa	producir ácidos grasos insaturados	
GnsA	Supresor de la mutación nula de secG	ABD18647.1	NINGUNO	Sobreexpresa	aumentar ésteres de ácidos grasos insaturados	<i>E. coli</i>
GnsB	Supresor de la mutación nula de secG	AAC74076.1	NINGUNO	Sobreexpresa	aumentar ésteres de ácidos grasos insaturados	<i>E. coli</i>
véase también la sección 2A – los elementos con :0 están insaturados (sin dobles enlaces) y con :1 están saturados (1 doble enlace)						

FIG. 5 (Cont.)

fabB	3-oxoacil-[proteína portadora de acilo] sintasa	BAA16180	EC:2.3.1.41	Sobreexpresa	modular la producción de ácidos grasos insaturados	<i>Escherichia coli</i>
fabK	trans-2-enoil-ACP reductasa II	AAF98273	1.3.1.9	Expresa	modular la producción de ácidos grasos insaturados	<i>Streptococcus pneumoniae</i>
fabL	enoi-(proteína portadora de acilo) reductasa	AAU39821	1.3.1.9	Expresa	modular la producción de ácidos grasos insaturados	<i>Bacillus licheniformis</i> DSM 13
fabM	trans-2,cis-3-decenoil-ACP isomerasa	DAA05501	4.2.1.17	Sobreexpresa	modular la producción de ácidos grasos insaturados	<i>Streptococcus mutans</i>
Producción de aldehídos grasos						
	tioesterasa			Expresa	producir	
Exportar						

FIG. 5 (Cont.)

	Exportador de ésteres de ceras (familia de FATP, proteína transportadora de ácidos grasos (de cadena larga))		NP_524723	NINGUNO	Expresa	exportar cera	<i>Drosophila melanogaster</i>
	proteína de transporte de ABC	supuesto transportador de alcanos	AAN73268	NINGUNO	Expresa	exportar productos	<i>Rhodococcus erythropolis</i>
CER5		transportador de ceras	At1g51500, AY734542, At3g21090, At1g51460	NINGUNO	Expresa	exportar productos	<i>Arabidopsis thaliana</i>
AIMRP5		asociado con resistencia a múltiples fármacos de <i>Arabidopsis thaliana</i>	NP_171908	NINGUNO	Expresa	exportar productos	<i>Arabidopsis thaliana</i>
AmiS2		transportador de ABC AmiS2	JC5491	NINGUNO	Expresa	exportar productos	<i>Rhodococcus sp.</i>
AIPGP1		GLICOPROTEÍNA DE ARABIDOPSIS/ THALIANA P	NP_181228	NINGUNO	Expresa	exportar productos	<i>Arabidopsis thaliana</i>
Acra		supuesta proteína de transporte de eflujo de múltiples fármacos, acra	CAF23274	NINGUNO	Expresa	exportar productos	<i>Candidatus protochlamydia amoebophila UWE25</i>

FIG. 5 (Cont.)

AcrB	probable proteína de transporte de eflujo de múltiples fármacos, acrB	CAF23275	NINGUNO	Expresa	exportar productos	<i>Candidatus protochlamydia amoebophila</i> UWE25
ToIC	proteína de membrana exterior [biogénesis de la envuelta celular]	ABD59001	NINGUNO	Expresa	exportar productos	<i>Francisella tularensis</i> subsp. <i>novicida</i>
AcrE	proteína de membrana que afecta la formación de septo y permeabilidad de membrana celular	YP_312213	NINGUNO	Expresa	exportar productos	<i>Shigella sonnei</i> Ss046
AcrF	proteína F de resistencia a acriflavina	P24181	NINGUNO	Expresa	exportar productos	<i>Escherichia coli</i>
tll1618	transportador de eflujo de múltiples fármacos	NP_682408.1	NINGUNO	Expresa	exportar productos	<i>Thermosynechococcus elongatus</i> BP-1
tll1619	transportador de eflujo de múltiples fármacos	NP_682409.1	NINGUNO	Expresa	exportar productos	<i>Thermosynechococcus elongatus</i> BP-1
tll0139	transportador de eflujo de múltiples fármacos	NP_680930.1	NINGUNO	Expresa	exportar productos	<i>Thermosynechococcus elongatus</i> BP-1
<u>5. Fermentación</u>						
	genes de punto de control de la replicación				aumentar la eficacia de producción	

FIG. 5 (Cont.)

	umuD	ADN polimerasa V, subunidad	YP_310132	3.4.21.-	Sobreexpresada	aumentar eficacia producción	la de	<i>Shigella sonnei</i> Ss046
	umuC	ADN polimerasa V, subunidad	ABC42261	2.7.7.7	Sobreexpresada	aumentar eficacia producción	la de	<i>Escherichia coli</i>
	NADH:NADPH transhidrogenasa (subunidades alfa y beta) (pntA, pntB)		P07001, P0AB70	1.6.1.2	Expresa	aumentar eficacia producción	la de	<i>Shigella flexneri</i>

FIG. 6

Gen	Nombre	Secuencia de Nucleótidos	Secuencia de Proteína
fabA	Beta-hidroxidecanoil tioéster deshidrasa	atgGTAGATA AACGCGAATC CTATACAAAA GAAGACCTTC TTGCCTCTGG TCGCGGTGAA CTGTTTGGCG CTAAAGGCC GCAATTGCCA GCACCGAACA TGCTGATGAT GGACCGTGTG GTCAAAATGA CCGAAACGGG TGGTAACTTC GACAAAGGGT ATGTTGAAGC AGAACTGGAT ATCAATCCGG ATCTGTGGTT CTTCGGATGC CACTTTATTG GCGATCCGGT TATGCCGGGA TGCCTGGGCC TGGACGCAAT GTGGCAGCTG GTAGGGTTCT ACCTCGGCTG GCTGGGCGGC GAAGGTAAAG GCCGCGCGCT GGGCGTTGGC GAAGTGAAAT TCACTGGTCA GGTACTGCCG ACAGCGAAAA AAGTGACCTA	MVDKRESYTK EDLLASGRGE LFGAKGPQLP APNMLMMDRV VKMTETGGNF DKGYVEAELD INPDLWFFGC HFIGDPVMPG CLGLDAMWQL VGFYLGWLGG EGKGRALGVG EVKFTGQVLP TAKKVTYRIH FKRIVNRRLI MGLADGEVLV DGRLIYTASD LKVGLFQDTS AF (SEQ ID NO: 94)

FIG. 6 (Cont.)

		CCGTATTAC TTTAAACGCA TTGTTAACCG TCGTCTGATT ATGGGCCTGG CGGATGGCGA AGTGCTGGTT GATGGTCGTC TGATCTATAC CGCCAGCGAC CTGAAAGTCG GTCTGTTCCA GGATACGTCT GCCTTCTGA (SEQ ID NO: 93)	
fabZ	(3R)-hidroximiristol- proteína portadora de acilo- deshidratasa	ttgACTACTA ACACTCATA TCTGCAGATT GAAGAGATTT TAGAACTTCT GCCGCACCGT TTCCCGTTCT TACTGGTGGA TCGCGTGCTG GATTTTGAAG AAGGTCGTTT TCTGCGCGCA GTAAAAAATG TCTCTGTCAA TGAGCCATTC TTCCAGGGCC ATTTCCCTGG AAAACCGATT TTCCCGGGTG TGCTGATTCT GGAAGCAATG GCACAGGCAA CAGGTATTCT GGCGTTTAAA	MTTNTHTLQI EEILELLPHR FPFLLVDRVL DFEGRFLRA VKNVSVNEPF FQGHFPGKPI FPGVLILEAM AQATGILAFK SVGKLEPGEL YYFAGIDEAR FKRPVVPGDQ MIMEVTFEKT RRGLTRFKGV ALVDGKVVCE ATMMCARSA A SEQ ID NO: 96)

FIG. 6 (Cont.)

		<p>AGCGTAGGAA AACTGGAACC GGGTGAGCTG TACTACTTCG CTGGTATTGA CGAAGCGCGC TTCAAGCGCC CGGTCGTGCC TGGCGATCAA ATGATCATGG AAGTCACTTT CGAAAAACG CGCCGCGGCC TGACCCGTTT TAAAGGGGTT GCTCTGGTCG ATGGTAAAGT AGTTTGCGAA GCAACGATGA TGTGTGCTCG TAGCCGGGAG GCCTGA SEQ ID NO:95)</p>	
cysM	<p>cisteina sintasa B (O-acetilserina sulfhidrolasa B)</p>	<p>gtgAGTACAT TAGAACAAAC AATAGGCAAT ACGCCTCTGG TGAAGTTGCA GCGAATGGGG CCGGATAACG GCAGTGAAGT GTGGTTAAAA CTGGAAGGCA ATAACCCGGC AGGTTGCGGTG AAAGATCGTG CGGCACTTTC GATGATCGTC GAGGCGGAAA</p>	<p>MSTLEQTIGN TPLVKLQRMG PDNGSEVWLK LEGNNPAGSV KDRAALSMIV EAEKRGEIKP GDVLIEMATSG NTGIALAMIA ALKGYRMKLL MPDNMSQERR AAMRAYGAEL ILVTKEQGME GARDLALEMA NRGEGKLLDQ FNNPDNPYAH YTTTGPEIWQ</p>

FIG. 6 (Cont.)

	AGCGCGGGGA	QTGGRITHFV
	AATTAAACCG	SSMGTGTIT
	GGTGATGTCT	GVSRFMREQS
	TAATCGAAGC	KPVTIVGLQP
	CACCAGTGGT	EEGSSIPGIR
	AACACCGGCA	RWPTEYLPGI
	TTGCGCTGGC	FNASLVDEVL
	AATGATTGCC	DIHQDAENT
	GCGCTGAAAG	MRELAVREGI
	GCTATCGCAT	FCGVSSGGAV
	GAAATTGCTG	AGALRVAKAN
	ATGCCCGACA	PDAVVVAIIC
	ACATGAGCCA	DRGDRYLSTG
	GGAACGCCGT	VFGEEHFSQG
	GCGGCGATGC	AGI
	GTGCTTATGG	(SEQ ID
	TGCGGAACTG	NO: 98)
	ATTCTTGTC	
	CCAAAGAGCA	
	GGGCATGGAA	
	GGTGCGCGCG	
	ATCTGGCGCT	
	GGAGATGGCG	
	AATCGTGGCG	
	AAGGAAAGCT	
	GCTCGATCAG	
	TTCAATAATC	
	CCGATAACCC	
	TTATGCGCAT	
	TACACCACCA	
	CTGGGCCGGA	
	AATCTGGCAG	
	CAAACCGGCG	
	GGCGCATCAC	
	TCATTTTGTC	
	TCCAGCATGG	
	GGACGACCGG	
	CACTATCACC	
	GGCGTCTCAC	
	GCTTTATGCG	

FIG. 6 (Cont.)

		<p>CGAACAATCC AAACCGGTGA CCATTGTCGG CCTGCAACCG GAAGAGGGCA GCAGCATTCC CGGCATTTCG CGCTGGCCTA CGGAATATCT GCCGGGGATT TTCAACGCTT CTCTGGTGGA TGAGGTGCTG GATATTCATC AGCGCGATGC GGAAAACACC ATGCGCGAAC TGGCGGTGCG GGAAGGAATA TTCTGTGGCG TCAGCTCCGG CGGCGCGGTT GCCGGAGCAC TGCGGGTGGC AAAAGCTAAC CCTGACGCGG TGGTGGTGGC GATCATCTGC GATCGTGGCG ATCGCTACCT TTCTACCGGG GTGTTTGGGG AAGAGCATT TAGCCAGGGG GCGGGGATTT AA (SEQ ID NO:97)</p>	
maoC	Aldehído deshidrogenasa /enoil CoA hidratasa	atgCAGCAGT TAGCCAGTTT	MQQLASFLSG TWQSGRGRSR

FIG. 6 (Cont.)

condensadas	CTTATCCGGT ACCTGGCAGT CTGGCCGGGG CCGTAGCCGT TTGATTCACC ACGCTATTAG CGGCGAGGCG TTATGGGAAG TGACCAAGTGA AGGTCTTGAT ATGGCGGCTG CCCGCCAGTT TGCCATTGAA AAAGGTGCCC CCGCCCTTCG CGCTATGACC TTTATCGAAC GTGCGGCGAT GCTTAAAGCG GTCGCTAAAC ATCTGCTGAG TGAAAAAGAG CGTTTCTATG CTCTTTCTGC GCAAACAGGC GCAACGCGGG CAGACAGTTG GGTTGATATT GAAGGTGGCA TTGGGACGTT ATTTACTTAC GCCAGCCTCG GTAGCCGGGA GCTGCCTGAC GATACGCTGT GGCCGGAAGA TGAATTGATC CCCTTATCGA AGAAGGTGG ATTTGCCGCG	LIHHAISGEA LWEVTSEGLD MAAARQFAIE KGAPALRAMT FIERAAMLKA VAKHLLSEKE RFYALSAQTG ATRADSWVDI EGGIGTLFTY ASLGSRELPD DTLWPEDELI PLSKEGGFAA RHLLTSKSGV AVHINAFNFP CWGMLEKLAP TWLGGMPAII KPATATAQLT QAMVKSIVDS GLVPEGAI SL ICGSAGDLLD HLDSQDVVTF TGSAAATGQML RVQPNIVAKS IPFTMEADSL NCCVLGEDVT PDQPEFALFI REVVREMTTK AGQKCTAIRR IIVPQALVNA VSDALVARLQ KVVVGDPAQE GVKMGALVNA EQRADVQEKV NILLAAGCEI RLGGQADLSA AGAFFPPTLL YCPQPDETPA VHATEAFGPV ATLMPAQNR HALQLACAGG
-------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

FIG. 6 (Cont.)

CGCCATTTAC	GSLAGTLVTA
TGACCTCAAA	DPQIARQFIA
GTCAGGCGTG	DAARTHGRIQ
GCAGTGCATA	ILNEESAKES
TTAACGCCTT	TGHGSPLPQL
TAACTTCCCC	VHGGPGRAGG
TGCTGGGGAA	GEELGGLRAV
TGCTGGAAAA	KHYMQRTAVQ
GCTGGCACCA	GSPTMLAAIS
ACGTGGCTGG	KQWVRGAKVE
GCGGAATGCC	EDRIHPFRKY
AGCCATCATC	FEELQPGDSL
AAACCAGCTA	LTPRRTMTEA
CCGCGACGGC	DIVNFACLSG
CCAACTGACT	DHFYAHMDKI
CAGGCGATGG	AAAESIFGER
TGAAATCAAT	VVHGYFVLSA
TGTCGATAGT	AAGLFVDAGV
GGTCTTG TTC	GPVIANYGLE
CCGAAGGCGC	SLRFIEPVKP
AATTAGTCTG	GDTIQVRLTC
ATCTGCGGTA	KRKTLLKKQRS
GTGCTGGCGA	AEEKPTGVVE
CTTGTTGGAT	WAVEVFNQHQ
CATCTGGACA	TPVALYSILT
GCCAGGATGT	LVARQHGDV
GGTGACTTTC	D
ACGGGGTCAG	(SEQ ID
CGGCGACCGG	NO:100)
ACAGATGCTG	
CGAGTTCAGC	
CAAATATCGT	
CGCCAAATCT	
ATCCCCTTCA	
CTATGGAAGC	
TGATTCCCTG	
AACTGCTGCG	
TACTGGGCGA	
AGATGTCACC	
CCGGATCAAC	

FIG. 6 (Cont.)

CGGAGTTTGC
GCTGTTTATT
CGTGAAGTTG
TGCGTGAGAT
GACCACAAA
GCCGGGCAA
AATGTACGGC
AATCCGGCGG
ATTATTGTGC
CGCAGGCATT
GGTTAATGCT
GTCAGTGATG
CTCTGGTTGC
GCGATTACAG
AAAGTCGTGG
TCGGTGATCC
TGCTCAGGAA
GGCGTGAAA
TGGGCGCACT
GGTAAATGCT
GAGCAGCGTG
CCGATGTGCA
GGAAAAGTG
AACATATTGC
TGGCTGCAGG
ATGCGAGATT
CGCCTCGGTG
GTCAGGCGGA
TTTATCTGCT
GCGGGTGCCT
TCTTCCCGCC
AACCTTATTG
TACTGTCCGC
AGCCGGATGA
AACACCGGCG
GTACATGCAA
CAGAAGCCTT
TGGCCCTGTC
GCAACGCTGA
TGCCAGCACA

FIG. 6 (Cont.)

AAACCAGCGA
CATGCTCTGC
AACTGGCTTG
TGCAGGCGGC
GGTAGCCTTG
CGGGAACGCT
GGTGACGGCT
GATCCGCAA
TTGCGCGTCA
GTTTATTGCC
GACGCGGCAC
GTACGCATGG
GCGAATTCAG
ATCCTCAATG
AAGAGTCGGC
AAAAGAATCC
ACCGGGCATG
GCTCCCCACT
GCCACAACCTG
GTACATGGTG
GGCCTGGTCG
CGCAGGAGGC
GGTGAAGAAT
TAGGCGGTTT
ACGAGCGGTG
AAACATTACA
TGCAGCGAAC
CGCTGTTCAG
GGTAGTCCGA
CGATGCTTGC
CGCTATCAGT
AAACAGTGGG
TGC GCGGTGC
GAAAGTCGAA
GAAGATCGTA
TTCATCCGTT
CCGCAAATAT
TTTGAGGAGC
TACAACCAGG
CGACAGCCTG

FIG. 6 (Cont.)

TTGACTCCCC
GCCGCACAAT
GACAGAGGCC
GATATTGTTA
ACTTTGCTTG
CCTCAGCGGC
GATCATTCT
ATGCACATAT
GGATAAGATT
GCTGCTGCCG
AATCTATTTT
CGGTGAGCGG
GTGGTGCATG
GGTATTTTGT
GCTTTCTGCG
GCTGCGGGTC
TGTTTGTGCA
TGCCGGTGTC
GGTCCGGTCA
TTGCTAACTA
CGGGCTGGAA
AGCTTGCGTT
TTATCGAACC
CGTAAAGCCA
GGCGATACCA
TCCAGGTGCG
TCTCACCTGT
AAGCGCAAGA
CGCTGAAAAA
ACAGCGTAGC
GCAGAAGAAA
AACCAACAGG
TGTGGTGGAA
TGGGCTGTAG
AGGTATTCAA
TCAGCATCAA
ACCCCGGTGG
CGCTGTATTC
AATTCTGACG
CTGGTGGCCA

FIG. 6 (Cont.)

	GGCAGCACGG TGATTTTGTC GATTAA (SEQ ID NO:99)	
--	------------------------------------------------------	--

Fuente	Nº de Registro de Genbank
Shigella sp. D9	ZP_05432652
Citrobacter youngae ATCC 29220	ZP_04561391.1
Salmonella enterica	YP_001570967.1
Escherichia fergusonii ATCC 35469	YP_002382254.1
Klebsiella pneumoniae NTUH-K2044	YP_002918743.1
Enterobacter cancerogenus ATCC 35316	ZP_03281954.1
Cronobacter turicensis	CBA29728.1
Erwinia pyrifoliae Ep1/96	YP_002649242.1
Pectobacterium carotovorum subsp. carotovorum PC1	YP_003018119.1
Dickeya dadantii Ech703	YP_002987184.1
Edwardsiella ictaluri 93-146	YP_002932813.1
Providencia alcalifaciens DSM 30120	ZP_03317956.1
Yersinia kristensenii ATCC 33638	ZP_04624337.1
Photorhabdus asymbiotica	YP_003041580.1
Pantoea sp. At-9b	ZP_05728924.1
Actinobacillus succinogenes 130Z	YP_001344737.1
Mannheimia succiniciproducens MBEL55E	YP_088386.1
Pasteurella multocida subsp. multocida str. Pm70	NP_245421.1
Haemophilus somnus 129PT	YP_719117.1
Proteus mirabilis HI4320	YP_002150544.1
Sodalis glossinidius str. 'morsitans'	YP_454706.1
Candidatus Blochmannia pennsylvanicus str. BPEN	YP_277927.1
Aggregatibacter aphrophilus NJ8700	YP_003007342.1
Vibrio cholerae MZO-3	ZP_01958381.1
Baumannia cicadellinicola str. Hc (Homalodisca coagulata)	YP_588853.1

FIG. 6 (Cont.)

Vibrionales bacterium SWAT-3	ZP_01815187.1
Aliivibrio salmonicida LFI1238	YP_002262988.1
Aeromonas salmonicida subsp. salmonicida A449	YP_001141819.1
Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis	NP_871303.1
Glaciecola sp. HTCC2999	ZP_03560821.1
Alteromonas macleodii ATCC 27126	ZP_04714556.1

FIG. 7

Gen	Nombre	Secuencia de nucleótidos	Secuencia de Proteína
fabB	B-cetoacil sintasa/ 3-oxoacil-[proteína transportadora de acilo] sintasa	atgAAACGTG CAGTGATTAC TGGCCTGGGC ATTGTTTCCA GCATCGGTAA TAACCAGCAG GAAGTCCTGG CATCTCTGCG TGAAGGACGT TCAGGGATCA CTTTCTCTCA GGAGCTGAAG GATTCCGGCA TGCGTAGCCA CGTCTGGGGC AACGTAAAAC TGGATAACCAC TGGCCTCATT GACCGCAAAG TTGTGCGCTT TATGAGCGAC GCATCCATTT ATGCATTCTT TTCTATGGAG CAGGCAATCG CTGATGCGGG CCTCTCTCCG GAAGCTTACC AGAATAACCC GCGCGTTGGC CTGATTGCAG GTTCCGGCGG CGGCTCCCCG CGTTTCCAGG TGTTCCGCGC TGACGCAATG CGCGGCCCGC	MKRAVITGLG IVSSIGNNQQ EVLASLREGR SGITFSQELK DSGMRSHVWG NVKLDTTGLI DRKVVRFMSD ASIYAFLSME QAIADAGLSP EAYQNNPRVG LIAGSGGGSP RFQVFGADAM RGPRGLKAVG PYVVTKAMAS GVSACLATPF KIHGYNYSIS SACATSAHCI GNAVEQIQLG KQDIVFAGGG EELCWEMACE FDAMGALSTK YNDTPEKASR TYDAHRDGFV IAGGGGMVVV EELEHALARG AHIYAEIVGY GATSDGADMV APSGEGAVRC MKMAMHGVDT PIDYLNSHGT STPVGDVKEL AAIREVFGDK SPAISATKAM TGHSLGAAGV QEAIYSLML EHGFIAPSIN IEELDEQAAG

FIG. 7 (Cont.)

GCGGCCTGAA	LNIVTETTDR
AGCGGTTGGC	ELTTVMSNSF
CCGTATGTGG	GFGGTNATLV
TCACCAAAGC	MRCLKD (SEQ
GATGGCATCC	ID NO:102)
GGCGTTTCTG	
CCTGCCTCGC	
CACCCCGTTT	
AAAATTCATG	
GCGTTAACTA	
CTCCATCAGC	
TCCGCGTGTG	
CGACTTCCGC	
ACACTGTATC	
GGTAACGCAG	
TAGAGCAGAT	
CCAAC TGGGC	
AAACAGGACA	
TCGTGTTTGC	
TGGCGGCGGC	
GAAGAGCTGT	
GCTGGGAAAT	
GGCTTGCGAA	
TTCGACGCAA	
TGGGTGCGCT	
GTCTACTAAA	
TACAACGACA	
CCCCGGAAAA	
AGCCTCCCGT	
ACTTACGACG	
CTCACCGTGA	
CGGTTTCGTT	
ATCGCTGGCG	
GCGGCGGTAT	
GGTAGTGGTT	
GAAGAGCTGG	
AACACGCGCT	
GGCGCGTGGT	
GCTCACATCT	
ATGCTGAAAT	

FIG. 7 (Cont.)

CGTTGGCTAC
GCGCAACCT
CTGATGGTGC
AGACATGGTT
GCTCCGTCTG
GCGAAGGCGC
AGTACGCTGC
ATGAAGATGG
CGATGCATGG
CGTTGATAACC
CCAATCGATT
ACCTGAACTC
CCACGGTACT
TCGACTCCGG
TTGGCGACGT
GAAAGAGCTG
GCAGCTATCC
GTGAAGTGTT
CGGCGATAAG
AGCCCGGCGA
TTTCTGCAAC
CAAAGCCATG
ACCGGTCACT
CTCTGGGCGC
TGCTGGCGTA
CAGGAAGCTA
TCTACTCTCT
GCTGATGCTG
GAACACGGCT
TTATCGCCCC
GAGCATCAAC
ATTGAAGAGC
TGGACGAGCA
GGCTGCGGGT
CTGAACATCG
TGACCGAAAC
GACCGATCGC
GAACTGACCA
CCGTTATGTC
TAACAGCTTC

FIG. 7 (Cont.)

		GGCTTCGGCG GCACCAACGC CACGCTGGTA ATGCGCAAGC TGAAAGATTA A (SEQ ID NO:101)	
fabF	3-oxoacil-[proteína portadora de acilo] sintasa II	gtgTCTAAGC GTCGTGTAGT TGTGACCGGA CTGGGCATGT TGTCTCCTGT CGGCAATACC GTAGAGTCTA CCTGGAAAGC TCTGCTTGCC GGTCAGAGTG GCATCAGCCT AATCGACCAT TTCGATACTA GCGCCTATGC AACGAAATTT GCTGGCTTAG TAAAGGATTT TAACTGTGAG GACATTATCT CGCGCAAAGA ACAGCGCAAG ATGGATGCCT TCATTCAATA TGGAATTGTC GCTGGCGTTC AGGCCATGCA GGATTCTGGC CTTGAAATAA CGGAAGAGAA CGCAACCCGC ATTGGTGCCG CAATTGGCTC CGGGATTGGC	MSKRRVVVTG LGMLSPVGNT VESTWKALLA GQSGISLIDH FDTSAYATKF AGLVKDFNCE DIISRKEQRK MDAFIQYGIV AGVQAMQDSG LEITEENATR IGAAIGSGIG GLGLIEENHT SLMNGGPRKI SPFFVPSTIV NMVAGHLTIM YGLRGPSISI ATACTSGVHN IGHAARI IAY GDADVMVAGG AEKASTPLGV GGFGAARALS TRNDNPQAAS RPWDKERDGF VLGDGAGMLV LEEYEHAKKR GAKIYAELVG FGMSSDAYHM TSPPEAGAGA ALAMANALRD AGIEASQIGY VNAHGTSTPA GDKAEQAVK TIFGEAASRV

FIG. 7 (Cont.)

	GGCCTCGGAC TGATCGAAGA AAACCACACA TCTCTGATGA ACGGTGGTCC ACGTAAGATC AGCCCATTCT TCGTTCCGTC AACGATTGTG AACATGGTGG CAGGTCATCT GACTATCATG TATGGCCTGC GTGGCCCGAG CATCTCTATC GCGACTGCCT GTACTIONCCGG CGTGCACAAC ATTGGCCATG CTGCGCGTAT TATCGCGTAT GGCGATGCTG ACGTGATGGT TGCAGGTGGC GCAGAGAAAG CCAGTACGCC GCTGGGCGTT GGTGGTTTTG GCGCGGCACG TGCATTATCT ACCCGCAATG ATAACCCGCA AGCGGCGAGC CGCCCGTGGG ATAAAGAGCG TGATGGTTTC GTACTGGGCG ATGGTGCCGG TATGCTGGTA CTTGAAGAGT	LVSSTKSMTG HLLGAAGAVE SIYSILALRD QAVPPTINLD NPDEGCDLDF VPHEARQVSG MEYTL CNSFG FGGTNGSLIF KKI (SEQ ID NO:104)
--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------

FIG. 7 (Cont.)

ACGAACACGC
GAAAAAACGC
GGTGCGAAAA
TTTACGCTGA
ACTCGTCGGC
TTTGGTATGA
GCAGCGATGC
TTATCATATG
ACGTCACCGC
CAGAAAATGG
CGCAGGCGCA
GCTCTGGCGA
TGGCAAATGC
TCTGCGTGAT
GCAGGCATTG
AAGCGAGTCA
GATTGGCTAC
GTTAACGCGC
ACGGTACTTC
TACGCCGGCT
GGCGATAAAG
CTGAAGCGCA
GGCGGTGAAA
ACCATCTTCG
GTGAAGCTGC
AAGCCGTGTG
TTGGTAAGCT
CCACGAAATC
TATGACCGGT
CACCTGTTAG
GTGCGGCGGG
TGCAGTAGAA
TCTATCTACT
CCATCCTGGC
GCTGCGCGAT
CAGGCTGTTC
CGCCAACCAT
CAACCTGGAT
AACCCGGATG
AAGGTTGCGA

FIG. 7 (Cont.)

		TCTGGATTTC GTACCGCACG AAGCGCGTCA GGTTAGCGGA ATGGAATACA CTCTGTGTAA CTCCTTCGGC TTCGGTGGCA CTAATGGTTC TTTGATCTTT AAAAAGATCT AA (SEQ ID NO:103)	
fadJ	enoil-CoA hidratasa y epimerasa e isomerasa/3- hidroxiacil-CoA deshidrogenasa condensadas	atgGAAATGA CATCAGCGTT TACCCTTAAT GTTTCGTCTGG ACAACATTGC CGTTATCACC ATCGACGTAC CGGGTGAGAA AATGAATACC CTGAAGGCGG AGTTTGCCTC GCAGGTGCGC GCCATTATTA AGCAACTCCG TGAAAACAAA GAGTTGCGAG GCGTGGTGTT TGTCTCCGCT AAACCGGACA ACTTCATTGC TGGCGCAGAC ATCAACATGA TCGGCAACTG CAAAACGGCG CAAGAAGCGG AAGCTCTGGC GCGGCAGGGC	MEMTSAFTLN VRLDNIAVIT IDVPGEKMNT LKAEFASQVR AIIKQLRENK ELRGVVFVSA KPDNFIAGAD INMIGNCKTA QEAEALARQG QQLMAEIHAI PIQVIAAIHG ACLGGGLELA LACHGRVCTD DPKTVLGLPE VQLGLLPGSG GTQRLPRLIG VSTALEMILT GKQLRAKQAL KLGLVDDVVP HSILLEAAVE LAKKERPSSR PLPVRERILA GPLGRALLFK MVGKKTEHKT QGNYPATERI LEVVETGLAQ GTSSGYDAEA

FIG. 7 (Cont.)

CAACAGTTGA	RAFGELAMTP
TGGCGGAGAT	QSQUALRSIFF
TCATGCTTTG	ASTDVKKDPG
CCCATTCAGG	SDAPPAPLNS
TTATCGCGGC	VGILGGGLMG
TATTCATGGC	GGIAYVTACK
GCTTGCCTGG	AGIPVRIKDI
GTGGTGGGCT	NPQGINHALK
GGAGTTGGCG	YSWDQLEGKV
CTGGCGTGCC	RRRHLKASER
ACGGTCGCGT	DKQLALISGT
TTGTACTGAC	TDYRGFAHRD
GATCCTAAAA	LIIEAVFENL
CGGTGCTCGG	ELKQQMVAEV
TTTGCCTGAA	EQNCAHTIF
GTACAACCTG	ASNTSSLPIG
GATTGTTACC	DIAAHATRPE
CGGTTCAGGC	QVIGLHFFSP
GGCACCCAGC	VEKMPLVEII
GTTTACCGCG	PHAGTSAQTI
TCTGATAGGC	ATTVKLAKKQ
GTCAGCACAG	GKTPIVVRDK
CATTAGAGAT	AGFYVNRILA
GATCCTCACC	PYINEAIRML
GGAAAACAAC	TQGERVEHID
TTCGGGCGAA	AALVKFGFPV
ACAGGCATTA	GPIQLLDEVG
AAGCTGGGGC	IDTGTKIIPV
TGGTGGATGA	LEAAYGERFS
CGTTGTTCCG	APANVVSSIL
CACTCCATTC	NDDRKGRKNG
TGCTGGAAGC	RGFYLYGQKG
CGCTGTTGAG	RKSKKQVDPA
CTGGCAAAGA	IYPLIGTQGQ
AGGAGCGCCC	GRISAPQVAE
ATCTTCCC GC	RCVMLMLNEA
CCTCTACCTG	VRCVDEQVIR
TACGCGAGCG	SVRDGDIGAV
TATTCTGGCG	FGIGFPPFLG
GGGCCGTTAG	GPFYIDSLG

FIG. 7 (Cont.)

GTCGTGCGCT	AGEVVAIMQR
GCTGTTCAAA	LATQYGSRFT
ATGGTCGGCA	PCERLVEMGA
AGAAAACAGA	RGESFWKTTA
ACACAAAAC	TDLQ
CAAGGCAATT	(SEQ ID
ATCCGGCGAC	NO:106)
AGAACGCATC	
CTGGAGGTTG	
TTGAAACGGG	
ATTAGCGCAG	
GGCACCAGCA	
GCGGTTATGA	
CGCCGAAGCT	
CGGGCGTTTG	
GCGAACTGGC	
GATGACGCCA	
CAATCGCAGG	
CGCTGCGTAG	
TATCTTTTTT	
GCCAGTACGG	
ACGTGAAGAA	
AGATCCCGGC	
AGTGATGCGC	
CGCCTGCGCC	
ATTAAACAGC	
GTGGGGATTT	
TAGGTGGTGG	
CTTGATGGGC	
GGCGGTATTG	
CTTATGTCAC	
TGCTTGTAAG	
GCGGGGATTC	
CGGTCAGAA	
TAAAGATATC	
AACCCGCAGG	
GCATAAATCA	
TGCGCTGAAG	
TACAGTTGGG	
ATCAGCTGGA	

FIG. 7 (Cont.)

GGGCAAAGTT
CGCCGTCGTC
ATCTCAAAGC
CAGCGAACGT
GACAAACAGC
TGGCATTAAAT
CTCCGGAACG
ACGGACTATC
GCGGCTTTGC
CCATCGCGAT
CTGATTATTG
AAGCGGTGTT
TGAAAATCTC
GAATTGAAAC
AACAGATGGT
GGCGGAAGTT
GAGCAAATT
GCGCCGCTCA
TACCATCTTT
GCTTCGAATA
CGTCATCTTT
ACCGATTGGT
GATATCGCCG
CTCACGCCAC
GCGACCTGAG
CAAGTTATCG
GCCTGCATTT
CTTCAGTCCG
GTGGAAAAAA
TGCCGCTGGT
GGAGATTATT
CCTCATGCGG
GGACATCGGC
GCAAACCATC
GCTACCACAG
TAAAACTGGC
GAAAAACAG
GGTAAAACGC
CAATTGTCGT
GCGTGACAAA

FIG. 7 (Cont.)

GCCGGTTTTT
ACGTCAATCG
CATCTTAGCG
CCTTACATTA
ATGAAGCTAT
CCGCATGTTG
ACCCAAGGTG
AACGGGTAGA
GCACATTGAT
GCCGCGCTAG
TGAAATTTGG
TTTTCCGGTA
GGCCAATCC
AACTTTTGGA
TGAGGTAGGA
ATCGACACCG
GACTAAAAT
TATTCCTGTA
CTGGAAGCCG
CTTATGGAGA
ACGTTTTAGC
GCGCCTGCAA
ATGTTGTTTC
TTCAATTTTG
AACGACGATC
GCAAAGGCAG
AAAAAATGTC
CGGGGTTTCT
ATCTTTATGG
TCAGAAAGGG
CGTAAAAGCA
AAAAACAGGT
CGATCCCGCC
ATTTACCCGC
TGATTGGCAC
ACAAGGCAG
GGCGAATCT
CCGCACCGCA
GGTTGCTGAA
CGGTGTGTGA

FIG. 7 (Cont.)

		<p>TGTTGATGCT GAATGAAGCA GTACGTTGTG TTGATGAGCA GGTTATCCGT AGCGTGCGTG ACGGGGATAT TGGCGCGGTA TTTGGCATTG GTTTTCCGCC ATTTCTCGGT GGACCGTTCC GCTATATCGA TTCTCTCGGC GCGGGCGAAG TGGTTGCAAT AATGCAACGA CTTGCCACGC AGTATGGTTC CCGTTTTACC CCTTGCGAGC GTTTGGTCGA GATGGGCGCG CGTGGGGAAA GTTTTTGGA AACAACTGCA ACTGACCTGC AATAA (SEQ ID NO:105)</p>	
xerC	tirosina recombinasa específica del sitio	<p>atgACCGATT TACACACCGA TGTAGAACGC TACCTACGTT ATCTGAGCGT GGAGCGCCAG CTTAGCCCGA TAACCCTGCT TAACTACCAG CGTCAGCTTG AGGCGATCAT</p>	<p>MTDLHTDVER YLRYLSVERQ LSPITLLNYQ RQLEAIINFA SENGLQSWQQ CDVTMVRNFA VRSRRKGLGA ASLALRLSAL RSFFDWLVSQ NELKANPAKG VSAPKAPRHL</p>

FIG. 7 (Cont.)

CAATTTTGCC	PKNIDVDDMN
AGCGAAAACG	RLLDIDINDP
GCCTGCAAAG	LAVRDRAMLE
CTGGCAGCAA	VMYGAGLRLS
TGTGATGTGA	ELVGLDIKHL
CGATGGTGCG	DLESGEVWVM
CAATTTTGCT	GKGSKERRLP
GTACGCAGTC	IGRNAVAWIE
GCCGTAAAGG	HWLDLRDLFG
GCTGGGAGCA	SEDDALFLSK
GCAAGTCTGG	LGKRISARNV
CGTTACGGCT	QKRFAEWGIK
TTCTGCGCTA	QGLNNHVHPH
CGTAGCTTTT	KLRHSFATHM
TTGACTGGCT	LESSGDLRGV
GGTCAGCCAG	QELLGHANLS
AACGAACTCA	TTQIYTHLDF
AAGCTAACCC	QHLASVYDAA
GGCGAAAGGT	HPRAKR GK
GTTTCGGCAC	(SEQ ID
CGAAAGCGCC	NO:108)
GCGTCATCTG	
CCGAAAAACA	
TCGACGTCGA	
CGATATGAAT	
CGGCTGCTGG	
ATATTGATAT	
CAATGATCCC	
CTCGCTGTAC	
GCGACCGTGC	
AATGCTGGAA	
GTGATGTACG	
GCGCGGGTCT	
GCGTCTTTCT	
GAGCTGGTGG	
GGCTGGATAT	
TAAACACCTC	
GACCTGGAGT	
CTGGTGAAGT	
GTGGGTTATG	

FIG. 7 (Cont.)

GGGAAAGGCA GCAAAGAGCG CCGCCTGCCG ATTGGTCGCA ACGCTGTGGC GTGGATTGAG CACTGGCTTG ATTTGCGCGA CCTGTTTGGT AGCGAAGACG ACGCGCTTTT TCTGTCGAAA CTGGGCAAGC GTATCTCCGC GCGTAATGTG CAGAAACGCT TTGCCGAATG GGGCATAAAA CAAGGGCTGA ATAATCACGT TCATCCGCAT AAATTACGTC ACTCGTTCGC CACGCATATG CTGGAGTCGA GCGGCGATCT TCGTGGTGTG CAGGAGCTGC TGGGTCATGC CAACCTCTCC ACCACGCAA TCTATACTCA TCTTGATTTT CAACACCTTG CCTCGGTGTA CGATGCGGCG CATCCACGCG CCAAACGGGG GAAATAA (SEQ ID NO:107)

FIG. 7 (Cont.)

yqeF	aciltransferasa predicha	atgAAAGACG TTGTGATTGT CGGGGCGTTA CGGACACCTA TCGGCTGCTT TCGTGGTGCG TTAGCGGGTC ATTCCGCCGT GGAACTTGGT AGTCTGGTCG TGAAAGCGTT AATAGAACGT ACCGGCGTTC CTGCATATGC GGTGGATGAA GTAATTCTTG GTCAGGTGTT GACTGCAGGG GCAGGGCAGA ATCCGGCAAG GCAATCGGCT ATTAAAGGTG GTCTGCCTAA TAGCGTTTCT GCAATCACTA TTAATGACGT TTGCGGTTCC GGGCTTAAAG CACTGCATCT GGCTACTCAG GCGATACAGT GTGGCGAGGC TGATATTGTC ATCGCCGGTG GCCAGGAAAA CATGAGCCGC GCACCACATG TTCTGACTGA TAGCCGCACC GGTGCACAGC	MKDVVIVGAL RTPIGCFRGA LAGHSAVELG SLVVKALIER TGVPAYAVDE VILGQVLTAG AGQNPARGSA IKGGLPNSVS AITINDVCGS GLKALHLATQ AIQCGEADIV IAGGQENMSR APHVLTDSRT GAQLGNSQLV DSLVDHGLWD AFNDYHIGVT AENLAREYGI SRQLQDAYAL SSQQKARAAI DAGRFKDEIV PVMTQSNQOT LVVDTDEQPR TDASAEGLAR LNPSFDSLGS VTAGNASSIN DGAAAVMMMS EAKARALNLP VLARIRAFAS VGVDPALMGI APVYATRCL ERVGWQLAEV DLIEANEFA AQALSVGKML EWDERRVNVN GGAIALGHPI GASGRILVS LVHEMVKRNA RKGLATLCIG GGQGVALTIE RDE
-------------	-------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

FIG. 7 (Cont.)

	TTGGCAATAG CCAGTTGGTT GACAGTCTTG TGCATGATGG GTTGTGGGAT GCCTTCAATG ATTATCATAT TGGTGTCCACC GCCGAAAATC TGGCTCGCGA ATATGGCATC AGCCGTCAGT TGCAGGATGC TTACGCACTT AGCTCGCAAC AAAAAGCGCG AGCGGCGATT GACGCCGGAC GATTTAAAGA TGAGATCGTC CCGGTAATGA CCCAAAGTAA CGGGCAGACG TTGGTTGTTG ATACCGATGA ACAGCCACGC ACTGACGCCA GCCCAGAAGG CTTAGCCCGT TTAAATCCTT CATTGATAG TCTCGGTTCT GTGACAGCGG GTAATGCATC ATCCATAAAC GATGGCGCAG CTGCGGTAAT GATGATGAGC GAAGCCAAAG CACGAGCGTT	(SEQ ID NO:110)
--	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------

FIG. 7 (Cont.)

GAATTTACCC
GTGCTGGCCC
GCATTCGCGC
ATTTGCCAGC
GTTGGTGTAG
ATCCGGCATT
GATGGGAATT
GCGCCGGTGT
ATGCGACCCG
CCGTTGCCTG
GAGCGTGTAG
GCTGGCAGTT
GGCTGAAGTC
GATCTTATCG
AGGCTAATGA
AGCGTTTGCT
GCACAGGCGC
TTTCGGTTGG
CAAGATGCTT
GAGTGGGATG
AGCGTCGGGT
CAATGTCAAT
GGTGGCGCGA
TCGCACTCGG
TCACCCGATA
GGCGCTTCCG
GTTGCCGAAT
CCTGGTTTCT
CTGGTTCATG
AAATGGTGAA
ACGTAATGCC
CGCAAAGGAC
TGGCAACGCT
TTGTATCGGC
GGGGGCCAGG
GTGTGGCATT
GACCATTGAA
CGTGACGAAT
AG
(SEQ ID NO:109)

FIG. 7 (Cont.)

murQ	Componente de PTS predicho		
		atgCAATTTG	MQFEKMITEG
		AAAAGATGAT	SNTASAEIDR
		TACTGAAGGC	VSTLEMCRII
		TCGAACACCG	NDEDKTVPLA
		CCTCGGCTGA	VERVLPDIAA
		AATTGACCGC	AIDVIHAQVS
		GTATCGACGC	GGRLIYLGA
		TGGAAATGTG	GTSGRLGILD
		CCGGATTATC	ASECPPTYGV
		AACGATGAAG	KPGLVVGLIA
		ATAAAACCGT	GGEYAIQHAV
		ACCGCTTGCC	EGAEDSREGG
		GTTGAGCGCG	VNDLKNINLT
		TACTGCCGGA	AQDVVVGIAA
		TATCGCCGCG	SGRTPYVIAG
		GCGATCGATG	LEYARQLGCR
		TTATCCACGC	TVGISCNPGS
		CCAGGTAAGC	AVSTTAEFAI
		GGCGGCGGGC	TPIVGAEVVT
		GTCTGATTTA	GSSRMKAGTA
		CCTCGGTGCG	QKLVLNMLST
		GGAACATCCG	GLMIKSGKVF
		GTCGTCTGGG	GnlmVDVVAT
		GATTCTGGAT	NEKLHVRQVN
		GCCAGCGAAT	IVKNATGCSA
		GTCCGCCAC	EQAEAALIAc
		CTACGGCGTG	ERNCKTAIVM
		AAACCGGGTC	VLKNLDAAEA
		TGGTGGTTGG	KKRLDQHGGF
		TTTGATTGCT	IRQVLDKE
		GGCGGCGAAT	(SEQ ID
		ATGCCATTCA	NO:112)
		GCACGCGGTG	
		GAAGGCGCGG	
		AAGATAGCCG	
		GGAAGGCGGT	
		GTTAATGATC	
		TGAAAAATAT	
		TAATTTAACG	
		GCACAGGATG	

FIG. 7 (Cont.)

TGGTGGTTGG
CATTGCTGCC
AGCGGTCGCA
CGCCGTATGT
GATTGCCGGA
CTGGAATACG
CACGCCAGCT
CGGCTGCCGC
ACAGTGGGAA
TTTCCTGTAA
TCCGGGGAGC
GCCGTTTCAA
CCACCGCTGA
GTTTGCCATT
ACACCGATTG
TAGGTGCCGA
AGTTGTTACC
GGTTCCTCGC
GGATGAAAGC
AGGTACAGCG
CAGAAACTGG
TGCTCAATAT
GCTTTCACC
GGGCTGATGA
TTAAATCCGG
CAAAGTG TTC
GGCAACCTGA
TGGTCGATGT
GGTCGCCACC
AACGAAAAAC
TGCATGTGCG
ACAGGTCAAT
ATTGTTAAAA
ACGCCACCGG
ATGTAGCGCA
GAGCAAGCGG
AAGCGGCGTT
AATTGCTTGC
GAGCGCAACT
GTAAAACGGC

169/170
 FIG. 7 (Cont.)

		CATTGTGATG GTGCTGAAAA ATCTCGATGC CGCAGAAGCT AAAAAACGCC TGGATCAACA CGGCGGCTTT ATTCGTCAGG TTTTAGACAA GGAATAA (SEQ ID NO:111)	
--	--	----------------------------------------------------------------------------------------------------------------------------------------------------------	--

Fuente	Nº de registro Genbank
Shigella boydii CDC 3083-94	YP_001881145.1
Escherichia fergusonii ATCC 35469	YP_002382013.1
Salmonella enterica subsp. arizonae	YP_001569590.1
Citrobacter sp. 30_2	ZP_04562837.1
Klebsiella pneumoniae subsp. pneumoniae MGH 78578	YP_001336360.1
Pectobacterium carotovorum subsp. carotovorum WPP14	ZP_03831287.1
Enterobacter cancerogenus ATCC 35316	ZP_03283474.1
Pantoea sp. At-9b	ZP_05730617.1
Cronobacter turicensis	CBA32510.1
Dickeya dadantii Ech586	ZP_05723897.1
Erwinia tasmaniensis Et1/99	YP_001907100.1
Serratia proteamaculans 568	YP_001479594.1
Edwardsiella ictaluri 93-146	YP_002934130.1
Sodalis glossinidius str. 'morsitans'	YP_455303.1
Yersinia aldovae ATCC 35236	ZP_04620215.1
Providencia stuartii ATCC 25827	ZP_02961167.1
Photorhabdus asymbiotica	YP_003040275.1
Proteus mirabilis HI4320	YP_002151524.1

FIG. 7 (Cont.)

Candidatus Blochmannia pennsylvanicus str. BPEN	YP_278005.1
Glaciecola sp. HTCC2999	ZP_03561088.1
Vibrio cholerae V51	ZP_04919940.1
Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis	NP_871411.1
Tolumonas auensis DSM 9187	YP_002892770.1
Actinobacillus pleuropneumoniae serovar 1 str. 4074	ZP_00134992.2
Aggregatibacter aphrophilus NJ8700	YP_003007711.1
Pseudoalteromonas tunicata D2	ZP_01135065.1
Vibrionales bacterium SWAT-3	ZP_01816638.1
Pasteurella multocida subsp. multocida str. Pm70	NP_245276.1
Mannheimia succiniciproducens MBEL55E	YP_088783.1
Haemophilus somnus 129PT	YP_718877.1
Shewanella loihica PV-4	YP_001094535.1
Aliivibrio salmonicida LFI1238	YP_002262558.1