

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 003 877**

51 Int. Cl.:

A01H 5/00 (2008.01)

C12N 15/82 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **12.01.2017** **PCT/GB2017/050068**

87 Fecha y número de publicación internacional: **20.07.2017** **WO17122011**

96 Fecha de presentación y número de la solicitud europea: **12.01.2017** **E 17701911 (4)**

97 Fecha y número de publicación de la concesión europea: **11.12.2024** **EP 3358941**

54 Título: **Planta modificada**

30 Prioridad:

13.01.2016 GB 201600631

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
11.03.2025

73 Titular/es:

**SUN PHARMACEUTICAL INDUSTRIES
AUSTRALIA LIMITED (100.00%)
1 Acacia Place
Notting Hill, Victoria 3168, AU**

72 Inventor/es:

**WINZER, THILO;
WALKER, TRACY CAROL;
MEADE, FERGUS;
LARSON, TONY ROBERT y
GRAHAM, IAN ALEXANDER**

74 Agente/Representante:

FERNÁNDEZ POU, Felipe

Observaciones:

Véase nota informativa (Remarks, Remarques o
Bemerkungen) en el folleto original publicado por la
Oficina Europea de Patentes

ES 3 003 877 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

Planta modificada

5 Campo de la invención

La descripción se refiere a plantas modificadas del género *Papaver* que comprenden uno o más genes que codifican una 3-O-desmetilasa de codeína (CODM) en donde las plantas modificadas tienen un contenido de alcaloide modificado cuando se comparan con plantas no modificadas en donde el contenido de alcaloide tiene un fenotipo de alta codeína; e incluyen el uso de ácidos nucleicos que codifican CODM y variantes de CODM para su uso como biomarcadores en la mejora de plantas.

Antecedentes de la invención

15 La variación genética en una población de plantas naturales ofrece una rica fuente de rasgos específicos; sin embargo, los métodos tradicionales de mejoramiento de plantas mediante la selección de un fenotipo específico son lentos y requieren mucho trabajo. La mutagénesis del germoplasma a través de mutágenos químicos o radiación crea una amplia gama de alelos modificados o mutantes de delección con fenotipo alterado, a menudo inalcanzable cuando se usan técnicas tradicionales de mejoramiento de plantas. Los métodos tales como la inducción de lesiones locales en genomas (TILLING), permiten la detección de cambios de una sola base después de la mutagénesis química y se han convertido en una herramienta importante para el mejoramiento de plantas. El TILLING, en combinación con el mejoramiento asistido por marcadores, ha dado como resultado la identificación de nuevas variedades de plantas con características deseadas. TILLING es un método genético inverso de alto rendimiento para obtener series alélicas de una población sometida a mutagénesis química. Una diana seleccionada se amplifica a partir del ADN agrupado mediante el uso de cebadores de PCR marcados. Después de la amplificación, los productos de PCR se desnaturalizan y vuelven a hibridarse. Si una mutación está presente en el ADN agrupado, se formará un heterodúplex. Después, se usa una nucleasa específica para escindir una hebra del heterodúplex y los productos se someten a electroforesis en un gel de acrilamida desnaturalizante. Las mutaciones se detectan mediante la observación de bandas escindidas.

30 *Papaver somniferum*, la amapola, es una fuente de alcaloides opiáceos clínicamente útiles tales como morfina, codeína, tebaína, noscapina y papaverina. Los alcaloides opiáceos tienen una amplia aplicación clínica que abarca desde analgésicos hasta supresores de la tos y antiespasmódicos. Si bien los alcaloides opiáceos pueden extraerse del látex cosechado de las vainas verdes de la amapola o de la paja de amapola que es la planta madura seca, la demanda de alcaloides opiáceos excede el suministro natural disponible, lo que requiere una costosa síntesis química. Esto se aplica particularmente a la codeína que se obtiene principalmente de la morfina a través de un proceso semisintético, ya que los niveles de codeína encontrados en las plantas representan típicamente menos del 5 % del nivel de morfina; sin embargo, la codeína ha sido el opiáceo más consumido en el mundo alcanzando 361 toneladas en 2013.

40 La codeína puede extraerse directamente de la planta y también puede obtenerse a través de un proceso semisintético mediante la metilación de la morfina. En la planta de amapola se sintetizan codeína y morfina a partir de (*R*)-reticulina a través de la tebaína intermedia, lo que implica varias etapas enzimáticas. La tebaína se convierte después en oripavina por la 3-O-desmetilasa de codeína (CODM), que después se transforma subsecuentemente en varias etapas en morfina, o la tebaína se transforma en codeína, que, a través de la actividad de la CODM, se convierte en morfina por desmetilación.

50 Se ha informado un aumento del contenido de codeína en la amapola después de la silenciamiento génico inducido por virus dirigido a la expresión de CODM. De manera similar, para aumentar el contenido de codeína en la amapola, las plantas se han sometido a mutagénesis y se han cribado para detectar alta codeína como se describe en el documento US8541647. La mutagénesis dio como resultado plantas de *Papaver somniferum* con alta codeína que contienen poca o ninguna oripavina, morfina o tebaína. Sin embargo, se desconoce el genotipo de estas plantas. Aunque la mutagénesis química es una herramienta importante para alterar rasgos específicos, es un proceso de cribado laborioso y que consume mucho tiempo. Sería conveniente identificar marcadores génicos que permitan la selección asistida por marcadores de plantas con un fenotipo requerido.

La presente descripción se refiere a la delección de una o más copias de la CODM que da como resultado plantas con alta codeína en la paja de amapola y niveles indetectables de morfina.

60 Descripción de la invención

De acuerdo con un aspecto de la invención, se proporciona una especie de planta de *Papaver* en donde la planta se modifica mediante la delección de la totalidad o parte de tres genes enlazados que codifican la 3-O-desmetilasa de codeína que comprende la secuencia de nucleótidos establecida en la SEQ ID NO: 7, o una secuencia de nucleótidos que es 99 % idéntica a la SEQ ID NO: 7, caracterizada porque la actividad de dicha 3-O-desmetilasa de codeína es indetectable y en donde la planta tiene un contenido de codeína de al menos 45 % del contenido total de alcaloide de

látex o paja seca extraído de la planta modificada, cuando se compara con una planta de *Papaver* de tipo silvestre de la misma especie.

5 En una modalidad preferida de la invención, dicha expresión reducida es el resultado de la delección de la totalidad o parte de un gen de 3-O-desmetilasa de codeína codificado por una molécula de ácido nucleico que comprende la secuencia de nucleótidos establecida en la SEQ ID NO: 1.

10 En una modalidad preferida de la invención, dicha expresión reducida es el resultado de la delección de la totalidad o parte de un gen de 3-O-desmetilasa de codeína codificado por una molécula de ácido nucleico que comprende la secuencia de nucleótidos establecida en la SEQ ID NO: 2.

15 En una modalidad preferida de la invención, dicha expresión reducida es el resultado de la delección de la totalidad o parte de un gen de 3-O-desmetilasa de codeína codificado por una molécula de ácido nucleico que comprende la secuencia de nucleótidos establecida en la SEQ ID NO: 3.

En una modalidad preferida de la invención, dicha expresión reducida es el resultado de la delección de la totalidad o parte de un gen de 3-O-desmetilasa de codeína codificado por una molécula de ácido nucleico que comprende la secuencia de nucleótidos establecida en la SEQ ID NO: 4.

20 En una modalidad preferida de la invención, dicha expresión reducida es el resultado de la delección de la totalidad o parte de un gen de 3-O-desmetilasa de codeína codificado por una molécula de ácido nucleico que comprende la secuencia de nucleótidos establecida en la SEQ ID NO: 5.

25 En una modalidad preferida de la invención, dicha expresión reducida es el resultado de la delección de la totalidad o parte de un gen de 3-O-desmetilasa de codeína codificado por una molécula de ácido nucleico que comprende la secuencia de nucleótidos establecida en la SEQ ID NO: 6.

30 En una modalidad preferida de la invención, dicha expresión reducida es el resultado de la delección de la totalidad o parte de cada gen de 3-O-desmetilasa de codeína codificado por una molécula de ácido nucleico que comprende la secuencia de nucleótidos establecida en la SEQ ID NO: 1, 2, 3, 4, 5, 6 y 7, de manera que dicha planta de *Papaver* no tiene actividad enzimática de 3-O-desmetilasa de codeína.

35 En una modalidad preferida de la invención, dicha planta se modifica mediante la delección de una molécula de ácido nucleico que comprende o que consiste de la secuencia de nucleótidos como se establece en la SEQ ID NO: 44.

Preferentemente, la planta se modifica mediante la delección de una molécula de ácido nucleico que comprende o que consiste de una secuencia de nucleótidos como se establece en la SEQ ID NO: 44.

40 En una modalidad preferida de la invención, dicha especie de *Papaver* se selecciona del grupo que consiste en: *Papaver somniferum*, *P. setigerum*, *P. orientale*, *P. pseudo-orientale*, *P. lasiothrix*, *P. cylindricum*, *P. fugax*, *P. triniifolium*.

En una modalidad preferida de la invención, dicha especie de *Papaver* es *Papaver somniferum*.

45 El "contenido total de alcaloide" se define con relación a los alcaloides principales: morfina, codeína, noscapina, oripavina y tebaína. La paja seca se refiere a los tallos, vástagos y hojas secos de las amapolas menos las semillas.

50 En una modalidad preferida de la invención, el contenido de codeína es al menos 50 %, 55 %, 60 %, 65 %, 70 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 % o 98 % del contenido total de alcaloide de látex o paja seca extraído de dicha planta modificada.

En una modalidad alternativa de la invención, el contenido de codeína es al menos 0,8 % de la paja seca de dicha planta modificada.

55 Preferentemente, el contenido de codeína está entre 0,8 % y 10,0 % de la paja seca; o entre 0,8 % y 3,0 % de la paja seca o entre 1,0 y 1,5 % de la paja seca. Preferentemente, el contenido de codeína es aproximadamente 1,1 % de la paja seca.

60 No cubierta por las reivindicaciones, se proporciona una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos establecida en la SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13.

Preferentemente, dicha molécula de ácido nucleico aislada comprende o consiste en una secuencia de nucleótidos como se establece en la SEQ ID NO: 44, o un fragmento de esta, en donde dicho fragmento comprende uno o más genes de 3-O-desmetilasa de codeína. Preferentemente, dicha molécula de ácido nucleico es ADN genómico.

65 Preferentemente, la molécula de ácido nucleico de la invención es ADNc.

Preferentemente, dicha molécula de ácido nucleico comprende una secuencia de nucleótidos en donde dicha secuencia se degenera como resultado del código genético a la secuencia de nucleótidos establecida en la SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13.

5 Preferentemente, dicha molécula de ácido nucleico comprende o consiste en una secuencia de nucleótidos en donde dicha secuencia se degenera como resultado del código genético a la secuencia de nucleótidos como se establece en la SEQ ID NO: 44.

10 Preferentemente, dicha molécula de ácido nucleico codifica un polipéptido de 3-O-desmetilasa de codeína como se representa en la SEQ ID NO: 15, 16 o 17.

No cubierto por las reivindicaciones, se proporciona un vector de expresión que comprende una molécula de ácido nucleico de acuerdo con la invención, adaptada para la expresión en una célula huésped microbiana.

15 No cubierto por las reivindicaciones, se proporciona un vector que comprende una molécula de ácido nucleico o fragmento de esta, que comprende una secuencia de nucleótidos como se establece en la SEQ ID NO: 44, o una variante de secuencia polimórfica de esta, en donde dicho fragmento comprende uno o más genes de 3-O-desmetilasa de codeína

20 No cubierta por las reivindicaciones, se proporciona una célula microbiana transformada con una molécula o vector de ácido nucleico de acuerdo con la invención.

Preferentemente, dicha célula microbiana es una célula bacteriana o una célula fúngica.

25 No cubierto por las reivindicaciones, se proporciona un cultivo celular microbiano que comprende una célula microbiana de acuerdo con la invención.

No cubierto por las reivindicaciones, se proporciona un fermentador que comprende un cultivo celular microbiano de acuerdo con la invención.

30 No cubierto por las reivindicaciones, se proporciona un proceso para la modificación de la codeína que comprende:

- i) proporcionar un cultivo celular microbiano transgénico de acuerdo con la invención, que expresa una molécula de ácido nucleico que codifica una 3-O-desmetilasa de codeína en cultivo;
- 35 ii) cultivar la célula microbiana en condiciones que desmetilan la codeína a morfina; y opcionalmente
- iii) aislar la morfina de la célula o cultivo celular microbiano.

Si las células microbianas se usan como organismos en el proceso de acuerdo con la invención, se cultivan o se cultivan de la manera con la que el experto está familiarizado, en dependencia del organismo huésped. Por regla general, los microorganismos se cultivan en un medio líquido que comprende una fuente de carbono, habitualmente en forma de azúcares, una fuente de nitrógeno, habitualmente en forma de fuentes orgánicas de nitrógeno como extracto de levadura o sales como sulfato de amonio, oligoelementos como sales de hierro, manganeso y magnesio y, si es apropiado, vitaminas, a temperaturas comprendidas entre 0 °C y 100 °C, preferentemente, entre 10 °C y 60 °C, mientras se gasifica en oxígeno.

45 El pH del medio líquido puede mantenerse constante, es decir, regulado durante el período de cultivo, o no. Los cultivos pueden cultivarse por lotes, semicontinualmente o continuamente. Los nutrientes pueden proporcionarse al comienzo de la fermentación o pueden darse semicontinualmente o continuamente. La morfina producida puede aislarse de los organismos como se describió anteriormente mediante procesos conocidos por el experto, por ejemplo, mediante extracción, destilación, cristalización, si es apropiado precipitación con sal, y/o cromatografía. Para este fin, los organismos pueden alterarse ventajosamente de antemano. En este proceso, el valor del pH se mantiene ventajosamente entre pH 4 y 12, preferentemente, entre pH 6 y 9, especialmente preferentemente, entre pH 7 y 8.

50 El medio de cultivo a usar debe cumplir adecuadamente con los requisitos de las cepas en cuestión. En el libro de texto "Manual of Methods for General Bacteriology" de la American Society for Bacteriology (Washington DC, Estados Unidos, 1981) pueden encontrarse descripciones de medios de cultivo para varios microorganismos.

60 Como se describió anteriormente, estos medios que pueden emplearse de acuerdo con la invención habitualmente comprenden una o más fuentes de carbono, fuentes de nitrógeno, sales inorgánicas, vitaminas y/o elementos traza.

Las fuentes de carbono preferidas son azúcares, tales como mono-, di- o polisacáridos. Los ejemplos de fuentes de carbono son glucosa, fructosa, manosa, galactosa, ribosa, sorbosa, ribulosa, lactosa, maltosa, sacarosa, rafinosa, almidón o celulosa. Los azúcares también pueden añadirse a los medios a través de compuestos complejos como la melaza u otros subproductos del refinado del azúcar. También puede ser ventajosa la adición de mezclas de una variedad de fuentes de carbono. Otras posibles fuentes de carbono son aceites y grasas como, por ejemplo, aceite de soja, aceite de girasol, aceite de cacahuete y/o grasa de coco, ácidos grasos como, por ejemplo, ácido palmítico, ácido

esteárico y/o ácido linoleico, alcoholes y/o o polialcoholes como, por ejemplo, glicerol, metanol y/o etanol, y/o ácidos orgánicos como, por ejemplo, ácido acético y/o ácido láctico.

Las fuentes de nitrógeno son habitualmente compuestos o materiales de nitrógeno orgánicos o inorgánicos que comprenden estos compuestos. Los ejemplos de fuentes de nitrógeno comprenden amoníaco en forma líquida o gaseosa o sales de amonio tales como sulfato de amonio, cloruro de amonio, fosfato de amonio, carbonato de amonio o nitrato de amonio, nitratos, urea, aminoácidos o fuentes de nitrógeno complejas tales como licor de maíz fermentado, harina de soya, proteína de soya, extracto de levadura, extracto de carne y otros. Las fuentes de nitrógeno pueden usarse individualmente o como una mezcla.

Los compuestos de sales inorgánicas que pueden estar presentes en el medio comprenden las sales de cloruro, fósforo y sulfato de calcio, magnesio, sodio, cobalto, molibdeno, potasio, manganeso, zinc, cobre y hierro.

Los compuestos inorgánicos que contienen azufre tales como, por ejemplo, sulfatos, sulfitos, ditionitos, tetratiónatos, tiosulfatos, sulfuros, o bien compuestos de azufre orgánicos tales como mercaptanos y tioles, pueden usarse como fuentes de azufre para la producción de productos de química fina que contienen azufre, en particular de metionina.

El ácido fosfórico, el dihidrógeno fosfato de potasio o el hidrógeno fosfato de dipotasio o las sales que contienen sodio correspondientes, pueden usarse como fuentes de fósforo.

Los agentes quelantes pueden añadirse al medio para mantener los iones metálicos en solución.

Los agentes quelantes particularmente adecuados comprenden dihidroxifenoles tales como catecol o protocatecuato y ácidos orgánicos tales como ácido cítrico.

Los medios de fermentación usados de acuerdo con la invención para cultivar microorganismos habitualmente también comprenden otros factores de crecimiento tales como vitaminas o promotores del crecimiento, que incluyen, por ejemplo, biotina, riboflavina, tiamina, ácido fólico, ácido nicotínico, pantotenato, α -cetoglutarato y piridoxina. Los factores de crecimiento y las sales se derivan con frecuencia de componentes de medios complejos tales como extracto de levadura, melaza, licor de maíz fermentado y similares. Además, es posible añadir precursores adecuados al medio de cultivo. La composición exacta de los compuestos del medio depende en gran medida del experimento particular y se decide individualmente para cada caso específico. Puede encontrarse información sobre la optimización de los medios en el libro de texto "Applied Microbiol. Physiology, A Practical Approach" (Editores P.M. Rhodes, P.F. Stanbury, IRL Press (1997) págs. 53-73, ISBN 0 19 963577 3). Los medios de crecimiento también pueden obtenerse a partir de proveedores comerciales, por ejemplo, Standard 1 (Merck) o BHI (infusión de cerebro y corazón, DIFCO) y similares.

Todos los componentes del medio se esterilizan, ya sea por calor (20 min a 1,5 bar y 121 °C) o por esterilización por filtración. Los componentes pueden esterilizarse juntos o, si es necesario, por separado. Todos los componentes de los medios pueden estar presentes al comienzo del cultivo o añadirse continuamente o por lotes, según se desee.

La temperatura de cultivo es normalmente entre 15 °C y 45 °C, preferentemente, de 25 °C a 40 °C, y puede mantenerse constante o puede alterarse durante el experimento. El pH del medio debe estar en el intervalo de 5 a 8,5, preferentemente, aproximadamente 7,0. El pH para el cultivo puede controlarse durante el cultivo tras añadir compuestos básicos como hidróxido de sodio, hidróxido de potasio, amoníaco y amoníaco acuoso o compuestos ácidos como ácido fosfórico o ácido sulfúrico. La formación de espuma puede controlarse mediante el empleo de antiespumantes tales como, por ejemplo, ésteres de poliglicol de ácidos grasos. Para mantener la estabilidad de los plásmidos es posible añadir al medio sustancias adecuadas que tengan un efecto selectivo, por ejemplo, antibióticos. Las condiciones aeróbicas se mantienen mediante la introducción de oxígeno o mezclas de gas que contienen oxígeno, tales como, por ejemplo, aire ambiente en el cultivo. La temperatura del cultivo es normalmente de 20 °C a 45 °C y preferentemente, de 25 °C a 40 °C. El cultivo continúa hasta que la formación del producto deseado es máxima. Este objetivo normalmente se logra dentro de 10 a 160 horas.

Después, el caldo de fermentación puede procesarse adicionalmente. La biomasa puede, de acuerdo con lo requerido, eliminarse total o parcialmente del caldo de fermentación por métodos de separación tales como, por ejemplo, centrifugación, filtración, decantación o una combinación de estos métodos o dejarse completamente en dicho caldo. Es ventajoso procesar la biomasa después de su separación.

Sin embargo, el caldo de fermentación también puede espesarse o concentrarse sin separar las células, mediante el uso de métodos conocidos tales como, por ejemplo, con la ayuda de un evaporador rotatorio, un evaporador de película delgada, un evaporador de película descendente, mediante ósmosis inversa o nanofiltración. Finalmente, este caldo de fermentación concentrado puede procesarse para obtener la morfina presente en el mismo.

No cubierta por las reivindicaciones, se proporciona una célula vegetal transformada o transfectada con un vector que comprende una molécula de ácido nucleico de acuerdo con la invención.

Preferentemente, dicho vector se adapta para la expresión en plantas.

Preferentemente, la molécula de ácido nucleico en el vector está bajo el control de, y unida operativamente a, un promotor apropiado u otros elementos reguladores para la transcripción en una célula huésped tal como una célula microbiana, (por ejemplo, bacteriana, levadura) o vegetal. El vector puede ser un vector de expresión bifuncional que funciona en múltiples huéspedes. En el caso del ADN genómico, este puede contener su propio promotor u otros elementos reguladores y, en el caso del ADNc, este puede estar bajo el control de un promotor apropiado u otros elementos reguladores para la expresión en la célula huésped.

Por "promotor" se entiende una secuencia de nucleótidos aguas arriba del sitio de inicio de la transcripción y que contiene todas las regiones reguladoras necesarias para la transcripción. Los promotores adecuados incluyen promotores constitutivos, específicos de tejido, inducibles, de desarrollo u otros para la expresión en células vegetales comprendidas en plantas en dependencia del diseño. Dichos promotores incluyen promotores virales, fúngicos, bacterianos, animales y derivados de plantas capaces de funcionar en células vegetales. Los promotores constitutivos incluyen, por ejemplo, el promotor de CaMV 35S (Odell y otros (1985) Nature 313, 9810-812); actina de arroz (McElroy y otros (1990) Plant Cell 2: 163-171); ubiquitina (Christian y otros (1989) Plant Mol. Biol. 18 (675-689); pEMU (Last y otros (1991) Theor Appl. Genet. 81: 581-588); MAS (Velten y otros (1984) EMBO J. 3. 2723-2730); promotor de ALS (Solicitud de Estados Unidos núm. de serie 08/409,297), y similares. Otros promotores constitutivos incluyen los de las Patentes de Estados Unidos núms. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; y 5,608,142.

Los promotores microbianos o vegetales regulados químicamente pueden usarse para modular la expresión de un gen en una célula vegetal o microbiana mediante la aplicación de un regulador químico exógeno. En dependencia del objetivo, el promotor puede ser un promotor inducible químicamente, donde la aplicación del químico indujo la expresión génica, o un promotor de represión química, donde la aplicación del químico reprime la expresión génica. Los promotores inducibles químicamente se conocen en la técnica e incluyen, pero no se limitan al promotor In2-2 de maíz, que se activa con protectores de herbicidas de bencenosulfonamida, el promotor GST de maíz, que se activa con compuestos electrofílicos hidrófobos que se usan como herbicidas preemergentes, y el promotor PR-1a del tabaco, que se activa por el ácido salicílico. Otros promotores regulados químicamente de interés incluyen promotores de respuesta a esteroides (ver, por ejemplo, el promotor inducible por glucocorticoides en Schena y otros (1991) Proc. Natl. Acad. Sci. USA 88: 10421-10425 y McNellis y otros (1998) Plant J. 14(2): 247-257) y promotores inducibles por tetraciclina y reprimibles por tetraciclina (ver, por ejemplo, Gatz y otros (1991) Mol. Gen. Genet. 227: 229-237, y las patentes de Estados Unidos núms. 5,814,618 y 5,789,156.

Cuando se desea una expresión potenciada en tejidos particulares, pueden utilizarse promotores específicos de tejidos. Los promotores específicos de tejido incluyen los descritos por Yamamoto y otros. (1997) Plant J. 12(2): 255-265; Kawamata y otros (1997) Plant Cell Physiol. 38(7): 792-803; Hansen y otros (1997) Mol. Gen. Genet. 254(3): 337-343; Russel y otros (1997) Transgenic Res. 6(2): 157-168; Rinhart y otros (1996) Plant Physiol. 112(3): 1331-1341; Van Camp y otros (1996) Plant Physiol. 112(2): 525-535; Canevascni y otros (1996) Plant Physiol. 112(2): 513-524; Yamamoto y otros (1994) Plant Cell Physiol. 35(5): 773-778; Lam (1994) Results Probl. Cell Differ. 20: 181-196; Orozco y otros (1993) Plant Mol. Biol. 23(6): 1129-1138; Mutsuoka y otros (1993) Proc. Natl. Acad. Sci. USA 90 (20): 9586-9590; y Guevara-García y otros (1993) Plant J. 4(3): 495-50.

De interés particular en el presente contexto son los constructos de ácido nucleico que operan como vectores vegetales. Los procedimientos y vectores específicos usados anteriormente con gran éxito en plantas se describen por Guerineau y Mullineaux (1993) (Plant transformation and expression vectors. In: Plant Molecular Biology Labfax (Croy RRD ed) Oxford, BIOS Scientific Publishers, pp 121-148). Los vectores adecuados pueden incluir vectores derivados de virus vegetales (ver, por ejemplo, el documento EP194809). Si se desea, pueden incluirse en el constructo marcadores genéticos seleccionables, tales como aquellos que confieren fenotipos seleccionables tales como resistencia a herbicidas (por ejemplo, kanamicina, higromicina, fosfotricina, clorsulfurón, metotrexato, gentamicina, espectinomycin, imidazolinonas y glifosato).

Preferentemente, dicha célula vegetal es de la familia *Papaver*.

Preferentemente, dicha célula vegetal es una célula de *Papaver somniferum*.

De acuerdo con un aspecto de la invención, se proporciona un proceso para la extracción de codeína y/o alcaloides de codeína relacionados de una planta de *Papaver* que comprende las etapas:

- i) cosechar una planta o material vegetal preparado a partir de una planta de acuerdo con la invención;
- ii) formar una mezcla de reacción de material vegetal de partículas;
- iii) extracción con solvente de la mezcla de reacción para proporcionar una fracción enriquecida en alcaloide; y
- iv) concentrar dicha fracción enriquecida en alcaloide para proporcionar una fracción enriquecida en codeína.

En un método preferido de la invención, dicho material vegetal comprende paja de amapola y/o látex de amapola. No cubierto por las reivindicaciones, es un método para producir una planta de *Papaver* con contenido elevado de codeína, que comprende las etapas de:

- 5 i) mutagénesis de semillas de tipo silvestre de una planta de *Papaver*;
- ii) cultivo de la semilla en i) para producir la primera y las subsecuentes generaciones de plantas;
- iii) obtener semillas de la primera generación de plantas y las subsecuentes generaciones de plantas;
- iv) determinar si la semilla y/o el material vegetal obtenido a través de la germinación de la semilla de dichas primera y subsecuentes generaciones de plantas tiene una secuencia de nucleótidos alterada y/o expresión
- 10 alterada de dicho polipéptido, que comprende las etapas de:
 - a. obtener una muestra y analizar la secuencia de ácido nucleico de una molécula de ácido nucleico seleccionada del grupo que consiste en:
 - 15 1. una molécula de ácido nucleico que comprende una secuencia de nucleótidos como se representa en la SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 o 14;
 2. una molécula de ácido nucleico que se hibrida con una hebra complementaria de dicha molécula de ácido nucleico en a) bajo condiciones de hibridación rigurosas y que codifica un polipéptido con actividad de 3-O-desmetilasa de codeína; y opcionalmente
 - 20 b. comparar la secuencia de nucleótidos de la molécula de ácido nucleico en dicha muestra con una secuencia de nucleótidos de una molécula de ácido nucleico de la planta de tipo silvestre original.

Preferentemente, dicha molécula de ácido nucleico se analiza mediante un método que comprende las etapas de:

- 25 i) extraer el ácido nucleico de dichas plantas mutadas;
- ii) amplificación de una parte de dicha molécula de ácido nucleico mediante una reacción en cadena de la polimerasa;
- iii) formar una preparación que comprende el ácido nucleico amplificado y el ácido nucleico extraído de la semilla de tipo silvestre para formar un ácido nucleico heterodúplex;
- 30 iv) incubar dicha preparación con una nucleasa de una hebra que corta en una región del ácido nucleico heterodúplex para identificar el error de apareamiento en dicho heterodúplex; y
- v) determinar el sitio del error de apareamiento en dicho heterodúplex de ácido nucleico.

- 35 Preferentemente, dicha molécula de ácido nucleico comprende la secuencia de nucleótidos como se establece en la SEQ ID NO: 44.

Los métodos para analizar la expresión de genes se conocen en la técnica. Por ejemplo, técnicas moleculares tales como la reacción en cadena de la polimerasa [PCR], es decir, que permiten la medición cuantitativa y cualitativa de la expresión del ARNm e incluyen métodos de protección contra nucleasas, transferencia de Northern, extensión de cebadores cuantitativa, microarreglos cuantitativos tales como el análisis de la expresión génica en serie [SAGE], ELISA para la medición de los niveles de proteínas. Estos son ejemplos de métodos disponibles para el experto para analizar la expresión génica.

- 45 No cubierta por las reivindicaciones, es una planta obtenida mediante el método de acuerdo con la invención.

No cubierta por la reivindicación, es el uso de una molécula de ácido nucleico que comprende o que consiste de la secuencia de nucleótidos establecida en la SEQ ID NO 1 y/o SEQ ID NO: 2 y/o SEQ ID NO: 3 y/o SEQ ID NO: 4 y/o SEQ ID NO: 5 y/o SEQ ID NO: 6 y/o SEQ ID NO: 7, o una molécula de ácido nucleico que es 99 % idéntica a la secuencia de nucleótidos establecida en la SEQ ID NO: 1, SEQ ID NO: 2, y/o SEQ ID NO: 3 y/o, SEQ ID NO 4 y/o, SEQ ID NO 5, y/o SEQ ID NO 6 y/o SEQ ID NO 7 y codifica un polipéptido con actividad de 3-O-desmetilasa de codeína como medio para identificar la presencia o ausencia de un gen que codifica dicha 3-O-desmetilasa de codeína en una planta de *Papaver*.

- 55 Preferentemente, dicha molécula de ácido nucleico comprende o consiste en la totalidad o parte del conjunto de secuencias de secuencia de nucleótidos establecida en la SEQ ID NO: 44.

No cubierto por las reivindicaciones, es un método para determinar la presencia o ausencia de una molécula de ácido nucleico que codifica un polipéptido con actividad de 3-O desmetilasa de codeína en una planta de *Papaver* que comprende:

- i) obtener una muestra de una planta de *Papaver*;
- ii) extraer el ADN genómico de la planta; y
- iii) analizar el ADN genómico para determinar la presencia de una molécula de ácido nucleico que comprende o que consiste de una secuencia de nucleótidos establecida en la SEQ ID NO 1 y/o SEQ ID NO: 2 y/o SEQ ID NO: 3. y/o SEQ ID NO: 4 y/o SEQ ID NO: 5 y/o SEQ ID NO: 6 y/o SEQ ID NO: 7.
- 65

Preferentemente, dicho ADN genómico comprende o consiste en una secuencia de nucleótidos como se establece en la SEQ ID NO: 44.

No cubierto por las reivindicaciones, es un método para obtener una planta de *Papaver* con mayor contenido de codeína que comprende:

- i) cruzar una planta de *Papaver* de acuerdo con la invención con una planta de *Papaver* de genotipo diferente para obtener semillas;
- ii) germinar las semillas obtenidas y permitir que la planta[as] germinada[as] se autopolinice(n);
- iii) germinar las semillas obtenidas en ii) y cribar las plantas germinadas para aumentar el contenido de codeína y/o la presencia y/o expresión de uno o más genes de 3-O-desmetilasa de codeína que comprenden una secuencia de nucleótidos establecida en la SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6 o SEQ ID NO: 7 en la progenie de plantas germinadas;
- iv) cruzar las plantas identificadas en iii) anteriormente con dicha planta de *Papaver* de diferente genotipo en la etapa i) para obtener semillas; v) repetir las etapas ii)-iv) al menos una vez para proporcionar una población de plantas de *Papaver* con mayor contenido de codeína en dicha planta de diferente genotipo.

Preferentemente, la etapa iii) comprende plantas que comprenden la SEQ ID NO: 44.

No cubierta por las reivindicaciones, es una planta de *Papaver* obtenida mediante el método de acuerdo con la invención.

A lo largo de la descripción y las reivindicaciones de esta descripción, las palabras "comprender" y "contener" y variaciones de las palabras, por ejemplo, "que comprende" y "comprende", significan "que incluyen pero no se limitan a", y no pretenden (y no lo hacen) excluir otros restos, aditivos, componentes, números enteros o etapas. "Que consiste esencialmente" significa que tiene los números enteros esenciales pero incluye los números enteros que no afectan materialmente la función de los números enteros esenciales.

A lo largo de la descripción y las reivindicaciones de esta descripción, el singular abarca el plural a menos que el contexto lo requiera de cualquier otra manera. En particular, donde se usa el artículo indefinido, la descripción debe entenderse que contempla la pluralidad así como también la singularidad, a menos que el contexto lo requiera de cualquier otra manera.

Los elementos, números enteros, características, compuestos, restos o grupos químicos descritos junto con un aspecto, modalidad o ejemplo particular de la invención deben entenderse aplicables a cualquier otro aspecto, modalidad o ejemplo descrito en la presente descripción, a menos que sean incompatibles con el mismo.

Una modalidad de la invención se describirá ahora solamente a manera de ejemplo y con referencia a las siguientes figuras y tablas: Figura 1 A-F: Contenido absoluto y relativo de alcaloide en material de cápsula F3 derivado de cruces con el cultivar de morfina HM6. Figura A, Figura B, cápsulas F3 homocigotas para las mutaciones/polimorfismos R158K y E193K en sus respectivas copias de CODM; Figura C, Figura D, cápsulas F3 homocigotas para CODM Q254*; Figura E, Figura F, cápsulas de control del cultivar HM6. DW, peso seco; % total, cantidades relativas de los alcaloides respectivos de alcaloides principales totales (morfina, codeína, oripavina y tebaína);

Figura 2: Contenido absoluto (A) y relativo (B) de alcaloide en cápsulas de un BC₂F₂ población que segrega las mutaciones/polimorfismos R158K y E193K vinculados en sus respectivas copias de CODM. Para configurar la población BC₂F₂ de la generación M7 de un individuo con alta codeína que alberga las mutaciones/polimorfismos se cruzó tres veces con el cultivar HM6 de Sun Pharmaceutical Industries (Australia) como progenitor recurrente seguido de una ronda de autopolinización. La BC₂F₂ se cultivó bajo vidrio para la tipificación genética y la elaboración de perfiles de metabolitos de las cápsulas. La asignación a las clases genotípicas homocigoto (Hom), heterocigoto (Het) y tipo silvestre (WT) se llevó a cabo como se describe en los métodos; lotes de semillas o S_185882, Δ S_185894, + S_185895;

Figura 3: Contenido absoluto y relativo de alcaloide en cápsulas M3 del mutante de alta codeína R80624_G07 sometido a mutagénesis por neutrones rápidos en comparación con las plantas de control del cultivar de tipo silvestre HN4 no sometido a mutagénesis. DW, peso seco;

Figura 4: Amplificación de fragmentos CODM de individuos M4 de la línea mutante de codeína/noscapina sometida a mutagénesis por neutrones rápidos y de plantas de control no sometidas a mutagénesis por neutrones rápidos; A, par de cebadores 178/179, tamaño de fragmento esperado: 1296 pb; carriles 1 y 12, marcador de tamaño (GeneRuler DNA Ladder Mix, Thermo Fisher Scientific); carriles 2-7, plantas mutantes M4 Sd-701292, Sd-701293, Sd-701294, Sd-701302, Sd-701303, Sd-701304; carriles 8-10, plantas de control Sd-763815, Sd-763818, Sd-863819; carril 11, sin control de plantilla. B, par de cebadores 178/181, tamaño de fragmento esperado: 1347 pb; carriles 1 y 10, marcador de tamaño (como arriba); carriles 2-4, plantas mutantes M4 Sd-701302, Sd-701303, Sd-701304; carriles 5-8, plantas de control Sd-763815, Sd-763818, Sd-863819, Sd-106305; carril 9, control sin plantilla. C, par de cebadores 180/179, tamaño de fragmento esperado: 1016 pb; carriles 1 y 10, marcador de tamaño (como arriba); carriles 2-4, plantas

mutantes M4 Sd-701302, Sd-701303, Sd-701304; carriles 5-8, plantas de control Sd-763815, Sd-763818, Sd-863819, Sd-106305; carril 9, control sin plantilla. D, par de cebadores 180/168, tamaño de fragmento esperado: 1659 pb; carril 1 y 10, marcador de tamaño (como arriba); carriles 2-4, plantas mutantes M4 Sd-701302, Sd-701303, Sd-701304; carriles 5-8, plantas de control Sd-763815, Sd-763818, Sd-863819, Sd-106305; carril 9, control sin plantilla. E, par de cebadores 167/168, tamaño de fragmento esperado: 2330 pb; carril 1, marcador de tamaño (como arriba); carriles 2-8, pérdida de plantas mutantes M4 Sd-701292, Sd-701293, Sd-701294, Sd-701302, Sd-701303, Sd-701304; carriles 8-14, plantas de control Sd-763815, Sd-763818, Sd-863819, Sd-764849, Sd-764851, Sd-764857, Sd-106305; carril 15, control sin plantilla. F, par de cebadores 175/176, tamaño de fragmento esperado: 1122 pb; carril 10, marcador de tamaño (como arriba); carriles 1-6, plantas mutantes M4 Sd-701292, Sd-701293, Sd-701294, Sd-701302, Sd-701303, Sd-701304; carriles 7-9, plantas de control Sd-763815, Sd-763818, Sd-863819; carriles 11-14, plantas de control Sd-764849, Sd-764851, Sd-764857, Sd-106305; carril 15, sin control de plantilla. G, par de cebadores 175/177, tamaño de fragmento esperado: 1122 pb; carril 15, marcador de tamaño (como arriba); carriles 1-7, plantas mutantes M4 Sd-701292, Sd-701293, Sd-701294, Sd-701302, Sd-701303, Sd-701304; carriles 7-13, plantas de control Sd-763815, Sd-763818, Sd-863819, Sd-764849, Sd-764851, Sd-764857, Sd-106305; carril 14, sin control de plantilla.

Figura 5: Ilustra el arreglo genómico de los genes CODM en la amapola. La estructura y posición de los tres genes CODM se muestran por encima de la línea negra central, que representa 426,6 kb de secuencia genómica. Los exones se representan mediante cuadros grises sólidos y los intrones mediante líneas negras finas. Las flechas indican la orientación de 5' a 3' de cada gen. El número anterior a la estructura génica indica la posición de inicio de cada gen (primer base del codón de inicio de la transcripción) dentro de la secuencia genómica que comprende los genes CODM (SEQ ID NO: 44).

Identificadores de secuencia de las secuencias suministradas (una descripción general se muestra en la Tabla 1):

- SEQ ID NO 1: secuencia genómica de variante del gen CODM
- SEQ ID NO 2: secuencia genómica de variante del gen CODM
- SEQ ID NO 3: secuencia genómica de variante del gen CODM
- SEQ ID NO 4: secuencia genómica de variante del gen CODM
- SEQ ID NO 5: secuencia genómica de variante del gen CODM
- SEQ ID NO 6: secuencia genómica de variante del gen CODM
- SEQ ID NO 7: secuencia genómica de variante del gen CODM
- SEQ ID NO 8 secuencia codificante deducida de la variante del gen CODM SEQ ID NO 1
- SEQ ID NO 9 secuencia codificante deducida de la variante del gen CODM SEQ ID NO 2
- SEQ ID NO 10 secuencia codificante deducida de la variante del gen CODM SEQ ID NO 3
- SEQ ID NO 11 secuencia codificante deducida de la variante del gen CODM SEQ ID NO 4
- SEQ ID NO 12 secuencia codificante deducida de la variante del gen CODM SEQ ID NO 5
- SEQ ID NO 13 secuencia codificante deducida de la variante del gen CODM SEQ ID NO 6
- SEQ ID NO 14 secuencia codificante deducida de la variante del gen CODM SEQ ID NO 7
- SEQ ID NO 15 secuencia de aminoácidos deducida de la variante del gen CODM SEQ ID NO 8
- SEQ ID NO 16 secuencia de aminoácidos deducida de la variante del gen CODM SEQ ID NO 9
- SEQ ID NO 17 secuencia de aminoácidos deducida de la variante del gen CODM SEQ ID NO 10 y 13
- SEQ ID NO 18 secuencia de aminoácidos deducida de la variante del gen CODM SEQ IDNO 11, 12 y 14.
- SEQ ID NO: 44 conglomerado genómico que comprende genes CODM

Tabla 1: Descripción general de copias/alelos de CODM amplificadas a partir de BC₂F₂ homocigoto para las mutaciones/polimorfismos R158K y E193K. *Las posiciones de las variaciones de nucleótidos observadas se refieren a la posición dentro de las secuencias genómicas respectivas (SEQ ID NO 1-7)

SEQ ID NO CODM genómico	Posición* y tipo de variación										SEQ ID NO de la secuencia de aminoácidos deducida
	177 [A->G] N11D	701 [C->A] R158K	1106 [G->A] E193K	798 [indel T] intrónica	116 [G->A] sinón.	1186 [T->C] sinón.	1601 [G->T] sinón.	1737 [T->C] UTR3'	1638 [G->A] UTR3'	1982-3 [indel TT] UTR3'	
SEQ ID NO 1	✓						✓			Inserc. TT	15
SEQ ID NO 2			✓			✓	✓				16
SEQ ID NO 3		✓					✓			Inserc. TT	17
SEQ ID NO 4				✓ (del T)							18
SEQ ID NO 5							✓		✓	Inserc. TT	18
SEQ ID NO 6		✓						✓			17
SEQ ID NO 7								✓		Inserc. TT	18
synon. = sinónimo; inserc. = inserción											

Tabla 2: Contenido de alcaloide en cápsulas de material de alta codeína cultivado en campo derivado de la familia de semillas M2 sometida a mutagénesis por EMS S-103234 que contiene las mutaciones vinculadas R158K y E193K en diferentes copias de CODM. El mutante de alta codeína se propagó mediante autopolinización y las generaciones M2 a M7 se evaluaron en el campo en temporadas consecutivas. caps SP, cápsula autopolinizada; surco de OP, surco de polinizado sin cubrir; M, morfina; C, codeína; O, oripavina; T, tebaína; pM, pseudo-morfina.

Temporada	Generación	ID de semillas		Núm. parcela	M	C	O	T	
2007/8			Surco de OP	Control HM2	3,1117				
	M2	S-103234	Surco de OP	MUT 437	2,9664	0,2557	0,0150	0,2019	
	M2	S-103234	caps SP	MUTS 437 G	4,4067	0,5943	0,0060	0,0769	
					M	C	O	T	
2008/9			Surco de OP	Control HM3	3,67				
	M3		Surco de OP	RMUU 314	3,7286	0,6138	0,0011	0,2387	
	M3		caps SP	RMUUS 314 E	4,2614	0,6006	0,0136	0,0813	
					M	C	O	T	
2009/10			Surco de OP	Control HM1	2,92	0,33	0,03	0,12	
	M4		Surco de OP	MUV 107	2,352	1,391	0,051	1,105	
	M4		caps SP	MUVS 107 A	2,315	2,316	0,024	1,209	
2010	M5		propagación fuera de temporada de MUVS 107 A						
					M	C	O	T	M+pM
2010/11			Surco de OP	Control HM4	2,352	0,144	0,015	0,105	2,569
			Surco de OP	Control de HT4	1,163	0,541	0,23	1,862	1,361
	M6		Surco de OP	MUWT 1964	2,073	0,709	0,017	0,429	2,356
	M6		caps SP	MUWTS 1964 C	2,154	2,001	0,059	0,696	2,266
					M	C	O	T	
2011/12			Surco de OP	Control HM5	3,373	0,129	0,017	0,116	
			Surco de OP	Control de HT5	1,001	0,189	0,709	2,473	
	M7		Surco de OP	MA 1983	3,005	1,691	0,040	1,316	
	M7		caps SP	MAS 1983 C	3,766	2,089	0,076	1,379	

Materiales y métodos

Mutagénesis por EMS para el cribado genético directo y inverso

Antes de la mutagénesis, las semillas M0 de cultivar de *P. somniferum* de Sun Pharmaceutical Industries (Australia) de alta morfina 2 (HM2) se remojaron en una solución de cloruro de potasio al 0,1 % (p/v) durante aproximadamente 24 horas. Después, la semilla se embibió durante tres (tratamiento B) o cinco horas (tratamiento C) con metanosulfonato de etilo (EMS) 200 mM (en tampón de hidrógeno fosfato de sodio 100 mM, pH 5,0, complementado con dimetilsulfóxido 700 mM). La semilla M1 tratada se lavó dos veces durante 15 minutos cada uno con solución de tiosulfato de sodio 100 mM y después dos veces durante 15 minutos cada uno con agua destilada. Se usó un volumen de 10 ml de las soluciones respectivas por cada 1000 semillas y todas las etapas se llevaron a cabo en una plataforma oscilante. La semilla M1 lavada se secó durante toda la noche en papel de transferencia Whatman 3MM (Whatman/GE Healthcare Life Sciences, Little Chalfont, Reino Unido) y se sembró a la mañana siguiente sobre una mezcla de compost John Innes no 2, vermiculita y perlita (4:2:1), después se cubrió con una capa fina de compost. Las plántulas se trasplantaron de tres a cuatro semanas después de la germinación a macetas más grandes y se cultivaron en el invernadero hasta la madurez. Se recolectaron a mano cápsulas secas de la población M1 una vez que las cápsulas se habían secado hasta aproximadamente 10 % de humedad en la planta. La semilla M2 se separó manualmente de la cápsula, lo que proporcionó las familias de semillas M2 que se usaron en el cribado directo basado en el campo y el cribado genético inverso descritos más abajo.

Cribado directo basado en el campo de líneas sometidas a mutagénesis por EMS

La semilla M2 se sembró en el campo en Tasmania en la temporada de cultivo 2007/2008. Las líneas de semillas M2 de la población sometida a mutagénesis por EMS se sembraron a mano en surcos de 3 m de largo. Después de la germinación, las plantas se entresacaron a 75 mm de separación. Típicamente, se autopolinizaron 4 plantas (designadas A, B, C, D) por línea de M2 (dos brotes inmaduros por planta seleccionada) asegurando una bolsa de lino sobre el brote. Los surcos polinizados sin cubrir se recolectaron a ~ 11 % de humedad y se analizaron (como se describe más abajo) para determinar el contenido de morfina (M), codeína (C), oripavina (O) y tebaína (T). El análisis de las cápsulas autopolinizadas seleccionadas de plantas M2 indicó contenidos más altos de codeína. Las semillas M3 a M7 derivadas de una cápsula M2 autopolinizada seleccionada se sembraron en las siguientes temporadas de cultivo en Tasmania para confirmar este fenotipo en cápsulas polinizadas sin cubrir y autopolinizadas. Típicamente, se autopolinizaron 4 plantas (designadas A, B, C, D) por generación como se describió anteriormente. Los cultivares comerciales de Sun Pharmaceutical Industries (Australia) se cultivaron junto a las líneas M2 a M7 como una comparación. En dependencia de la temporada, se incluyeron los cultivares de alta morfina de Sun Pharmaceutical Industries (Australia) HM1, HM2, HM3, HM4 y HM5, y los cultivares de alta tebaína de Sun Pharmaceutical Industries (Australia) HT4 y HT5. Estos cultivares producen los principales morfínicos morfina, codeína, oripavina y tebaína, pero carecen de la capacidad biosintética para producir noscapina.

Análisis de paja de amapola de cribado directo basado en el campo

Las cápsulas de amapola se cosecharon de las líneas mutantes respectivas a mano una vez que las cápsulas se secaron a aproximadamente 10 % de humedad en la planta. La semilla se separó manualmente de la cápsula, y después el material de la paja de la cápsula (paja de amapola) se envió a la instalación de extracción de Sun Pharma en Port Fairy, Australia.

Las muestras de paja de amapola se trituraron en un molino de bolas Retsch Modelo MM400 en un polvo fino. Muestras de dos gramos de paja de amapola molida ($2 \pm 0,003$ g) se extrajeron en 50 ml de una solución de ácido acético al 10 %. La suspensión de extracción se agitó en un agitador orbital a 200 rpm durante un mínimo de 10 minutos y después se filtró para proporcionar un filtrado transparente. El filtrado final se pasó a través de un filtro de 0,22 μ m antes del análisis.

Las soluciones se analizaron mediante el uso de un sistema UPLC Waters Acquity (Waters Corporation, Milford, MA, Estados Unidos) equipado con una columna Waters Acquity BEH C18, 2,1 mm $\times$ 100 mm con empaque de 1,7 micras. La fase móvil usó un perfil de gradiente con el eluyente A que consiste en ácido trifluoroacético al 0,1 % en agua desionizada y el eluyente B que consiste en acetonitrilo al 100 %. Las condiciones de gradiente de fase móvil usadas son las enumeradas en la Tabla 3, el número de curva de gradiente según lo determinado mediante el uso de un paquete de programa informático de cromatografía Waters Empower. El régimen de flujo fue de 0,6 ml por minuto y la columna se mantuvo a 45 grados centígrados. El volumen de inyección fue de 1 μ l y los alcaloides se detectaron mediante el uso de un detector UV a 285 nm.

La pérdida por secado (LOD) de la paja se determinó mediante secado en un horno a 105 grados centígrados durante 16-20 horas. Las concentraciones de alcaloide para morfina, pseudomorfina, codeína, tebaína, oripavina y noscapina se cuantificaron mediante el uso de calibración estándar externa y los resultados se calcularon sobre una base de peso seco. La reticulina se cuantificó mediante el uso de la calibración de tebaína (suponiendo la respuesta de tebaína) y los resultados se calcularon sobre una base de peso seco (% de peso con relación a tebaína; %WRT).

Tabla 3: Programa de flujo de gradiente

TIEMPO (minutos)	Eluyente A %	Eluyente B %	Flujo (ml/min)	Número de curva
0,00	95,0	5,0	0,60	INICIAL
0,80	90,0	10,0	0,60	6
3,40	75,0	25,0	0,60	3
3,60	95,0	5,0	0,60	6
4,00	95,0	5,0	0,60	11

Los tiempos de retención típicos son los siguientes:

Compuesto	Tiempo de retención (minutos)
Morfina	1,14
Pseudomorfina	1,26
Codeína	1,69
Oripavina	1,80

(continuación)

Compuesto	Tiempo de retención (minutos)
Reticulina	2,05
10-Hidroxitebaína	2,32
Tebaína	2,53
Noscapina	3,16

Establecimiento de una población de TILLING para el cribado genético inverso

Un subconjunto de las poblaciones de EMS de la Banda C se cultivaron en el invernadero para establecer una población de TILLING. Se cultivó un hermano M2 por familia de semillas M2 para la extracción del ADN. El ADN se extrajo de plántulas de cuatro semanas de edad como se describe más abajo. En total, se obtuvo ADN de 4146 plantas M2 (Tabla 4). Las plántulas M2 se sacrificaron después de la toma de muestras de ADN. Cualquier mutación identificada en una planta M2 se recuperó mediante la resiembra y el recribado de los hermanos M2 respectivos.

El ADN para la población de TILLING se cuantificó mediante el uso del colorante Hoescht 33258 y se normalizó a 10 ng/μl en TE mediante el uso de la plataforma TECAN Freedom Evo® 100 (TECAN, Suiza). Las muestras de ADN se agruparon cuatro veces antes del cribado.

Tabla 4: Poblaciones de TILLING

Tratamiento con EMS	Núm. de familias de semillas M2	Identificadores de familias de semillas M2	Núm. de líneas de semillas M2 germinadas (% de germinación)
B	4779	S-100001 a S-102800, S-102831 a S-103227, S-103647 a S-105228	3807 (79 %)
C	419	S-103228 a S-103646	339 (56 %)

Preparación de ADN genómico

Para la extracción del ADN genómico se recogieron 30-50 mg de material de hojas de plántulas cultivadas en invernadero de 4-6 semanas de edad. El ADN genómico se extrajo mediante el uso del kit de plantas BioSprint 96 en la estación de trabajo BioSprint 96 (Qiagen, Crawley, Reino Unido) de acuerdo con el protocolo del fabricante. El ADN extraído se cuantificó espectrométricamente en el NanoDrop® (Thermo Scientific, Reino Unido) y se normalizó a 5 o 10 ng/ul.

Clonación de los genes CODM a partir de ADN genómico

Las secuencias de genes CODM de longitud completa se amplificaron a partir de la línea homocigota del mutante MAS1983 C mediante el uso de una enzima de PCR de alta fidelidad. Las reacciones de PCR que contienen tampón HF Phusion 1x, dNTP 0,2 mM, 0,5μM cebador directo (id cebador 167) y reverso (id cebador 168) (Tabla 5) y 0,4 U de polimerasa de ADN Phusion (ThermoFisher) en un volumen total de 20 μl se ejecutaron en un termociclador Tetrad (Bio-Rad) bajo las siguientes condiciones: 98 °C durante 30 segundos, seguido de 10 ciclos de 98 °C durante 10 segundos, 65 °C que disminuye a 60 °C en 0,5 °C por ciclo durante 30 segundos, 72 °C durante 1 minuto, seguido de 28 ciclos de 98 °C durante 10 segundos, 60 °C durante 30 segundos, 72 °C durante 1 minuto, seguido de 72 °C durante 10 minutos, después una retención a 7 °C. Los productos de PCR se comprobaron en gel de agarosa al 1 %, se extrajeron mediante el kit de extracción de PCR Wizard SV (Promega) de acuerdo con las instrucciones del fabricante, y la concentración se determinó espectrométricamente en el NanoDrop® (Thermo Scientific).

El producto de PCR de CODM purificado se clonó mediante el uso del kit de clonación por PCR CloneJET (Thermo Scientific) de acuerdo con el protocolo del fabricante, y se transformó en células competentes SoloPack (StrataClone) de acuerdo con el protocolo del fabricante.

Se seleccionaron diez colonias individuales positivas para el inserto de CODM mediante PCR de colonias. Los plásmidos se extrajeron mediante el uso del kit QIAprep Spin Miniprep (Qiagen) y se usaron para la secuenciación de Sanger. Los cebadores para la amplificación y la secuenciación se muestran en la Tabla 5. Entre los diez clones secuenciados, se identificaron siete secuencias del gen CODM diferentes (SEQ ID 1-7, Tabla 5). Las secuencias transcritas y codificantes se dedujeron mediante la comparación con la secuencia de ADNc de CODM publicada (GenBank: GQ500141) que dio como resultado las SEQ ID 8-14 (secuencias codificantes) y SEQ ID 15-18 (secuencias de aminoácidos).

Amplificación de fragmentos de CODM a partir de ADN genómico

Las amplificaciones por PCR de fragmentos CODM se llevaron a cabo en ADN genómico de individuos M4 derivados del material mutante sometido a mutagénesis por neutrones rápidos de alta codeína/noscapina R80624_G07 (hermanos mutantes M4: Sd-701292, Sd-701293, Sd-701294, Sd-701302, Sd-701303, Sd-701304; plantas de control Sd-763815, Sd-763818, Sd-863819, Sd-764849, Sd-764851, Sd-764857, Sd-106305) así como también en material no sometido a mutagénesis por neutrones rápidos derivado del cultivar HM2 de morfina de Sun Pharmaceutical Industries (Australia).

Las reacciones de PCR que contenían tampón GoTaq Flexi 5x, dNTP 0,2 mM, cebador directo e inverso 1,0 μ M (Tablas 5 y 6), 1 U de polimerasa de ADN GoTaq Flexi (Promega, M8305), $MgCl_2$ 1,5 mM y 10 ng de ADN molde en un volumen total de 10 μ l se ejecutaron en un termociclador Tetrad (Bio-Rad) bajo las siguientes condiciones: 94 °C durante 2 minutos, seguido de 10 ciclos de 94 °C durante 15 segundos, 65 °C que disminuye a 60 °C en 0,5 °C por ciclo durante 30 segundos, 72 °C durante 1-1,5 minutos (en dependencia de la longitud del fragmento), seguido de 25 ciclos de 94 °C durante 15 segundos, 60 °C durante 30 segundos, 72 °C durante 1-1,5 minutos (en dependencia de la longitud del fragmento), seguido de 72 °C durante 10 minutos, después una retención a 7 °C. Los productos de PCR se analizaron en geles de agarosa al 1-1,5 %. Los cebadores y las combinaciones de cebadores usados para la amplificación se muestran en las Tablas 5 y 6.

Cribado genético inverso de 3-O-desmetilasa de codeína (CODM)

Amplificamos CODM a partir de placas agrupadas mediante el uso de F9/R5 CODM (Tabla 5). Este par de cebadores amplificaría las tres copias de CODM conocidas simultáneamente. Se configuraron reacciones de PCR de 10 μ l con 0,1 μ l de polimerasa de ADN GoTaq Flexi (Promega, M8305), 0,6 μ l de $MgCl_2$ (25 mM), 2 μ l de tampón incoloro GoTaq Flexi, 0,2 μ l de dNTP (New England Biolabs, N0447), 0,5 μ l de cada cebador (10 μ M), 4,1 μ l de agua de grado de biología molecular HyClone (Thermo Scientific, SH3053801) y 2 μ l (20 ng) de ADN molde.

Las condiciones de PCR fueron las siguientes: 95 °C 2 min, 10 ciclos de (94 °C durante 20 s, temperatura de hibridación de 67-62 °C (-0,5 °C por ciclo) durante 30 s, 72 °C durante 75 s) 37 ciclos de (94 °C durante 20 s, 62 °C durante 30 s, 72 °C durante 75 s, extensión final a 72 °C durante 5 min.

Tabla 5: Cebadores usados para clonación, amplificación de fragmentos y secuenciación

ID_cebador	Propósito del cebador	secuencia (5'→3')
167	Clonación de CODM a partir de ADN genómico	GTGGTTGCTCAGACCCTTCGT
168	Clonación de CODM a partir de ADN genómico	AGCACCACCAAGCCTTC
169	Secuenciación	AGAAACCAACCAACGTACCCTAGC
175	Secuenciación, amplificación de fragmentos de CODM	AACCATGGAGTCGACGCTTT
176	Amplificación de CODM	TCTCTGGTGTGACCAAGCTC
177	Amplificación de CODM	TCTCTGGTGTGACCAAGCTA
178	Secuenciación, amplificación de fragmentos de CODM	CTCACGCTTGACAGAAATTCC
179	Secuenciación, amplificación de fragmentos de CODM	CTCGACGCTACGGTAAATCC
180	Amplificación de fragmentos de CODM	GTTAACCATGGAGTCGACGC
181	Amplificación de fragmentos de CODM	AAATGTCGCGATTGAGAGCC
CODMseqF1	Secuenciación	TCCGGGTTTGTGAATTGCAT
CODMseqR1	Secuenciación	ATTCCAGTGTCTCCCTGCAT
pJET1.2 Dir	Secuenciación	CGACTCACTATAGGGAGAGCGGC
pJET1.2 Inv	Secuenciación	AAGAACATCGATTTTCCATGGCAG
CODM_F9	TILLING	CGACGCTTTACTGATGGACA
CODM_R5	TILLING	CATCCACGAAATTTAAGCACAA

Tabla 6: Combinación de cebadores para la amplificación de fragmentos de CODM genómicos mostrados en la Figura 4

Fragmento	ID de cebador directo	ID de cebador inverso	Tamaño de fragmento esperado	Figura
1	178	179	1296	Figura 4A
2	178	181	1347	Figura 4B
3	180	179	1016	Figura 4C
4	180	168	1659	Figura 4D
5	167	168	2330	Figura 4E
6	175	176	1122	Figura 4F

7	175	177	1122	Figura 4G
---	-----	-----	------	-----------

Esto fue seguido inmediatamente por la formación de heterodúplex como sigue: 99 °C durante 10 min, 70 °C durante 20 s (-0,3 °C por ciclo) × 70, viejo a 4 °C. Después de la formación de heterodúplex, los productos de PCR se diluyeron cinco veces con dH₂O. 2 µl de este producto de PCR diluido se trató con 2 µl de concentración de trabajo de enzima de escisión de ADNdc (Advanced Analytical Technologies Inc. (AATI), FS-CLKENZ) en la reacción de digestión de heterodúplex. Las mutaciones se detectaron mediante el uso del Kit de descubrimiento de mutaciones (AATI, DNF-920-K0500T) en la máquina Fragment Analyzer™ (AATI) de acuerdo con las instrucciones del fabricante para el kit de detección de mutaciones. La máquina se equipó con un cartucho de matriz de 96 capilares (fluorescencia), 55 cm de longitud efectiva/80 cm de longitud total, DI 50 µm (AATI, A2300-9650-5580). Las condiciones de funcionamiento fueron las siguientes:

Condición completa:

Preejecución: 12 kV 30 s
 Inyección de voltaje de marcador (35 pb y 5000 pb): 7,5 kV 15 s
 Inyección de voltaje de muestra: 7,5 kV 45 s
 Separación: 12 kV 75 min

Una mutación en un grupo resultó en la formación de heterodúplex después de la PCR y la escisión subsecuente. Las mutaciones se detectaron si se observaron dos fragmentos pequeños de ADN cuyos tamaños totales correspondían al tamaño del producto de PCR de longitud completa en un carril.

Generación de una población F₃ para la mutación de truncamiento Q254* de CODM

Una mutación de transición G:C a A:T que introduce un codón de parada prematuro en el residuo de aminoácido Q254 (Q254*) se identificó en una copia de CODM en la familia de semillas M2 sometida a mutagénesis por EMS S-104530. La línea mutante se cruzó con el cultivar HM6 de Sun Pharmaceutical Industries (Australia) y la generación F₁ se autopolinizó para generar la F₂ segregante que se genotipó para Q254*. El material F₃ derivado de una planta mutante F₂ homocigota se probó en el campo en Tasmania, se genotipó por QC y se evaluó el perfil de alcaloides.

Generación de una población BC₂F₂ que segrega las mutaciones R158K y E193K de CODM

Los mutantes de alta codeína identificados en la criba directa con EMS que se encontró que contenían las mutaciones R158K y E193K de CODM (SEQ ID NO 2, 3, 6, 9, 10, 13, 16, 17) se cruzaron tres veces con el cultivar HM6 parental recurrente de Sun Pharmaceutical Industries (Australia) seguido de una ronda de autopolinización para generar una población BC₂F₂ segregando R158K y E193K en sus respectivas copias de CODM. Los polimorfismos/mutaciones se rastrearon a través de las generaciones mediante genotipado. Además de configurar cruces BC₁, también se autopolinizó material F₁ para generar poblaciones F₂ segregantes. El material F₃ derivado de plantas F₂ homocigotas para R158K y E193K se evaluaron en el campo en Tasmania. La población BC₂F₂ segregante se cultivó, genotipó y fenotipó en el invernadero.

Genotipado de la mutación CODMb Q254*

La mutación Q254* de CODMb se genotipó por medio de PCR específica de alelo. Se configuraron reacciones de PCR de 10 µl en pocillos de una placa de PCR de 96 pocillos (Starlabs, 11402-9800) con 0,5 U de polimerasa de ADN GoTaq Flexi (Promega, M8305), MgCl₂ 1,5 mM, 1 × de tampón incoloro GoTaq Flexi, dNTP 200 mM (New England Biolabs, N0447), 500 nM de cada cebador (Tabla 7) y 20 ng de ADN molde.

Tabla 7: Secuencias de cebadores para ensayos de PCR específicos de alelo para genotipar la mutación Q254*

Marcador	Cebador_alelo de tipo silvestre (5'→3')	Cebador_alelo mutante (5'→3')	Cebador_común (5'→3')
CODM_Q254*	GAAATCCACCTCTCTCTTTCTTTATTGG	GATTGAAATCCACCTCTCTCTTTCTTTATTAA	TGTCAGATGGAATAATACTGAAAGTTGCA

Las condiciones de PCR fueron las siguientes: 95 °C durante minutos, 35 ciclos de 95 °C durante 30 segundos, 59 °C durante 30 segundos, 72 °C durante 30 segundos, seguido de una extensión final a 72 °C durante 5 minutos.

Los productos de PCR se visualizaron en gel de agarosa al 1,5 % p/v.

Genotipado de mutaciones/polimorfismos R158K y E193K

Los ensayos de genotipado KASP (LGC Genomics) se desarrollaron para rastrear las mutaciones que conducen a las sustituciones de aminoácidos E193K y R158K en CODMa/c y CODMb, respectivamente. Los cebadores específicos de alelo y comunes usados en los ensayos se muestran en la Tabla 8.

Las reacciones de PCR que contenían 12,5 ng de ADN, 3 % de DMSO, 1x de mezcla maestra (KASP4,0 ROX bajo, LGC Genomics) y 1x de mezcla de cebadores (LGC Genomics) en un volumen total de 5 µl se ejecutaron en una máquina de PCR Hydrocycler (LGC) bajo las siguientes condiciones: 94 °C durante 15 minutos, seguido de 10 ciclos de 94 °C durante 20 segundos y 65 °C que disminuye a 57 °C durante 1 minuto, seguido de 38 ciclos de 94 °C durante 20 segundos y 57 °C durante 1 minuto. Después de la terminación de la PCR KASP, la lectura fluorescente de los productos de PCR se realizó en un ViiA7 (Applied Biosystems). El análisis de agrupamiento se realizó con el programa informático Applied Biosystems ViiA(TM) 7 v.1.2.1. A pesar de investigar la misma mutación/polimorfismo en dos variantes CODM diferentes (Tabla 1), típicamente se observaron tres clases genotípicas distintas para el marcador CODM_R158K en poblaciones segregadas. Las clases se asignaron a homocigoto para el alelo mutante (AA), homocigoto para el alelo de tipo silvestre (GG) y heterocigoto (AG) en base a los controles (clase homocigota: ADN de individuos homocigotos para SEQ ID NO: 3 y 6, clase heterocigota: ADN de individuos homocigotos para SEQ ID NO: 3 y 6 mixtos con cantidades iguales de ADN HM6, clase de tipo silvestre: ADN HM6).

Tabla 8: Secuencias de cebadores de ensayos de genotipado KASP para las mutaciones/polimorfismos R158K y E193K

ID KASP	Cebador_alelo de tipo silvestre FAM (5'→3')	Cebador_alelo mutante HEX (5'→3')	Cebador_común (5'→3')
CODM_E193K	CAACGGTTGTCTTTGATGTTGG	ATCAACGGTTGTCTTTGAGATGTTGA	CCCATCTTCAAAATAAGTCTGTGTCATACCTT
CODM_R158K	AGTTCTGGAAACAAATGAGGCTTCC	GAGTTCTGGAAACAAATGAGGCTTCT	GTTTAGCATGTTAAGTCTTCTCTCTCCATT

Mutagénesis por delección por neutrones rápidos y cribado directo basado en invernadero

La mutagénesis por neutrones rápidos (FNM) se llevó a cabo por el Centro de Investigación de Energía de HAS, Departamento de Protección de Radiación, H-1121 Budapest, Konkoly Thege út 29-33, X. epelet, Hungría. Aproximadamente 40 g de semilla del cultivar de alta noscapina 4 (HN4) de Sun Pharmaceutical Industries (Australia) se expusieron a 20 Gy (tiempo de haz 50 min). La generación M1 se autopolinizó. Se cultivó una planta M2 por familia de semillas M2. Se recogió látex de plántulas M2 de 9 semanas de edad y se analizó. Las plantas M2 que mostraron una composición de alcaloide alterada en el látex en comparación con los controles HN4 se cultivaron hasta la madurez para confirmar el fenotipo en el material de cápsula M2 seca. La semilla M3 se recogió de plantas M2 con fenotipos confirmados. Después se cultivaron cinco plantas M3 por planta mutante M2 para confirmar la heredabilidad del fenotipo en el material de cápsula M3 seco maduro. Las generaciones M1 a M3 se cultivaron todas en invernadero.

Análisis de látex de hojas y cápsulas secas de material cultivado en invernadero

El látex se recogió de plantas de 9 semanas de edad de los pecíolos cortados, con una sola gota dispersa en 500 ml de ácido acético al 10 %. Esto se diluyó 10x en ácido acético al 1 % para dar una solución de alcaloide en ácido acético al 2 % para un análisis adicional. Las cápsulas se cosecharon de las mismas plantas usadas para el análisis de látex y las cápsulas individuales se trituraron hasta obtener un polvo fino en un molino de bolas (modelo MM04, Retsch, Haan, Alemania). Las muestras de paja de amapola molida se pesaron después con precisión a $10 \pm 0,1$ mg y se extrajo en 0,5 ml de una solución de ácido acético al 10 % con agitación suave durante 1 h a temperatura ambiente. Las muestras se aclararon después mediante centrifugación y una submuestra de 50 ml se diluyó 10x en ácido acético al 1 % para dar una solución de alcaloide en ácido acético al 2 % para un análisis adicional.

Todas las soluciones se analizaron mediante el uso de un sistema Waters Acquity UPLC (Waters Ltd., Elstree, Reino Unido) equipado con una columna Waters Acquity BEH C18, 2,1 mm x 100 mm con empaque de 1,7 micras. La fase móvil usó un perfil de gradiente con el eluyente A que consistió en bicarbonato de amonio 10 mM de pH 10,2 y el eluyente B metanol. Las condiciones de gradiente de fase móvil usadas son las enumeradas en la tabla más abajo con un gradiente lineal. El régimen de flujo fue de 0,5 ml por minuto y la columna se mantuvo a 60 °C. El volumen de inyección fue de 2 ml y los picos eluidos se ionizaron en modo APCI positivo y se detectaron dentro de una precisión de masa de 5 ppm mediante el uso de un Thermo LTQ-Orbitrap. Las corridas se controlaron mediante el programa informático Thermo Xcalibur (Thermo Fisher Scientific Inc., Hemel Hempstead, Reino Unido).

TIEMPO (minutos)	Eluyente A %	Eluyente B %	Flujo (ml/min)
0,0	98	2	0,50
0,2	98	2	0,50
0,5	60	40	0,50
4,0	20	80	0,50
4,5	20	80	0,50

Todo el análisis de datos se llevó a cabo mediante el uso del lenguaje de programación R y secuencias de comandos personalizados. Se usaron estándares de morfina, oripavina, codeína, tebaína y noscapina para identificar y cuantificar estos alcaloides mediante el uso de la masa exacta, el tiempo de retención y las áreas de picos para el ion pseudomolecular. Otros picos de alcaloide supuestos se cuantificaron por sus áreas de iones pseudomoleculares con relación al estándar de tebaína. Para los alcaloides supuestos, se usó el paquete rcdk de Bioconductor (Guha, (2007) J. Stat. Software 6(18)) para generar fórmulas pseudomoleculares a partir de masas exactas dentro de las restricciones elementales C = 1 100, H = 1 200, O = 0 200, N = 0 3 y precisión de masa < 5 ppm. El acierto con el error de ppm más bajo dentro de estas restricciones se usó para asignar una fórmula supuesta.

Construcción y cribado de bibliotecas de cromosomas artificiales bacterianos (BAC).

Se usó el cultivar de alta noscapina cv. 1 para generar una biblioteca de cromosomas artificiales bacterianos (BAC) como se describe en Winzer y otros. ((2012), Science 336:1704-8).

Un fragmento de 703 pb ubicado en la región del promotor de CODM (Raymond M, "Isolation and characterization of latex-specific promoters from Papaver somniferum", Tesis de maestría, Instituto Politécnico y Universidad Estatal de Virginia, Blacks-burg, VA) se amplificó con los cebadores AAAATCCGCCCTCCATGC (directo) y CCGACTTTGGCCCACTTGT (inverso) mediante el uso de un kit de síntesis de PCR de digoxigenina (DIG) (Roche Applied Science, Indianapolis, IN) de acuerdo con las instrucciones del fabricante para obtener la sonda de detección marcada con DIG. El cribado de la biblioteca BAC se realizó como se describió en Winzer y otros. ((2012), Science 336:1704-8) y dio como resultado la identificación de 10 clones BAC positivos, específicamente BAC33_D07, BAC86_F04, BAC89_C05, BAC109_H06, BAC129_K11, BAC152_G13, BAC158_A11, BAC185_L02, BAC195_N12 y BAC230_D02.

Secuenciación BAC y ensamble de secuencia.

Cada BAC se secuenció en dos plataformas de secuenciación: La secuenciación de extremos emparejados se realizó en la plataforma Illumina HiSeq (Illumina, Inc, San Diego, CA).

Además, cada clon BAC se secuenció mediante el uso de un secuenciador MinION (Oxford Nanopore Technologies Ltd, Oxford, Reino Unido). Los BAC purificados se fragmentaron mínimamente mediante el uso de tratamientos de 20 segundos con NEBNext® Fragmentasa de ADNds® (New England Biolabs Inc., Ipswich, MA), y bibliotecas de secuenciación preparadas mediante el uso del kit de secuenciación de Oxford Nanopore Technologies SQK007 con el paquete de expansión de codificación nativa EXP-NDB002. Brevemente, las roturas de una sola hebra en el ADN se repararon mediante el uso de mezcla de reparación de ADN NEBNext® FFPE antes de la reparación final y dA con cola mediante el uso del módulo NEBNext® Ultra™ II End Repair/dA-Tailing. Se ligaron secuencias de código de barras únicas a fragmentos para cada BAC, antes de agrupar 3-4 BAC por biblioteca. Los adaptadores de secuenciación (que incluyen un adaptador de horquilla para permitir la secuenciación 2D) se ligaron a los extremos de los fragmentos, junto con una proteína de unión a adaptador. Los fragmentos donde se unió la proteína de anclaje se purificaron mediante el uso de perlas de estreptavidina MyOne™ C1 (Invitrogen/Thermo Fisher Scientific, Reino Unido). Las bibliotecas se ejecutaron después en celdas de flujo MinION R9 mediante el uso de un protocolo de secuenciación de 48 horas, y la llamada de bases y el desmultiplexado se realizaron mediante el uso de la plataforma EPI2ME de Metrichor. Las lecturas que pasaron las comprobaciones de calidad 2D se dividieron de acuerdo con el código de barras para un análisis adicional.

Ensamble de secuencias

Las lecturas MiniON crudas de cada clon BAC se ensamblaron con el programa informático CANU v1.3 (Koren y otros (2016) bioRxiv, doi.org/10.1101/071282) y los límites de inserción se identificaron con las posiciones de las secuencias vectoriales y la referencia cruzada de clones BAC superpuestos. Estos conjuntos iniciales se corrigieron adicionalmente con el programa informático NANOPOLISH. Las lecturas cortas de extremo emparejado de Illumina de alta calidad se mapearon después con el programa informático BWA (Li y Durbin (2009) Bioinformatics, 25: 1754-60) a los conjuntos de MinION resultantes, que se usaron como referencia de andamiaje. Se generaron referencias de consenso con la alineación mapeada y los índices se corrigieron de acuerdo con los análisis de variación de la alineación. El consenso corregido se usó después como una nueva referencia en un proceso reiterativo hasta que no se pudo lograr ninguna mejora adicional. Este método de ensamble ha demostrado dar una identidad de bases del 99,2 % a una secuencia BAC previamente ensamblada y publicada (BAC164_F07, número de acceso de la base de datos: JQ659012). Los BAC solapados se combinaron después para dar un fragmento del genoma CODM de 426 KB continuo.

Ejemplo 1

Existen múltiples copias de CODM en el genoma de la amapola

Las copias de CODM se clonaron a partir del ADN genómico de un BC₂F₂ de alta codeína individual genotipado como homocigoto para las mutaciones/polimorfismos R158K y E193K (ver Ejemplo 3). Se obtuvieron al menos siete secuencias génicas genómicas CODM diferentes (SEQ ID NO 1-7, Tabla 1) que dieron lugar a siete secuencias transcritas deducidas y cuatro secuencias de aminoácidos deducidas. Por lo tanto, hay al menos cuatro copias de CODM en el genoma de la amapola diploide (suponiendo que las secuencias identificadas representan el número máximo de variantes alélicas teóricamente posibles) o hasta siete copias (suponiendo que no hay variación alélica entre las secuencias CODM identificadas). Las mutaciones/polimorfismos R158K y E193K cosegregaron juntas a través de generaciones subsecuentes cuando se generó la población BC₂F₂ que indica un enlace cercano de al menos tres copias de CODM (SEQ ID NO 2, 3 y 6; tres copias ya que R158K ocurre en dos copias diferentes de CODM).

Ejemplo 2

Una mutación de inactivación en una copia de los genes CODM conduce a un aumento limitado de la codeína

El cribado genético inverso de los genes CODM produjo una mutación del codón de parada prematuro en CODM (Q254*) que probablemente hace que la proteína truncada resultante sea no funcional. Las plantas CODM Q254* se cruzaron con el cultivar de morfina HM6 y el efecto de la mutación se analizó en material F3 cultivado en el campo en Tasmania. El material F3 homocigoto mostró un aumento de 2 veces de la codeína ($0,46 \pm 0,16$ %) en comparación con HM6 ($0,18 \pm 0,10$ % de DW, Figura 1C y E). Igualmente, la cantidad relativa de codeína de los alcaloides mayores totales (morfina, codeína, oripavina y tebaína) es con un 12 % más de dos veces mayor que la de HM6 (5 %, Figura 1D y F). El aumento limitado en la codeína por esta mutación de inactivación de CODM supuesta se explica por tener múltiples copias funcionalmente redundantes de CODM (Ejemplo 1).

Ejemplo 3

Los polimorfismos/mutaciones de sustitución de aminoácidos en tres copias de CODM, se asocian con niveles elevados de codeína y niveles disminuidos de oripavina, lo que indica una actividad disminuida de CODM

El cribado directo basado en el campo del material HM2 sometido a mutagénesis por EMS condujo al descubrimiento de una planta mutante M2 con niveles de codeína más altos que el material HM2 no sometido a mutagénesis. La planta M2 se originó de la familia de semillas M2 S-103234 (Tabla 2). El aumento del contenido de codeína se confirmó en las generaciones autopolinizadas M3 a M7 cultivadas en temporadas de cultivo consecutivas en el campo en Tasmania, lo que demuestra que el rasgo de alta codeína es heredable. La secuenciación de los genes CODM de este material reveló mutaciones/polimorfismos en tres copias de CODM que conducen a sustituciones de aminoácidos, R158K (en SEQ ID NO 3 y 6) y E193K (en SEQ ID NO2). Los ensayos de marcadores KASP se diseñaron para rastrear la mutación en cruces asistidos por marcadores. Mediante el uso de estos marcadores, la generación M7 (MAS1983 C) se cruzó con el cultivar HM6 de alta morfina del progenitor recurrente. Las mutaciones/polimorfismos R158K y E193K se segregaron juntas, lo que indica un estrecho enlace entre las respectivas copias de CODM en las que residen.

Las plantas F3 homocigotas para ambas mutaciones se cultivaron en el campo en Tasmania en la temporada 2014/15 junto con el progenitor HM6 recurrente (Figura 1A y D). El análisis de este material F3 mostró un aumento de 5 veces de la codeína ($1,10 \pm 0,27$ % de DW) en comparación con el recurrente HM6 parental ($0,18 \pm 0,10$ % de DW, Figura 1A y E). La cantidad relativa de codeína alcanzó el 18 % en comparación con el 5 % en HM6 (Figura 1B y F).

Después de otros cruces con el progenitor recurrente HM6, la generación BC₂F₁ se autopolinizó y la BC₂F₂ segregante se cultivaron hasta la madurez en el invernadero para evaluar el impacto de las mutaciones en los niveles de codeína. Las plantas BC₂F₂ homocigotas para las mutaciones contenían en promedio el doble de codeína que aquellas plantas homocigotas para los alelos de tipo silvestre ($0,80 \pm 0,37$ % de DW frente a $0,37 \pm 0,21$ % de DW; Figura 2A). Las cápsulas de plantas heterocigóticas contenían niveles intermedios. Consistente con un impacto en la actividad CODM en las plantas mutantes, los niveles de oripavina fueron aproximadamente cinco veces menores en material BC₂F₂ mutante homocigoto en comparación con el material homocigoto para los alelos de tipo silvestre (Figura 2A).

La cantidad relativa de codeína fue más de dos veces mayor en cápsulas de plantas mutantes homocigotas que en las de plantas homocigotas para los alelos de tipo silvestre (19,4 % frente a 8,9 %, Figura 2B).

La cosegregación de las mutaciones/polimorfismos R158K y E193K de CODM en sus respectivas copias CODM con mayor contenido de codeína muestra que estas mutaciones están vinculadas - si no son causales - con el rasgo de alta codeína.

Ejemplo 4

La ausencia completa de actividad CODM en una planta sometida a mutagénesis por neutrones rápidos se debe a una delección de todos los genes CODM que conduce a una pérdida completa de morfina y oripavina y una ganancia de codeína

El cribado directo basado en invernadero de material sometido a mutagénesis por neutrones rápidos del cultivar de morfina y noscapina HN2 de alto rendimiento produjo una planta mutante M2 (R80624_G07) que carecía completamente de morfina y oripavina pero había alcanzado altos niveles de codeína, lo que indica una pérdida completa de la actividad CODM. La planta había retenido su capacidad para sintetizar noscapina. La planta mutante M2 se autopolinizó y la heredabilidad del fenotipo de codeína y noscapina se confirmó en cápsulas M3 (Figura 3). Las plantas M3 contenían en promedio 1,9 % de peso seco cada una de codeína y noscapina, lo que representó en cada caso casi el 44 % del total de alcaloides. La generación M3, también, se autopolinizó. Se cultivó la progenie M4 y se aisló el ADN. El uso de varios cebadores en varias combinaciones no amplificó ninguno de los genes CODM en las plantas mutantes M4, mientras que la amplificación fue exitosa en ADN de plantas de control de síntesis de morfina no sometidas a mutagénesis por neutrones rápidos (Figura 4). La ausencia completa de amplificación de los genes CODM sugiere que la pérdida completa de la actividad CODM observada en el material de cápsula derivado del mutante de alta codeína R80624_G07 es una consecuencia de una delección de todos los genes de CODM de su genoma. Esto a su vez indica que todas las copias de CODM están estrechamente vinculadas.

Ejemplo 5

Los genes CODM se agrupan en el genoma de la amapola

La secuenciación y el ensamble BAC revelaron que los genes CODM ocurren como un grupo de al menos tres genes CODM dentro del genoma de la amapola (Figura 5). Este hallazgo es consistente con el hecho de que todos los genes CODM se eliminan en una planta sometida a mutagénesis por neutrones rápidos que acumula altos niveles de codeína pero carece de morfina (Figura 4 y ejemplo 4). Se ha demostrado que la mutagénesis por neutrones rápidos es capaz de inducir delecciones que varían de varios cientos a decenas de miles de pares de bases (O'Rourke y otros (2013) *Frontiers in Plant Science* 4: Artículo 210). La falta de amplificación de fragmentos CODM en la planta de alta codeína sometida a mutagénesis por neutrones rápidos (Figura 4) indica que todo el conglomerado génico CODM se ha eliminado.

REIVINDICACIONES

1. Una especie de planta de *Papaver* en donde la planta se modifica mediante la delección de la totalidad o parte de tres genes enlazados que codifican la 3-O-desmetilasa de codeína que comprenden la secuencia de nucleótidos establecida en la SEQ ID NO: 7, o una secuencia de nucleótidos que es 99 % idéntica a la SEQ ID NO: 7, caracterizada porque la actividad de dicha 3-O-desmetilasa de codeína es indetectable y en donde la planta tiene un contenido de codeína de al menos 45 % del contenido total de alcaloide de látex o paja seca extraído de la planta modificada.
2. La planta de *Papaver* de acuerdo con la reivindicación 1, en donde la planta se modifica mediante la delección de una molécula de ácido nucleico que comprende o que consiste de una secuencia de nucleótidos como se establece en la SEQ ID NO: 44.
3. La planta de *Papaver* de acuerdo con la reivindicación 1 o 2, en donde el contenido de codeína es al menos 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 96 %, 97 % o 98 % del contenido total de alcaloide de látex o paja seca extraído de dicha planta modificada.
4. La planta de *Papaver* de acuerdo con una cualquiera de las reivindicaciones 1 a 3, en donde el contenido de codeína es al menos 0,8 % de la paja seca de dicha planta modificada.
5. La planta de *Papaver* de acuerdo con la reivindicación 3 o 4, en donde el contenido de codeína está entre 0,8 % y 10,0 % de la paja seca.
6. La planta de *Papaver* de acuerdo con la reivindicación 5, en donde el contenido de codeína está entre 0,8 % y 3,0 % de la paja seca.
7. La planta de *Papaver* de acuerdo con la reivindicación 6, en donde el contenido de codeína está entre 3,0 % y 10,0 % de la paja seca.
8. La planta de *Papaver* de acuerdo con la reivindicación 6, en donde el contenido de codeína está entre 1,0 % y 1,5 % de la paja seca.
9. La planta de *Papaver* de acuerdo con la reivindicación 1 o 2, en donde dicha planta tiene un contenido de codeína que es al menos 55 % del contenido total de alcaloide de látex o paja seca extraído de dicha planta modificada o tiene un contenido de codeína que es al menos 0,8 % de la paja seca.
10. La planta de *Papaver* de acuerdo con una cualquiera de las reivindicaciones 1 a 9, en donde dicha planta tiene niveles indetectables de morfina.
11. La planta de *Papaver* de acuerdo con una cualquiera de las reivindicaciones 1 a 9, en donde dicha planta carece de morfina y oripavina.
12. La planta de *Papaver* de acuerdo con una cualquiera de las reivindicaciones 1 a 11, en donde dicha planta se modifica mediante mutagénesis por neutrones rápidos,
13. La planta de *Papaver* de acuerdo con una cualquiera de las reivindicaciones 1 a 12, en donde dicha planta es *Papaver somniferum*.
14. Un proceso para la extracción de codeína y/o alcaloides de codeína relacionados de una planta de *Papaver* que comprende las etapas:
 - i) cosechar una planta o material vegetal preparado a partir de una planta de acuerdo con una cualquiera de las reivindicaciones 1 a 13;
 - ii) formar una mezcla de reacción de material vegetal de partículas;
 - iii) extracción con solvente de la mezcla de reacción para proporcionar una fracción enriquecida en alcaloide; y
 - iv) concentrar dicha fracción enriquecida en alcaloide para proporcionar una fracción enriquecida en codeína.

Figura 1 (1 de 3)

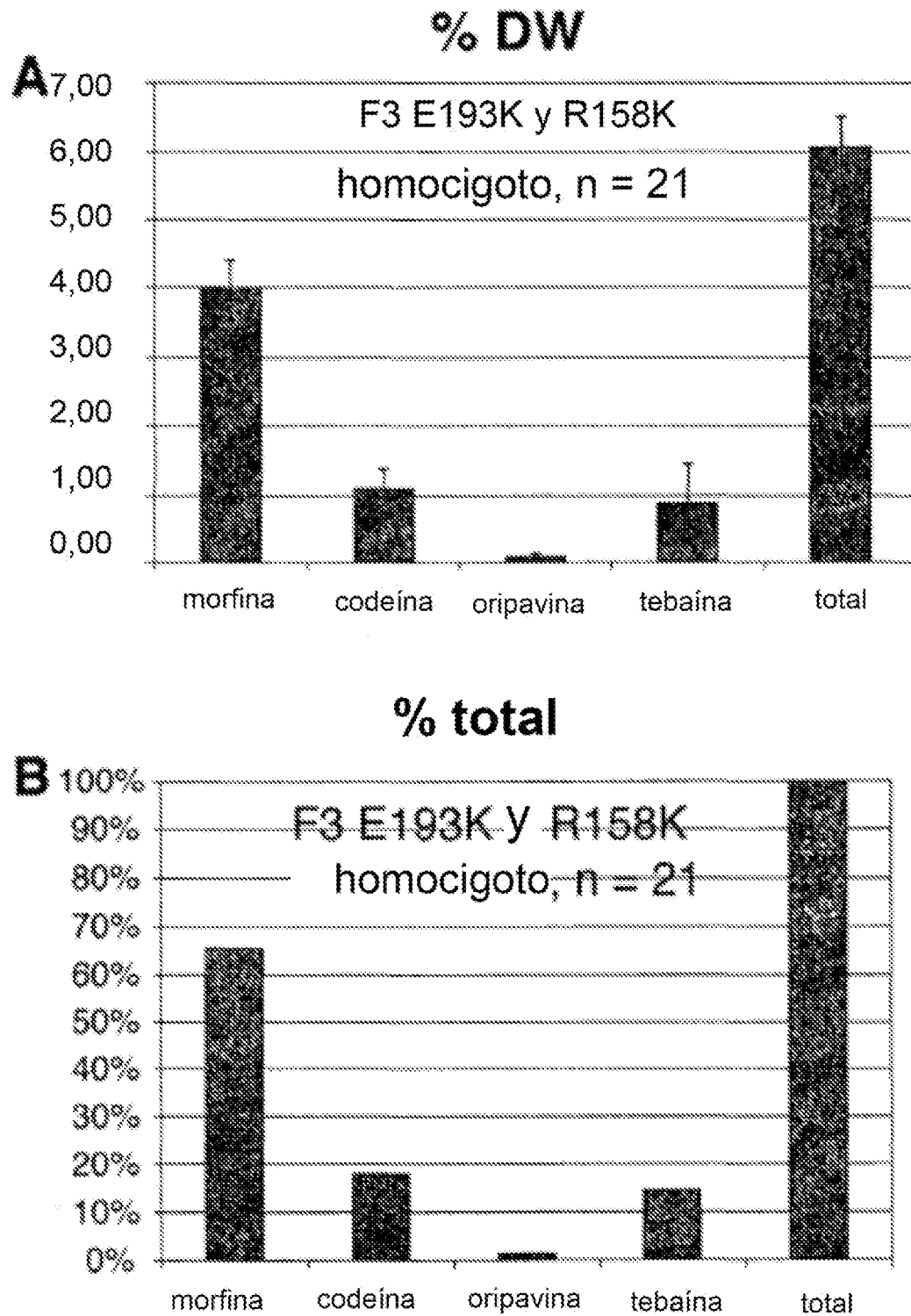


Figura 1 (2 de 3)

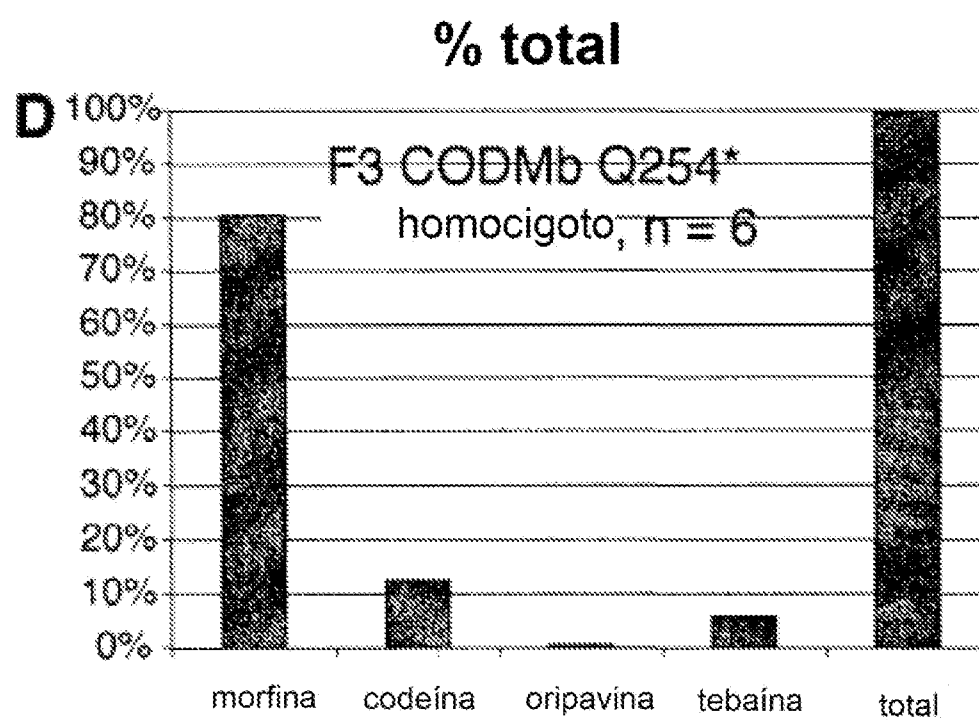
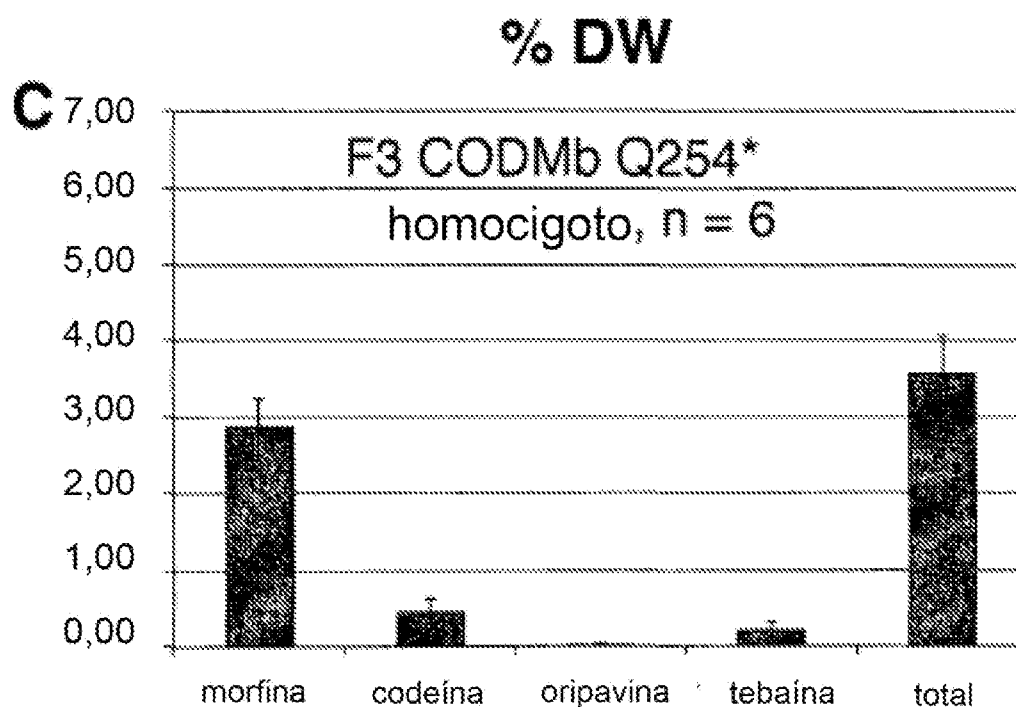


Figura 1 (3 de 3)

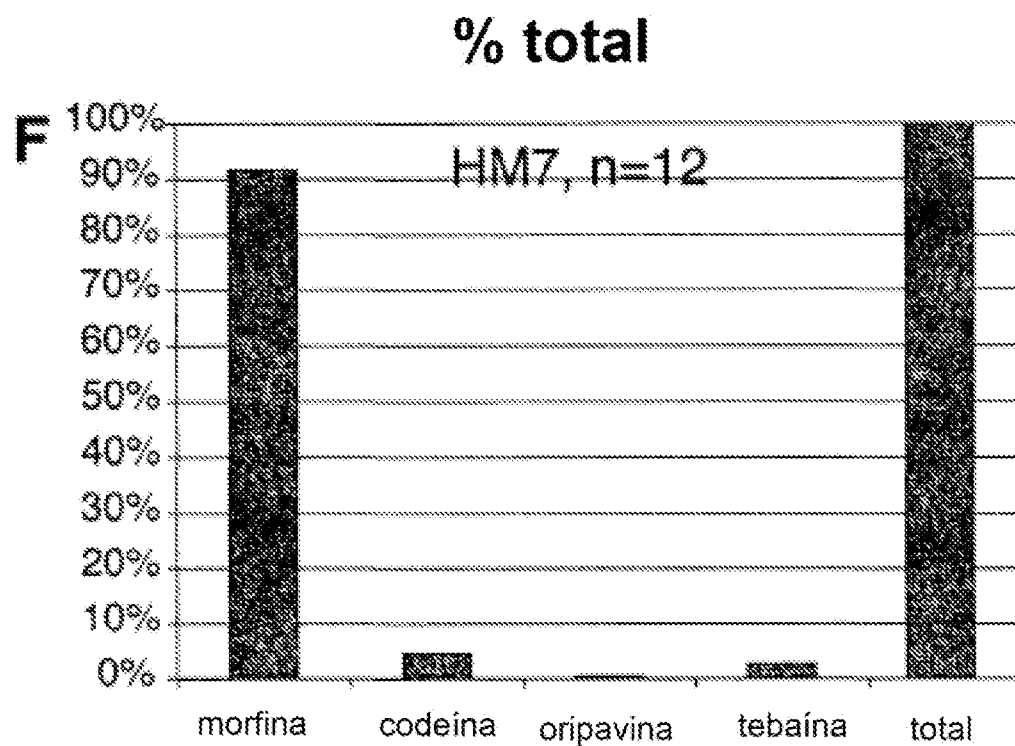
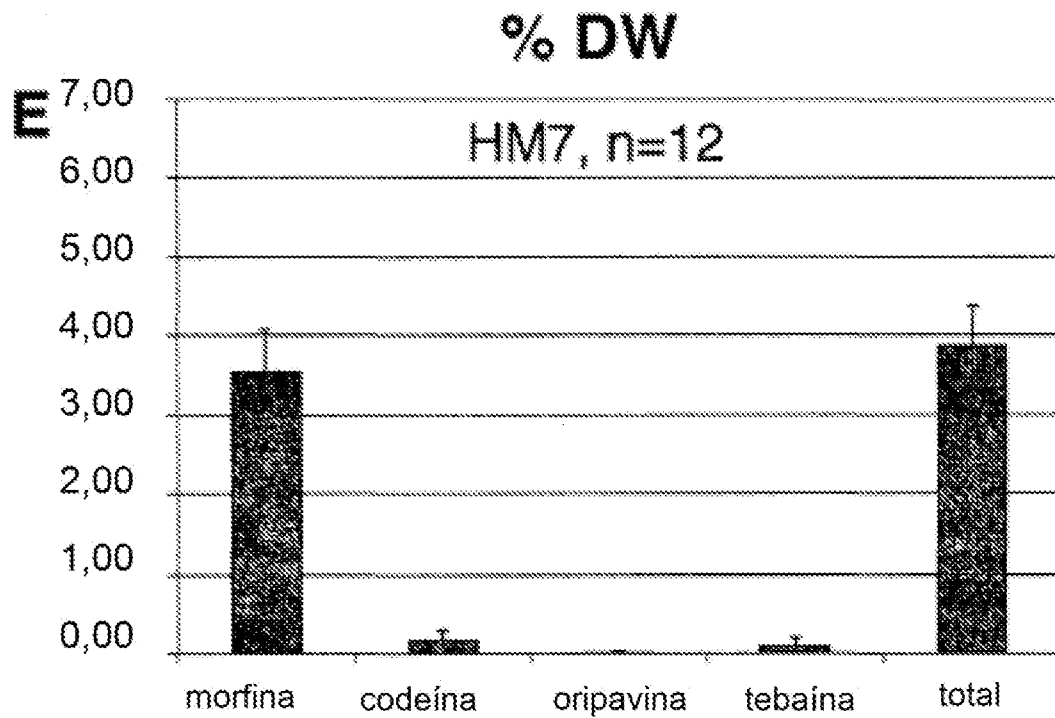


Figura 2A

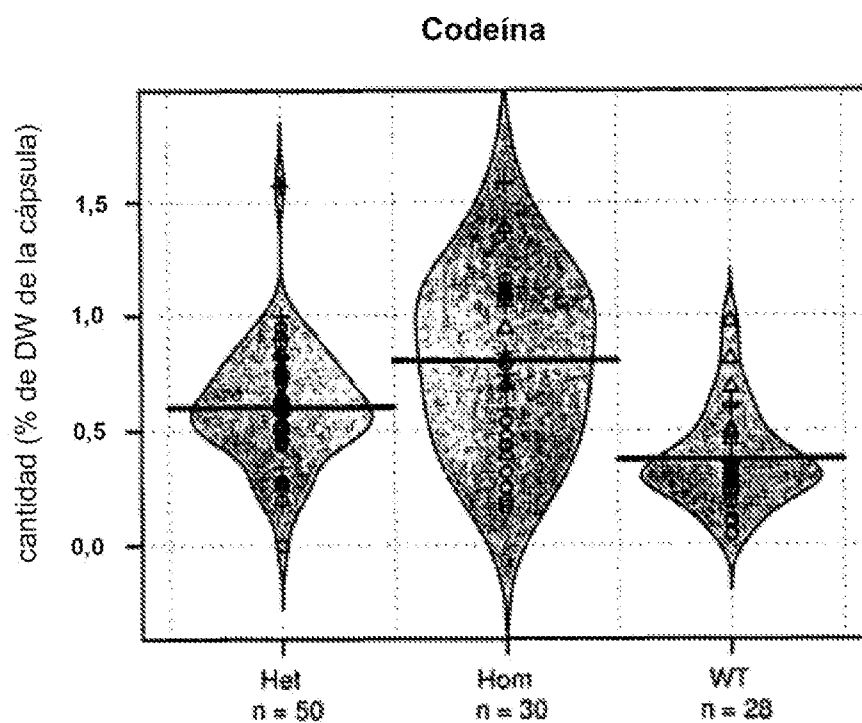
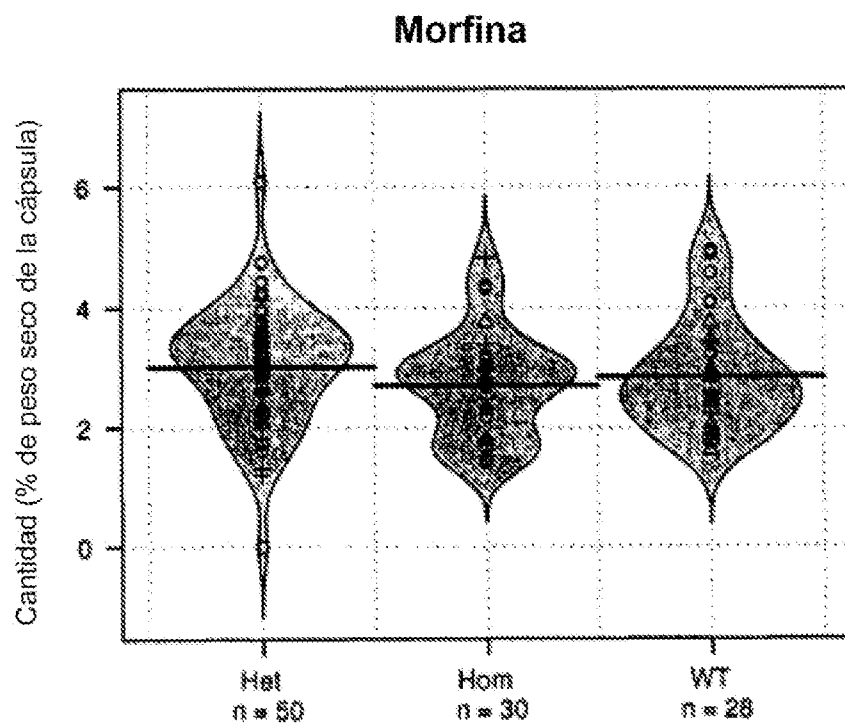


Figura 2 A continuación

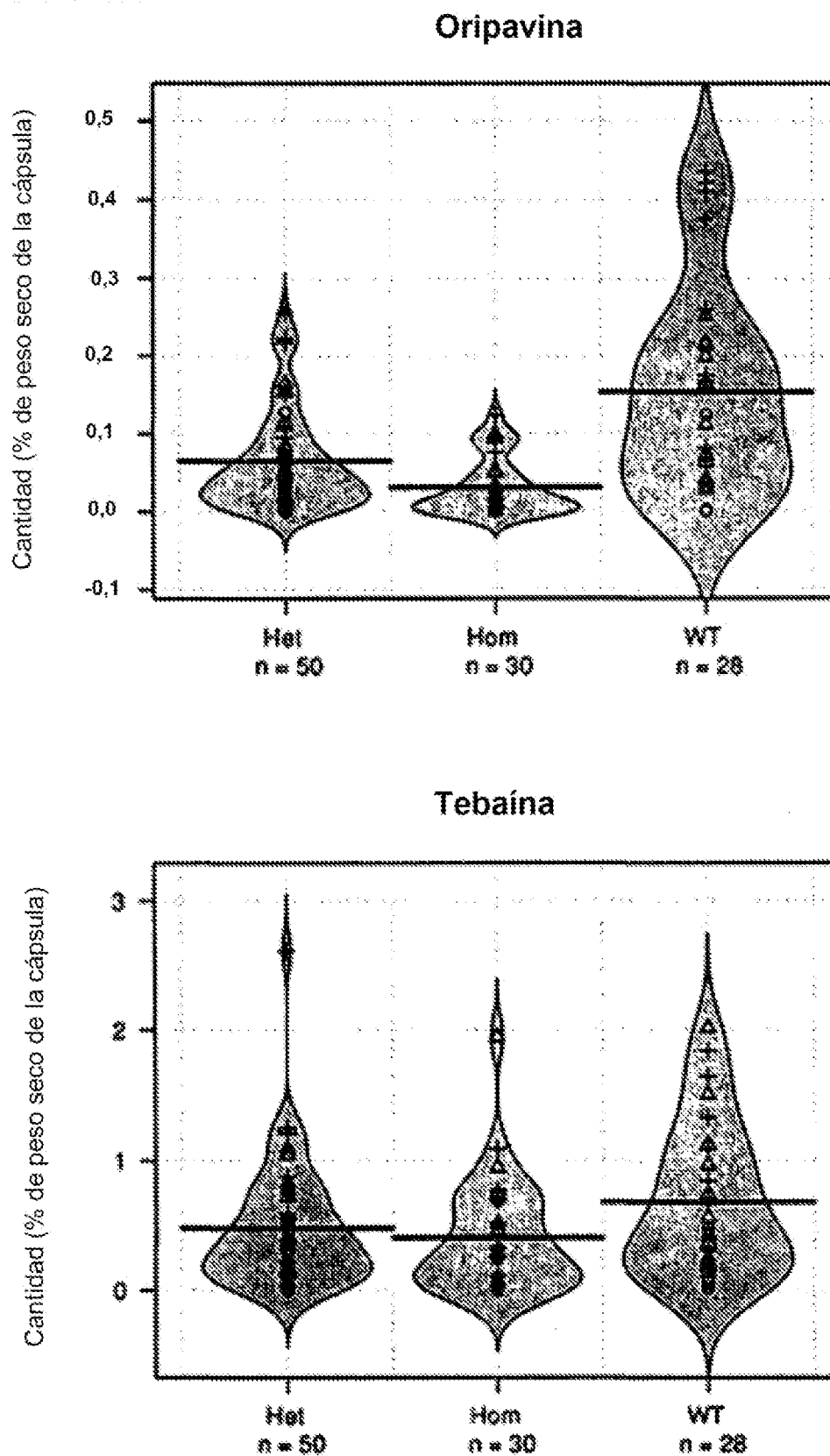


Figura 2 A continuación

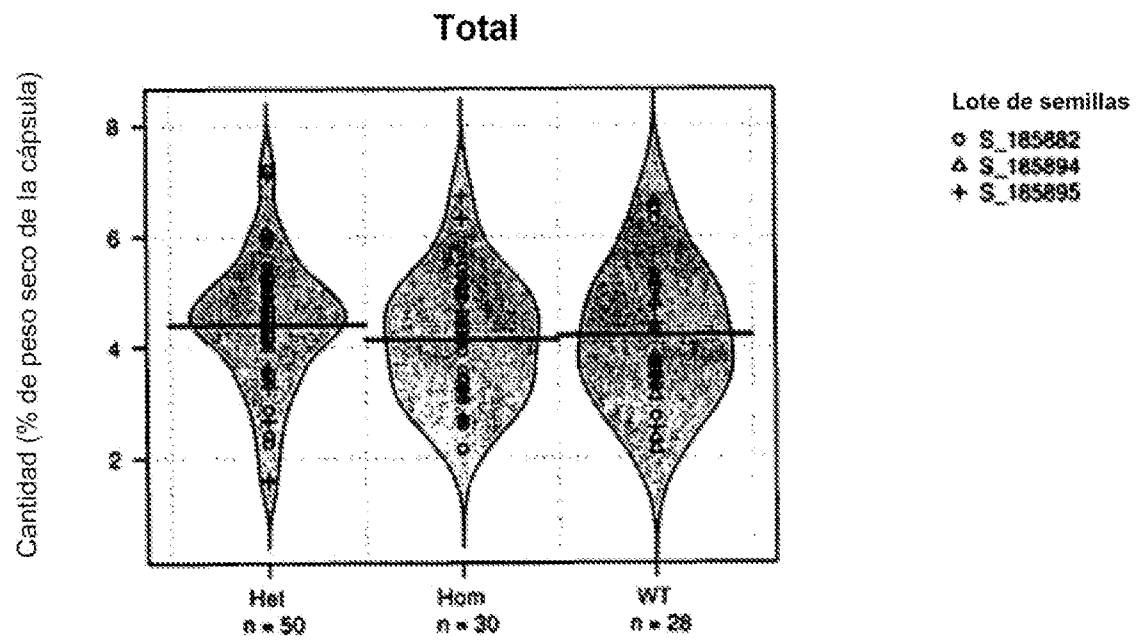


Figura 2b

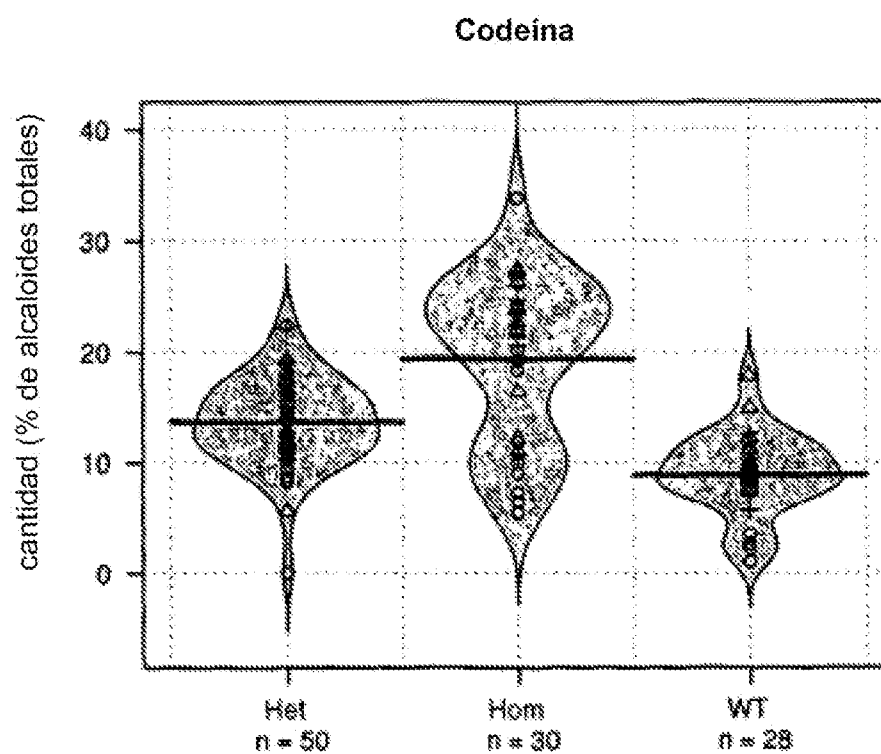
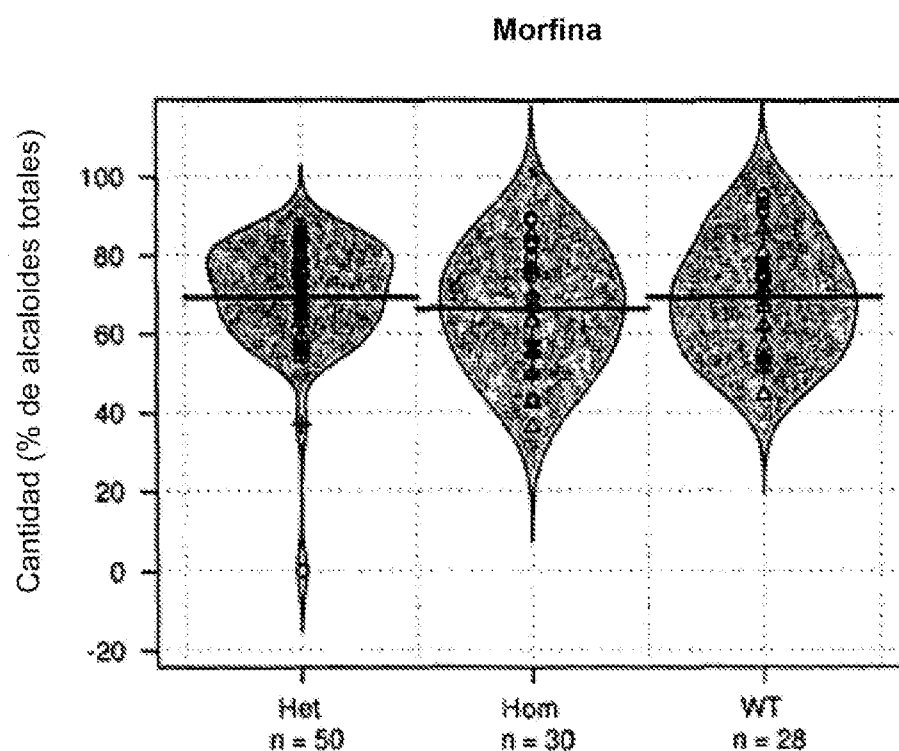


Figura 2 B continuación

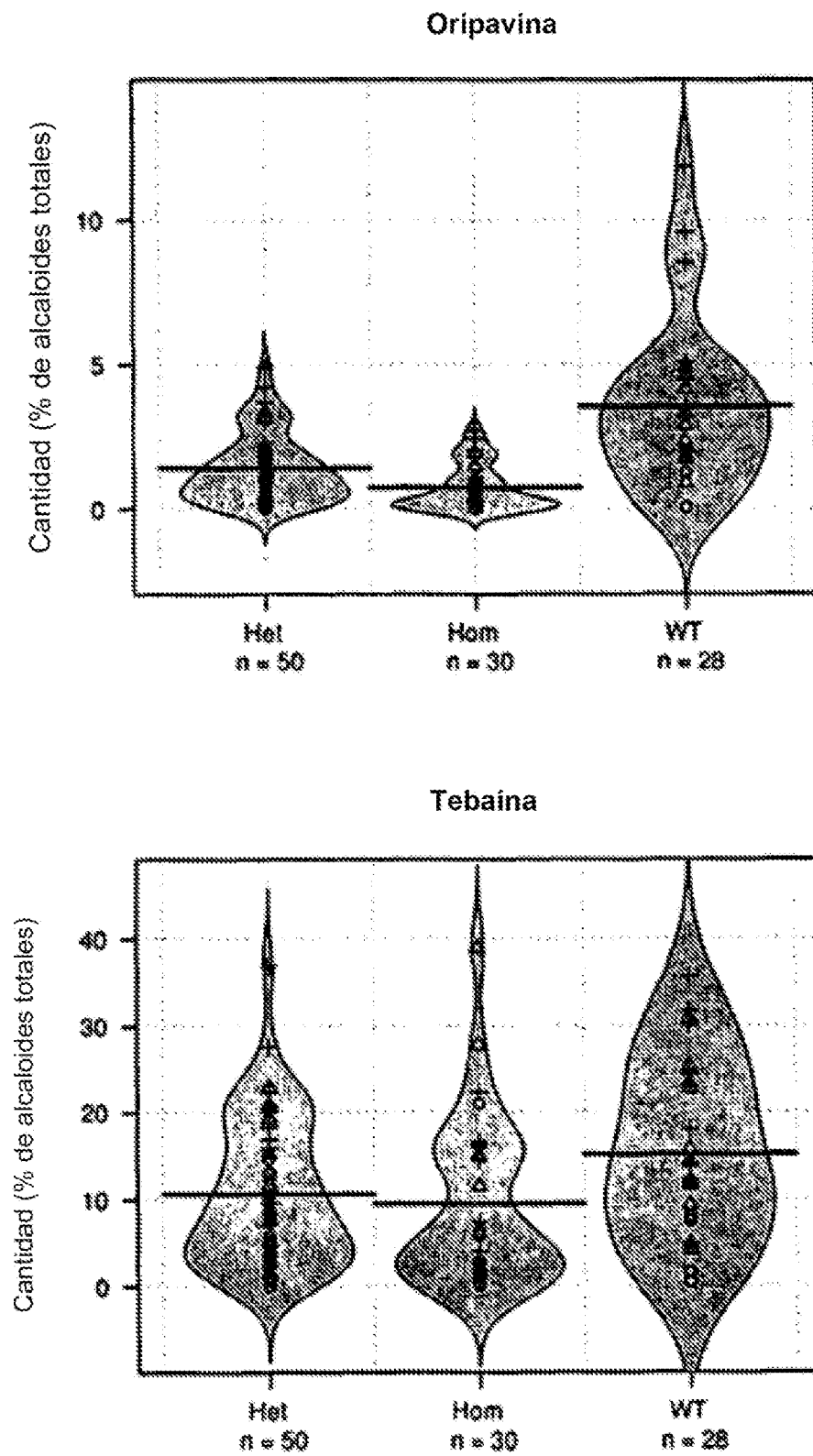


Figura 2B continuación

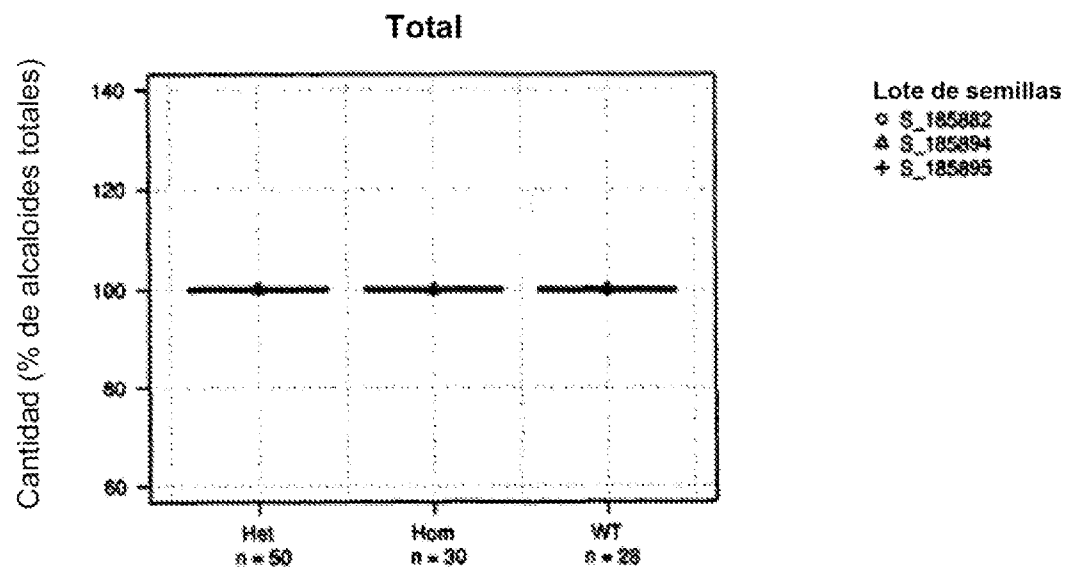


Figura 3

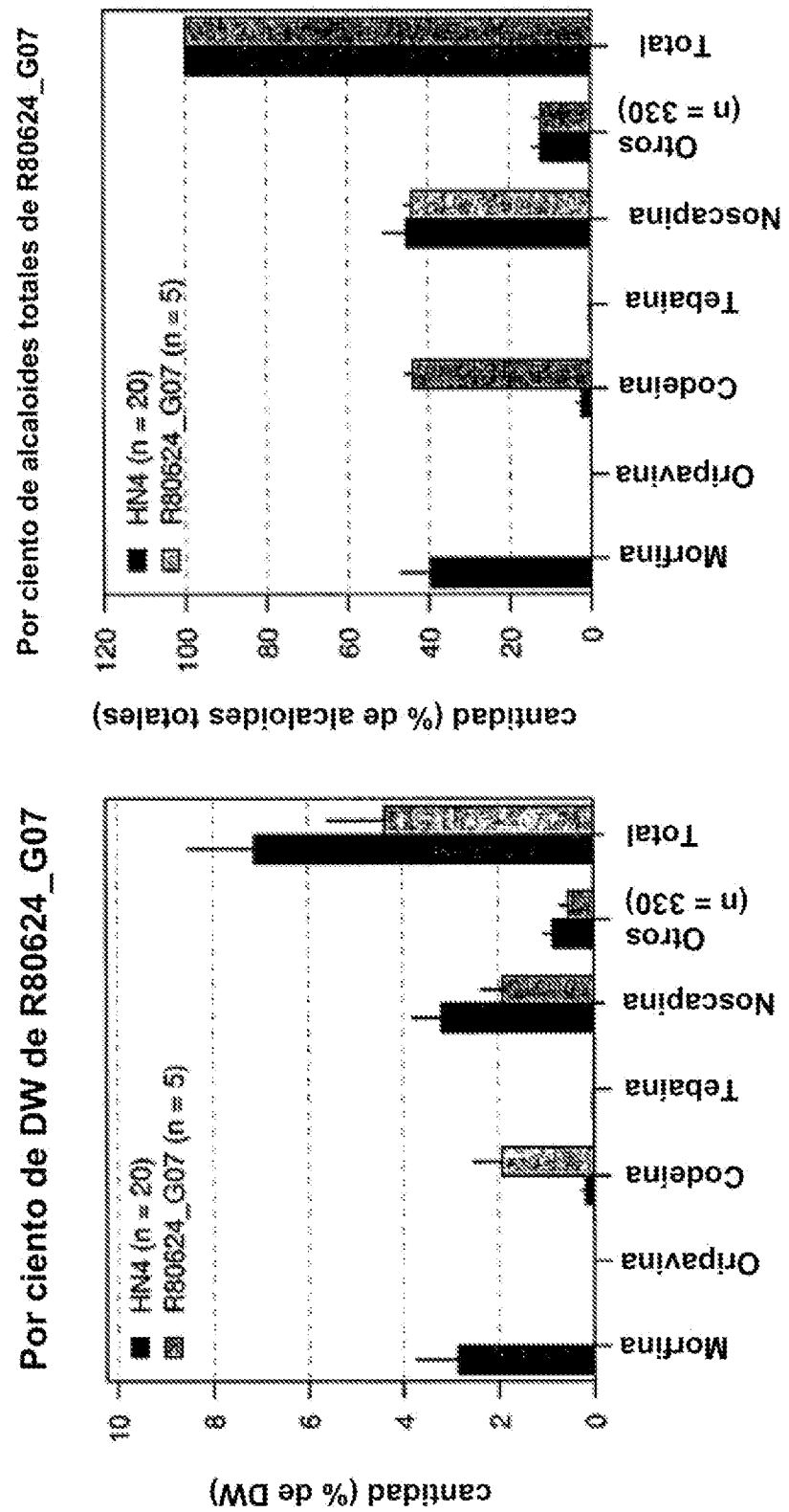


Figura 4

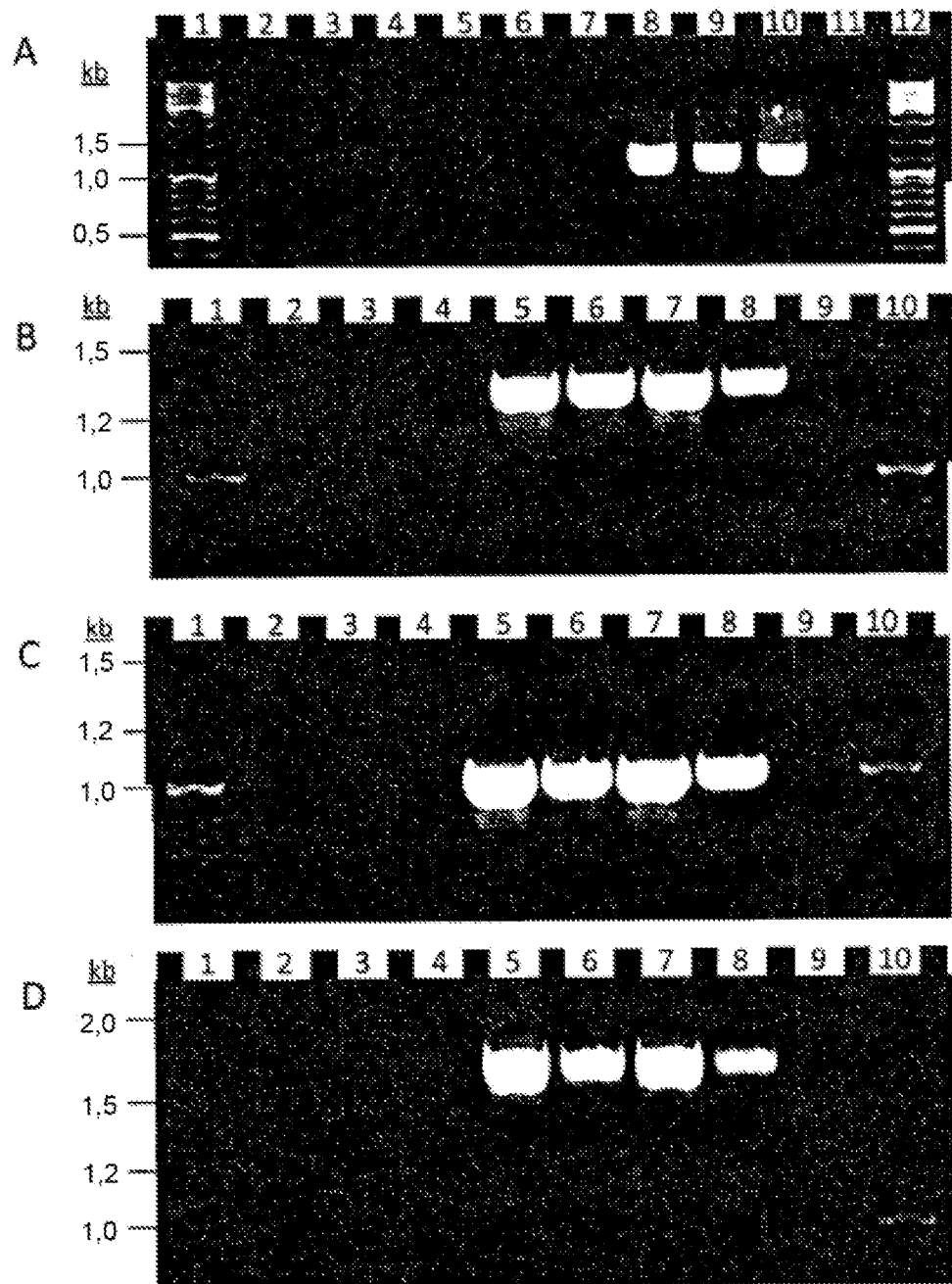


Figura 4:

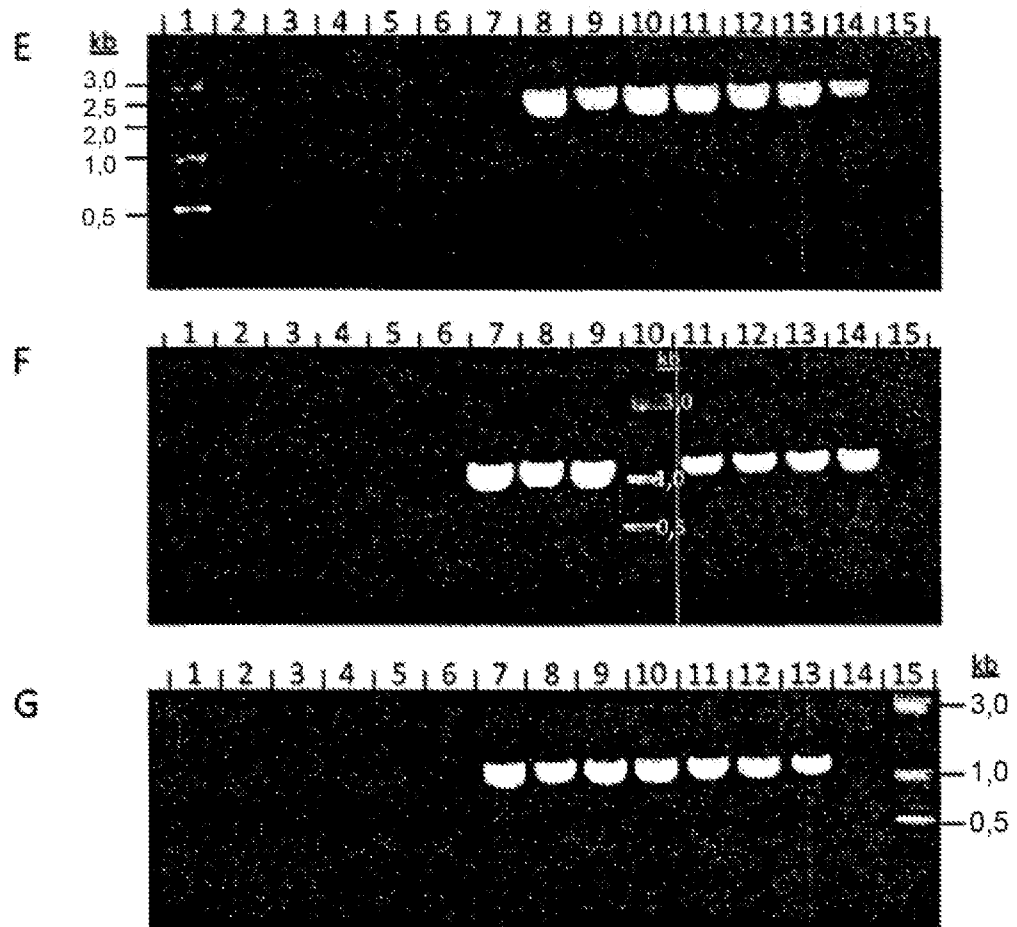


Figura 5

