

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 645 742**

51 Int. Cl.:

A01H 5/00 (2006.01)

C12N 9/00 (2006.01)

C12N 15/82 (2006.01)

C12N 9/04 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **25.04.2012** **PCT/US2012/034980**

87 Fecha y número de publicación internacional: **01.11.2012** **WO12149009**

96 Fecha de presentación y número de la solicitud europea: **25.04.2012** **E 12776946 (1)**

97 Fecha y número de publicación de la concesión europea: **02.08.2017** **EP 2701487**

54 Título: **Enzimas que codifican polinucleótidos de la senda biosintética de lignina del yute**

30 Prioridad:

29.04.2011 US 201161480668 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
07.12.2017

73 Titular/es:

**BANGLADESH JUTE RESEARCH INSTITUTE
(100.0%)**

**Manik Mia Avenue
Dhaka 1207, BD**

72 Inventor/es:

**ALAM, MAQSUDUL;
KHAN, HASEENA;
ZAMAN, MAHBOOB;
UDDIN, MOHAMMED, KAMAL;
HAQUE, MOHAMMED, SAMIUL;
ISLAM, MOHAMMED, SHAHIDUL y
AZAM, MUHAMMAD, SHAFIUL**

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 645 742 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

Enzimas que codifican polinucleótidos de la senda biosintética de lignina del yute

5 Campo de la invención

La invención se relaciona con la identificación y caracterización de varias porciones de la senda biosintética de lignina del yute. Más específicamente, la invención se relaciona con poli nucleótidos de las plantas de yute que codifican las enzimas responsables por la síntesis de lignina, y métodos para utilizar estos polinucleótidos y enzimas para la regulación de genes y la manipulación de la producción de lignina para dar fibras con un contenido de lignina deseado y otras características

Antecedentes de la invención

15 La lignina es un nombre colectivo para un heteropolímero aromático complejo de monolignoles (alcohol hidroxicinámico), usualmente derivado de la fenilalanina en un proceso multietapa. (Whetten, R. y Sederoff, R., (1995) Lignin Biosynthesis, Plant Cell, 7, pp. 1001-1013). Estos polímeros, depositados primariamente en las paredes celulares, aseguran la necesaria resistencia mecánica de los tallos de las plantas y, más importante, la hidrofobicidad de los tejidos vasculares de la planta. (Vanholme, R. et al. (2010) Lignin biosynthesis and structure, Plant Physiol, 153, pp. 895-905). Debido a su naturaleza hidrófoba, la lignina sirve como un componente principal de los tejidos vasculares y juega un papel esencial en el transporte de agua. Además de su papel estructural y orientado al transporte, la lignina es un componente clave del sistema de defensa de una planta. (Goujon, T. et al. (2003) Genes involved in the biosynthesis of lignin precursors in Arabidopsis thaliana, Plant Physiology and Biochemistry, 41, pp. 677-687). De manera no sorprendente, las condiciones ambientales influyen la cantidad de la lignina depositada. (Boerjan, W. et al. (2003) Lignin biosynthesis, Annu Rev Plant Biol, 54, pp. 519-546). Por ejemplo, la biosíntesis de lignina se induce en respuesta a diversas condiciones de estrés tal como heridas, tensión biótica, e infección por patógenos. La lignina limita la invasión de patógenos y protege los polisacáridos de la pared celular contra la degradación microbiana. (Vanholme et al., 2010).

30 Una gran parte de nuestro actual entendimiento de la biosíntesis de lignina viene del entendimiento completo de esta senda de *A. thaliana* y *P. trichocarpa*. (Goujon, et al., 2003; Shi, et al. (2010) Towards a systems approach for lignin biosynthesis in Populus trichocarpa: transcript abundance and specificity of the monolignol biosynthetic genes, Plant Cell Physiol, 51, pp. 144-163). Estos son tres monómeros básicos de monolignol: p-cumarilo, coniferilo, y alcoholes sinapilo. Estos monolignoles se incorporan en las tres unidades de lignina, o bloques de construcción: p-hidroxifenilo (H), guaiacilo (G), y siringilo (S). Ver la figura 1. Estos monolignoles difieren en el número de grupos metoxi. El P-hidroxifenilo (H) no tiene grupo metoxi, el guaiacilo (G) tiene un grupo metoxi, y el siringilo (S) tiene dos grupos metoxi. (Goujon et al., 2003). Sin embargo, además de estos tres monolignoles, algunos otros fenilpropanoides, tales como los aldehídos de hidroxicinámico, los ésteres de hidroxicinámico, y los acetatos de hidroxicinámico también se pueden incorporar. (Boerjan et al., 2003).

40 Después de la biosíntesis de estos bloques de construcción de lignina básicos, ellos se transportan a zonas lignificantes. En las zonas lignificantes, ocurre la polimerización mediante acoplamiento basado en radicales libres oxidativos mediante peroxidasas o lacasas y se forma una estructura similar a malla al reticular la concelulosa y hemicelulosa. (Boerjan et al., 2003; Vanholme, R. y otros (2008) Lignin engineering, Curr Opin Plant Biol, 11, págs. 278-285). La lignificación ocurre en diferentes fases durante el engrosamiento secundario de la pared celular cuando se completa la formación de la matriz de polisacáridos. La deposición de lignina está influenciada por la naturaleza de la matriz de polisacáridos. En la pared celular primaria, esta se encuentra como estructuras esféricas; mientras que, en la pared celular secundaria, esta forma lamelas. (Boerjan et al., 2003).

50 Sin importar el papel indispensable de la lignina en la vida de una planta, este es un factor limitante principal en el uso efectivo/eficiente de costos del material vegetal en la pulpa e industrias de biocombustibles. La lignina también limita el uso de biomasa para la producción de fibras, química, y de energía. La remoción de la lignina es un proceso muy costoso y estas industrias se beneficiarían del acceso a la biomasa que tiene menos lignina, o una lignina que sea fácil de degradar. En las últimas pocas décadas, se ha logrado algún entendimiento de la senda biosintética de la lignina, aunque porciones del proceso no se han entendido completamente.

A pesar de la importancia de la síntesis de lignina en el bienestar total de la planta de yute, así como también su impacto en varios aspectos de la calidad de la fibra, actualmente no existe información disponible que detalle la biosíntesis de lignina en yute. Por lo tanto, existe la necesidad de identificar, aislar y utilizar genes y enzimas provenientes de la planta de yute que estén involucrados en la biosíntesis de la lignina. La presente invención se dirige esta necesidad.

Resumen de la invención

Un aspecto de la invención es una molécula de ácido nucleico aislada que tiene al menos un 90% de identidad de secuencia con una secuencia de ácido nucleico seleccionada del grupo que consiste de: SEQ ID NO: 1, 3, 5, 7, 9 y 15.

5 En ciertas realizaciones, la presente invención se relaciona con una cualquiera de las secuencias de polinucleótido anteriormente mencionadas o las secuencias de polipéptido, en donde dicha secuencia tiene al menos 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con cualquiera de las secuencias identificadas por una SEQ ID NO.

10 Un aspecto de la invención es un vector de expresión que comprende una molécula de ácido nucleico aislada de la presente invención.

Un aspecto de la invención es un anticuerpo aislado o fragmento de unión de antígeno del mismo que se une específicamente a una molécula de polipéptido de la presente invención.

15 Un aspecto de la invención es una célula vegetal transfectada por un vector de la presente invención.

Un aspecto de la invención es un material derivado de una planta transgénica de la presente invención.

20 Un aspecto de la invención es una semilla de una planta transfectada por un vector de la presente invención.

Un aspecto de la invención es un método para hacer una planta transgénica, que comprende las etapas de transfectar al menos una célula vegetal con un vector de la presente invención, y hacer crecer al menos una célula vegetal en una planta.

25 Un aspecto de la invención es un método para mejorar el crecimiento, la producción de fibra, la resistencia de la fibra, la resistencia a enfermedades, o la utilización de agua en una planta de yute, que comprende incorporar en una planta de yute una secuencia de ácido nucleico no nativa de la presente invención

30 Breve descripción de los dibujos

Figura 1: Senda Propuesta de biosíntesis de monolignol del yute.

35 Figuras 2a y 2b: alineamiento de la secuencia de proteínas de ColCAD1, ColCAD2, ColCAD3, ColCAD4, ColCAD5, ColCAD6, y ColCAD7 con secuencias de proteínas vegetales CAD.

Figura 3: Alineamiento de la secuencia de proteínas de ColCCoAOMT1, ColCCoAOMT2 y ColCCoAOMT3 con las secuencias de proteína CCoAOMT vegetal.

40 Figura 4: alineamiento de la secuencia de proteínas de Col4CL1, Col4CL4 y Col4CL6 con las secuencias de proteínas 4CL vegetales.

Figura 5: alineamiento de la secuencia de proteínas de Col6HCT1 con las secuencias de la proteína 6HCT vegetales.

45 Figura 6: alineamiento de la secuencia de proteínas de ColC3H con las secuencias de la proteína C3H vegetales.

Figura 7: alineamiento de la secuencia de proteínas de ColC4H1 y ColC4H2 con las secuencias de la proteína C4H de los vegetales.

50 Figura 8: alineamiento de la secuencia de proteína del ColPAL1 y ColPAL2 con las secuencias de la proteína PAL de vegetal.

Figura 9: alineamiento de la secuencia de proteína de ColCCR2 con las secuencias de proteína CCR vegetal.

55 Figura 10: alineamiento de la secuencia de proteína de ColCCR3 con las secuencias de proteína CCR vegetal.

Figura 11: alineamiento de la secuencia de proteína de ColF5H con las secuencias de la proteína F5H vegetal.

60 Figura 12: alineamiento de la secuencia de proteína de ColCOMT con las secuencias de proteína COMT.

Figura 13: gel de ADN de ColCAD2.

Figura 14: gel de ADN de ColCCoAOMT1.

65 Figura 15: gel de ADN de Col4CL1.

Figura 16: gel de ADN de ColCCR3.

Figura 17: gel de ADN de ColF5H.

Descripción detallada de la invención

Diez de las familias de enzimas conocidas están asociadas con la biosíntesis de monolignol. (Goujon et al., 2003). Las familias son la PAL (fenilalanina amonía-liasa), C4H (cinamato-4-hidroxilasa), 4CL (4-cumarato: CoA ligasa), HCT (p-hidroxicinamoil-CoA: shikimato/quinato p-hidroxicinamoil transferasa), C3H (4 -cumarato 3-hidroxilasa), CCoAOMT (cafeoil-CoA O-metiltransferasa), CCR (cinamoil-CoA reductasa), F5H (ferulato 5-hidroxilasa), COMT (ácido cafeico O-metiltransferasa) y CAD (cinamil alcohol deshidrogenasa). Un esquema propuesto de la senda de biosíntesis del monolignol en el yute se muestra en la Figura 1.

La senda de biosíntesis de lignina en el yute debe su complejidad, en parte, a la presencia de varias enzimas multifuncionales, y a las enzimas constituyentes que abarcan varias familias de genes diferentes. La primera enzima de la senda fenilpropanoide es PAL (fenilalanina amonía-liasa), que origina la desaminación de la fenilalanina, produciendo ácido cinámico. La segunda enzima de la senda, C4H (cinamato 4-hidroxilasa), convierte el ácido cinámico en ácido 4-hidroxicinámico, que es seguido por las subsecuentes etapas de hidroxilación y metilación en la medida en que la senda se ramifica. La enzima 4CL cataliza la ligación CoA de ácidos hidroxicinámicos, generando precursores fenólicos activados de la biosíntesis de lignina. (Hu et al. (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees, Nat Biotech, 17, pp. 808-812).

La siguiente enzima en la senda de monolignol (HCT) cataliza la producción de ésteres de p-cumaroilo-shikimato/quinato, que son los sustratos para C3H. Se demostró que HCT transfiere el grupo acilo de p-cumaroil-CoA a shikimato o quinato. (Hoffman et al. (2005) Plant Biosystems, v. 139, No. 1, pp. 50-53). Las etapas de hidroxilación en C3 y C5 se realizan mediante dos enzimas P450 de citocromo, 4-cumarato 3-hidroxilasa (C3H) y ferulato 5-hidroxilasa (F5H), respectivamente. Las etapas de metilación se realizan mediante CCoAOMT (cafeoil-coenzima A (CoA) O-metiltransferasa) y COMT (cafeico-O-metiltransferasa). El CCoAOMT es una enzima bifuncional que convierte la cafeoil-CoA a feruloil-CoA y 5-hidroxiferuloil-CoA a sinapoil-CoA y juega un papel en la síntesis de polisacáridos feruloilados. (Inoue et al., 1998). Se ha demostrado que CCoAOMT está involucrado en la biosíntesis de lignina en los elementos traqueales diferenciales de la *Zinnia elegans* (Ye, Z. H. and Varner J. E. (1995) Differential expression of two O-methyltransferases in lignin biosynthesis in *Zinnia elegans*, Plant Physiol. 108, pp. 459-467). La CCoAOMT está involucrada en el refuerzo de la pared celular vegetal, y también está involucrada en la respuesta a heridas o la infestación de patógenos mediante la formación creciente de polímeros de ácido ferúlico unidos a la pared celular.

Las enzimas adicionales involucradas en la senda de biosíntesis de monolignol son la cinamoil coenzima A reductasa (CCR) y la cinamoil alcohol deshidrogenasa (CAD). El CCR cataliza la reducción de hidroxicinamoil CoA ésteres para producir cinamaldehydos, mientras que el CAD cataliza su reducción a alcoholes cinamilo. (Goujon et al., 2003).

Una de las últimas enzimas involucradas en la senda del monolignol es la cinamil alcohol deshidrogenasa (CAD), que cataliza la NADPH dependiente de la conversión de coniferaldehído, 5-hidroxi-coniferaldehído, y sinapaldehído a los alcoholes correspondientes (Kim, S. J. et al. (2004) Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in *Arabidopsis*, Proc. Natl. Acad. Sci. USA, 101, pp. 1455-60). En *Arabidopsis*, los mutantes únicos de los genes CAD AtCAD-C y AtCAD-D se encontraron que tenían actividades de CAD inferiores, y el doble mutante, obtenido mediante el cruce de dos mutantes, tenía 40% de disminución en el contenido de lignina en el tallo, lo que demuestra que estos son los genes CAD principales involucrados en la síntesis de lignina del tallo. (Sibout, R. et al. (2005) Cinnamyl Alcohol Dehydrogenase-C and -D are the primary genes involved in lignin biosynthesis in the floral stem of *Arabidopsis*, Plant Cell, 17, pp. 2059-76).

Dos enzimas son específicas para la senda de biosíntesis de monolignol. Ellas son el ácido cafeico O-metiltransferasa (COMT) y el cinamoil coenzima A reductasa (CCR). El COMT se identificó primero en angiospermas. El COMT es capaz de convertir el ácido cafeico en ácido ferúlico, así como también convertir el ácido 5-hidroxiferúlico a ácido sinápico. (Dixon, R. A., et al. (2001) The biosynthesis of monolignols: a "metabolic grid," or independent pathways to guaiacyl and syringyl units? Phytochemistry, 57, pp. 1069-1084). La regulación hacia abajo del gen COMT en maíz (*Zea mays*) ha mostrado que origina una reducción significativa de la actividad COMT (una caída del 70 al 85%), dando como resultado la modificación sin el contenido de lignina y la composición, indicando que esta enzima es una enzima clave para la síntesis de lignina.

El ácido ferúlico generado por el COMT se puede hidroxilar mediante ferulato 5 hidroxilasa (F5H), que es un citocromo monooxigenasa dependiente de P450, para formar ácido 5-hidroxi-ferúlico. El F5H también es capaz de hidroxilar coniferaldehído y alcohol coniferilo para formar 5-hidroxi-coniferaldehído y alcohol 5-hidroxi-coniferilo, respectivamente. (Meyer, K. et al. (1996) Ferulate-5-hydroxylase from *Arabidopsis thaliana* defines a new family of cytochrome P450-dependent monooxygenases, Proc. Natl. Acad. Sci. USA, 93, pp. 6869-74). El F5H se cree que es

una etapa limitante de la velocidad en la biosíntesis de lignina de siringilo, una propuesta soportada por la observación de que el mutante de *Arabidopsis* deficiente en la expresión F5H también está afectado en el nivel de la acumulación de ésteres de sinapato en silículas y semillas. (Ruegger, M. et al. (1999) Regulation of ferulate-5-hydroxylase expression in *Arabidopsis* in the context of sinapate ester biosynthesis, *Plant Physiol.*, 119, pp. 101-10).

La segunda enzima específicamente involucrada en la biosíntesis de lignol, CCR, cataliza la conversión de feruloil CoA y 5-hidroxi-feruloil CoA en coniferaldehído y 5-hidroxi-coniferaldehído, respectivamente. Esta etapa conduce directamente a la biosíntesis de unidades de lignina G (coniferaldehído) y S (5-hidroxi-coniferaldehído). (Ma et al., 2005). En tabaco, la regulación hacia abajo del gen CCR que utiliza construcciones contrasentido, produjo plantas con desarrollo anormal y crecimiento reducido, así como también morfología anormal de la hoja y vasos colapsados. También hubo una reducción asociada en el nivel de los compuestos de lignina G. (Ralph, J. et al. (1998) NMR characterization of altered lignins extracted from tobacco plants down-regulated for lignification enzymes cinnamylalcohol dehydrogenase and cinnamoyl-CoA reductase, *Proc. Natl. Acad. Sci USA*, 95, pp. 12803-8).

Identificación computacional de los genes y transcriptos

De manera notoria, hemos determinado las secuencias de las enzimas de yute involucradas en la biosíntesis de lignina. La senda de la biosíntesis de lignina ha sido bien caracterizada en cada enzima y cada enzima se codifica por una familia de genes en la mayoría de las especies de plantas. Un total de 106 secuencias de genes de *A. thaliana* y *P. trichocarpa* fueron recuperadas del NCBI y la base de datos del genoma de *P-trichocarpa* (http://genome.jgi-psfz.org/Poptr1_1/). (Goujon et al., 2003; Shi et al., 2010). Los genes de la biosíntesis de monolignol de yute se identificaron de los modelos de gen del montaje del genoma de *Corchorus olitorius* y los datos de transcriptoma de *C. olitorius* y *C. capsularis* utilizando el programa BLASTN con el corte evalúe en 1e-20. (Altschul, S.F., et al. (1990) Basic local alignment search tool, *J Mol Biol*, 215, pp. 403-410). Los contigios gADN resultantes fueron sometidos a una predicción de modelo de gen que utiliza el software AUGUSTUS (Stanke, M. et al. (2004) AUGUSTUS: a web server for gene finding in eukaryotes, *Nucleic Acids Research*, 32, W309-W312). Los modelos del gen y los isotigos de los datos de transcriptoma de *C. olitorius* y *C. capsularis* fueron buscados contra el NCBI nr base de datos (no redundante) para confirmación adicional. Para el *C. olitorius*, los isotigos fueron mapeados sobre los modelos de gen predicho utilizando GMAP (con un 95% de valor de corte). (Wu, T.D. and Watanabe, C.K. (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences, *Bioinformatics*, 21, pp. 1859-1875).

Un alineamiento de la secuencia de aminoácido de proteínas putativas codificadas por los genes ColCAD con otras proteínas CAD disponibles en la base de datos NCBI, utilizando el programa CLUSTAL W, se muestra en las Figuras 2a y 2b. La siguiente es una lista de las proteínas alineadas con las proteínas ColCAD putativas, con los números de acceso del GeneBank en paréntesis: PtcCADL4 (proteína similar a deshidrogenasa de alcohol cinámico de *Populus trichocarpa*, CADL4, gi224138226); RcoCAD (alcohol deshidrogenasa de *Ricinus communis*, putativa, gi25558709); FraCAD (*Fragaria x ananassa*, cinamil alcohol deshidrogenasa, gi13507210) (Chandler et al. (2002) Cloning, expression and immunolocalization pattern of a cinnamyl alcohol dehydrogenase gene from strawberry (*Fragaria x ananassa*), *J. Exp. Bot.*, 53 (375), pp. 1723-1734); GhiCAD5 (*Gossypium hirsutum*, cinnamyl alcohol deshidrogenasa 5, gi268528129); PtcCAD (*Populus trichocarpa*, gi183585165) ((2010) Towards a systems approach for lignin biosynthesis in *Populus trichocarpa*: transcript abundance and specificity of the monolignol biosynthetic genes, *Plant Cell Physiol.*, 51 (1), pp. 144-163); GhiCAD3 (*Gossypium hirsutum*, GhiCAD3 (*Gossypium hirsutum*, gi229368450) (Genes of phenylpropanoid pathway cloning and expression in developing cotton fibre); y and GhiCAD (*Gossypium hirsutum*, gi166865124) ((2009) Molecular and biochemical evidence for phenylpropanoid synthesis and presence of wall-linked phenolics in cotton fibers, *J Integr Plant Biol*, 51 (7), pp. 626-637).

Un alineamiento de la secuencia de aminoácido de las proteínas putativas codificadas por los genes ColCCoAOMT con otras proteínas CCoAOMT disponibles en la base de datos NCBI, que utilizan el programa CLUSTAL W, se muestra en la Figura 3. La siguiente es una lista de las proteínas alineadas con las proteínas ColCCoAOMT putativas, con los números de acceso de GeneBank en paréntesis: PtrCCoAOMT (*Populus tremuloides*, gi3023436); GhiCCoAOMT2 (*Gossypium hirsutum*, gi229368460); y GhiCCoAOMT1 (*Gossypium hirsutum*, gi253509567).

Un alineamiento de la secuencia de aminoácido de las proteínas putativas codificadas por los genes Col4CL con otras proteínas 4CL disponibles en la base de datos NCBI, que utilizan el programa CLUSTAL W, se muestra en la Figura 4. La siguiente es una lista de las proteínas alineadas con las proteínas Col4CL putativas, con los números de acceso de GeneBank en paréntesis: Ccap4CL1 (*Corchorus capsularis*, gi294514718); Rco4CL (*Ricinus communis*, gi255565415); y Ptc4CL (*Populus trichocarpa*, gi224074401).

Un alineamiento de la secuencia de aminoácido de las proteínas putativas codificadas por los genes Col6HCT con otras proteínas Col6HCT disponibles en la base de datos NCBI, que utilizan el programa CLUSTAL W, se muestra en la Figura 5. La siguiente es una lista de las proteínas alineadas con las proteínas Col6HCT putativas, con los números de acceso de GeneBank en paréntesis: CycarHCT (*Cynara cardunculus*, gi: 73671233) (2007) Isolation and functional characterization of a cDNA coding a hydroxycinnamoyltransferase involved in phenylpropanoid biosynthesis in *Cynara cardunculus*, *BMC Plant Biol.* 7, 14); and PtcHCT (*Populus trichocarpa*, gi183585181).

Un alineamiento de la secuencia de aminoácido de las proteínas putativas codificadas por los genes ColC3H con otras proteínas C3H disponibles en la base de datos NCBI, que utilizan el programa CLUSTAL W, se muestran en la Figura 6. La siguiente es una lista de las proteínas alineadas con las proteínas ColC3H putativas, con los números de acceso de GeneBank en paréntesis: EglC3H (*Eucalyptus globulus*, gi: 295413824); PtcC3H (*Populus trichocarpa*, gi: 224139664); y PalxPgrC3H (*Populus alba* X *Populus grandidentata*, gi166209291).

Un alineamiento de la secuencia de aminoácidos de las proteínas putativas codificadas por los genes ColC4H con otras proteínas C4H disponibles en la base de datos NCBI, que utilizan el programa CLUSTAL W, se muestran en la Figura 7. La siguiente es una lista de las proteínas alineadas con las proteínas ColC4H putativas, con el número de acceso de GeneBank en paréntesis: GarC4H (*Gossypium arborium*, gi9965897) y GarC4H (*Gossypium arborium*, gi9965899).

Un alineamiento de la secuencia de aminoácido de las proteínas putativas codificadas por los genes ColPAL y otras proteínas PAL disponibles en la base de datos NCBI, que utilizan el programa CLUSTAL W, se muestra en la Figura 8. La siguiente es una lista de las proteínas putativas alineadas con las proteínas ColPAL putativas, con los números de acceso de GeneBank en paréntesis: JcoPAL (*Jatropha curcas*, gi113203757) y PtrPAL (*Populus trichocarpa*, gi: 183585195).

Un alineamiento a la secuencia de aminoácido de las proteínas putativas codificadas por los genes ColCCR2 con otras proteínas CCR disponibles en la base de datos NCBI, que utilizan el programa CLUSTAL W, se muestra en la Figura 9. La siguiente es una lista de las proteínas alineadas con las proteínas ColCCR2 putativas, con los números de acceso de GeneBank en paréntesis: AthCCR (*Arabidopsis thaliana*, gi: 15237678); CofCCR (*Camellia oleifera* gi228480464); y AlyCCR (*Arabidopsis lyrata*, gi: 297793385).

Un alineamiento de secuencia de aminoácido de las proteínas putativas codificadas por los genes ColCCR3 con otras proteínas CCR disponibles en la base de datos NCBI, que utiliza el programa CLUSTAL W. La siguiente es una lista de las proteínas alineadas con las supuestas proteínas ColCCR3 putativas, con los números de acceso de GeneBank en paréntesis: RcoCCR (*Ricinus communis*, gi: 255556687) y AthCCR (*Arabidopsis thaliana*, gi: 15226955).

Una alineación de la secuencia de aminoácidos de las proteínas putativas codificadas por los genes ColF5H con otras proteínas F5H disponibles en la base de datos NCBI, que utilizan el programa CLUSTAL W, se muestra en la Figura 11. La siguiente es una lista de las proteínas alineadas con las proteínas ColF5H putativas, con los números de acceso de GeneBank en paréntesis: EglF5H (*Eucalyptus globules*, gi: 255970299) y PtcF5H (*Populus trichocarpa*, gi: 6688937).

Un alineamiento de la secuencia de aminoácido de las proteínas putativas codificadas por los genes ColCOMT con otras proteínas COMT disponibles en la base de datos NCBI, que utiliza el programa CLUSTAL W, se muestra en la Figura 12. La siguiente es una lista de las proteínas alineadas con las proteínas ColCOMT putativas, con los números de acceso de GeneBank en paréntesis: GhiCOMT (*Gossypium hirsutum*, gi: 253509569) y EcaCOMT (*Eucalyptus camaldulensis*, gi: 262474806).

Análisis del motivo de las regiones promotoras

Para cada uno de los modelos de gen predichos, ambas cadenas de las regiones corriente arriba de 2000 bp fueron extraídas e investigadas para secuencias de motivos cis contra la base de datos PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) (Lescot, M., et al. (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences, Nucleic Acids Res, 30, pp. 325-327). Si cualquier porción de las secuencias seleccionadas se encontrará como traslapando un gen cercano, la parte de la región corriente arriba se excluye de análisis adicional. Se compiló una lista de motivos importantes que son conocidos por estar involucrados en la respuesta a varios procesos y tensiones de desarrollo Tabla 1.

Tabla 1: Lista de motivos encontrados en la región promotora de los genes biosintéticos de monolignol de yute

		Acido	Capacidad de respuesta unión	herida	Mejorador de transcripci ón	Ácido abscísico	Etileno	Auxina	Elemento de respuesta	Capacidad de respuesta
PAL	Contig_310473_g31	*								
	Contig_301819_g4		*							
	Contig_302040_g24		*			*			*	
	Contig_309920_g2			*	*	*				
C4H	Contig_311419_g56									
	Contig_307848_g46	*					*			
	Contig_384302_g42									
4CL	Contig_331398_g112									
	Contig_299467_g73							*	*	
	Contig_337813_g88				*					
	Contig_306165_g86								*	*
	Contig_309926_g132						*	*		
	Contig_304308_g125	*		*	*				*	
	Contig_305998_g57	*	*							
HCT	Contig_308231_g128					*				
C3H	Contig_597141_g180									*
CCoAO MT	Contig_306979_g168		*						*	
	Contig_308402_g160									
	Contig_296218_g172	*	*		*				*	
	Contig_311951_g178				*					
CCR	Contig_310406_g194							*		
	Contig_297048_g201		*							
	Contig_294439_g187	*								
	Contig_309320_g191		*							
FSH	Contig_304192_g207		*							
	Contig_595270_g204					*				
COMT	Contig_299952_g235		*							
CAD	Contig_335778_g278						*			

	Contig_95812_g280										
	Contig_321021_g281	*									
	Contig_808760_g282	*									
	Contig_808760_g284					*					
	Contig_356907_g287							*			
	Contig_356907_g288				*		*				
	Contig_355903_g289					*					
	Contig_306706_g300		*		*					*	
	Contig_304576_g307	*	*		*					*	

Los polinucleótidos de la presente invención fueron aislados por secuenciamiento de alto rendimiento de las bibliotecas de cADN que comprenden tejido de plantas de yute recolectados de *Corchorus olitorius* L. Algunos de los polinucleótidos de la presente invención pueden ser secuencias parciales, porque, ellas no representan un gen de longitud completa que codifique un polinucleótido de longitud completa. Tales secuencias parciales se pueden extender al analizar y secuenciar varias bibliotecas de ADN que utilizan cebadores y / o sondas y técnicas de hibridación y / o PCR bien conocidas. Las secuencias parciales se pueden extender hasta un marco de lectura abierto que codifica un polipéptido, un polinucleótido de longitud completa, un gen capaz de expresar un polipéptido, o se identifica otra porción útil del genoma.

La identificación del ADN genómico y el ADN de especies heterólogas se puede lograr mediante técnicas de hibridación de ADN/ADN estándar, bajo condiciones apropiadamente estrictas, utilizando todo o parte de una secuencia de polinucleótidos como una sonda para seleccionar una biblioteca apropiada. Alternativamente, las técnicas PCR que utilizan cebadores de oligonucleótidos que son diseñados con base en el ADN genómico conocido, cADN, o secuencias de proteínas se pueden utilizar para amplificar e identificar las secuencias genómicas y de cADN.

Los polipéptidos de la presente invención se pueden producir al insertar una secuencia de polinucleótido de la presente invención que codifica un polipéptido deseado en un vector de expresión y expresar el polipéptido en un huésped apropiado. Cualquiera de una variedad de vectores de expresión conocidos por aquellos expertos en la técnica se podría emplear. La expresión se puede lograr en cualquier célula huésped apropiada que se ha transformado o transfectado con un vector de expresión que contiene una molécula de polinucleótido que codifica un polipéptido recombinante. Las células huéspedes adecuadas incluyen procariotas, levaduras y células eucariotas superiores.

Los polipéptidos que comprenden las sendas biosintéticas de lignina purificados de yute, o producidas mediante métodos recombinantes, se pueden utilizar para generar anticuerpos monoclonales, fragmentos de anticuerpos o derivados como se define aquí, de acuerdo con métodos conocidos. Los anticuerpos que reconocen y unen fragmentos de los polipéptidos que comprenden las sendas biosintéticas de la lignina de la invención, también se contemplan, siempre y cuando los anticuerpos sean específicos para los polipéptidos que comprenden la senda biosintética de la lignina.

Las construcciones genéticas de la presente invención también pueden contener un marcador de selección que sea efectivo en las células vegetales, para permitir la detección de las células transformadas que contienen la construcción inventiva. Tales marcadores, que son bien conocidos en la técnica, típicamente confieren resistencia a una o más toxinas o crean señales visuales por su presencia bajo microscopio fluorescente. Alternativamente, la presencia de la construcción deseada en las células transformadas se puede determinar por medio de otras técnicas bien conocidas en el arte, tales como el Southern y los Western blots. La construcción genética de la presente invención se puede enlazar a un vector que tiene al menos un sistema de replicación, por ejemplo, *E. Coli* o levadura (*Saccharomyces cerevisiae*), por medio de la cual después de cada manipulación, la construcción resultante se puede clonar y secuenciar.

Las construcciones genéticas de la presente invención se pueden utilizar para transformar una variedad de plantas, tales como monocotiledóneas (por ejemplo, arroz) y dicotiledóneas (por ejemplo, yute, *Arabidopsis*). En una realización preferida, las construcciones genéticas inventivas se emplean para transformar el yute. Como se discutió

anteriormente, la transformación de una planta con una construcción genética de la presente invención se puede utilizar para producir contenido de lignina modificado en la planta.

Las técnicas para incorporar establemente construcciones genéticas en el genoma de plantas objetivo son bien conocidas en el arte e incluyen introducción mediada por *Agrobacterium tumefaciens*, electroporación, inyección en el tejido meristemático u órganos reproductivos, inyección en embriones inmaduros, y similares. La selección de la técnica dependerá de la planta/tejido/huésped objetivo a ser transformado.

El término "planta" incluye plantas completas, órganos/estructuras vegetativas de brotes (por ejemplo, hojas, tallos, y tubérculos), raíces, flores, y órganos/estructuras florales (por ejemplo, brácteas, sépalos, pétalos, estambres, carpelos, anteras y óvulos), semillas (incluyendo embriones, endospermas y recubrimiento de las semillas) y fruta (el ovario maduro), tejido vegetal (por ejemplo, tejido vascular, tejido del suelo, y similares) y células (por ejemplo, células de guarda, células de huevo, tricomas y similares) y progenia de las mismas. La clase de plantas que se puede utilizar en el método de la invención es generalmente amplio como la clase de plantas superiores e inferiores susceptibles a técnicas de transformación, que incluyen angiospermas (plantas monocotiledóneas y dicotiledóneas), gimnospermas, helechos, briofitas y algas multicelulares. Esta Incluye plantas de una variedad de niveles ploidies, que incluyen aneuploide, poliploide, diploide, haploide y hemigigoto.

Los geles de ADN de las reacciones de PCR que utilizan cebadores delanteros e inversos para varias enzimas de yute se muestran en las Figuras 13-17. En la Figura 13, el gel de ADN es de CAD2 de *Corchorus olitorius*. El carril 1 es el producto de PCR de CAD2 que utiliza el cADN como una plantilla. El cebador delantero y el cebador inverso son las SEQ ID NO. 53 y 54, respectivamente. El carril 2 es 1 Kb + escalera. En la Figura 14, el gel de ADN es de CCoAOMT1 de *Corchorus olitorius*. El carril 1 es el 1 Kb + escalera, y el Carril 2 es el producto de PCR de CCoAOMT1 que utiliza cADN como una plantilla. El cebador delantero y el cebador inverso son la SEQ ID NO. 55 y 56, respectivamente. En la Figura 15, el gel de ADN de 4CL1 de *Corchorus olitorius*. El carril 1 es un 1 Kb + escalera, y el carril 2 es el producto PCR de 4CL1 utilizando cADN como plantilla. El cebador delantero y el cebador inverso son la SEQ ID NO. 57 y 58, respectivamente. En la Figura 16, el gel de ADN es de CCR3 de *Corchorus olitorius*. El carril 1 es el 1 Kb + escalera, y el Carril 2 es el producto de PCR de CCR3 utilizando cADN como una plantilla. El cebador delantero y el cebador inverso son las SEQ ID NO. 59 y 60, respectivamente. En la Figura 17, el gel de ADN es de F5H de *Corchorus olitorius*. El carril 1 es el 1 Kb + escalera, y el Carril 2 es el producto de PCR de F5H que utiliza el cADN como una plantilla. El cebador delantero y el cebador inverso son las SEQ ID NO. 61 y 62, respectivamente.

Definiciones

Una célula que ha sido "transformada" o "transfectada" por ADN exógeno o heterólogo cuando dicho ADN se ha introducido dentro de la célula. El ADN de transformación puede o no estar integrado (legado covalentemente) al genoma de la célula. En procariotes, levaduras y células de mamíferos, por ejemplo, el ADN que se transforma puede ser mantenido en un elemento episomal tal como un plásmido. Con respecto a las células eucarióticas, una célula establemente transformada es aquella en la cual el ADN de transformación se ha integrado en el cromosoma de tal manera que este es heredado por las células de la hija a través de la replicación del cromosoma. En la práctica de la presente invención contempla una amplia variedad de células vegetales establemente transformadas.

Un "casete de expresión" se refiere a una construcción de ácido nucleico, que cuando se introduce en una célula huésped, da como resultado la transcripción y/o traducción de un ARN y/o el polipéptido, respectivamente. El casete de expresión puede incluir un ácido nucleico que comprende una secuencia promotora, con o sin una secuencia que contiene señales de poliadenilación de mARN, y uno o más sitios de enzimas de restricción ubicados corriente abajo del promotor que permiten la inserción de las secuencias de gen heterólogo. El casete de expresión es capaz de dirigir la expresión de una proteína heteróloga cuando el gen que codifica la proteína heteróloga está operablemente ligado al promotor mediante la inserción en uno de los sitios de restricción. El casete de expresión recombinante permite la expresión de la proteína heteróloga en una célula huésped cuando el casete de expresión que contiene la proteína heteróloga se introduce en la célula huésped. Los casetes de expresión se pueden derivar de una variedad de fuentes que dependen de la célula huésped a ser utilizada para expresión. Por ejemplo, un casete de expresión puede contener componentes derivados de fuentes virales, bacterianas, de insectos, de plantas o de mamíferos. En el caso de ambas expresiones de transgenes e inhibición de genes endógenos (por ejemplo, mediante contra sentido, o supresión de sentido), la secuencia de polinucleótido insertada requiere ser idéntica y puede ser "sustancialmente idéntica" a una secuencia del gen de la cual esta se deriva. Preferiblemente, el casete de expresión recombinante permite la expresión en una etapa temprana de infección y/o permite la expresión en sustancialmente todas las células de un organismo, tal como una planta. Ejemplos de los casetes de expresión adecuados para la transformación de plantas se pueden encontrar en las Patentes US Nos. 5,880,333 y 6,002,072; las Publicaciones de Patentes Internacionales Nos. WO/1990/002189 y WO/2000/026388; Ainley and Key (1990) Plant Mol. Biol., 14, pp. 949-967; and Birch (1997) Annu. Rev. Plant Physiol. Plant Mol. Biol., 48, pp. 297-326, todas las cuales se incorporan aquí mediante referencia.

El término "célula huésped" se refiere a una célula de cualquier organismo. Las células huéspedes preferidas se derivan de las plantas, bacterias, levaduras, hongos, insectos, u otros animales. El término "célula huésped

recombinante" (o simplemente "célula huésped") se refiere a una célula en la cual se ha introducido un vector de expresión recombinante. Se debe entender que el término "célula huésped" pretende referirse no solo a la célula objeto particular a la progenie de tal célula. Ya que ciertas modificaciones pueden ocurrir en las generaciones exitosas debido a la mutación o influencias ambientales, tal progenie puede, de hecho, no ser idéntica a la célula parental, pero están aún incluidas dentro del alcance del término "célula huésped" como se utiliza aquí. Los métodos para introducir las secuencias de polinucleótidos en varios tipos de células huésped son bien conocidos en la técnica. Se suministran células huésped o progenie de células huéspedes transformadas con los casetes de expresión recombinantes de la presente invención. Las células huésped pueden ser células vegetales. Preferiblemente, las células vegetales son células de yute.

El término "operablemente ligado" o "operablemente insertado" significa que las secuencias regulatorias necesarias para la expresión de las secuencias codificantes son colocadas en una molécula de ácido nucleico en las posiciones apropiadas con relación a la secuencia codificante con el fin de posibilitar la expresión de la secuencia codificante. Esta misma definición es algunas veces aplicada a la disposición de otros elementos de control de la transcripción (por ejemplo, mejoradores) en un casete de expresión. Las secuencias de control transcripcional y traduccional son secuencias regulatorias del ADN, tales como promotores, mejoradores, señales de poliadenilación, terminadores, y similares, que suministran la expresión de la secuencia codificante en una célula huésped.

Los términos "promotor", "región promotora" o "secuencia promotora" se refieren de manera general a las regiones regulatorias transcripcionales de un gen, que se pueden encontrar en el lado 5' o 3' de la región codificante, o dentro de la región codificante, o dentro de los intrones. Típicamente, un promotor es una región regulatoria de ADN capaz de unir ARN polimerasa en una célula e iniciar la transcripción de una secuencia codificante corriente abajo (dirección 3'). La secuencia promotora 5' típica se une a su terminal 3' mediante el sitio de iniciación de transcripción y se extiende corriente arriba (dirección 5') para incluir el número mínimo de bases o elementos necesarios para iniciar la transcripción a niveles detectables por encima del fondo. Dentro de la secuencia promotora está un sitio de iniciación de la transcripción (convenientemente definido por el mapeo con nucleasa S1), así como también dominios de unión de proteína (secuencias de consenso) responsables por la unión de la ARN polimerasa.

El término "construcción de ácido nucleico" o "construcción de ADN" es algunas veces utilizado para referirse a una secuencia codificante o a secuencias operablemente ligadas a secuencias regulatorias apropiadas e insertadas en un casete de expresión para transformar una célula. Este término se puede utilizar intercambiamente con el término "ADN transformante" o "transgén". Tal construcción de ácido nucleico puede contener una secuencia codificante para un producto de gen de interés, junto con un gen marcador seleccionable y/o un gen reportero. El término "gen marcador seleccionable" se refiere a un gen que codifica un producto que, cuando se expresa, confiere un fenotipo seleccionable tal como resistencia a antibióticos sobre una célula transformada. El término "gen reportero" se refiere a un gen que codifica un producto que es fácilmente detectable por métodos estándar, directa o indirectamente.

Una región "heteróloga" de una construcción de ácido nucleico es un segmento identificable (o segmentos) de la molécula de ácido nucleico dentro de una molécula mayor que no se encuentra en asocio con la molécula mayor en la naturaleza. Cuando la región heteróloga codifica un gen vegetal, el gen usualmente estará flanqueado por el ADN que no flanquea el ADN genómico vegetal en el genoma del organismo fuente. En otro ejemplo, la región heteróloga es una construcción donde la secuencia codificante misma no se encuentra en la naturaleza (por ejemplo, un cADN donde la secuencia codificante genómica contiene intrones, o secuencias sintéticas que tienen codones diferentes del gen nativo). Las variaciones alélicas o los eventos mutacionales de ocurrencia natural no dan origen a una región heteróloga de ADN como se definió aquí. El término "construcción de ADN" también se utiliza para referirse a una región heteróloga, particularmente una construida para el uso en la transformación de una célula.

El término "vector" pretende referirse a una molécula de ácido nucleico capaz de transportar otro ácido nucleico al cual este se ha ligado. Un tipo de vector es un "plásmido", que se refiere a un bucle de ADN de doble cadena circular en el cual segmentos de ADN adicionales se pueden ligar. Otro tipo de vector es un vector viral, donde los segmentos de ADN adicionales se pueden ligar en el genoma viral. Ciertos vectores son capaces de replicación autónoma en una célula huésped en la cual ellos se introducen (por ejemplo, vectores bacterianos que tienen un origen bacteriano de replicación y vectores mamíferos episomales). Otros vectores se pueden integrar en el genoma de una célula huésped a la introducción en la célula huésped, y de esta manera se replican a lo largo del genoma huésped. Más aún, ciertos vectores son capaces de dirigir la expresión de los genes a los cuales ellos están operablemente ligados. Dichos vectores se denominan aquí como "vectores de expresión recombinantes" (o simplemente, "vectores de expresión"). En general, los vectores de expresión de utilidad en las técnicas de ADN recombinantes están a menudo en la forma de plásmidos. En la presente especificación, "plásmido" y "vector" se pueden utilizar intercambiamente como el plásmido está en la mayoría de las formas del vector comúnmente utilizadas. Sin embargo, la invención pretende incluir tales otras formas de vectores de expresión, tales como vectores virales (por ejemplo, retrovirus defectuoso de replicación, adenovirus y virus adenoasociados), que cumplen funciones equivalentes.

El "Porcentaje de identidad de secuencia" se determina al comparar dos secuencias alineadas óptimamente sobre una ventana de comparación, donde la porción de la secuencia de polinucleótido en la ventana de comparación

puede comprender adiciones o supresiones (por ejemplo, espacios) en comparación con la secuencia de referencia (que no comprende adiciones o supresiones) para el alineamiento óptimo de las dos secuencias. El porcentaje se calcula al determinar el número de posiciones en la cual la base de ácido nucleico idéntica o el residuo de aminoácido ocurre en ambas secuencias para producir el número de posiciones coincidentes, dividiendo el número de posiciones coincidentes por el número total de posiciones en la ventana de comparación, y multiplicando el resultado por 100 para producir el porcentaje de la identidad de secuencia.

El término "identidad sustancial" de las secuencias de polinucleótido significa que un polinucleótido comprende una secuencia que tiene al menos 25% de identidad de secuencia comparado con una secuencia de referencia determinada utilizando los programas descritos aquí; preferiblemente BLAST utilizando parámetros estándar, como se describió. Alternativamente, la identidad porcentual puede ser cualquier entero del 25% al 100%. Las realizaciones más preferidas incluyen secuencias de polinucleótidos que tienen al menos: 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% o 99% de identidad de secuencia comparado con la secuencia de referencia. Estos valores se pueden ajustar apropiadamente para determinar la identidad correspondiente de las proteínas codificadas por las dos secuencias de nucleótidos al tener en cuenta la degeneración del codón, la similitud del aminoácido, el posicionamiento del marco de lectura, y similares.

El término "identidad sustancial" de las secuencias de aminoácido (y de los polipéptidos que tienen estas secuencias de aminoácidos) normalmente significa identidad de secuencia de al menos el 40% comparado con la secuencia de referencia como se determinó utilizando los programas descritos aquí; preferiblemente BLAST utilizando parámetros estándar, como se describió. La identidad porcentual preferida de los aminoácidos puede ser cualquier entero de 40% a 100%. Las realizaciones más preferidas incluyen secuencias de aminoácidos que tienen al menos 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% o 99% de identidad de secuencia comparado con la secuencia de referencia. Los polipéptidos que son "sustancialmente idénticos" comparten secuencias de aminoácidos como se anotó anteriormente, excepto las posiciones de residuo que no son idénticas pueden diferir mediante los cambios de aminoácidos conservadores. Las sustituciones conservadoras de aminoácidos se refieren a la intercambiabilidad de residuos que tienen cadenas laterales similares. Por ejemplo, un grupo de aminoácidos que tienen cadenas laterales alifáticas es la glicina, alanina, valina, leucina e isoleucina; un grupo de aminoácidos que tienen cadenas laterales de hidroxilo alifáticas es la serina y la treonina; un grupo de aminoácidos que tienen cadenas laterales que contienen amida es la asparagina y la glutamina; un grupo de aminoácidos que tiene cadenas laterales aromáticas es la fenilalanina, tirosina y triptófano; un grupo de aminoácidos que tienen cadenas laterales básicas es la lisina, arginina y la histidina; y un grupo de aminoácidos que tienen cadenas laterales que contienen azufre es la cisteína y la metionina. Los grupos de sustitución de aminoácidos conservadores preferidos son: valina-leucina-isoleucina, fenilalanina-tirosina, lisina-arginina, alanina-valina, ácido aspártico-ácido glutámico y asparagina-glutamina.

Listado de secuencias

SEQ ID NO: 1

Nombre del Organismo: *Corchorus olitorius*

```

ATGGGCAGTTTACCAGAGCAAGAGCACCCCAAGAGGCTTTGGATGGGCAGCTAGAGACACTTC
TGGCCACCTTTCTCCCTTCAAATTCCTCAGAAGGGCAACAGGAGAAAAGATGTGGCATTCAAAGT
GCTTTATTGTGGGATATGCCACTCAGATCTTCATATGATCAAGAATGAATGGGGCACTGCCATCTA
CCCTATGGTCCCTGGGCATGAGATTGTTGGAGAAGTAACAGAGGTAGGAAGCAAGTAGAAAAAG
TTCAAAGTTGGAGACAAAGTTGGAGCTGGAGTCTTGTTAATTCGTGCCGCTCTTGCGATAACTGT
GCTAATAATCTGAAAACTACTGCCACAAGCCGTTTTCACCTATGCTGCAAAAACTACGATGGA
ACCATTACCTATGGAGGCTACTCCGACCATGGTTGCCGATGAGCACTTCATAATCCGAATTCCA
GACACTCTGCTCTCGACGCCGCGCTCCCTGCTCTGCGCCGAATCAGATTTATAGTCCTTTGA
GATATTCCAACCTCGACAAACCGGTTTCCATATTGGTGTGGTTGGCCTTGGTGGTTAGGCCATA
TGGCTGTCAAATTTGCCAAGGCTATGGGGCCAAAGTCAAGTGATTAGCACTCTCCCAACAAG
AAGAAGGAAGCTTTGGAATCTTGGTGCTGATTCTTTTGTATTAGCGAGAGGATCAGCT
CCAGACTGCCATGGGAACAATGGATGATCATTGATACAGTCTGCTCCACACCTTTACTGCC
ATTGATCGGTTTGTAAAGTCTCATGCAAAAGCTTATTTGGTTGGTCTCCAGACAAACCTTGAG
CTACATGTCTTCCCTATGATCATAGGAGGAAGACGGTGGCAGGAAGTGGCGTGGGAGGGATAG
AGGAGACTCAGGAGATGATGAATTTGACGCCAAATATGACTTGAAACAGACATTGAAGTTATA
CCCGTTGACTATGTCAACACTGCCATGGAAACGCTTGTCAAAGGCGATGTTAAATACAGATTGTG
ATCGACATTGGGAACACACTGAAGGCTACATCATCTTAA

```

Tipo: ADN

Longitud: 1089
Nombre de la Secuencia: cADN completo CAD1 completo
SEQ ID NO: 2
Nombre del Organismo: *Corchorus olitorius*

5

MGSLPEQEHKPEAFGWAARDTSGHLSPFKFSRRATGEKDVAFKVLVCG
ICHSDLHMIKNEWGTAIYPMVPGHEIVGEVTEVGSKEKFKVGDVGA
GVLVNSCRSCDNCANNLENYCPQAVFTYAAKNYDGTITYGGYSDTMV
ADEHFIIRIPDTLPDAAAPLLCAGITVYSPLRYFLDKPGFHIGVVGLG
GLGHMAVKFAKAMGAKVTIVISTPNKKKEALENLGADSLISAEQDQL
QTAMGMTDGIIDTVSAPHPLPLIGLLKSHAKLILVGLDPKPLELHVFP

MIIGRKTIVAGSGVGGIEETQEMMNFAAKYDLKPDIEVIPVDYVNTAME
RLVKGDVKYRFVIDIGNTLKATSS

TIPO: PRT
Longitud: 362
Nombre de la secuencia: péptido CAD1
SEQ ID NO: 3
Nombre del Organismo: *Corchorus olitorius*

10

15

ATGAGCAGATTGCCAGAGGAAGAGCACCTAACAGGCTTTGGATGGGCAGCCAGAGACACTT
CTGGAGTTCTCTCCCTTCAAAATTTCCAGAAAGGCAACAGGGGAGAAGGATGTGGCATTCAAG
GTGCTTTATTGTGGGATATGCCATTCTGATCTTCATGGTCAAGAATGAATGGGCGTTTCCATC
TACCTCTTGTCCCTGGGCACGAGATTGTTGGAGAAGTGACAGAAGTGGGAAGCAAGGTACAAA
AGTTCAAAGTTGGAGACAGAGTTGGTGTGGCTGCATGTTGGGTCATGTCTCTGTGATAGC
TGCAACCAACATCTCGAGAACTATTGCCCAAAATGATACTTATGGGGCAAAGTACTATGAT
GGAACATTACATATGAGGTTACTCTGACACTATGGTTCAGATGAACACTTCATTGTCCGAATT
CCTGAAAATTGGCACTTGATGCTGCTCTCTCTTTGTGCTGGAATCACAGTTTATAGCCCAT
GAAATATTATGACTTGACAAGCCTGGTTGTCATGTGGTGTGGTGGCCTTGGTGGCTAGGCC
ATATGGCTGTCAAATTTGCCAAGGCTATGGGGCCAAGGTACAGTAATTAGCACTTCTCCTAATA
AGAAGAAGGAAGCTTTGAAAATCTGGTGTGATTCAATTTTGGTCAGCAAGACCAAGATCAG
ATTCAGCCGCCATGGACACATTGGATGGAATCATTGATACAGTGTCACTCAACATCCTATCTG
CCATTGCTTGGGATGTTAAAGACTAATGGAAGCTTGTCTGTTGGTGTGACCGGAGAAACCACTT
GAGTTGCCGGCGTCCCTTTACTCGGAAAGAGGAGGTAGTAGCAGGAAGCATGATTGGGGGAA
TGAAGGAGACACAAGAGATGATTGATTTGCAGTAAACACAACATTAAACAGACATTGAAGTT
ATAGCTATGGATTATGTAACACTGCCATGGACCGCTTCTCAAAGCTGATGTCAAATACAGATT
GTCATTGACATTGGCAACACATTGAAACCAACCCCTTAA

Tipo: ADN
Longitud: 1086
Nombre de la secuencia: cADN completo CAD2
SEQ ID NO: 4
Nombre del organismo: *Corchorus olitorius*

20

MSRLPEEEHPNKAFGWAARDTSGVLSPFKFSRRATGEKDVAFKVLVCG
ICHSDLHMVKNEWGVSIIPLVPGHEIVGEVTEVGSKVQKFKVGDVGV

GCMVGSCHSCDSTNNLENYCPKMILTYGAKYYDGTITYGGYSDTMV
ADEHFIVRIPENLPLDAAAPLLCAGITVYSPLKYYGLDKPGLHVGVVGL
GGGLHMAVKFAKAMGAKVTIVISTPNKKKEALENLGADSLVSKDQD
QIQAAAMDTLDGIIDTVSAQHPIPLPLGMLKTNGKLVLVGAPEKPLEPA
FPLLGRRLVAGSMIGGMKETQEMIDFAAKHNIKPDIEVIAMDYVNTA
MDRLKADVKYRFVIDIGNTLKPTP

25

Tipo: PRT
Longitud: 361
Nombre de la secuencia: péptido CAD2
SEQ ID NO: 5
Nombre del organismo: *Corchorus olitorius*

30

ATGATCAAGAACGACTGGGGCTATTCCATCTATCCTCTCGTCCCTGGACATGAAATAGTTGGGGAA
GTAACAGAGGTGGTTCTAAGGTAAGCAAGTTCAAAGTTGGAGACAAGTTGGAGTTGGATACA
TGGTTGGATCATGCCGTTCTTGCGATGATTGCTCCGATAATCTTGAAAACTACTGCCGAAAATGA
TACCTACTTGTGGGGCAAAGTACCATGATGGAACATTACATATGGAGGTTTCTCTGACACTATGG
TTGCCGATGAACACTTCGTTGCCGAATCCGGACAACATGCCCTTGATGCGGGCGCTCTCTTCT
TTGTGCCGGAACACAGTTTACAGTCCAATGAAATATTATGGAATCGACAAGCCTGGTTTGCATTT
GGGTGTTGTTGGATTGGGAGGGCTTGCCATGTTGCTGTGAAATTTGCAAAAGCTATGGGGGCCA
AGGTGACAGTGATCAGCACCTCTCTAGTAAGAAGCAGGAAGCTTTGAAATCTTGTTGCTGAT
TCGTTTTGGTTAGCCGGGACGAAGATCAGCTTAAGGCTGCCAAGGGCACAATGAATGGTATAGT
GGATACGGTATCTGCCAAACATGATCTGCAGCCATTACTTGGACTGTTGAAGATCATGGAAAGCT
TGTTCTAATTGGTGTCCGGTAAAGCCATATGAGTACCAGCAGCTTTGATCTGGGGAGGAA
GCTAGTAGGAGGAAGTAACGTCGGAGGATTGGAAGAGACTCAAGAGATGATTGATTTGACGCA
AAACACAATGTGACAGCAACGTTGAAGTATTCCGATGGATTATGTGAACACTGCCTTTGAGCG
ACTTGCAAAAGCCGATGTTAGATATCGATTTGCTGTTGACATTGGCAACACTTGAAGACTTCGTC
TTAA

Tipo: ADN

Longitud: 921

5 Nombre de la secuencia: cADN completo CAD3

SEQ ID NO: 6

Nombre del organismo: *Corchorus olitorius*

MIKNDWGYSIYPLVPGHEIVGEVTEVSGKVSFKVGDKVGVMVGS
CRSCDDCSDNLENYCPKMIPTCGAKYHDGITYGGFSDTMVADEHFVV
RIPDNMPLDAAAPLLCAGTTVYSPMKYYGLDKPGLHLGVVGLGLGH
VAVKFAKAMGAKVTVISTSPSKKQEALILGADSLVSRDEDQLKAAKG
TMNGIVDTVSAKHDLQPLLGLLNHKGKLVIGVPVKPYELPAASLILGR
KLVGGSNVGGLLEETQEMIDFAAKHNVTANVEVIMDYVNTAFERLAK
ADVRYRFVVDIGNTLKTSS

Tipo: PRT

Longitud: 306

15 Nombre de la secuencia: péptido CAD3

SEQ ID NO: 7

Nombre del organismo: *Corchorus olitorius*

ATGGATTCTCAAACCAAATCTGATAACTGCCTTGGGTGGGCCCGAGAGATCCATCTGGAGTTCTG
TCACCCCTACACATTTAGTCGCCGGCCTCTTGGAAAGTAATGATGTTCCATAAAGATTACCCACTGTG
GAGTTGTTATGCTGATTTCAATTTGGTCTAGGAATAAGCATGGAGATAGTATGTACCTGTGGTCC
CCGGGCATGAAATTGCTGGTGTGTCAAGGAGGTGGCTCGAATGTTACCGCTTCAAGTTGGT
GATCCTGTGGAGTGGGAACCTATGTTAACTCATGCAGAGATTGTGAGTACTCAATGATGGCCTT
GAAGTTCATTGTGAAAAAATAGTTCTTACCTTAATTGTATTGATGAGGATGGAACAGTCACTAAA
GGCGGTTATTCTAACCATATCATTGTCCACGAAAGGTACTGCTAAGAATACCTAGCAATTATCCTT
TGGCTTCAGCAGCACCTTTGCTTTGTGCTGGCATCACTGTTATGCACCAATGATGCGTCATAACAT
GAATCAGCCTGGCAATCATTAGGAGTGATTGGGCTCGGTGGCCTTGGTCACATGGCAGTGAAGT
TTGGGAAGGCTTTTGGTTGCATGTAACAGTTTTAAGCACAAGCATATCTAAGAAAGATGAGGCG
TTGAGTCTTCTGGTGAGACAACCTTTGTTGCTCTCTGACCAAGAGCAGATGAAGGGCCTATCG
AAGTCATTGGACTTTATAGTTGACACTGCATCTGGTGATCATCCCTTGATCCCTACATGTCACTCC
TGAAGATTGCTGGTGTATGTCTTGTGGGTTCCCAAGTGAAGTCAAATCAGCCCTGCAAGTC
TCAATCGGGTATGAGAACATCTCAGGAAGCGTAACGGGGGAGTAAAAGTGATCCAGGAGAT
GATAGACTTTTGTGCTGCTCATAAAGTTTACCCGAGATAGAAGTAATCCCAATTCAATATGCAAA
TGAAGCTTTGAGAGGCTAGAAAAGAGGGATGTGAAGTACAGGTTTGTGATTGACATCGAGAAC
AGCCTGAAATGA

Tipo: ADN

Longitud: 1065

20 Nombre de la secuencia: cADN completo CAD4

SEQ ID NO: 8

25 Nombre del organismo: *Corchorus olitorius*

MDSQTKSDNCLGWAARDPSGVLSPTYFSRRPLGSNDVSIKITHCGVCY
ADFIWSRNKHGDSMPVVPGEIAGVVKEVGSNVHRFKVGDVPVGVGT
YVNSCRDCEYCNDGLEVHCEKIVLTFNCIDEDGTVTKGGYSNHIIVHER
YCLRIPSNYPLASAAPLLCAGITVYAPMMRHMNMNQPGKSLGVIGLGG
GHMAVKFGKAFGLHVTVLSTISKKDEALSLGADNFVSSDQEQMKG
LSKSLDFIVDTASGDHPFDPYMSLLKIAGVYVLVGFSEVKFSPASLNLG
MRTISGSVTGGVKVIQEMIDFCAAHKVYPQIEVIPIQYANEALERLEKR
DVKYRFVIDIENSLK

Tipo: PRT

Longitud: 354

5 Nombre de la secuencia: péptido CAD4

SEQ ID NO: 9

Nombre del organismo: *Corchorus olitorius*

ATGGCTAAGTCACCTGAAGAAGAGCACCTGTAAAGCCTTGGCTGGGCTGCCAGGGACAC
TTCTGGTCATCTCTCTCTTCAACTCTCCAGAAGGGCAAATGGTGAAGGAGATGTGAGGTTAA
GGTGTTATATTGTGGAATATGCCATTCTGACCTCATTTTATCAAGAATGAATGGAACTTCTCCATC
TATCTTTGGTTCCTGGGCATGAAATGTGGGGGAAGTGACAGAAGTTGGCAAAAAAGTTCAAAA
GGTTAAGATTGGAGACAAAGTGGGTGGGATGCATTATTGGTGTGTGACACTGTGAGAGTT
GTGCCAATGACCTTGAAATTAAGTGTCTCTAAAGCAATTGCAACCTACAATGGGACTTACTATGATG
GAACCATGACATACGGAGGCTACTCCGATTCAATGGTTGCCGATGAACGATACGTCTGAGATTG
CTGATGGCATGGCCCTTGACTCTGCTGCCCCATTGCTGTGCTGGAATTAAGTTTACAGTCCATT
GAAGTATTTGGATTAGGTGAAGCTGGTAATCACATTGGCATTGTTGGCCTGGTGGCCTTGCCCA
TGAGCTGTTAAGTTGCCAAGCTTTGGGGTCCAAAGTTACAGTAATTAGCACTTCCCCTGGTAA
GAAGAAGGAAGCCTTGGAATCTTTGGGGCTGATTGTTCTTGGTTAGCCGTGACCAAGACGAGA
TGCAGGCTGCCATGGGCACATTGGATGGGATCATAGACACAGTCTCTGCCGTTATCAATATGTC
CATTGCTGGTCTTCTGAAATCCCATGGAAGCACTCATTATGGTGGGCGCACCAATCGAACCACTTG
AGTTACCTGTCTTTCTTGATCATGGGAAGGAAGACGATGGCTGGGAGTGAATTGGAGGAATG
AAAGAGACACAAGAGATGATTGATTTTGAGCAAAACACAAATAAAGCAGACATCGAAGTGAT
TCCGATGGATTATGTGAACAAAGCTATGGAAGACTTGAAGGGGTGATGTTAGATACAGATTG
TAATTGACATTGGAACACCTTGGCTACTACCAAGCCTTAG

10

Tipo: ADN

Longitud: 1092

Nombre de la secuencia: cADN completo CAD5

15 SEQ ID NO: 10

Nombre del organismo: *Corchorus olitorius*

MAKSSPEEEHPVKAFGWAARDTSGHLSPFNFSRRATGEGDVRFKVLYC
GICHSDLHFIKNEWNFISIYPLVPGHEIVGEVTEVGKKVQVKIGDKVGV
GCIIGACHTCESCANDLENYCPKAIATYNGTYDGTMTYGGYSDSMVA
DERYVVQIPDGMALDSAAPLLCAGITVYSPLKYFGLGEAGNHIGIVGLG
GLGHVAVKFAKALGSKVTIVSTSPGKKKEALELLGADSFLVSRDQDEM
QAAMGTLDGIIDTVSAVHPIMPLGLLKGSHGLIMVGAIEPLELPVFS
LIMGRKTMAGSGIGGMKETQEMIDFAAKHNIKADIEVIPMDYVVKAM
ERLEKGDVRYRFVIDIGNTLATTKP

20

Tipo: PRT

Longitud: 363

Nombre de la secuencia: péptido CAD5

SEQ ID NO: 11

25 Nombre del organismo: *Corchorus olitorius*

ATGGCTATGGAAACACCTAACCACACTCAGACGGTAGCAGGGTGGGCTGCTCATAATTCCTCAGG
CAAGATCGTCCCTTACACCTTCAAAGAAGGGAAAAATGGCGTGAACGATGTGACCATTAAGTGA
TGTATTGTGGGATCTGCCATACTGATCTCCACCATGTTAAGAACGATTGGGGTATCACCATGTATC
CTGTAGTTCCTGGGCGATGAAATTACTGGGGTGATCACCAGATTGAAACAATGTGAAGAATTC
AAAGTGGGAGACAGGGTAGGTGTGGTGTCTTGGCAGCATCTGTTTGAATGCGAGTTCTGTA
AAAGCTCGCAAGAGAACTACTGTGACCAATCCAGTTCACTTACAATGGCATCTTTGGGATGGTA
GCGTTACTTATGGCGGCTATTCCCAATGTTAGTCGCCGATACCGGTACGTTGTTCTGTGTGCCG
ATAACCTGCCGATGGACGCCGAGCGCCACTGTTGTGTGCCGGGATACCGTTTTAGCCCCATGA
AAGATAGCCAACTGCTCGAGTCACCGGGCAAAAAGTGGGCATAGTTGGTTAGGCGGTCTCGGT
CATGTGCTGTCAAATGGCAAAGGCATTTGGTCATCATGTGACCGTGATAAGCACTTCTCCATCA
AAAGAAAAGGAAGCTAAACAGCGTTTGGGCGCAGATGTTTCATAGTTAGCACCAACACCGAACA
AATGCAGAGAGGAAAGCAACGCTGGATGTTATTTGGACACAGTTTCAGCTAAACACTCGCTCG
GACCAATCTTGAAGTCTTAAAGTGAATGGTACTTTAGTGGTTGTGGGAGCACCAGACAAGCCA
ATCGACCTTCTTCAATTCATTATTTGGGAAAAGCAGTGAAGGGGAGCATGACAGGGGG
GATGAAAAGAGACACAAGAAATGATGGATGTGTGTGGCAACACACATTACATGTGATATAGAA
GTAATTAACCTGATCAGATAAATGAAGCCCTGATCGACTTCCAAAATGATGTTGATACAGA
TTTGAATCGACATTGCTGGAAGTCCAAGCTTTAA

Tipo: ADN

Longitud: 1080

5 Nombre de la secuencia: cADN completo CAD6

SEQ ID NO: 12

Nombre del organismo: *Corchorus olitorius*

MAMETPNHTQTVAGWAAHNSSGKIVPYTFKRRENGVNDVTIKVMYC
GICHTDLHHVKNDWGITYPVVPGHEITGVITKIGNNVKNFKVGDV
GVGCLAASCLECEFKSSQENYCDQIQFTYNGIFWDGSVTYGGYSQML
VADHRYVVRVDPNLPMDAAAPLLCAGITVFSPMKDSQLLESPPKKVGI
VGLGGLGHVAVKMAKAFGHVTVISTSPSKEKEAKQRLGADDFIVSTN
TEQMQRKRTLVDLDTVSAKHSGLPILELLKVNGTLVVVGAPDKPIDL
PSFPLIFGKRAVKGSMTGGMKETQEMMDVCGKHNITCDIEVIKPDQIN
EALDRLSKNDVRYRFVIDIAGRSKL

Tipo: PRT

Longitud: 359

15 Nombre de la secuencia: péptido CAD6

SEQ ID NO: 13

Nombre del organismo: *Corchorus olitorius*

ATGGGCAGTCTTGAAACTGAGAGAACCAACACAGGCTGGGCTGCCAGAGACCTTCTGGAGTCTT
GTCTCTTACACTTACACTCTCAGAAACACAGGTCAGAAGATGTTTTATCAAGGTTATTTGCTGT
GGAATCTGCCACACTGATCTTCATCAAGCCAAAATGATCTTGCCATGTCAAACACCCAATGGTT
CCTGGGCATGAAGTGGTGGGAGGTGTTGGAAGTAGGATCACAGGTGACCAAACTCAGAGTAG
GAGAGATAGTTGGCGTGGTGTATTGTTGGTGTGCAGAACTGCCGCCATGCAACACTGAC
AATGAACAATACTGCAACAAGAAGATTGGTCTTACAATGATGTCTACACTGATGGCAAAACCACT
CAAGGTGGCTTTGCTGCCCTCATGGTCTGATCAAAAGTTTGTGGTGAAAATCCCTGATGGAATG
GCAGCAGAACAGGTGGCTCCACTCTGTGTGCTGGTGTGACAGTTTACAGCCCACTGAAACACTTT
GGTCTAATGGAGAGTGGTTAAGAGGAGGATTTTGGGCTTGAGGAGTAGGTACATGGGA
GTGAAGATAGCCAAAGCAATGGGACACCATGTGACTGTTATCAGCTCTTCAGACAAGAAAAAGT
TGAGGCTTGAGCATCTTGGTCTGATGATTATGTAGTCAGCTCTGATGCTGAAAGCATGCAAA
AGATTGCTGATTCACCTGACTATATCATCGATACCGTGCCTGTTTTATCCCCCTGAGCCTTACCTT
TCAGTGTGAAACTTGATGAAAGTTGATCTTGACTGGTGTATCAATACTCTCTCAGTTTGTTA
CCCCATGGTCATGCTTGGGAGAAAGGTAATTACAGGGAGTTTCGTTGGGAGCATGAAGGAAAC
AGAGGAGATGCTTGATTCTGTAAGAGAGAAGGATTTAAACTCAATGATTGAAGTTGTGAAGATGG
ATTATATCAACACAGCCATGGAGAGGCTCGAGAAGAATGATGTTCTGCTACAGGTTCTGCTGGAT
GTTGCCGGAAGCAACTTGAGTAA

Tipo: ADN

Longitud: 1071

20 Nombre de la secuencia: cADN completo CAD7

SEQ ID NO: 14

25 Nombre del organismo: *Corchorus olitorius*

MGSLETERTTTGWAAARDPSGVLSPYTYTLRNTGPELVFIKVICCGICHT
DLHQAKNLDGMSNYPMPVPGHEVVGEVLEVGSQLTKFRVGEIVGVGCI
VGCCRNCRPCNTDNEQYCNKKIWSYNDVYTDGKPTQGGFAASMVAD
QKFVVKIPDGMAAEQVAPLLCAGVTVYSPLKHFGLMESGLRGILGLG
GVGHMGVKIAKAMGHHVTVISSDKKKVEALEHLGADDYVVSSDAES
MQKIADSLDYIIDTPVFHPLPYLSVLKLDGKLITGVINTPLQFVTPM
VMLGRKVITGSFVGSMEETEEMLDFCKEKDLNSMIEVVKMDYINTAM
ERLEKNDRVYRFVVDVAGSKLE

Tipo: PRT

5 Longitud: 356

Nombre de la secuencia: péptido CAD7

SEQ ID NO: 15

Nombre del organismo: *Corchorus olitorius*

10

ATTATCTGGGCATATACCCCATATTTGAAGGCATGGGATGTCCTCCATGACCATCTGCCTCTAAT
TTCAGATTCCTCATCAATGTAACCTTCAACTTTGCTGCATCAACATCATGGATAACATTGAGCTTGA
ACCACTTGTCATAGATGCCGAACCGAGACCGGACTTTTGTAGTACGTGAGCGAACCATTTGATAG

CCCGGAGCATTAGGGTGGTGGCTCGAGGAGGGCTTGCTCAAAAACCTGCATGATGCAAAATCCT
GATGTTCCCTGGGGCACATACCAAGTATCCTTCAAACTGCCACCCCTGACGAGTAATTGTATCCAT
GTATGGCAATCTCGGTTTCGGGGCTTTGTTTGTCTAGTAGAGGATGCGGCTGTCTGTGGAGTA
CCCAACATCTATGAACGCCCTCAATGAAGCTATAACGCTGCGGCTCAGGCACGTTGTAGGGCTCT
GGATATGGTAGTATGTTTCGGTTTCAGGGGAAGAGAGATGAACCTTTCGTAAGATCAACCTGTTG
TCCGAAGCAAAGGCTTGATACAAAAGTGACTGACTAGATAAATTAATATCAAGTAGAAGCAAGT
CATTTTGGTTAAGGCTGGTACAAATGTCATGTATGAAAAAGTGAATGTTAATGATTGATGGGTTT
AATTAGTATGGGACATTGATCTTTGCATATGAGTAATTAAATATAGAAATGAGAAGGAATTTGTT
AATGATTCAATGCTTATTGTATCTGAGAGTGGCAGATGATCATAATAGCTTAATTTTGTAGGTTTG
ATTCCTAGTGGAAACATATATAGCTTTTAAATAGCTAGGCCAGAAATTTTGCAAAATATCA
GCAATAAAGAGACACAATAAGATAAAGATTCTATCGATTAAACAGCTCAAGAATGGACATCAGCAA
GCAAAACAGCAGCAAAACATTTTCATATGAAGTCATGAGTCACGGGCTTAAACATAGCATAGCAAGG
GGGAAACAATGTAAGGACAAAAGCAGTAGTTAGAGGAACAAAGAAAGCAAAATCCAATACCAT
GAATCAGATTCCCGAGCAGATAAAACAACTAAATGGGCTAGAAAGCGAGGGATCGACGATTATCA
GTTCTTTAATTATGCTAGCGCGTGATCGATGAACCTTATTCTAACATCTAGACTACCCAGTTGGGAT
TCTAATATCTGGACTTCTAGTTGAAAGTGAACCTAGCTACTGACAAGGAATTTGCCACATCGAT
CACAAAACGGTATCTGACATCAGATTTTGCAAGCTGTCCATTGCTGTGTTAATCTCATCCATCCGA
ATCAGCTCAATGCTGCACTAATATTGCTTGGCAAAAAGCTAGCATCTCTGTGTTCTTTTCAT
CCCTCAATATCACTTCTCCAACAAGCTTCGGGCTAAAACAAAGGAAAGATGGGCAATTCAG

GGGCTTATTGGGCAATCCCAAAGTGACCAAGTTTTCATTACCTTCAGCAGATAAGCAGTGGAAG
CAGGGGATGAACCTGCAGACACCGTGCAATGATATAGTCCATGGTGCCAACGGCCGACTTCATTT
TGCAAGGATCATTGGAAGAAGAAATGAGTCAGCACCAAGTCTAATTAATGCTTCTTCTCTCTTT
GGGGAGCTACTAATGACAGTAACCTTCAACCCAAAGGCCTTACCGATTTTACAGCAACATGACCA
AGCCCAACCAAGTCTGCAACTCCCAATGCTTGCCTGCCTCAGTCATTCATAGTATTTTATAGGGC
TGTCACCTGTGATCCCGGCACATAATAGTGGTGACCCGCATCCAGTGGCATGTTATCAGGAAACC
GAACCACAAAGCGCTGGTCAACAACATCATCTGAATAACCAACATAATTTCTGTCCCATCCAC
ATAATTAGAGTTATAGGTAATATCATCTGAGGGCAATAGTTCTCCAGTCTCTGCTGGCAGCACTC
ACATTTCTTGCAAGGAGCAACCATGACCCCAACCCCAACCCGGTCTCCCTCTTGAAATTTTGTCAAC
TTATTTCCCACTTTTGTGCAACCAACAATTTTATGCCCAGGAACAACAGGGTAACGGGTGAAA
CCCCATTCAATTTAAGGGTATGCAAGTCGAATGGCACTCCACAGTAGAGTATTTTCAAGGTG
ACATCTTCTCACCATTTTCCCTCTGGAAGACTGGAAGGAGAGAGAGTTCGGAGGAATCCCGA

15

GCAGCATAACCAAAAGCCTTTTGTGGTGCTCTTCTCTGGTGATTTGACATTTCTTTTTCGACTCT
GTTTTTTGTTTTACTTTACTCTTTTGTCTCTTCACTGTTGAAGAAGTGGTCTTATTGTTTTAA
GTTTGCAAAATCTAACAGATGGGTATAGGACAATTTATAGGAGGAGATGGTGGTCTCTTAGACAC
TGTTTACGTACACCCCTTGATACCTTTATTAGTTATAGA

Tipo: ADN

20 Longitud: 2492

Nombre de la secuencia: cADN parcial CAD8

SEQ ID NO: 16

Nombre del organismo: *Corchorus olitorius*

ATGGCTCCAACCTCAGGCAGAACAGCAAACCTCAAGCTAGTAGGCACCAGGAAGTTGGCCACAAGA
GTCTCTTGCAAGTGATAAACTCTACCAATATATACTTGAGACCAGTGTTATCCTAGGGAGCCTG
AAGCCATGAAGAGCTCAGGGAGCTTACTGCAAAACATCCATGGAATCTCATGACTACTTCAGCT
GATGAAGGGCAGTTTTTGAACATGTTGCTTAAGCTTATTAATGCCAAGAACACCATGGAAATTGGT
GTTTACACTGGCTATTCACTCCTTGCCACTGCCCTTGCTCTGCCTGAAGATGGCAAGATTTTGGCCA
TGGACATCAACCGTGAACAACTATGAATTGGGTTTACCAGTAATCCAAAAAGCCGGTGTGCTCACA
AGATTGACTTCAAGAAGGCCCTGCTCTCTGTTCTTGACCAATGATTGAAGCTGGGACATACC
ATGGAAACATTCGATTTTCATCTTGTGATGCTGACAAGGACAACATTAACCTACCACAAGAGGC
TGATTGAGCTAGTTAAGGTTGGGGAGTCATCGGCTACGACAACACCTATGGAACGGTCCGCTG
GTGGCGCCTCCGATGCTCCATTAAGGAAGTATGCTTGTATTACAGAGACTTTGCTTGGAGCTT
AACAAAGCCTAGCTGCTGATCCAAGGATTGAAATCTGCCAACTTCTGTTGGCGATGGAATTACC
CTCTGCCGTCGGATTAAGTGA

5

Tipo: ADN

Longitud: 744

Nombre de la secuencia: cADN completo CCoAOMT1

SEQ ID NO: 17

10 Nombre del organismo: *Corchorus olitorius*

MAPTQAEQQTQASRHQEVGHKSLLQSDKLYQYILETSVYPREPEAMKE
LRELTAKHPWNLMTTSADGQFLNMLLKLINAKNTMEIGVYTGYSLLA
TALALPEDGKILAMDINRENYELGLPVIQKAGVAHKIDFKEGPALPVLD
QMIEAGTYHGTDFIFVDADKDNYYINYHKRIELVKVGGVIGYDNTLW

NGSVVAPPDAPLRKYVLYYRDFVLELNKALAADPRIEICQLPVGDGITL
CRRIK

15 Tipo: PRT

Longitud: 247

Nombre de la secuencia: péptido CCoAOMT1

SEQ ID NO: 18

Nombre del organismo: *Corchorus olitorius*

20

ATGGCAACCAATACCAAGAGCAGCAAACCTCAGGCTGGCAGACACCAGGAGTTGGCCACAAGA
GTCTTTTGCAAAGTGATGCTCTTACCAGTATATCCTTGAGACAAAGTGATCCAAGAGAGCCTG
AACCCTGAAGGAGCTCAGGGAATTGACTGCCAAGCATCCATGGAATCTGATGACAACATCAGCA
GATGAAGGGCAATTCCTGAACATGCTTCTGAAGCTGATCAATGCCAAGAACACCATGGAAATTGG
TGTTTACTGCTACTCTCTTAGCCACAGCCCTGCTCTTCTGACGATGGCAAGATCTTGCC
ATGGACATTAAACAGGGAAAACCTACGAGTTGGGCTGCCTGTAATCCAAAAGCAGGCGTTGCACA
CAAAATTGAATTCAGAGAGGGCCCTGCTTGCCTGTTCTTGACAACTAGTCGAAGATGAAAAGA
ATCATGGATCATATGACTTCATCTTCGTGGATGCTGACAAAGACAACATACATAAACTACCACAAGA
GGTTAATAGACCTTGTAAGTTGGAGGCTTAATCGGCTACGACAACACTCTATGGAATGGCTCT
GTTGTGCCCCACCTGATGCTCCTCTCAGGAAATATGTCAGGTATTACAGAGACTTTGTTTGGAA
CTCAACAAGGCTCTCGCCGCCGACCCAGGATTGAGATCTGTATGCTTCCCGTCGGCGATGGAATC
ACCCTTTGGCGTCAATCAATGA

Tipo: ADN

25 Longitud: 744

Nombre de la secuencia: cADN completo CCoAOMT2

SEQ ID NO: 19

Nombre del organismo: *Corchorus olitorius*

30

MAPTQAEQQTQASRHQEVGHKSLLQSDKLYQYILETSVYPREPEAMKE
LRELTAKHPWNLMTTSADGQFLNMLLKLINAKNTMEIGVYTGYSLLA
TALALPEDGKILAMDINRENYELGLPVIQKAGVAHKIDFKEGPALPVLD
QMIEAGTYHGTDFIFVDADKDNYYINYHKRIELVKVGGVIGYDNTLW

NGSVVAPPDAPLRKYVLYYRDFVLELNKALAADPRIEICQLPVGDGITL
CRRIK

Tipo: PRT
 Longitud: 247
 Nombre de la secuencia: péptido CCoAOMT2
 SEQ ID NO: 20
 Nombre del organismo: *Corchorus olitorius*

ATGGGTTCAACTGGTGAACCAATTCACTCCAACCTCAAGTCTCCGATGAGGAAGCAAACTTGTT
 GCCATGCAATTGGCTAGTGCTCAGTTCTTCCATGGTCCTCAAATCTGCCATAGAAGTTGACCTAC
 TTGAAGTCATGGCCAAGGCTGGACCTGGTGTCTTTCTGTCCCAACAGAAGTAGCTTCCCAATTGC
 CCACCAAGAACCTGATGCACCCGTATGCTCGACCGTATCTTGGGCTCCTTGCTAGTTACTCCAT
 TTAACTTGCTCCTAAGGAATCTTCTGATGGCAAAGTTGAGAGGCTCTATGGCCTTGGCCCTGTG
 TGTAAATACCTGGTCAAGAATGAAGATGGTGTCTCTTCCGCCCTTAATCTCATGAATCAAGAC
 AAGGTCCTAATGGAGAGCTGGTACTACTTGAAGATGCAGTGTGGAAGGTGGAATCCATTCAA
 CAAGGCTATGGCATGACCGCTTCGAGTACCATGGCACTGACCTAGATTCAACAAGGTTTCAA
 CAGGGGAATGTCTGATCACTCAACTATCACCATGAAGAAGATTCTCGAGACCTACGATGGATTG
 AGGGGCTCAAAACATTGGTTGACGTTGGTGGTGGTGGTGGCCACGCTTAACATGATCGTCTCC
 AAGCACCCTCCATTAAGGGCATTAACTTTGATTTGCCTCATGTATTGAGGATGCTCCAGCTCTTC
 CTGGTGTGAGCATGTTGGTGGAGATATGTTGTAAAGTGTCCAAAAGGAGATGCCATTTTCATGA
 AGTGGATATGTCATGATTGGAGCGATGAACACTGCGTAAATCTTGAAGAAGTGCTATGAAGCT
 TTGCCAGACAATGGGAAAGTATCGTTGCCAATGCATTCTTCTGATTACCCAGATGTAGCCTT
 GCCACAAAGCTAGTTGTTTATCGATTGATCATGTTGGCTCAACCCCTGGTGGGAAAGAAAG
 GACAGAGAAGGAATTTGAAGCCTTGGCAAAGGGGCGAGTTTTCAAGGTTTCCAAGTAAAGTGTT
 GTGCTTTTGGCACTTACATCATGAGGTTCTCAAACCTGTTAA

Tipo: ADN
 Longitud: 1098
 Nombre de la secuencia: cADN completo CCoAOMT3
 SEQ ID NO: 21
 Nombre del organismo: *Corchorus olitorius*

MGSTGETQFTPTQVSDEEANLFAMQLASASVLPMLKSAIELDLLEV
 AKAGPGAFLSPTEVASQLPTKNPDAPVMLDRILRLASYSILTCSLRNL
 DGKVERLYGLGPVCKYLVKNEDGVALSALNLMNQDKVLMESWYYLKD
 AVLEGGIPFNKAYGMTAFEYHGTDPFNKVFNRGMSDHSTITMKKILE
 TYDGFELKTLVDVGGGVGATLNMIVSKHPSIKGINFDLPHVIEDAPAL
 PGVEHVGGDMFVSVPKGDAIFMKWICHWDSEHCVKFLKKCYEALPD
 NGKVIVAEICLPDYPDASLATKLVVHIDCIMLAHNPGGKERTEKEFEAL
 AKGAGFQGFQVKCCAFGTYYIMEFLKTV

Tipo: PRT
 Longitud: 365
 Nombre de la secuencia: péptido CCoAOMT3
 SEQ ID NO: 22
 Nombre del organismo: *Corchorus olitorius*

ATGGCACCCCAAGCACCAGAATTGCCACAAGAAGATATCATTTTCGATCAAACTTCTGATATCT
 ATATTCTAAGCATCTTCTTTACATTATATTGCTTCGAGAACATTCGAAGTTGCTTCTAAGCCT
 TGTTTGATCAATGGGACGACGGGTGAGATTATACCTACGAAGAAGTTGAATCAGCTCGCCG
 AGTCGCCGCTGGGCTTCAAACTCGGCGTTCAACAACGTCAAGTTATAATGCTTTATTACCCAAT
 ACTCCGAGTTTGTCTGCTTTCTCGGCGCTTCACTCTCGGCGCCGTTTGACGGCGCGCAACC
 CGTTTTTACCGCCCCGGAGTTGCCAAGCAAGCAAAGCTTCAACGCCAGGATCATACTACTC
 AAGCTTCGTACGTGACAAAGTGAAGGAATTCGCCAGGAAAATGATGCAAGGTGATGTGATT
 GACTCAGCTCCGGAAGGATGTTTACATTCTCCGAGTTAACTCAAGCTGACGAGAAGCATCTCCG
 GAAGTCGAAATCAATCCGACGATGTCGTGGCACTTCTTATTCTCGGGAACACCGGGTCCCA
 AAAGGTGTGATGTTAACTCACAAGGTTTGGTTACGAGCGTGGTCAACAGGTGACGGTGAAAA
 CCCAAATTTGACTTCCACAGCGATGACGTATCTTATGACTCTGCCCATGTTCCATATCTATGCC
 TGAATCGATCATGCTTTCGGGCTTCGGGCGGAGCTGCGATTTTGATCATGCAGAAAGTTTGACA
 TCGGATTATTGTTGGATTGATTGAGAAATACAAATACGATTGCTCGATGGTGCCACCCATAG
 TTTTGCTATTGCCAAGTCATCGGAACTGAAAAATACGACTTGTCTCGATAAGGATGGTGAAAT
 CCGGCGCCGCGCGTGGTAAAGAGCTGGAAGATGCTGTGAGAGCAAGTTTCTGTGTGCCAA
 ACTCGGTGAGGATATGGGATGACAGAAGCAGGACGATTCTAGCAATGTGCTTGGGATTGCCA
 AGGAACCATTTGAAATCAAATCCGGTGCTGTGGGACGGTGTAGAAATGCAGAGATGAAAAATC
 GTTGACCCAGACACCGGTGCCTCACTTCAAGAAACCAAGCTGGAGAGATTGCAATAGAGGGGA
 TCAGATCATGAAAGGTTACCTAAATGACCCAGAGGCCACAGCTAGGACATTGACAAAGATGGCT
 GGTATACATACCGGTGATATCGGTTACATTGATGACGACGATGAACCTTATCGTTGATCGTTTGA

 AGGAATTGATCAATATAAGGGTTCCAGGTTGCTCTGCTGAGCTTGAAGCTATGCTTATTGCC
 ACCCTGAGATTATTGATGCTGCTGCTGCAATGAAGGATGAGGTAGCTGGAGAAGTTCTGTT
 GCATTTGTGTGAAATCAGAGAAATCAGGAATCACTGAGGATGAAATCAAGCAATATATTTCAA

 GCAGGTTGTGTTCTACAAGAGAATAAGCCGTGTGTTCTTCATGGAATCAATCCAAAGGCACCATC
 AGGCAAGATTTTGAGAAAGGAATTGAGAGCTAAATGGCTTCTGGAACACTGA

- 5 Tipo: ADN
 Longitud: 1635
 Nombre de la secuencia: cADN completo 4CL1
 SEQ ID NO: 23
 Nombre del organismo: *Corchorus olitorius*

10

MAPQAPELPQEDIIFRSKLPDIYIPKHLPLHSYCFENISKVASKPCLINGT
 TGQIYTYEEVELTARRVAAGLHKLGVQQRQVIMLLPNTPEFVLSFLGA
 SFLGAVCTAANPFFTAPEVAKQAKASNARIITQASYVDKVKEFAQEND
 VKVMCIDSAPGCLHFSELTQADENDLPEVEINPDDVVALPYSSGTTGL
 PKGVM LTHKGLVTSVAQQVDGENPNLYFHSDDVILCTLPFHIYALNS
 IMLCGLRAGAAIIMQKFDIGLLDLIYKITYKIAPMVPPIVLAIKSSET
 EKYDLSSIRMVKSGAAPLGKELEDAVRKFPKALGQGYGMTEAGPVL
 AMCLGFAKEPFEIKSGACGTVVRNAEMKIVDPDTGASLPRNQAGEICI
 RGDQIMKGYLNDPEATARTIDKDGWLHTGDIGYIDDDDELFIVDRLKE
 LIKYKGFQVAPAELEAMLIAHPEIIDAAVVAMKDEVAGEVPVAFVVKSE
 KSGITEDEIKQYISKQVVFYKRISRFFMESIPKAPSGKILRKELRAKLAS
 GNY

- 15 Tipo: PRT
 Longitud: 544
 Nombre de la secuencia: péptido 4CL1
 SEQ ID NO: 24
 Nombre del organismo: *Corchorus olitorius*

TGTGTTAGGGATATGGGCTGACTGAGACTGGAGGAGGAGCGACTAGGGTGATAGGGCCTGAGG
 AGTCAGCACGGTATGGGACGGTCGGCTCGCCTTGCAAGAAATATGGAAGCCAAGATAGTTGACCTT
 GAACTGGAGAGGCCCTGCTCTGGGCGAGAGGGGAATTATGGTTGCGAGGGCCAACAGTAA
 TGAAGGGTTATATAGGAGATGAGAAGGCAACTGCTGAAACCTTGGATTGAGAAGGCTGGTTAAA
 AACTGGTGATATATGTTATTTGACTCTGAGGGGTTTCTCTATATTGTAGATAGATTGAAGGAATT
 GATCAATACAAGGCATATCAGGTTCTCCGCTGAATTGGAACAGTTACTTCATCCCATCTAAA
 ATTGCCGATGCGAGCTGTGATTCCGTACCTGATGAAGAAGCAGGGCAGATTCCCATGGCCTATGTT

 GTAAGAAATGCCGGAAGTAGCATCACCGAGGCAGAAGTCATGGATTTCGTTGCAAAACAGGTTGC
 ACCATACAAGAAGATCCGACGTGTTGCTTTATCGATTCTATTCCAAATCTCCGGCAGGAAAGAT
 CTTAAGGAGGGAGCTGATTAAACATTCTCTTCCAGTGGTTATCAAAGTTATAACAATCTGGAAC
 ATACAACAGTCAACAGAGGAGTTGCCGACTGCTGAGAAAGACCTCTAGTGAACATGAAAACCT
 ACTTGCCATCTGCTGAAGGCTTTGATTCACTGTTGGTTATTGTTACAGTTCAAAAACCTCAGATT
 GAAAGTAAATTCTTTTAGATTTTAGACAGCTTGGGATTCTAAATGAGTTCCTTACTGGCGGCTT
 GATCACCATAGCAACAGACAATACTAAATGTATCTATTTGTGTTCAAGTTTTCAGGGGAA
 AAAGAAATTAGTTGAGTGTGTTGAACATGACATGAGATGAGGGACA

Tipo: ADN

Longitud: 964

5 Nombre de la secuencia: cADN parcial 4CL2

SEQ ID NO: 25

Nombre del organismo: *Corchorus olitorius*

TCTCGGACTTAAATGTTCTTCCACCCTTACAGGTTTGAGTTTAAACCCCAAGTCAAGATGGTTTC
 AGCCTTCAATATCCCTCTTATTTCAAGAATACAACAGAACAAATTTTCATGACAAGCAAAGGACCAA
 GGGGAGGAATAGTAAGAAGGACATAATTTAAGCATAAGGACTACCAATAAGTAATAACATGGT
 GATTTTTAGTAGTAACATTTTGACTCACTACATTAATTAATAGCAAACTACAATATCATCATGATT
 ATAAATCCAAATAAAAACTACAATACTAGAAGGAAGAAATGAAAATTTCAAGGCAGGGTAGCA
 GATTAGCAGCCAGCTTAGCTCTGAGATCTTTCTCAATATCTTCTGAAGGAGACTTGGGAATT
 GCATGAACAAAGTAAATCTTATGCAATCTCTGTAGAAAACCACTGTTTTCATGAATTTCTTGA
 CAGCTTCTCAGTAAGTTCAAACCATTTGATCTAACCAAAATGCAACAGGAACTTCCCAGCAA
 CTTCACTTTTTGCGGAACCTACAGCTGCATCTGCAATTGATGGATGGCTTCAAGGAGAGATTCAA
 GCTCAGCTGGTGGCACTTGAAGCCTTTGAATTTGATGATTTCCTTTACCTATCAACAATGAAAAT
 CTCATCATCTTCATCAACATAACCAATGTCACCTGTATGAAGCCAAACCTCCACATCTATGGTGGCT
 GCTGTGGCCGACATCATTTCAATAACCTAAAATTTGTTCAATAGAAATGTTGAACAAATCAGGA
 TTGACAGAATTTCTTTATCAGATCAACAGTTTTAAAGATGATAATTAGTTAACCTTTTCATGATT
 GGGATCCGCAATGCAATTCGCCAGTTGATTGTAGCCGAGGGAGCAGCCGTTTCAGGGTCA
 ATGACCTTAAGCTCAGCATTTCTAACCAAGTGCACATGAACCAAGACTTGGTTGGAAGGGTTGC
 TTAGCAATCTAGGCACATTGATAGAAGTGGTCCAGCTTCTGTATCCCATATCCCTATATTTGTG
 TGGAAATGACATAGTTAATGCATTGTAAGAGTAGAGTGTGAACACTACATGGTATGATTCTATA
 TAAATGTAATATGGACAGACATAAC

10

Tipo: ADN

Longitud: 1154

Nombre de la secuencia: cADN parcial 4CL3

15 SEQ ID NO: 26

Nombre del organismo: *Corchorus olitorius*

ATGGAGAAATCTGGGTATGGAAGAGATGGGATTTACAGGTCTCTAAGGCCACCTGTAATCTTCC
 CAAGGATCCAAATCTTTCCATGATTCTTTCTGTTTGAAGAACATTTCTTACCCTCACCATCCAG
 CCTCATGATGCTGATTCCAATGAAATCTAACTTTTCTCAGTTCAAATCCACTGTCATCAAGCTT
 TCCCATGCTTTTCTCAATCTGGGTATCAACAAAAAGACCGTGTCTTGATTTTGACCAAAATCGA
 TCCAATCCCTCTCTGTTTCTCGCTGTAACCGCCATTGGCGCTATTGCTACAACCGCCAAACCCATT
 TACACTGTCAATGAACTCTGAAACAAATCAAGATTCTGACTCCCAAGCTTCTTTACTGTTCTG
 AATTGTTGACAAAGTTAAGGATTCAAGCTTCTGTTATATTGCTTGGCTCTAAACAGAACAAACC
 CCCATCTCATCTGATGTAATAATGTCCAAAAATCTTATCTTTTCATGACCTTCTCGATTTAGCGG
 GGAGCTGACAGAGCTTCCCGCGTTTCTGTTAAGCAAACTGATACGGCGTCACTTTTATACTCCT
 CCGGCACGACAGGGTAAGTAAAGTGTGTTTACGCATAGGAATTCATTGCAGCGGCTTG
 ATGATAACCAAGGACCAAGAACTTCCCGCGATAAGCACCGGTTTTCTGTGTTTGGCCCTG
 TTCCATGCTTTGGATTGGCGTTATTGCGTTTTCACAGCTGCAGATAGGAACACTTTGGTTTCTA
 TGGCAAAAGTTCGATTTTGGGTTGTTTGAAGATGCAGAGAGTATAAGCCACCCATTTGTGG
 GTTGTGCGCCAATTGTGCTTGCATGGCTAAGCAGAGTGTGTTAAGAGTTTGATCTTTCTCA
 GTGAGGCAAAATGGCTCTGGCGTGTCTCTTGGGAAGGATTGATGGAGGAATGTGCAAAAAA
 TTTTCTCAGGCTGTGGTTATGCAGGGTTTGAATGACTGAACTTGTGGCATTGTCTCAGTGGA
 GAATCTACAGTTGTGTCGACATACTGTTGAGTGGAAATGCTTTCAAGCATTGAAGCTCA
 AATAATCAGTACTGAGAGCTAAAGCCTCTTCTCCCAATCAATTAGGGGAAATAGGGTTCGAGG
 GCCTAATATGATGCAAGTTACTACAACAATCCAGAGCAACAACTAAACAATAGATAAAAAAGG

GTTGGGTACATACAGGAGATCTTGGATACTTTGATGAAGATGGGAATCTTTATGTTGTGACCGAA
 TTAAGAGTTGATCAAAATAAAGGATTCCAGATTGCACGACCGAACTTGAAGGACTACTTGAT
 CTCATCTGAAATATTGGATGCTGTTGTCTATCCGATCTCTGACGCTGAAGCTGGTGAAGTCCGG
 TTGCATATGTTGTTCTGCTCCCTAACAGCTCACTGACTGAGGAGGATGTCCAAATTTATAGCTAA
 ACAGGTGGCACCCTTCAAAAGACTAAGGAGAGTTACATTATAACGAGTGTCCCAAGTCCGCTT
 CAGGAAAAATCTGAGGAGAGAGCTTATAGCGGAAGTAAGATCCAAGATGTGA

- 5 Tipo: ADN
 Longitud: 1644
 Nombre de la secuencia: cADN completo 4CL4
 SEQ ID NO: 27
 Nombre del organismo: *Corchorus olitorius*

10

MEKSGYGRDGIYRSLRPPVILPKDPNLSMISFLFRNISSYPHHPALIDAD
 SNEILTFSQFKSTVIKLSHAFLNLGINKKDRVLIFAPNSIQFPLCFFAVTA
 IGAIATTANPIYTVNELSKQIKDSTPKLLVTVPFLDKVKDFKLPVILLGP
 KQNKPPSPSDVKNVPKILSFHDLDLAGSVTELPVSVKQTDASLLYS
 SGTGVSKGVVLTNRNFIAAALMITKDQELAGDKHRVFLCVLPFHV
 GLAVIAFSQLQIGNTLVSMKFDGFLFKNAEKYKATHLWVVPPIVLA
 MAKQSVVKKFDLSSVRQIGSGAAPLGKDLMECAKNFPQAVVMQGF
 MTETCGIVSVENPTVGVHRHTGSAGMLVSSIEAQIISTESLPLPPNQLG
 EIWVRGPNMMQGYNNPEATKLTIDKKGWVHTGDLGYFDEDGNLYV
 VDRIKELIKYKGFQIAPAELEGLLVSHPEILDVVIYPDAEAGEVPVAY
 VVRSPNSSLTEEDVQNFIKQVAPFKRLRRVTFITSVPKSASGKILRREL
 IAEVRSKM

- 15 Tipo: PRT
 Longitud: 547
 Nombre de la secuencia: péptido 4CL4
 SEQ ID NO: 28
 Nombre del organismo: *Corchorus olitorius*

20

TCCAACCAACGGCGGCCACCGCGCCACGGACACCTTATGCCTCTGTATCAATTCCAACAAC
 GCCCTATCTCAAATTTGTCATCAGCAAAACCGCAGCCCTGCTCTCAAGGAACAAAAAGCAG
 CTGTTGAGTGAGTAAATGTGAACAATGGCAAAACGCATAACACACGTCGCTGTTTTTCAAGTA
 AAGATTGGGGTTCTCCATCTACTGCTGCGCTACGCTCGTGATCAAGCTTTTGGGTCAAAAC
 GACTCCTTTGGGTAACCTGTTGTTCTGAAGAGAATGGTAAAGCAACAGGGTCATCAGGGTCAA
 TGGAAACTTCGGGGATATCGTTTTTATTCCCTCTGATAACACCGTGAAATGCAAGCAGTTTTCTG
 GGGGATCATCGATGGTTACAACCTTGAATCTTGGCCAATTTAGGGAAGTTTTGTTTTGGGTG
 TCTCATCTTAAGCTGTGCGACATATTGGGATTGCGTGATAATAATCTTAGCCCGAGCAGCTTTGAA
 TTGTTTGAAGATTTGTTGATGTGTAAGGTTGGCAGTTGTGGAACCTGCTCAATCATGGA
 AGCTCCCATGAAAGAGAAGACGAATTGAGCAGTTGGGAGGAGAATCATGATGACATCGCCCT

TCTCGATGCCTAAATTTGACAAACGAGCGGTCTTTCGAGAAATCAAGTGAGTTTCAGAGAAAG
TGTAGGTTTTGGCAGAGGATATCCAGAGATCAACATGGCTTGTCAGGGAAAGAGGAGTTTT
TCAAAGCAGTAAGTATGGAGAGGAGGTGATTGGAGATGGGAATGTCAGGCAATTTGGACCTGA
AAATGTGATGATTAGTTTGAAGGGAAGTTTTGGTGTTCGAAAGCTCTGCTGAGGCTTTTGT
GGTTCAACAGGATGATCAGCTATGGAGATCATGGTGGTAATTTGGTTTATGAGTACTTGTGGG
ATTTTTGGTTAGAGGGAAGTGAATGAAAGGAAAGAAAGAGAAATAATATAAGGTAATAAG
AAGTGGCAAGCAAGGGCCAAGGAGAGGTAGGTGGAAGAGGAA

Tipo: ADN

Longitud: 1089

5 Nombre de la secuencia: cADN parcial 4CL5

SEQ ID NO: 29

Nombre del organismo: *Corchorus olitorius*

ATGGCAACGAGCTTGAACCTCCATTTAGCTTCCCAACCTCCGAATCCAAACAACCTGTTTTCTG
ACTGGTATTCACCGAAACAGGAATCTATACGACGAGCATGCTCCGATCCATCCCACTAATC
CATATCTTGATGTTGTTCTTCTTCTCTCACCAACATCAAGGGTCACTGCTTTCATTGATTCC
TCATCTGGGTTTTCAATATCTTACTCAAAGCTTACCTTGGTCCAATCTATGGCTTCTGGTCTCCA
CCACCTGGGTGTTTCCAAAGGTGACGTGGTCTTGTCTTTGTGCCAATCTCTTCACTATCTTATTA
TTTTCTCAGTGTTTTATATTTAGGCGCAATCGTTACCCCTATGAATCCACTGAGTAGCATTTCCGA
GATCAAGAAACAGATTGCTGATTCTAATGTCGTTTCTGCTTCTTCTTGAACGTTTGACAAA
CTGGAGAAGTTGGGTGTTTATGCAATTGGGTACCGGAAACATGAATGGATTGAGAAAAGGT
TGATTTTTTACCTTTTATAAGCTTATGGCGGGCAATCTGGTAATAAGGCCCAAGGCCAGTGAT
TAAGCAGCAAGACACTGCGGCGATAATGATTCATCGGGAACACAGGAACGAGTAAGGGAGTT
GTATTACACATGGGAATTTATAGCAATGATTGAGCTTTTGTAAATTTGAAGCTTCACAGTATG
AATATCAGGTTCAAGAGATTGATATTTAGTGTCTTACCAATGTTCCATATATATGGGCTATCACT
GTTTGTGGTTGATTGTTATCGTTGGGTCTACAGTTGTTGTCATGAGGAAATTAATGCTGGTGA
ATTGGTAAAGTAATTGATAAGTTTGGGATCACCCACTTCCAGTTGTTCCACCTATACCTACAGCA
TTGACAATAAGCGCAAGGGTGTGTTGAAAAAATCAAGAGCTTGAACAGGTTTCTTGGGT
GCTGCTCTATAAGCAGGAAATCCATAGAGATTTTGTTCAGGCTTCCCTCATGTTGATTTCATC
AGGGCTATGGGATGACAGAATCACTGAGTAGGAACCTCGGGCTTCAACACCGGAAACATCA
TAAATATCTTCAATAGGACTTCTAGCACCAAAATGCAAGCTAAAGTGGTGGATCTGAATCTGG
TTCTCTATGCTCTGTTGATTACGGCGAGCTTGGTTAAGAGGACCTGCAATTATGCAACGATA

10

CTTGAATAATGTTGAAGCCACCCTGATGTCAATCCACAGAGATGGTTGGCTACGTACTGGTGACAT
TGCTGTTTTGATGAAGATGGCTATATGTTTATCTGACCGCTTAAAGAGATTATAAAGTACAA
AGGCTATCAGATAGCTCTGCGGATTTAGAGGCCATATTAATACCATCCTGAGATACTTGACGC
TGCTGTAACCTGGAGCCAGTGATGAAGCATGTGGTGAGATTCTGTGGCATTGTGGTGAGGAGGC
ATGGTTGCACACTGACCATGGCGCTGTCATGGAATTCGTGGCTAAGCAGTTGCACCTTATAAGA
AAGTAAGAAAGGTGGTGTTCAAATCAATACCGAGGTCTGCTGAGGAAAGATCTCCGAAGA
GAACCTCAAGAAGTCTTATGTTCAAGGCTTTAA

Tipo: ADN

Longitud: 1689

15 Nombre de la secuencia: cADN completo 4CL6

SEQ ID NO: 30

Nombre del organismo: *Corchorus olitorius*

MATSLNSHFSFPTSESKTTRFPDWYSPETGIYTSKHASVSIPTNPYLDV
VSFIFSHQHQGVTAFIGDSSGFSISYKLLPLVQSMASGLHHLGVSKGD
VVLLLLPNSLHYPIIFFSVLYLGAIVTPMNPSSISEIKKQIADSNVRFAF
TLLETVDKLEKLGVAIGVPMNNLDSEKVDLFPFYKLMAGQSGNKA
RPVIKQDQDTAAIMYSSGTTGTSKGVVLTGHNFIAMIELFVKFEASQY
PGSEIVYLAALPMFHIYGLSLFVVGLLSLGSTVVVMRKFNAGELVKVID
KFGITHFPVPPILTALTISAKVCENNFKSLKQVSCGAAPISRSIEDF
VQAFPHVDFIQGYGMTESTAVGTRGFNTGKHHKYSSIGLLAPNMQAK
VVDLNSGSSMPPGDYGLWLRGPAIMQRYLNNVEATLMSIHRDGLWR
TGDIACFDEDDGYMYSLDRLEIKYKGYQIAPADLEAILITHPEILDAVT
GASDEACGEIPVAFVVRHGTCLTHGAVMDFAKQVAPYKKVRKVVF
SNSIPRSAAGKILRRELKFLCSRL

20

Tipo: PRT

Longitud: 562

Nombre de la secuencia: péptido 4CL6

SEQ ID NO: 31

Nombre del organismo: *Corchorus olitorius*

ATGATTGTGAACGTGAAAGAATCAACAATGGTACCTCCGGCGGAGGAGACTCCACGGGTTTGCTT
ATGGAACTCCAATGTGGACTTGGTGGTGCCAGGTTCCATACCCCGAGCGTCTACTTCTATAGGCC

5

ATCGGGTGATCCAACCTCTTTGACCCAAAGGTCATGAAGGAGGCTCTGGGCAAGGCCTTGGTGC
CGTTTTACCCCATGGCGGGCGGTTGAAGAGGGATGAAGATGGAAGGATTGAGATTGATGCAA
TGGTGCAAGGGTGCTCTTTGTTGAGGCTGAGACTAATGCTGCTGATGATTTTGGTATTGTC
TCCCACTTTGGAGCTCAGGCAGCTCATTCCAAGTGTGATTATTCTGGTGGCATCGAGACTTACCCG
CTCTGGTTTTGCAGGTCACTTATTTCAAATGTGGTGGAGCATCACTTGGTGTGGCATGCAACATC
ATGCGGCAGATGGCTTTTCTGGTCTCCATTTTATCAATACATGGTCCGATATGGCTCGTGGCTTGA
CCTCACAATTCACCATTCATCGATCGTACCCTGCTCCGTGCCGGGATCCACGCAACCTGCATTC
GAGCACATTGAATACCAACCACTCCTGCATTGAAATCTGCACCTGAATCCACAGGTTCTGAAGGT
GCAGCAGTCTCCATTTTCAAATTGACCCGAGAAGCAGCTAAATGCACCTAAAGCTAAGTCCAAGGAA
GATGGGAACACTATTGCTTAGCTCATATGAGATGTTGTGAGTCAATGATGAGATCAGTCTGC
AAAGCACGTGGACTTCTGATGATCAAGAGTCAAAATTGTACATTGCCACTGATGGAAGGGCTAG
GTTGCGCCCCCACTTCCACCTGGTACTTTGGAATGTTATTTTACCGCTACCCCAATTGCAGTG
GCCGGTGAGCTAATGTCAAAGCAACATGGTATGCTGCTGGGAAAATTCATGATGCTTGGTTCG
CATGGACAATGATTATCTAAAGTCAGCCCTCGATTACCTAGAACTTCAGCCTGATTATCTGCCCTT
GTTCTGGAGCACATACATTTAAGTGTCCGAATCTGGGATTACTAGTTGGTCAAGGCTGCCAATC
CACGATGCAGATTTTGGATGGGGCCGACCATATTTATGGGCTCTGGTGAATCCCTTATGAGGG
GTTATCTTTTGTGTACCAAGTCAACCAATGATGGGAGCTTATCAGTTGCCATCGCTCTGCAAAAC
GAACACATGAAACTGTTTGAGAAGATCTTTTATGATGACATATAA

Tipo: ADN

Longitud: 1299

Nombre de la secuencia: cADN completo 6HCT1

SEQ ID NO: 32

Nombre del organismo: *Corchorus olitorius*

15

MIVNVKESTMVPPAEETPRVCLWNSNVDLVVPRFHTPSVYFYRPSGAS
NFFDPKVMKEALGKALVPFYPMAGRLKRDEIDGRIEDCNGAGVLFVEA
ETNAVIDDFGDFAPTLRLQLIPTVDYSGGIETYP LLVLQVTFKCGGAS
LGVGMQHHAADGFSGLHFINTWSDMARGLDLTIPPFDRTLLRARDPP
QPAFEHIEYQPPPALKSAPESTGSEGAASIFKLTREQLNALKAKSKED
GNTIAYSSYEMLSGHVWRSVCKARGLPDDQESKLYIATDGRARLRPPL
PPGYFGNVIFTATPIAVAGELMSKPTWYAAGKIHDA LVRMDNDYLKSA

LDYLELQPDLSALVRGAHTFKCPNLGITSWSRLPIHDADFGWGRPIFM
PGGIPYEGLSFVLPSPNTDGSLSVAIALQTEHMKLF EKIFYDDI

Tipo: PRT

Longitud: 432

Nombre de la secuencia: péptido 6HCT1

SEQ ID NO: 33

Nombre del organismo: *Corchorus olitorius*

20

ATGCCAATTCGGATGCAGAAAGCAAGAATAAACTTTGCAAGTAAGGATTACTTGGGAGCAGT
GGCATTCAACAACATAACAAGGCTAGTATTTGGGAAGCGTTTATGAACTCTGAGGGCATAATTG
ACGAGCAAGGCAAGAATTCAAGGGCATTGTGTCAATGGAACAAAGATTGGTCATCCCTTGCC
ATGGCAGAGCACATTCATGGCTTCGTTGGATGTTCCCTTGAAGAGGAAGCATTGCCAAGCAT
GCAGCAAGGAGGGACAATCTTACCAGAACCATCATGGAGGAGCACACCGCTGCTCGCAAGAAGA
GCGTGGTGCTAAGCAGCATTTTGTGATGCTTGTCTCACATTGCAAGAAAAGTATGACCTCAGCG
ACGACACAGTTATTGGACTACTTTGGGACATGATTACAGCAGGCATGATACACAGCAATTGCA
GCAGAGTGGGCAATGGCAGAGTTAATCAAGAACCAAGAGTGACGAAAGGCACAGAGGAG
CTAGATCGTGTGGTAGGATTGAAAGGGTGATGTCGAAACTGATTTCTCAAGCTGCTTACCTT
CAAAGTGTAACCAAGGAGGCATTGAGATGACCCCCCACTCTCTAATGCTACCCCAAAAGCC
AACGCCAATGTCAAATCGGAGGTTATGACATCCCCAAGGGATCAAATGTGCATGTCAACGCTG
GGCAGTGGCCAAATGATCCGGCTGTATGGAAGGACCTGAAGTGTTCGGCCAGAGCGATTCTGG
AGGAGGATGTGGACATGAAGGGTCATGATTATCGCTGCTCTCTTTTGGTGGGGAGGAGGGT
ATGCCCTGGAGCACAACTTGGGATCAACCTGCTGATCATGTTGGGTCACTTACTGCACCATTT
TGTTGGACACCACAGAGGGAGTAAGGGCCGAGGAAATCGACATGGTGAAAATCCGGACTT
GTTGCTACATGAAGACTCTGTGACGGCTGTGGCCACTCTAGGCTGCCTCCGATCTCTACAAA
CGTAGCTGTGACATATAA

Tipo: ADN

Longitud: 1062

5 Nombre de la secuencia: cADN completo C3H

SEQ ID NO: 34

Nombre del organismo: *Corchorus olitorius*

MPISDAESKNKTLQVRDYLGAFAFNNITRLVFGKRFMNSEGIIDEQ GK
EFKGIVSNGTKIGASLAMA EHIPWLRWMFPLEEEFAKHAARRDNLTR

10

TIMEEHTAARKKSGGAKQHFDALLTLQEKYDLSDDTVIGLLWDMITA
GMDTTAIAAEWAMAEIKNPRVQKQAEELDRVVGFERVMSETDFSS
LPYLQSVTKEAFRMPPTPLMLPHKANANVKIGGYDIPKGSNVHNV
WAVANDPAVWKDPEVFRPERFLEEDVDMKGHDYRLLPFGAGRVRCP
GQLGLINLVSMGLHLLHHFVWTPPEGVKAEIDMAENPGLVAYMKT
PVQAVATPRLPDLYKRVAVDI

Tipo: PRT

Longitud: 353

15 Nombre de la secuencia: péptido C3H

SEQ ID NO: 35

Nombre del organismo: *Corchorus olitorius*

20

ATGGATCTTCTCTCCTGGAGAAGGCCCTTATTGGTCTTTCTGGGCTGTCTTATGCCATCGCCA
TCTCTAACTCCGGGAAAGCGTTACAAGCTCCCTCTGGTCTTTACCCGTTCCGCTCTCGGCAA
CTGGCTCAAGTGGGCGACGACTGAACACCGTAACCTAAGTACTGGCCAAGAAATATGGCG
ACATATTCTCTCCGAATGGGCGAGCGCAACCTGGTGGTGGTCTCGTCCGGAACTAGCCAAA
GAGGTGCTCCACCCAGGGAGTGGAATCGGGTCAAGAACCCGGAATGGTGTGTTGACATATT
CACCGGCAAGGGTCAAGACATGGTGTTCACGGTGTACGTTGTCCAGCAGTACCGGTTTGGATGG
AGGAGGAAGCGCTCGCGTTGTGGAGGATGTCAAGAAAAATCCGAGGCTGCCACCAATGGCAT
CGTTCTGAGGAGGAGGTGCGAGTCATGATGTACAACAATATGTACAGAATCATGTTGACCCGGA
GGTTCGAGAGCGAGGAGGATCCTCTGTTTGTAACTCAAGGCTTTGAACGGGAGAGGAGTCG
ATTGGCTCAGAGTTTTGAGTACAATTATGGAGATTTATTCCTATTTTGAGGCCTTTCTTGAGAGGT
TACTTGAAGATCTGCAAGGAGGTTAAAGAGGAGGTTGCAACTCTCAAGGACTACTTTGTGCA
AGAGAGAAAGAACTTGCAAGCACGAAGAGCATGAGCAACGAAGGATTGAAATGTGCCATAGAT
CATATTTGGATGCTCAGCAGAAAGGGGAGATCAACGAGGACAATGTTCTGTATATCTGCAGAA
CATCAATGTTGCCGAATTGAGACAACATTATGGTCGATCGAGTGGGCGATTGACAGCTGGTGA
ACCACCTGAAATCCAGAAGAAGCTGCGAGATGAACCTTGACACTCTCTTGACCCGGCCACCAG
ATCACCGAACCGGACACCTACAACCTCCCTTACCTTCAGGCTGTATCAAGGAGACTTTGAGGCTT
CGTATGGCCATTCTCTGCTCGTCCCCACATGAACCTCCACGATGTAAAGTTGCCGGCTATGATA
TCCCCCTGAGAGCAAACTTTGGTCAATGATGGTGGCTCGCCACAACCTGCCAGTGGA
AACCCCGAGAGTTTAGGCCGAGAGGTTCTTTGAAGAGGAATCTAAGGTTGAGGCCAATGGCAA
TGACTTCAGGTACCTCCATTTGGGGTTGGAAGAAGAGTGGCCCTGGAATTATTCTCGCTTTGCC

CATCCTTGGTATCACTTTGGGCCGCTTAGTACAGAATTTGAGCTCTTGCCTCCACCAGGACAGTCC
AAGATTGATACCTCGAGAAAGGTGGACAGTTCAGTTTGACATTTGAAGCATTCCACCATTGTT
TTGTAA

Tipo: ADN

Longitud: 1446

5 Nombre de la secuencia: cADN completo C4H1

SEQ ID NO: 36

Nombre del organismo: *Corchorus olitorius*

MDLLFLEKALIGLFVAVILAIAISKLRGKRYKLPPGPLPVPVFGNWLQVG
DDLNRNLTLDAKKYGDIFLLRMGQRNLVVVSSPELAKEVLHTQGVEF
GSRTRNVVDFITGKGQDMVFTVYVVQYRFGWEEEAARVVEDVKKN
PEAATNGIVLRRRLQLMMYNNMYRIMFDRRFESEEDPLFVKLALNGE
RSRLAQSFYNYGDFIPILRPFLRGYLIKKEVKERRLLQFKDYFVEERKK
LASTKSMSNEGLKCAIDHILDAQQKGEINEDNVLYIVENINVAAIETTL
WSIEWGIAELVNHPEIQKKLRDELDTLLGPGHQITEPDYKLPYLQAVI
KETLRLMAIPLLVPHMNLHDAKLAGYDIPAESKILVNAWWLANNPA
QWKNPQEFRRPERFFEEESKVEANGNDFRYLPFGVGRRSCPGLIALPIL
GITLGRVLQNFELLPPPGQSKIDTSEKGGQFSLHILKHSTIVL

10

Tipo: PRT

Longitud: 481

Nombre de la secuencia: péptido C4H1

15 SEQ ID NO: 37

Nombre del organismo: *Corchorus olitorius*

ATGGACCTCCTCTTCCTCGAAAAAGCTCTCATTTCCCTTTTCGTCAACCATAATTGTAGCCATTGTAGT
CTCCAAGCTCCGCGGCAAGCGTTACAACTTCCCTCGGCCCAATCCAGTCCCTGTTTCGGCAAC
TGGCTCCAAGTCGGCGATGACTTGAACCAACCGCACTCACCGACTTGGCCAAGATTGCGAGA
CATTTTCTTGCTCCGTATGGGACAGCGCAACTCGTCGTTGTCTTCCCGGAGCTGCCAAGA
AGTCCTCCATACCCAGGGAGTTGAGTTGGTTCAAGAACTAGAAATGTGGTGTGATATTTTCAC
AGGGAAGAGTCAGGACATGTTTTCACTGTCTATGGTGAACATTGGCGCAAATGAGGAGATCA
TGACCGTCCCTTTCTTCAACAAGGTTGTTCAACAGTATAGACATGGGTGGGAAGCGAGGTT
GCCGCCGTCGTTGAGGATGTGAAGAAGAACCAGAGTCGGCCACCACTGGGATTGTTCTGAGGA

20

AGAGATTGCAGCTTATGATGTACAACAATATGTACCGGATCATGTTTGATAGAAGTTTGAGAGT
GAAGATGATCCTTTGTTGTTAAGCTCAAGGCTTGAATGGTGAGAGGAGTAGATTGGCACAGAG
CTTTGATTACAACATATGGTGATTTCATCCCAATTTGAGGCCTTTCTTGAGAGGGTATTGAGGTTA
TGCAAGGAAGTCAAAGAAATGAGATTGCAACTCTTTAGGACCATTTCTTGAGGAGAGGAAGAA
GCTTTCAAGCACAAAAAGGCTGACAACAATGCTCTGAAGTGTGCCATTGATCACATTTGATGC
TCAGCAGAAAGGAGAAATCAATGAAGATAATGTTCTACATTGTTGAGAAATATCAATGTTGCTGC
CATTGAAACAACCTTTGTGGTCAATTGAATGGGAATTGCTGAGCTTGTGAACCATCTGAGATCCA
GCAGAAGCTCCGCAATGAAATCGACACTGTACTCGGACCAGGAGTGCAAGTTACCGAACCCGACA
CCCAACAAGCTCCATATCTCAGGCAGTGATCAAGGAGACCTCCGGCTCCGGATGGCCATCCCTC
TATTAGTGCCACACATGAACCTCATGATGCTAAGCTTGGTGGCTATGACATCCAGCTGAAAGCA
AGATCCTTGTTAACGCATGGTGGTTAGCCAACAACCCGGCTCAATGGAAGAACCCGGAAGAGTTC
AGGCCCAAGGTTTTTCGAGGAGGAAGCTAAGGTTGAAGCCAATGGAAATGACTTCAGGTACCT
TCCATTGGTGTGGAAGGAGGAGTTGCCAGGAATTATTCTTGCTTGCCAATCTTAGGAATCAC
ATTGGGACGTTTGGTGCAAACTTTGAGCTATTGCCTCCTCTGGACAGTCAAAGCTTGATACCTC
AGAGAAAGGAGGACAATTCAAGCTTGACATCTCAAGCATTCAACTATTGTTGCAAGGCCACGAG
TCTTTTAA

Tipo: ADN

Longitud: 1518

25 Nombre de la secuencia: cADN completo C4H2

SEQ ID NO: 38

Nombre del organismo: *Corchorus olitorius*

MDLLFLEKALISLFVTIIIVVSKLRGKRYKLPPGPIPVVFGNWLQVG
DDLNRHRLTDLAKKFGDIFLLRMGQRNLVVVSSPELAKEVLHTQGVEF
GSRTNRNVVDIFTGKGQDMVFTVYGEHWRKMRRIMTVPFFTNKVVQ
QYRHGWAEVAEVAVVEDVKKNPESATTGIVLRKRLQLMMYNMYRIM
FDRRFESEDDPLFVKLKALNGERSRLAQSFYDNYGDFIPILRPFRLGYLK
LCKEVKEMRLQLFRDHFLEERKKLSSTKRPDNNALKCAIDHILDAQQKG
EINEDNVLYIVENINVAIETTLWSIEWGIAELVNHPEIQKLRNEIDTV
LGGPGVQVTEPDTHKLPYLQAVIKETLRLRMAIPLLVPHMNLHDAKLGG
YDIPAESKILVNAWWLANNPAQWKNPEEFRPERFFEEAAKVEANGND

FRYLPFGVGRRSCPGIILALPILGITLGRVQNFELLPPPGQSKLDTSEKG
GQFSLHILKHSTIVAKPRVF

5 Tipo: PRT
Longitud: 505
Nombre de la secuencia: péptido C4H2
SEQ ID NO: 39
Nombre del organismo: *Corchorus olitorius*

10

GGATTGTTATCTCTCAGGTCATAATTTAAACATGGCTCATCTCTAAGAAAGCAATCTTGC
CATTATAACACTATCTTGATCTTTGTTCTTGCTACCAACAAAACTGTCCCTCACCTTTTGCCACT
CTTCTTGTCACCTACTTTTTATTGAGAAAGCTTCTAATCTTCCAATGCCCTCCCTCAGGTCCTCT
CTCTGTTCCCATATTTGGAACCTGGCTCCAGGTTGGAATGACCTAAACCAACCGGCTCCTAGCTTCC
TTGGCTGAAACCTATGGCCAGCTCTCTCTCTAAACCTGGCTCCAAGAATCTAGCAGTGGTTTCA
GACCCAGAGCTAGCTCCAGGTTCTTACACCCAAGGGGTCGAATTCGTTCCAGGCCACGCAA
CGTGGTGTGTTGACATTTTACGGGCAATGGGCAGGACATGGTGTTCACAGTTTACGGAGACCATT
GGCGAAAAATGCGCAGAATTATGACTTTGCCATTTTCTACTAAACAAAGTTGTCACAATTACAGTA
ACATGTGGGAGGAAGAGATGGAGCTTGTTAGTGACTTGAAGAGAGATGAGGAATTGGTGAA
GAGCAAAAGGATTGTTATCAGAAAAGCTCTGAGCTTATGCTTTATAACATCATGTATAGGATGAT
GTTTGATGCCAAGTTCGAGTCCATGGAAGACCTTTATCGTTGAGGCAACAGGTTCAATCCGA
GAGAAGCCGCTTGCTCAGAGTTTGTAGTACAATTATGGTGATTTCATTGCTCAGACCTTTT
TTGAGAGGTAAGTGAACAAGTGCAAGGATTGAGAGCAGGCGCTTGCCTTCTCAACAATA
TTATGTTAAGAAAAGAAGGAAATATGGGTGCTAATGGAGAGAAGCACAATCAGCTGTGCA
ATTGATTACATAATAGATGCTGAAATGAAGGAGAGATTAGTGAAGAAATGTGCTCTACATTGT
GGAAAACATCAATGTTGACGCTTGAACCACTCTGTGTCAATGGAGTGGCAATAGCTGAGG
TGGTGAACCAACCAATGTGCAAGCAAAAGATCCGCCAAGAAATCTCACAAGTCTCAAAGGAGAG
GCTGTACAGAAATCAAACTCTTGAATTGCTTACTTCAAGCACTGTCAAGGAGACACTACGG
TTACACACCCCAATTCCTCTGTTGGTCTCACATGAACCTGAAGAGGCAAAATGGGAGGGTTC

ACAATTCAAAGGAGTCCAAGTTGTAGTCAATGCCTGGTGGCTGGCCAAACAATCCAAATGGTG
GGAAAATCCAGAGGAATCAGGCCAGAGCGGTTTTGGAAGAAGAATCGGCCACACAAGCCGTC
GCCGGAGGAAAGTTGATTGAGTATTTGCCATTTGGAATGGGAAGCGTAGCTGCCCTGGTAT
CATACTGGCACTGCCAATCCTGGGGCTTATCATTGCCAAATGGTTACAAATTTTGAATCAAAGCT
CCCCAAGGAACACACAAGATTGATGTGAGTGAGAAAGGAGGCAATTAGTTACACATAGCAAA
C

15

Tipo: ADN
Longitud: 1580
Nombre de la secuencia: cADN parcial C4H3
SEQ ID NO: 40

20

Nombre del organismo: *Corchorus olitorius*

ATGCTGTGCAGAATCAACACTTTGTCAGGGTTATTCTGGAATCAGATTCGAAATTTTGAAGCA
ATTACCAAATTTCTGAACAGAACATAACTCCATGTTTGCCACTTCGTGGAACAATTACAGCTTCAG
GGGATCTGGTCCACTTTCTTACATTGACGACTTCTACCGGAAGACCAAATCCAAGGCGGTG
GACCTAACGGTAATCGTTAAACGCCGAGGAAGCTTCAATCGTCCGGGATCGAATCTGGGTTT
TTTACTTGCAGCCCAAGGAAGGCTTGTCTGTTAATGGGACAGCAGTTGGGTCTGGAATGGCT
TCTATGGTCTTTTTGAAGCCAACATTTAGCTGATTGTCCGAAGTTTGTCTGCAATTTTCGCTGA
AGTCATGAATGTTAACTGAATTTACCGACCATTTGACACATAAATTGAAGCATCATCTGGACA
AATCGAAGCTGCTGCAATTATGGAACATATTTGGATGGAAGTGGGTATGTTAAAGCGCGAAGA
AATTACATGAAATGGATCCATTGCAGAAACCAAGCAAGATCGTTATGCTTTAAGAACTTCCAC
AATGGCTTGGTCCACAGATTGAAGTGATCAGATTTGCAACAAAGTCAATTGAAAGAGAAATCAAT
TCTGTTAATGACAACCTTTGATCGATGTTTCTAGAAACAAGGCTCTTCATGGTGGAACTTCCAAG
GAACCTCTATTGGTGTCTATGGATAATGCTCGTTGGCTATTGCTTCAATTGGGAACTCATGTT
TGCTCAATTTCTGAACTGTTAATGATTTTACAACAATGGGCTCCATCAATCTGTCAGGTGGA
AGGAATCCCAAGTTGGATTATGATTCAAAGGTGCTGAAATGCCATGGCTTCTATTGTTCCGAA
CTTCAATCCTGCTAATCCTGTTACCAATCATGTCCAAGTCTGAGCAACATAATCAGGATGTTA
ATTCTTGGGATTGATTTCCGCAAGGAACTCAGAAGCTGTTGATATTTGAAGCTTATGCTTC
TACTACTTAGTTGCTCTGTGCAAGCTATTGATTGAGGCATTGGAAGAAACTTGAGGAACAC
GGTAAAGAACTGTGAGCCAGATTGCTAAGAAGGTTTGACCACTGGTCCAATGGTGAACCTC
ACCTTCAAGATTCTGTGAGAAGGATGCTCAAAGCCGTTGACCGGAATACGTTTTGCTTACA

TTGATGATCCTTGCACTGCTACTTACCCATTGATGCAGAAATTGAGACAAGTTCTTGTGAGCATG
CATTGACAAATGGTGAGAGTGAGAAGATGCAAGCACTTCAATTTCCAGAAAATTGACGATTT
GAAGAGGAATTGAAGACTTTGTTGCCTAAGGAAGTTGAGAGTGCAAGAGTGGCACTTGAGAATG
GGTCAATGTGGCAGTGCCAAACAGAAATCAAGGAATGCAAGATTACCCATTGTATAAATTTGTG
AGGGAAGAGCTTGGAAGTGGGCTTTGACTGGTGAAAAAGTTAGGTACCTGGTGAGGAATTTG
ACAAGGTTTTACAGCTATGTGCCAGGGGAACTCATTGATCCAATGCTTGAGTGCTCAAGGAAT
GGGACGGTGCCCTCTTCCCATCTGCTAG

5 Tipo: ADN
Longitud: 1677
Nombre de la secuencia: cADN completo PAL1
SEQ ID NO: 41
Nombre del organismo: *Corchorus olitorius*

10 MLVRINTLLQGYSGIRFEILEAITKFLNQNITPCLPLRGITASGDLVPLS
YIAGLLTGRPNKAVGPNGESLNAEEAFNRAGIESGFFTLQPKGLALV
NGTAVGSGMASMVLFEANILAVLSEVLSAIFAIVMNGKPEFTDHLTHK
LKHHPGQIEAAAIMEHILDGSGYVKAAKKLHEMDPLQKPKQDRYALRT
SPQWLGPQIEVIRFATKSIEREINSVNDNPLIDVSRNKALHGGNFQGT
IGVSMNARLAIASIGKLMFAQFSELVNDNFYNNGLPSNLSGGRNPSLD
YGFKAIEIAMASYCSELQFLANPVTNHVQSAEQHNQDVNSLGLISARK
TSEAVDILKMSSTYLVALCQAIDLRHLEENLRNTVKNVTSQIAKKVLT
GANGELHPSRFCEKDLLKAVDREYVFAYIDDPSCSATYPLMQLRQVLV
EHALTNGESEKNASTSIFQKIAAFEEELKTLTPKEVESARVALENGSNVA
VPNRIKECRSYPYKVFVREELGTGLLTGEKVRSPGEEFDKVFAMCQGK
LIDPMLECLKWDGAPLPIC

15 Tipo: PRT
Longitud: 558
Nombre de la secuencia: péptido PAL1
SEQ ID NO: 42
Nombre del organismo: *Corchorus olitorius*

20 ATGCTGGAGTTTTCGGCCAAGGCACTGAGTCGTGCCATACGTTGCCGCACACGCGACGAGGGC
GGCTATGCTGTGAGGATCAACACTCTTTGCAAGGATATTCTGGCATTAGATTGAGATCTTAGA
AGCTATAACCAAGTTCTCAATGTGAACATTACCCCATGTGTGCCACTTAGGGGATCAATTACTGC
ATCTGGTGACTTGGTGCCATTGTCTACATAGCCGGCTTTGACAGGCAGGCCAATTCCAAGGC
TCTTGGGCTAACGCTGAGGCCATGAACCCACAGAAGCTTTAGCCGGGCGGAATCCATGGTG
GGTTTTTCGAGTTGCAGCCCAAGAGGGCTTGTCTTAGTCAATGGTACAGCCGTTGGGTGAGGCT

TAGCCTCTTTGGTCCTTTATGAGGCCAATGTTTTAGCAGTCCTCTCAGAAGTTTTATCAGCAATTTT
 GCTGAAGTTATGCAAGGCAAACCTGAATTCAGTACCATTTAACTCACAAGTTGAAGCATCATCCA
 GGACAAATTGAAGCTGCTGCCATTATGGAACATATTTGGATGGTAGCTCTTACATTAAGCAGCA
 CAAAAATTGCATGAAATGGATCCTCTCAGAAACCAAAACAGGACAGATATGCTCTTAGAACTTC
 CCTCAATGGCTTGGCCCTCAAATTGAAGTAATCAGGTCAGCAACCAAAATGATTGAAAGGGAAT
 CAATTCAGTGAATGATAATCCTTTAATTGATGTTTCAAGAGACAAGGCTTTACATGGAGGGAATTT
 CCAAGGTACCCCAATTGGTGTTTCAATGGAACAACCTCGTTTAGCCATTGCTGCAATTGGTAACT
 CATGTTGCTCAATTTCTGAGCTTGTAAATGATTATTACAACAATGGTTGCCTCAAATCTGTCTG
 CTAGCCGCAACCCGAGTTTGATTATGGTTTCAAAGGAGCTGAAATTGCAATGGCTTCTTATTGCT
 CAGAGCTTCAATTCCTTGGTAATCCTGTCACTAACCATGTCCAAAGTGCTGAGCAACATAACCAAG
 ATGTCAACTCATTAGGGTTGATCTCAGCAAGAAAAACAGCTGAAGCTATTGATATATTGAAGCTGA
 TGTCTTCAACTTTCTTGAATGCTTTATGCCAAGCAATTGACTTGAGGCATTGGAAGAGAATTGAA
 GAACACTGTCAAGAACAAGTTAGTCAAATTGCCAAGAGGCTTTAACCATGGGATCAACCGGTG
 AACTTCATCCATCAAGATTCTGTGAAAAAGATCTTCTCAGAGTTGTCGATCGCAACATCTCTATGC
 CTATATTGATGATCCTTGCAAGTACCCATTAATGCAGAAGTTGAGACAAGTACTGGTAGA
 CCATGCCCTTGATGAATGGTGACAATGAGAAGAACTCAACCACCTCCATTTCCAGAAGATTGGTGC
 CTTTGAAGAAGAAATTGAAAAACCTTTTGCCTAAAGAAGTTGAGAGTGCTAGAACGAATTCGAGA
 ATGGAAATGCAGCAATTCCTAACAGAATCAAAGAATGCAGGCTCTATCATTGTACAAGTTTGTA
 GGGAAAGTTCTTGGGACTAGCTGTTAACTGGTGAAAAGGTGATCTCCAGGGGAAGAATGTGAC
 AAGGTTTTCTCAGCAATTTGTGAGGAAAGTTGATTGATCATTGTTCAATGCTTGAAGGAGTG
 AATGGTGCTCCTTGCCTATATGCTAA

5 Tipo: ADN
 Longitud: 1743
 Nombre de la secuencia: cADN complete PAL2
 SEQ ID NO: 43
 Nombre del organismo: *Corchorus olitorius*

MLEFSAKALSRAHTLPHTATRAAMLVRINTLLQGYSGRFEILEAITKFL
 NVNITPCVPLRGSITASGDLVPLSYIAGLLTGRPNKALGPNGEAMNPT
 EAFSRAGIHGGFFELQPKGLALVNGTAVGSGLASLVLYEANVLAVLSE
 VLSAIFAEVMMQKPEFTDHLTHLKHHPGQIEAAAI MEHILDGSSYIKA
 AQLKHEMDPLQKPKQDRYALRTSPQWLGPQIEVIRSATKMIEREINSV
 NDNPLIDVSRDKALHGGNFQGTPIGVSMNTRLAIAAIGKLMFAQFSE
 LVNDYYNNGLPNSLSASRNPSLDYGFKGAEIAMASYCSELQFLGNPVT
 NHVQSAEQHNQDVNSLGLISARKTAEADILKLMSSSTFLIALCQAIDLR
 HLEENLKNTVKNTVSQIAKRVLTMGSNGELHPSRFCEKDLLRVVDREH
 LYAYIDDPSCASYPMLQKLQVLVDHALMNGDNEKNSTTSIFQKIGAF
 EEELKTLKPKEVESARIEFENGNAAPNRIKECRSYPLYKFVREVLTSL
 TGEKVISPGEEDKVFSAICAGKLIDPLFQCLKEWNGAPLPIC

15 Tipo: PRT
 Longitud: 580
 Nombre de la secuencia: péptido PAL2
 SEQ ID NO: 44
 20 Nombre del organismo: *Corchorus olitorius*

TGAAATTTCAAACATCATATTTTACTAGCGAATTAATTTAGATATAGAGGAACATTAAATGGT
AGGCTAAATATACCCATTTAATATTGCTTGATCAATTAAGATCAAAATATTAATGTCGACGAGA
TTATTTTTCAGTCCAATATTCTATCTATAAAAAAGAAAGGAAAAAACAACATTTTGACATT
AAATCTTCTTGATATTATATCTTCTCAGCAGCTTAACACAAAATATCATGATGAGAGGCAACCT
AAGACTGAATCCGAAGTGATTGTCCTCTTGTTGCTGGCAGAGGAAGATGACCTTTCTCTTGCA
AGCTCTCACTGTTTCATATAGGCATTGCTCACTGGGTGAATCCAACCCCAAGTCCCTCAGCTT
TTGGTTTGTGAACCTGTAGGGCTTGTCTCTGGGTTCTCTCATCCGAGCACCTGCATGGATTITTA
ATTGTAATTATATCATAAATTAGTTAGAAAAAGGATTGAGAATTTCTATTTAATACTAATATTTT
CTATTAGCGTTTCGAGTCTATTAGAAAATTCATTCTTGATAATTAACACTACGAGTCCAGTACTCTAC
AATCAGACCTGACCGGTATGTATGAGTTTAAATCATACGTTGTTACCTAACCTCCATTTCATCATT
CAAAGACAAAAGCCTAAAGCTATCTTAGAGAGTGGGAATTTTCAAAGACAATTTCAAAGGAG
AAATGGGAGGACCCATTTTGGCCATTGTTGACGCCCCCATAGTTCATAGAAAATTTCAAAGAAAAG
AGGCCTTTTCTTCAACCATATTTCCCTTAATTGGGAAATCCCATGATTGAATCTTTGAGTAGGT
GTGGATAGAATAGAATAGAATATCATTTTCTTTTATCTAAGTCTCTTGAAATGTTGTTTGTAA
AGTGACATTAGGTAGGTGAAGTCTCTCTGCACCACTACTATTTATACTAACTCAGTATATAGTAT
ACCATTTACATTTAATTTTCAACTTGAAAAAGAAATAGAAAAAATTCAGTCAAAAAACCTACACT
TCTTCTGTAGTAGTAAGATTATGTCATTTGTCAAACTGGGTACTCACTGCCCTTCCCTTCCCT
ACTCTGAATTAAGAAAAACAAATGAAATTTGCTGCCATGCTTATCCACTTAATTATCTTCCAATA
ATCTTTTCAATGGCTCTACATCTTATTCTTCTGTTTATTTTCTCCAAATTTCTTCTCTCTCT

TTTATATAGGAAGTGGTGATGTCATGAGTGATGAGAAAAATCTTTAATAAAATCAATGTCATAAT
CTTTAATATTATTTTCTTTTCTAATAAGGTTTCACTTTTGATTTCATTAAAGTTTGGTTGGACCAT
GATTGCCCTCCACTTTTCTTCAATAATATTATTATATATCAATTATCTTTTATTTTCTTTTCCATTTT
CTCAAGAAATTCCTCTGTGAACTTTAAACAACTTTAGAGAAACCATATATATCTCAAGATGA
TGAAGGCCAATATTCTAATTCAAATCTACCTTGAAGTAAATCATTTTATATATATATTTCTT
CTACATTTTATGGACCATATATATTTAATCTGAAGATGAGAACATCAAGTTCTTCTGTCACTAT
CTTCAACAAACAAAGGTTGACATTATATTTAACCAAGTTGGGATAATATCTAAAGCCTTGGTT
TTTTATTTTGAATATTGTGCGACATTTAAGACTTAGAGCAAGCAATATATATTGAATACGTGTTG
ATGTTCAAATTATTGACAATTCAAATGACCCAATTAACCAATGAATAAGATTATCATCTAATCTAA
GTTTTCATTTCTTTTGTGTTTAAAGATTAGGTGTCCTCTTACTATAAGAAACGTGTTTCATAT
AAATAATTCTTTTCAAGAGAAAGTAAAGTTAGGTTAAAGAAAGTGAATGATCTTTCCCTCTAGAT
ATTTCTAGATGTTAAATAATGAATTCGGCTCCAACTTACCAGAAATTTAGTGAAGCCCTCTA
AATTTATTTTGTGTTCTATTTTGTCTATAAATAATTAATAAATACTAGACATTTGGCTAATCTT
TGGTACATGAACCCCTAAATATTCCTTATATCTTCTAATGATATCCATGCCATGTAATCTAAAA
GTGGACCCCTAATAATCCTCAGATATGTCATTTATATGTTTTTGTGACTATAATGACAACATATC
GAACCTGTTAATGTTTGTAGTATGATATTTTCTCCCTGTTTGCAAAATGGATAGAAAAAATTGC
AAGAGCAAGTTTACATTATCTCTGCAACCAAGGGGTGCCACTTACTACTACAGCACTAGA
AAATCACTAATAAAATTTGATTACTATACTAAAATTAATCTTAATTAGCAATTTTGTGTTAATTAAG
TATACTTAAAAAGGAAATTAGAGAAATTAATCTTGGTAGGGATGGGATACTCAGGGAAGAACT
TGGCAAGAATGTCCACCACTCTCCAGGTGGAGGACGCTCTCGGCGCAGAGGTAACGGCCGGA
GGCGGAGGGATTCTCGAAGACAAGAATGTGTCTAATGAACATCTCTAACATGAACATAGGCTT

GAACTGAATTGGCATAGGTCTTTCAGAGCCAGTTAAGTACTTAAGAATGTGAATAATGCTAGCAT
TTACAGTTGATTGTAGCAATGGACCAAGCAACCAAACTGGAAGTTATTGTGACTAAGTCAACCCCTT
TTTCTTGGCTGTTCCAGGCTGCCTGCTCTGCCACAGTCTCCATAACAATACCAATTCTGCAAT
TTTTGTTTAAACATGAAAATTTGTTAATTAAAGAAAGGAAATGTTAATTTGTGCTAATTATAGA
ATGAAAATGGTGGTTAACTTAAATCCCTTTGTTGTTCTATTTTGGTAGGGGAGATTATTAATTG
GTAATACTAACAAATAGATTGTAGCCATAAGTTTACACATTCATTGCTAATTAGCCAAAAACATGCA
TGGTGTGGGATTTTATTACTAAAAAAGGAAAGACCAATTTGGACATTTAAAAACAAGC
ATTTGATCAGGTGATAGACCAAGCCAGCAAAACAGACTTTTTGTCTGTTCCCTGCTCTTCTTAA
TTCCAGAAACAGAAACAGGAAATTTCTGGTCAACACAAGGTCAAGTTAATTTTGACTTTTCAATA
CTCTCACTAATCATTTATGTTATGACCCACCACTATTTTAACTGGGCTGTTTATTACTATGTTTT

5

AATTAATACGCGTGACTTGGATTCTTAATTTCAATTAATTATTGTGATTAGTGGATTCAATTCCT
GATAATCTCAAGTATTTACTTAATTAAGAAGGTG

Tipo: ADN
Longitud: 3455
Nombre de la secuencia: cADN parcial CCR1
SEQ ID NO: 45
Nombre del organismo: *Corchorus olitorius*

15

ATGTCGAAGCAAGGAGAGCGGTGTCGTTACCGGAGCCAGCGGCCATTGGCTCTGGCTCG
TCAAGCTCCTCTGGCTCGCGTTACACTGTCCAGGAACCGTCAGAAATATCAAGGATGAGAAA
GAAACGAAGCATCTAGAATCTTTGAAGGAGCAGAATCTCGTCTCCGTCTCTCCAGATCGATCTC
CTCGATTATGATCAATCTCCGCCGTATCGAAGGCTGCGCTGGCGTCTCCACCTCGCTCCCCCT
GCACCGTCGATCAAGTTCACGACCTCAGAAGGAGCTTTGGATCCTGCAATTAAGGAACACTCA
ATGTACTGACAGCTGCCAAGGAGCTCGGTGTTAAGCGCGTGGTGTACGCTCTCTGTCTCTCCA
TAACACCTAGTCCGAACCTGGCTGCAGATAAGATTAACAGAGGATTGCTGGACTGATATTGAC
TACTGCAAGCAAAATGAGTTATGGTATCCAAATTTCAAACACTGGCTGAGAAGGCAGCATGGGA
ATTTTCAAGGAGAAAGTTTGGATGTGGTGGTGGTGAACCTGGCACTGTGATGGGTCAAATA
TTCTCCAACCTTACTGCTAGCATGTGGATGTTGTGCGCTATTGCAAGGCTGCACAGAGACAT
ATCAAGACTTTTTATGGGATCTGTCCATTTCAAAGATGTTGATTAGCACATATTTGGTGTATGA
GAACCATCAGCAAGTGAAGGCACATGTCTTGAGGCCATATCTATTATGGTATTTTGTGCGC
CAAAGTGTGAACTTACCTGAATATAATGTGCCAGTTTCCGAGGGACACTCAACCTGGACT
ATTAAGGCTAAGAATGGAGTCAAAAGCTGATGATTGGGGTTGGAATTCATCCCATGGAAC
AGATAATCAAAGATGCTGTCGAGAGTTTAAAAAGCAAGGGCCTTATTTAA

Tipo: ADN

Longitud: 969

5 Nombre de la secuencia: cADN completo CCR2

SEQ ID NO: 46

Nombre del organismo: *Corchorus olitorius*

MSKQGEAVCVTGASGAIGSWLVKLLLRGYTVHGTVRNIKDEKETKHL
ESLEGAESRLRLFQIDLLDYDAISAAIEGCAGVFHLASPTVDQVHDPQ
KELLDPAIKGTNLVLTAAKELGVKRVVVTSSVSSITPSNWPADKIKTED
CWTIDIDYCKQNELWYPISKTLAEKAAWEFSKEKGLDVVVVNPVTVMG
PNIPPTLTASMWMLRLRLQGCTETQDFFMGSVHFKDVALAHILVYEN
PSASGRHMCLEAISHYGDFAKVAELYPEYNVPSLPRDTQPGLLRKN
GGQKLMDLGLEFIPMEQIHKDAVESLSKSGLI

10

Tipo: PRT

Longitud: 322

Nombre de la secuencia: péptido CCR2

15 SEQ ID NO: 47

Nombre del organismo: *Corchorus olitorius*

TGCGATCGTGCCAAACCAAACTGGAATCTCAAACCAACGGGGCTTCGATGAGACATCATGGA
CTGACTTGGAGTACTGTAAGAGTCGACAGAAATGGTATCCAGTTTCGAAGACGATGGCAGAGAAA
ACAGCATGGGAATTTGAGAGAAGCATGGATGGATGTTGTGGCTATAAACCCAGCCAGTGT
AGGCCCTCTCTTGCAACCAACTGAATGCCAGCTGCGCTGTGTTGCAATTGTTGGAAGGATC
CAAAGATACCCAGGAATACCATTTGGTTGGAGCTGTGCATGTTAAAGATGTCGCCAAGGCTCAGA
TTTGTGTTTGTGAGTCTCTCTGCTCTGTTGATGATCTTTGCACCAATGGCATTATCAGTTTGG
ACTTTCGCTGAAACTGTCTCCACCTCTCCCTCAGTATCTGTCACAGGTTTACTGGAGATACAC
AACCTGGCTTAGTTTCTGCAAGATGACGAAAGCATTAAATTGAGCTAGGTCTAATCTTCAACC
CAGTTGAAGAAGCCGTCGAGAGACCGTGGAGAGTCTGCAAGCCAAGGCTTCTTGAAGCAGCA
GCAACCATCAGAGTCTAGAAATTAAGTCTCTTTTATCCTTTCTGTCTCTCAAAACAGTTAG
AAGCTTTATAGTTTATAGTTTATAGTTTATAGTTTATAGTTTAAAGCATGGATTCAATCAGA
AGTTTTGTGCCTAAAGCTTCTATTATTGTTTCATTGTGATAAGTTATAACAAGTATAAATA

20

Tipo: ADN

Longitud: 787

Nombre de la secuencia: cADN parcial CCR3

SEQ ID NO: 48

25 Nombre del organismo: *Corchorus olitorius*

AIVPNPNWNPQTNGAFDETSWTDLEYCKSRQKWYPVSKTMAEKTAW
EFAEKHGMDEVVAINPATCLGPLLQPNLNASCALLQLLEGSKDTQEYH
WLGAVHVVDVAKAQILLFESPSASGRYLCTNGIYQFGTFAETVSHLPQ
YPPVHRFTGDTQPLVSVCKDAKRILIELGLIFTPVEEAVRETVESLQAKG
FLKQQQPSES

Tipo: PRT
Longitud: 202
Nombre de la secuencia: péptido CCR3
SEQ ID NO: 49
Nombre del organismo: *Corchorus olitorius*

ATGGCTCTTCTTTTCATCGTCCCTTCTTACTCCTCTTGGCCTGGTCTCTCGACTTCGGCGAAAACC
ATTCCCTCCGGGGCCCAAAGGGTTACCAATCATCGGAAACATGATGATGATGGACCAATTAACCC
ACCGTGGTCTGGCCAAGCTGGCCAACAAATACGGCGCATATTCATATGAAGATGGGTATTTA
CATATGGTCGCCGTTTCAAATCCCGACATGGCTCGCCAAAGTTTCAAGTTCAAGACAACATTTCT
CAAACAGACCAGCCACCATAGCCATAAGTTATTTAACTTATGACAGAGCTGATGGCTTTTCCC
ATTACGGACCGTTTTGGAGACAGATGCGTAAGCTTTGTGAATGAAGCTTTTCAGTCGGAAAAGA
GCCGAATCATGGGAGTCTGTAAGAGATGAAGTTGATTCATGGTTCGAACCGTGTGGGCAACAC
TGGGAAATCCATTAATGTTGGTGAGTTGATTTTAACTTACGAAAAACATTATTTACAGGGCGGC
TTTTGGTCCAGTTCACAAGAAGGACAAGACAGTTCATCGGAATCTTACAAGAGTTTTCGAACT
GTTCCGGTCTTTTAAATATGCCGATTTTATCCCTGGCTGACGTGGGTGATCCTCAAGACTTAAC
AACAGGCTTAAGAAGCGCTCGTCAAGCTTAGACAAGTTCATCGATACCATTAATTGATGAACATAT
CAGAAGAGGAACAACAAGAATGTTTCTGATGATGTTGATACCGATATGGTCGATGATTGCTT
GCTTTTTACAGTGAAGAAGCTAAAGTAAATGAATCGGAGGATCTTCAAAACGCCATCAGACTCACT
AGAGAAAATATCAAGCCATTATCATGGACGTTATGTTGGAGGGACAGAGACGGTGGCATCAGC
AATAGAATGGGCCTTATCTGAGCTGATGAGAAGCCAGAAGATATGAAGAGAGTCCAACAAGAG
TTGGCTGACGTGGTGGTCTTGACCGGAAAGTCAAGAATCCGATTTGACAAAACTAACCTTCTT
AAATGCATTTGAAAGAAACCTCCGCCTCCACCCACCAATCCCACTTCTCTCCATGAGACCGCG
AGGACGCGGAGGTGCGCGGATACAGAATCCCGCCAAGTCG

Tipo: ADN
Longitud: 1161
Nombre de la secuencia: cADN parcial F5H
SEQ ID NO: 50
Nombre del organismo: *Corchorus olitorius*

MALLFIVPFLLLGLVSRRLRRKPFPPGPKGLPIIGNMMMMDQLTHRGL
AKLANKYGGIFHMKMGYLHMAVSNPDMARQVLQVDNIFSNRPAT
IAISYLTDRADMAFAHYGPFWRQMRKLCVMKLF SRKRAESWESVRD
EVDSMVRTVSANTGKSINVGELIFNLTKNIIYRAAFGSSSQEGQDEFIGI
LQEF SKLF GAFNIADFI PWLTWVDPQGLNNRLKNARQALDKFIDTIIDE
HIQKRNNKNNVSDDDVDTDMVDDLAFYSEEAKVNESED LQNAIRLTRE
NIKAIIMDV MF GGTETVASAIEWALSELMRSPEDMKRVQQLADVVG
LDRKVEESDFDKLTLKCTLKLRLHPPILLHETAEDA EVGGYRIPAKS

Tipo: PRT
Longitud: 387
Nombre de la secuencia: péptido F5H
SEQ ID NO: 51
Nombre del organismo: *Corchorus olitorius*

ATGGGTCTCAACTGGTGAAACCAATTCTACTCCTCAACTCAAGTCTCCGATAGGGAAGCAAACTTGTC
GCCATGCAATTGGCTAGTGCTCAGTTCTTCCCATGGTCTCAAACTGCCATAGAACCTGACCTAC
TTGAAGTCATGGCCAACTGGACCTGGCTGGTCTTTCTGTGCCCAAGCAAGTAGCTCTCCAAATGG
CACCAAGCAACCTTAGCGACCCGTATGCTCGACCGATATCTCGGGCTAGTCTTGTATCTCCAT
TTTAACTGTCTCTTAAGGAATCTTCTGATGGCAAAAGTGTAGAGGCTCTATGGCCTTGGCCCTGTG
TGTAATACCTGGTCAAGAATGAAGATGGTGTGCTCTTTCGCCCTTAATCTCATGAATCAAGAC
AAGGTCCTAATGGAGAGCTGGTACTACTGAAAGATGCAGTGTGGAAGGTGAATTCATTCAA
CAAGGCCTATGGCATGACCGGCTGCAAGTACCATGGCACTGACCTAGATTACAAGGTTTTCAA
CAGGGGAATGTCGTGATCACTCAACTATCCATGAAGAAGATTCTCGAGACCTACAGTGATTGC
AAGGGCTCAAAACATTTGTTGACCTTGTTGGTGGTGTGTTGTCGCACGCTCAAGATGATGCTCC
AAGCAAGCTTCCATTAAAGGGCATCAATTGTTGATGCTCATGATGAGGATGCTCAAGCTCTTC
CTGTGTTGAGCATGTTGGTGAGATATGTTGTAAAGTGTCCAAAAGGAGATGCCATTTTCATGA
AGTGATATGTCATGATTGGAGCGATGAACACTGCGTAAAATCTTGAAGAAGTGCTATGAAGCT
TTGGCAGACAATGGGAAAGTCATGTTGCCGAATGCATCTTCTGATTACCAAGATGCTAGCCTT
GCCACAAAGCTAGTTGTTTCATATCGATTGATCATGTTGGTGTCCAAACCTGGTGGGAAAGAAAG
GACAGACAAAGAAATTTGAAGCTTTGGCAAAAGGGGACGTTTCCAAGTTTCCAAGTAAAGTGT
GCTGTTTGGCACTACATCATGAGGTTCTTCAAACTGTTTAA

Tipo: ADN

5 Longitud: 1098

Nombre de la secuencia: CaffeateO -metiltransferasa (COMT), cADN completo

SEQ ID NO: 52

Nombre del organismo: *Corchorus olitorius*

MGSTGETQFTPIQVSDEEANLFAMQLASASVLPMLVLSAIELDLLEVM
AKAGPGAFLSPTVEASQLPTKNPDAPVMLDRILRLASYSILTCSLRNLPL
DGKVERLYGLGPVCKYLVKNEDGVALSALNLMNQDKVLMESWYYLKD
AVLEGGIPFNKAYGMTAFEYHGTDPFRNKVFNRGMSDHSHTIMKKILE
TYDGFEGKLKTLVDVGGGVGATLNMIVSKHPSIKGINFDLPVHIEDAPAL
PGVHVGGMDMFVSPYDKDAIFMKWICHWDSEHCVKFLKKCEALPD
NGKHVIAECILPDPTASDLATKLHVHIDCIMLAHNPGGKERTKEFEAL
AKGAGFGQFQVKCCAFGTIIMEFLKTV

10

Tipo: PRT

Longitud: 365

Nombre de la secuencia: péptido COMT

SEQ ID NO: 53

15 Nombre del organismo: *Corchorus olitorius*

GAATGGGGCGTTTCCATCTA

Tipo: ADN

20 Longitud: 20

Nombre de la secuencia: cebador delantero CAD2

SEQ ID NO: 54

Nombre del organismo: *Corchorus olitorius*

25 TTGGCAACACATTGAAACCA

Tipo: ADN

Longitud: 20

Nombre de la secuencia: cebador inverso CAD2

30

SEQ ID NO: 55

Nombre del organismo: *Corchorus olitorius*

AAGAGCTCAGGGAGCTTACT

35

Tipo: ADN

Longitud: 20

Nombre de la secuencia: cebador delantero CCoAOMT1

SEQ ID NO: 56

40 Nombre del organismo: *Corchorus olitorius*

TCTGCCGTCGGATTAAAGTGA

Tipo: ADN
 Longitud: 20
 Nombre de la secuencia: cebador inverso CCoAOMT1
 SEQ ID NO: 57
 5 Nombre del organismo: *Corchorus olitorius*

 CTTCCAACGCCAGGATCATA

 Tipo: ADN
 10 Longitud: 20
 Nombre de la secuencia: cebador delantero 4CL1
 SEQ ID NO: 58
 Nombre del organismo: *Corchorus olitorius*

 15 CTATGCTTATTGCCACCCCT

 Tipo: ADN
 Longitud: 20
 Nombre de la secuencia: cebador inverso 4CL1
 20 SEQ ID NO: 59
 Nombre del organismo: *Corchorus olitorius*

 AATACCATTGGTTGGGAGCTG

 25 Tipo: ADN
 Longitud: 20
 Nombre de la secuencia: cebador delantero CCR3
 SEQ ID NO: 60
 Nombre del organismo: *Corchorus olitorius*
 30 GGATTCAATCAGAAGTTTGTGCCT

 Tipo: ADN
 Longitud: 20
 35 Nombre de la secuencia: cebador inverso CCR3
 SEQ ID NO: 61
 Nombre del organismo: *Corchorus olitorius*

 CCAGCCACCATAGCCATAAG
 40
 Tipo: ADN
 Longitud: 20
 Nombre de la secuencia: cebador delantero F5H
 SEQ ID NO: 62
 45 Nombre del organismo: *Corchorus olitorius*

 ACAGAATCCCCGCCAAGT

 Tipo: ADN
 50 Longitud: 20
 Nombre de la secuencia: cebador inverso F5H

REIVINDICACIONES

1. Una molécula de ácido nucleico aislada que tiene:
 - (a) Al menos 90% de identidad de secuencia con una secuencia de ácido nucleico seleccionada del grupo que consiste de: SEQ ID NO: 1, 3, 5, 7, 9, y 15; o
 - (b) Al menos 95% de identidad de secuencia con la secuencia de ácido nucleico seleccionada del grupo que consiste de: SEQ ID NO: 11 y 13.
2. La molécula de ácido nucleico aislada de la reivindicación 1, en donde dicha molécula tiene:
 - (a) al menos un 95% de identidad de secuencia con dicha secuencia de ácido nucleico seleccionada del grupo que consiste de SEQ ID NO: 1, 3, 5, 7, 9 y 15;
 - (b) al menos 98% de identidad de secuencia con dicha secuencia de ácido nucleico seleccionada del grupo que consiste de: SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 y 15;
 - (c) al menos 99% de identidad de secuencia con dicha secuencia de ácido nucleico seleccionada del grupo que consiste de: SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 y 15, o
 - (d) 100% de identidad de secuencia con la secuencia de ácido nucleico seleccionada del grupo que consiste de SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, y 15.
3. La molécula de ácido nucleico aislada de la reivindicación 1 o 2 donde dicha molécula se selecciona del grupo que consiste de la SEQ ID NO: 1, 3, 5, 7, 9, 11, 13 y 15.
4. Una molécula de polipéptido aislada que tiene:
 - (a) al menos 90% de identidad de secuencia con una secuencia de aminoácido seleccionada del grupo que consiste de la SEQ ID NO: 2, 4, 6, 8, y 10; o
 - (b) al menos 95% de identidad con una secuencia de aminoácido seleccionada del grupo que consiste de la SEQ ID NO: 12 y 14.
5. La molécula de polipéptido aislada de la reivindicación 4, en donde dicha molécula tiene:
 - (a) al menos 95% de identidad de secuencia con dicha secuencia de aminoácido seleccionada del grupo que consiste de la SEQ ID NO: 2, 4, 6, 8, y 10;
 - (b) al menos 98% de identidad de secuencia con dicha secuencia de aminoácido seleccionado del grupo que consiste de la SEQ ID NO: 2, 4, 6, 8, 10, 12, y 14
 - (c) al menos 99% de identidad de secuencia con dicha secuencia de aminoácido seleccionada del grupo que consiste de la SEQ ID NO: 2, 4, 6, 8, 10, 12, y 14; o
 - (d) 100% de identidad de secuencia con dicha secuencia de aminoácido seleccionado del grupo que consiste de la SEQ ID NO. 2, 4, 6, 8, 10, 12 y 14.
6. Un par de cebadores delantero e inverso útil para la amplificación de un cADN seleccionado del grupo que consiste de la SEQ ID NO. 53 y la SEQ ID NO: 54
7. Un vector de expresión que comprende una molécula de ácido nucleico aislada de la reivindicación 1.
8. Un anticuerpo aislado o un fragmento de unión de antígeno del mismo que se une específicamente a una molécula de polipéptido de la reivindicación 4.
9. Una planta transfectada que comprende el vector de expresión de la reivindicación 7.
10. Un material derivado de la planta transfectada de la reivindicación 9, en donde dicho material comprende el vector de expresión de la reivindicación 7.
11. Una semilla de la planta transfectada en la reivindicación 9 en donde dicha semilla comprende el vector de expresión de la reivindicación 7.

12. Un método para elaborar una planta transgénica, que comprende las etapas de:

transfectar al menos una célula vegetal con un vector de la reivindicación 7; y

5 hacer crecer dicha al menos una célula vegetal en una planta.

13. Un método para mejorar el crecimiento, el rendimiento de fibra, la resistencia de la fibra. La resistencia a la enfermedad, o la utilización de agua en una planta de yute, que comprende incorporar en una planta de yute una secuencia de ácido nucleico no nativa de la reivindicación 1.

10

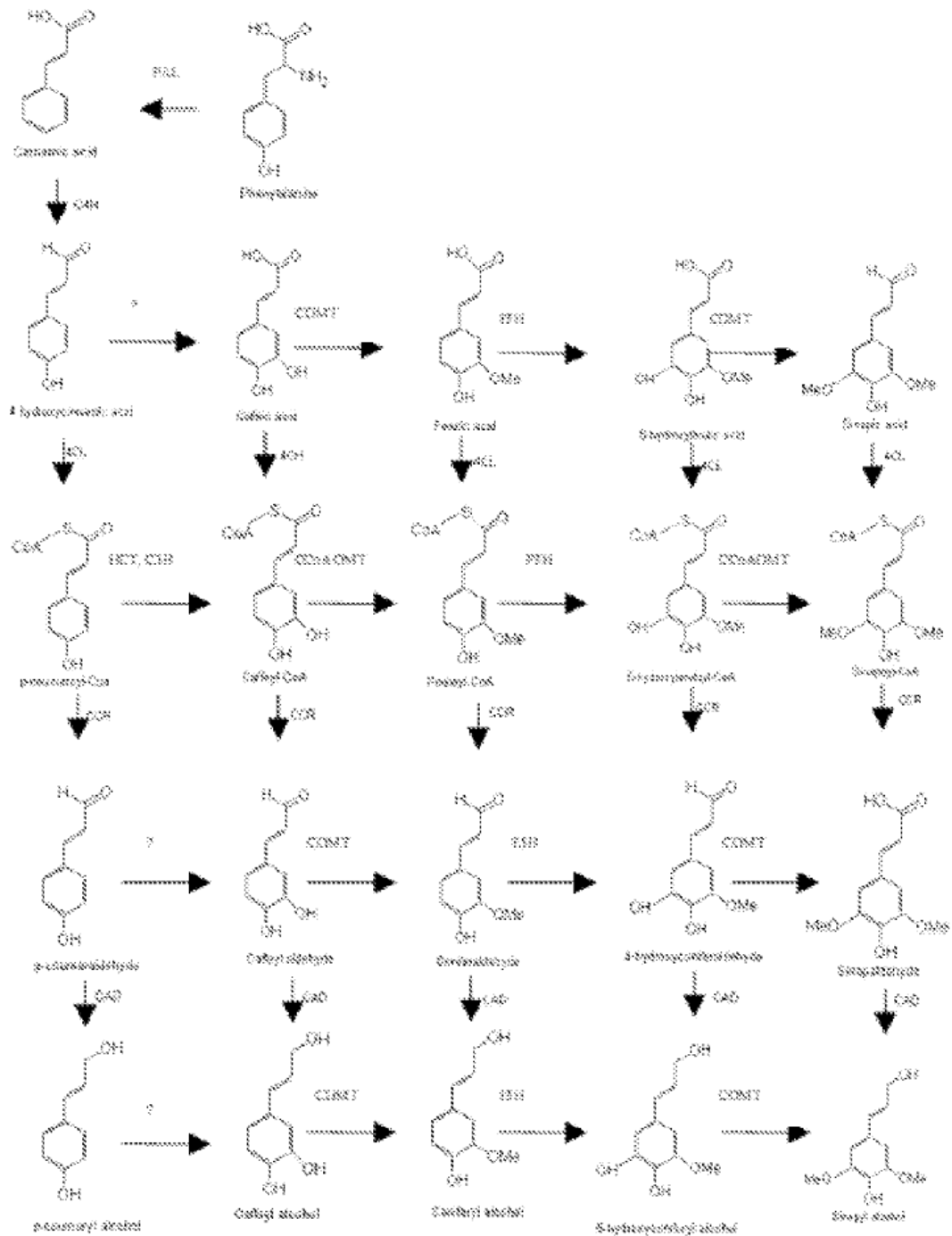


FIGURA 1

[illegible]


```

ColCCoACMT2          -MATN-TQEQQTQAGRRQEVGHYSLLQSDALVQYILETSVYPREPEPMKELRELTAKHP- 57
GhiCCoACMT2 gi:229368460 -MATNTTQEQQPAAGRRQEVGHYSLLQSDALVQYILETSVYPMPEPEPMKELRELTAKHP- 58
PtrCCoACMT gi:3023436   -MATN-GEEQQSQAQRHQEVGHYSLLQSDALVQYILETSVYPREPEPMKELRELTAKHP- 57
ColCCoACMT1          -MAFT-QAEQQTQASPRQEVGHYSLLQSDALVQYILETSVYPREPEPMKELRELTAKHP- 57
ColCCoACMT3          MGSTIGETQFTPTQVSEDEANLPMQLASASVLMVLKSAIELELLEWMAKAGPGAFLSPT 60
GhiCCoACMT1 gi:253509567 MGSTIGETQMTPTQVSEDEANLPMQLASASVLMVLKSAIELELLEWMAKAGPGAFLSPT 60
                        1 . . . . . * * : : * : : : * * : : *

ColCCoACMT2          ---WNLMITTSADEGQFLNMLIKLINAKNTMEIGVYT-----GYSLLATALALPDD 104
GhiCCoACMT2 gi:229368460 ---WNLMITTSADEGQFLNMLIKLINAKNTMEIGVYT-----GYSLLATALALPDD 108
PtrCCoACMT gi:3023436   ---WNIMITTSADEGQFLNMLIKLINAKNTMEIGVYT-----GYSLLATALALPDD 104
ColCCoACMT1          ---WNLMITTSADEGQFLNMLIKLINAKNTMEIGVYT-----GYSLLATALALPDD 104
ColCCoACMT3          EVASQLPTKNDAPVMLERILRLASYSILTCSLNLPDQKVERLYGLGPFVCKYLVKNE 120
GhiCCoACMT1 gi:253509567 EVASKLPTTNDAPVMLERILRLASYNVLTCSLRTFPQKVERLYGLGPFVCKYLVKNE 120
                        : : * . . . * : : : * : : : : : : : : : * . . . . * : *

ColCCoACMT2          GKILAMD--INRENYELG-LPVIQKAGVAHKIEFKEGPALPVLQMLVEDEKNHGSYDFIF 161
GhiCCoACMT2 gi:229368460 GKILAMD--INRENYELG-LPVERIAGVAHKIEFKEGPAMPVLQMLVEDEKNHGSYDFIF 162
PtrCCoACMT gi:3023436   GKILAMD--INRENYELG-LPVIQKAGVAHKIDFKEGPALPVLQMLVEDCKYHGSYDFIF 161
ColCCoACMT1          GKILAMD--INRENYELG-LPVIQKAGVAHKIDFKEGPALPVLQMLVEAGTYHGTDFIF 161
ColCCoACMT3          GVALSALNANCKVLMESWYLLKDAVLGGIPFNKAYGNTAFETHTGTPDRENKVFNRQM 180
GhiCCoACMT1 gi:253509567 GVTLSALSANCKVLMESWYLLKDAVLGGIPFNKAYGNTAFETHTGTPDRENKVFNRQM 180
                        * * : : * : : : : : : : * * : : : : : : : : : : : : : :

ColCCoACMT2          VDAD-----KINYINYNKRLIELVKVGGLIGYDNTLWG----- 195
GhiCCoACMT2 gi:229368460 VDAD-----KINYINYNKRLIELVKVGGLIGYDNTLWG----- 196
PtrCCoACMT gi:3023436   VDAD-----KINYINYNKRLIELVKVGGLIGYDNTLWG----- 195
ColCCoACMT1          VDAD-----KINYINYNKRLIELVKVGGLIGYDNTLWG----- 195
ColCCoACMT3          SDHSTIITMKRILETYDGFEG-LKTLVDVGGGTCATLNMIVSKHPSIKGINFELPHVIEDA 239
GhiCCoACMT1 gi:253509567 SDHSTIITMKRILETYDGFEG-LKTLVDVGGGTCATLNMIVSKYPTIKGINFELPHVIEDA 239
                        * . : . . . * * . * . * . * . . . . .

ColCCoACMT2          -----SVWAPPDAPLRKYVRYR-----DFVLELNKALADP--RIEIC 282
GhiCCoACMT2 gi:229368460 -----SVWAPPDAPLRKYVRYR-----DFVLELNKALADP--RIEIC 283
PtrCCoACMT gi:3023436   -----SVWAPPDAPLRKYVRYR-----DFVLELNKALADP--RIEIC 282
ColCCoACMT1          -----SVWAPPDAPLRKYVLYR-----DFVLELNKALADP--RIEIC 282
ColCCoACMT3          PALP-GVEHVGGGMFVSVPKGEAIEMKWI CHWSDEHCVKFLKNCYEALPDNGKVI VAE 298
GhiCCoACMT1 gi:253509567 PSCPVGVEHVGGGMFVSVPKGEAIEMKWI CHWSDEHCVKFLKNCYEALPDNGKVI VAE 299
                        * . * * : * : : : * : : : * * : : : *

ColCCoACMT2          MLPVGDGITLCRRIK----- 247
GhiCCoACMT2 gi:229368460 MLPVGDGITLCRRYK----- 248
PtrCCoACMT gi:3023436   MLPVGDGITLCRRIQ----- 247
ColCCoACMT1          QLPVGDGITLCRRIK----- 247
ColCCoACMT3          ILPDYPDASLATKLWVHIDCINLAHNPQKERTKEFEALANGAGFGQFQKCCAFGTYI 358
GhiCCoACMT1 gi:253509567 ILPDYPDPSLATKLWVHIDCINLAHNPQKERTAKEFEALANGAGFGQFQITCSAFGTNI 359
                        ** . . * . : :

ColCCoACMT2          -----
GhiCCoACMT2 gi:229368460 -----
PtrCCoACMT gi:3023436   -----
ColCCoACMT1          -----
ColCCoACMT3          MEFLKIV 365
GhiCCoACMT1 gi:253509567 MEFLKSV 366

```

FIGURA 3

```

Col14CL1      -----SAPQAPELSGEDIIFRSKLSDIYHK--HIFLHSCFEMTSKV 41
Comp4CL1 gi:254514718 -----SAPQAPELSGEDIIFRSKLSDIYHK--HIFLHSCFEMTSKV 41
Col14CL4      -----SEKSG---YGRDGIYRSIRPLVLEKSPHLEMTISLPMMSY 45
Roa4CL gi:255565415 -----SEKSG---YGRDGIYRSIRPLVLEKSPHLEMTISLPMMSY 45
Col14CL6      KATSLNHSFPFPISEKXITFFDQWYSFETGIYTSKHAESISTNSYLDVVSSTPSHQ--H 55
Pto4CL gi:224074401 -----SEKLTATASNTI-LPEWYSFETGIYTSKHAESIPLEASDFLDVVSSTPSHN-H 51
          .  *  *  .  :  *  *  :  *  *  :

Col14CL1      ASKPCLINGITGQYTHHEVELTARRRAGLHKLGVQGRQVINLLLENTFPEVLSILGAS 151
Comp4CL1 gi:254514718 ASKPCLINGITGQYTHHEVELTARRRAGLHKLGVQGRQVINLLLENTFPEVLSILGAS 151
Col14CL4      PHHPALLIADSEKILTPEQFFKSTWIKLSHAFINLGINRNDKALIFAPMSIQFELCTFAYT 155
Roa4CL gi:255565415 PHHPALLIADSEKILTPEQFFKSTWIKLSHAFINLGINRNDKALIFAPMSIQFELCTFAYT 155
Col14CL6      QQNTASTDESSGFTISYKILLFLVQSGASGLHKLGVKGDVLLLENTFPEVLSILGAS 158
Pto4CL gi:224074401 NGLTALIDSSGFTISYKILLFLVQSGASGLHKLGVKGDVLLLENTFPEVLSILGAS 151
          . . : * . . . 25 : . . 25 : : : * : : : : : : : : : * : : : * :

Col14CL1      FLGAVCTAARFFFTAFENAGAKASKARIIITQASYVDKVKFAQEN-DVWAGCISAPF 165
Comp4CL1 gi:254514718 FLGAVCTAARFFFTAFENAGAKASKARIIITQASYVDKVKFAQEN-DVWAGCISAPF 165
Col14CL4      ALGAIATIRAFPIYVREISNQINDSTKILLVTVFELSDKVKFAQEN-DVWAGCISAPF 165
Roa4CL gi:255565415 SIGAIATIRAFPIYVREISNQINDSTKILLVTVFELSDKVKFAQEN-DVWAGCISAPF 165
Col14CL6      YLGAIVTIRAFPIYVREISNQINDSTKILLVTVFELSDKVKFAQEN-DVWAGCISAPF 175
Pto4CL gi:224074401 YLGAIVTIRAFPIYVREISNQINDSTKILLVTVFELSDKVKFAQEN-DVWAGCISAPF 175
          : * : : : * : : : * : : : : : : : : : : : : : : : : : : :

Col14CL1      C-----CLHPELITQADEN--SLPEVEINPDQVVALFYSSTTGIFKGVDLTHRLVIT 211
Comp4CL1 gi:254514718 C-----CLHPELITQADEN--SLPEVEINPDQVVALFYSSTTGIFKGVDLTHRLVIT 212
Col14CL4      SVKQVPIKLSFHELLDLAGVTELPANVSFQGDIALYLSSTTGIFKGVDLTHRLVIT 215
Roa4CL gi:255565415 N-----SRKSPDHLVGLGCHSEFFINVFQNDIALYLSSTTGIFKGVDLTHRLVIT 215
Col14CL6      VDPLP-----FYKLRGQSGHAPFVFIQQDTAKINYSSTTGIFKGVDLTHRLVIT 211
Pto4CL gi:224074401 IYK-----GVGVKPIVFIQQDTAKINYSSTTGIFKGVDLTHRLVIT 212
          . . . * : : : ***** . ***** : : :

Col14CL1      SVAGQVGEENFKNITFHS-DDVILCTILRPHIVALSDELQIRAGAILISQNPDIQLL 275
Comp4CL1 gi:254514718 SVAGQVGEENFKNITFHS-DDVILCTILRPHIVALSDELQIRAGAILISQNPDIQLL 271
Col14CL4      ALNITISQ--ELAGIK-HRVFLCVLPLFHYVGLAVAFSGLQIGVILVSSKAPFELPL 277
Roa4CL gi:255565415 ASQNTISQ--ELAGIK-HRVFLCVLPLFHYVGLAVAFSGLQIGVILVSSKAPFELPL 277
Col14CL6      MIELVTEFASQVETPGSETVLAALRPHIVALSDELQIRAGAILISQNPDIQLL 291
Pto4CL gi:224074401 NVELVTEFASQVETPGSETVLAALRPHIVALSDELQIRAGAILISQNPDIQLL 272
          . . . . . * . . . ***** : . . . . . * : : :

Col14CL1      DLQKVKITIAFPPFIVLAIAKSE-TEKIDLSISQVYSGAFLKLELDAVRAKPPG 325
Comp4CL1 gi:254514718 DLQKVKITIAFPPFIVLAIAKSE-TEKIDLSISQVYSGAFLKLELDAVRAKPPG 325
Col14CL4      KAKKVKITIAFPPFIVLAIAKSE-TEKIDLSISQVYSGAFLKLELDAVRAKPPG 325
Roa4CL gi:255565415 KAKKVKITIAFPPFIVLAIAKSE-TEKIDLSISQVYSGAFLKLELDAVRAKPPG 325
Col14CL6      KVIDKVKITIAFPPFIVLAIAKSE-TEKIDLSISQVYSGAFLKLELDAVRAKPPG 325
Pto4CL gi:224074401 KVIDKVKITIAFPPFIVLAIAKSE-TEKIDLSISQVYSGAFLKLELDAVRAKPPG 325
          . . . . . ***** : : . . . . . ***** : : :

Col14CL1      AKLDQGVNTIAGFVLAELGFAKEFFIKSQAGTIVWAKKIVIPDTGASLPKQAG 355
Comp4CL1 gi:254514718 AKLDQGVNTIAGFVLAELGFAKEFFIKSQAGTIVWAKKIVIPDTGASLPKQAG 355
Col14CL4      AVWAGQGVNTIAGFVLAELGFAKEFFIKSQAGTIVWAKKIVIPDTGASLPKQAG 355
Roa4CL gi:255565415 ALNAGQGVNTIAGFVLAELGFAKEFFIKSQAGTIVWAKKIVIPDTGASLPKQAG 355
Col14CL6      VDTIQGVNTIAGFVLAELGFAKEFFIKSQAGTIVWAKKIVIPDTGASLPKQAG 355
Pto4CL gi:224074401 VDTIQGVNTIAGFVLAELGFAKEFFIKSQAGTIVWAKKIVIPDTGASLPKQAG 355
          . . ***** : : . . . . . ***** : : : * :

Col14CL1      EICIRGQVNTIAGFVLAELGFAKEFFIKSQAGTIVWAKKIVIPDTGASLPKQAG 445
Comp4CL1 gi:254514718 EICIRGQVNTIAGFVLAELGFAKEFFIKSQAGTIVWAKKIVIPDTGASLPKQAG 435
Col14CL4      EICIRGQVNTIAGFVLAELGFAKEFFIKSQAGTIVWAKKIVIPDTGASLPKQAG 435
Roa4CL gi:255565415 EICIRGQVNTIAGFVLAELGFAKEFFIKSQAGTIVWAKKIVIPDTGASLPKQAG 435
Col14CL6      EICIRGQVNTIAGFVLAELGFAKEFFIKSQAGTIVWAKKIVIPDTGASLPKQAG 435
Pto4CL gi:224074401 EICIRGQVNTIAGFVLAELGFAKEFFIKSQAGTIVWAKKIVIPDTGASLPKQAG 435
          * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :

Col14CL1      VASAELEKLIAPFIIIDAVVAKKIVIPDTGASLPKQAGTIVWAKKIVIPDTGASLPKQAG 505
Comp4CL1 gi:254514718 VASAELEKLIAPFIIIDAVVAKKIVIPDTGASLPKQAGTIVWAKKIVIPDTGASLPKQAG 515
Col14CL4      VASAELEKLIAPFIIIDAVVAKKIVIPDTGASLPKQAGTIVWAKKIVIPDTGASLPKQAG 515
Roa4CL gi:255565415 VASAELEKLIAPFIIIDAVVAKKIVIPDTGASLPKQAGTIVWAKKIVIPDTGASLPKQAG 515
Col14CL6      VASAELEKLIAPFIIIDAVVAKKIVIPDTGASLPKQAGTIVWAKKIVIPDTGASLPKQAG 515
Pto4CL gi:224074401 VASAELEKLIAPFIIIDAVVAKKIVIPDTGASLPKQAGTIVWAKKIVIPDTGASLPKQAG 515
          : * : : * : : * : : * : : * : : * : : * : : * : : * : : * : :

Col14CL1      YKRIEAVTFMEISIKAPSKKILARELAKLASQVY 544
Comp4CL1 gi:254514718 YKRIEAVTFMEISIKAPSKKILARELAKLASQVY 545
Col14CL4      YKRIEAVTFMEISIKAPSKKILARELAKLASQVY 547
Roa4CL gi:255565415 YKRIEAVTFMEISIKAPSKKILARELAKLASQVY 547
Col14CL6      YKRIEAVTFMEISIKAPSKKILARELAKLASQVY 547
Pto4CL gi:224074401 YKRIEAVTFMEISIKAPSKKILARELAKLASQVY 544
          : * : : * : : * : : * : : * : : * : : * : : * : : * : : * : :

```

FIGURA 4

```

Col16HCT1      MIWVKESTWVPSAEETPRVCLWKEWDLVVRPPTTSEVTFYRPSGASNTFFISKVAKEAL 50
PtcHCT q1:183585181 MIWVKESTWVPSAEETPRRGWKEWDLVVRPPTTSEVTFYRPTGASNTFFIAKVAKEAL 50
Cyc2a=HCT q1:73671233 MKIEVRESTWVRPAEETPRINLWKEWDLVVRPPTTSEVTFYRPGGASNTFFISKVAKEAL 50
* _;_*;***** ***** *****_*****_*****_**_*;_*;**

Col16HCT1      GKALVPPFPMWGRRLRRDSDGRHEIDQNGASVLFVEASTNAVIDDFGDSAPTLELRQLIPT 120
PtcHCT q1:183585181 GKALVPPFPMWGRRLRRDSDGRHEIDQNGEGVLFVEASTASVQADFGDSAPTLELRQLIPT 120
Cyc2a=HCT q1:73671233 GKALVPPFPMWGRRLRRDSDGRHEIDQDQDQVLFVEASTDVIDDFGDSAPTLELRQLIPA 120
_*;*****_**_*;_*;*****_*; ***** :_*; *****_**_*;*****_

Col16HCT1      VDTSGGIETVELINLVQVTFPKCGGSLGVGMQHAADGFGGLHFIWWSDMARGLDLETF 180
PtcHCT q1:183585181 VDTSGGIETVELINLVQVTFPKCGVSLGVGMQHAADGFGGLHFIWWSDMARGLDLETF 180
Cyc2a=HCT q1:73671233 VDTLGIETVELINLVQVTFPKCGVSLGVGMQHAADGFGGLHFIWWSDMARGLDLETF 180
***; **_*;*,*****_*****_*****_*****_*****_*****_*****_

Col16HCT1      SPIDRTLLARDDFPQSAHEHIEYQPPFALNEAPSTGEEG----LAWSPILTRDQLNA 235
PtcHCT q1:183585181 SPIDRTLLARDDFPQSAHEHIEYQPPFALNTVLETSEEE----LAWSPILTRDQLNT 235
Cyc2a=HCT q1:73671233 SPIDRTLLARDDFPQSAHEHIEYQPPFALNTAPTPTPTDSEFVEITVSPILTRDQVNA 240
*****;*****_*;*****_*;_*;_*;_*;_*;_*;_*;_*;_*;*****_*;_*;

Col16HCT1      LKAKKEGKNTIYSSVEALSGHYWRWVCRARLFDQDQETINVIATDGRARLAPPLPGY 285
PtcHCT q1:183585181 LKAKKEGKNTIYSSVEALSGHYWRWVCRARLFDQDQETINVIATDGRARLAPPLPGY 285
Cyc2a=HCT q1:73671233 LKCKKEGKNTIYSSVEALSGHYWRWVCRARLFDQDQETINVIATDGRARLAPPLPGY 290
**_*;**_*;_*;_*; *****_*;*****_*****_*****_*****_*****_*; **

Col16HCT1      FGWVIFTATPIAAGELQENPTVYAAQWINDALVRMKNQYLRSLDYLELQFQSLALVRG 335
PtcHCT q1:183585181 FGWVIFTATPIAAGELQENPTVYAAQWINDALVRMKNQYLRSLDYLELQFQSLALVRG 335
Cyc2a=HCT q1:73671233 FGWVIFTTPIAAGELQENPTVYAAQWINDALVRMKNQYLRSLDYLELQFQSLALVRG 360
*****_*;*****_*; *****_*;*****_*;*****_*;*****_*;*****_*;*****_

Col16HCT1      ANTPKCFMELGITTEWRLSINDAPPOWERFIFMCPGGIAYEGISFVLPSSTINDGSLVNAI 415
PtcHCT q1:183585181 ANTPKCFMELGITTEWRLSINDAPPOWERFIFMCPGGIAYEGISFVLPSSTINDGSLVNAI 415
Cyc2a=HCT q1:73671233 ANTPKCFMELGITTEWRLSINDAPPOWERFIFMCPGGIAYEGISFVLPSSTINDGSLVNI 420
***;*****_*****_*****_*****_*****_*;**_*;*****_*;

Col16HCT1      LQZEHMLFEMFYNDI- 435
PtcHCT q1:183585181 LQZEHMLFEMFYNDIX 435
Cyc2a=HCT q1:73671233 LQZEHMLFEMFYNDI-- 435

```

FIGURA 5


```

GarC4H gi:9965897      MDLLFLEKALLGLFVAVVLAITISKLRGKRFKLPDGLPVPVFGNWLQVGGDLNHRNLTD 60
GhiC4H2 gi:268528127 MDLLFLEKALLGLFVAVVLAITISKLRGKRFKLPDGLPVPVFGNWLQVGGDLNHRNLTD 60
ColC4H1               MDLLFLEKALLGLFVAVVLAITISKLRGKRFKLPDGLPVPVFGNWLQVGGDLNHRNLTD 60
ColC4H2               MDLLFLEKALLGLFVAVVLAITISKLRGKRFKLPDGLPVPVFGNWLQVGGDLNHRNLTD 60
*****:****:***:***:*****:*****:*****:*****:*****

GarC4H gi:9965897      LAKKFGDIFLLRMGQRLNVVVSPELAKVVLHTQGVFPGSRTRNVVFDI FTGKGQDMVFT 120
GhiC4H2 gi:268528127 LAKKFGDIFLLRMGQRLNVVVSPELAKVVLHTQGVFPGSRTRNVVFDI FTGKGQDMVFT 120
ColC4H1               LAKKFGDIFLLRMGQRLNVVVSPELAKVVLHTQGVFPGSRTRNVVFDI FTGKGQDMVFT 120
ColC4H2               LAKKFGDIFLLRMGQRLNVVVSPELAKVVLHTQGVFPGSRTRNVVFDI FTGKGQDMVFT 120
*****:*****:*****:*****:*****:*****:*****:*****

GarC4H gi:9965897      VYGEHWKMRIRINTVPFFITNKVQQYRFGWEDEAAARVVEVVKKNPEAATNGIVLRRRLQL 180
GhiC4H2 gi:268528127 VYGEHWKMRIRINTVPFFITNKVQQYRFGWEDEAAARVVEVVKKNPEAATNGIVLRRRLQL 180
ColC4H1               VY-----VYQQYRFGWEDEAAARVVEVVKKNPEAATNGIVLRRRLQL 181
ColC4H2               VYGEHWKMRIRINTVPFFITNKVQQYRFGWEDEAAARVVEVVKKNPEAATNGIVLRRRLQL 180
**                      *****:***_*:*****:****:***:*****:****

GarC4H gi:9965897      NMYNNMYRIMEDTRFESEEDPLFVRLKALNGERSRLAQSFEYNYGDFIPILRPFILRGYLE 240
GhiC4H2 gi:268528127 NMYNNMYRIMEDTRFESEEDPLFVRLKALNGERSRLAQSFEYNYGDFIPILRPFILRGYLE 240
ColC4H1               NMYNNMYRIMEDTRFESEEDPLFVRLKALNGERSRLAQSFEYNYGDFIPILRPFILRGYLE 221
ColC4H2               NMYNNMYRIMEDTRFESEEDPLFVRLKALNGERSRLAQSFEYNYGDFIPILRPFILRGYLE 240
*****:*****:*****:*****:*****:*****:*****:*****

GarC4H gi:9965897      ICKEVKDRRLQLFKDHFVEERKLGSTKSMNDGLKCAIDHILDAQQKGEINEENVLYIV 300
GhiC4H2 gi:268528127 ICKEVKDRRLQLFKDHFVEERKLGSTKSMNDGLKCAIDHILDAQQKGEINEENVLYIV 300
ColC4H1               ICKEVKDRRLQLFKDYFVEERKLGSTKSMNENGLKCAIDHILDAQQKGEINEENVLYIV 281
ColC4H2               ICKEVKDRRLQLFKDHFVEERKLGSTKSPDNNAIKCAIDHILDAQQKGEINEENVLYIV 300
:*****:*****:***:*****:***_*:***:*****:*****:*****

GarC4H gi:9965897      ENINVAIETILWSIEWGIAELVNHPFIQKLRHELDIVLGPNGQITEPDTHKLPYLQAV 360
GhiC4H2 gi:268528127 ENINVAIETILWSIEWGIAELVNHPFIQKLRHELDIVLGPNGQITEPDTHKLPYLQAV 360
ColC4H1               ENINVAIETILWSIEWGIAELVNHPFIQKLRDELDIVLGPNGQITEPDTHKLPYLQAV 341
ColC4H2               ENINVAIETILWSIEWGIAELVNHPFIQKLRNEIDIVLGPNGQITEPDTHKLPYLQAV 360
*****:*****:***:***:***:***:*****:*****:*****

GarC4H gi:9965897      IKETLRLMAIPLLVPHNNLHDAKLGGYDIPAESKILVNANWLANNPANWKNPEEFRPER 420
GhiC4H2 gi:268528127 IKETLRLMAIPLLVPHNNLHDAKLGGYDIPAESKILVNANWLANNPANWKNPEEFRPER 420
ColC4H1               IKETLRLMAIPLLVPHNNLHDAKLAGYDIPAESKILVNANWLANNPANWKNPEEFRPER 401
ColC4H2               IKETLRLMAIPLLVPHNNLHDAKLGGYDIPAESKILVNANWLANNPANWKNPEEFRPER 420
*****:*****:*****:*****:*****:*****:*****:*****

GarC4H gi:9965897      FFEERAKVEANGNDFRYLPGVGRRCPCGIILALPILGITLGRVQNFELLPPFGQSQID 480
GhiC4H2 gi:268528127 FFEERAKVEANGNDFRYLPGVGRRCPCGIILALPILGITLGRVQNFELLPPFGQSQID 480
ColC4H1               FFEERAKVEANGNDFRYLPGVGRRCPCGIILALPILGITLGRVQNFELLPPFGQSQID 461
ColC4H2               FFEERAKVEANGNDFRYLPGVGRRCPCGIILALPILGITLGRVQNFELLPPFGQSQID 480
*****:*****:*****:*****:*****:*****:*****:*****

GarC4H gi:9965897      TTEKGGQFSLHILKHSITIVAKPRQF 505
GhiC4H2 gi:268528127 TTEKGGQFSLHILKHSITIVAKPRQF 505
ColC4H1               TTEKGGQFSLHILKHSITIVL---- 481
ColC4H2               TTEKGGQFSLHILKHSITIVAKPRVF 505
*_******

```

FIGURA 7

44

```

AtbCCRgi_15237678      SLTD-----ERFVGVTCASGCIQKSLVHQILLACSEVHATVKNLQSEKETKHLELEQ  54
Alvgi_297793383      -MT-----ERFVGVTCASGCIQKSLVHLLHRCYEVHATVKNLQSEKETKHLELEQ  53
ColCCR2              -MK-----QCEAVGVTCASGCIQKSLVHLLHRCYEVHATVKNLQSEKETKHLELEQ  53
ColCCRgi_228480464    -MSENTKSGEDGQNVGVTCGSGTIGSKLVRLLDGRTVHATVMDLDEKETKHLELEQ  52
                        ::          : ,***** ,* *****  ** *** ,** ,*:::***** ,**

AtbCCRgi_15237678      LAIRLHLFENDLIQYDTVAAINSCSGVPHLASPCIVDSVQDPQKQLDPAVKGITINLT  114
Alvgi_297793383      LAIRLHLFEMDLQYDTVAADVSCSGVPHLASPCIVDSVQDPQKQLDPAVKGITINLT  113
ColCCR2              AESRLRLFGISLLDYDAISAAIEDGACGFHLASPCIVDQVHSPQKQLDPAVKGITINLT  113
ColCCRgi_228480464    AESRLRLFGISLLDYDSIVAATGSSGVPHLASPCIVDQVHSPQKQLDPAVKGITINLT  112
                        * ,** ,** ,***** ** ,* ,***** ** ,** ,** ,** ,** ,** ,** ,**

AtbCCRgi_15237678      AAKELGVRRVVVTSISAITPSPKWPADKINKEBCWAAEDVCRQNGDWYPIENTLAEKAL  174
Alvgi_297793383      AAKELGVRRVVVTSISAITPSPKWPADKINKEBCWADQVCRQNGDWYPIENTLAEKAL  173
ColCCR2              AAKELGVRRVVVTSISAITPSPKWPADKINKEBCWADIDVCRQNGDWYPIENTLAEKAL  173
ColCCRgi_228480464    AAKELGVRRVVVTSISAITPSPKWPADKINKEBCWADIVCRQNGDWYPIENTLAEKAL  172
                        **** ,* ,***** : ,***** ,* ,** ,* ,** ,* ,** ,* ,** ,* ,** ,* ,**

AtbCCRgi_15237678      WEFEKELGLDWWNPQTVMSPVPPSLNASSHMLLRLLQCTETTYENFFMGSVHPDVA  234
Alvgi_297793383      WEFEKELGLDWWNPQTVMSPVPPSLNASSHMLLRLLQCTETTYENFFMGSVHPDVA  233
ColCCR2              WEFEKELGLDWWNPQTVMSPVPPSLTASSHMLLRLLQCTETTYQDFMGSVHPDVA  233
ColCCRgi_228480464    WEFEKELGLDWWNPQTVMSPVPPSLNASSHMLLRLLQCTETTYENFFMGSVHPDVA  232
                        *** ,* ,***** ***** ** ,* ,** ,* ,** ,* ,** ,* ,** ,* ,** ,* ,**

AtbCCRgi_15237678      LAHILVYEDFYSKGRHLQVEAISHYGDFVAKVARELYRYNVKFLPRETQGLLRKNGAK  294
Alvgi_297793383      LAHILVYENPSAKGRHLQVEAISHYGDFVAKVARELYRYNVKFLPRETQGLLRKNGAK  293
ColCCR2              LAHILVYENPSAKGRHLQVEAISHYGDFVAKVARELYRYNVKFLPRETQGLLRKNGAK  293
ColCCRgi_228480464    LAHILVYENPSAKGRHLQVEAISHYGDFVAKVARELYRYNVKFLPRETQGLLRKNGAK  292
                        ***** ,* ,** ,** ,* ,***** ,* ,** ,* ,** ,* ,** ,* ,** ,* ,** ,* ,**

AtbCCRgi_15237678      KLNLGLGFISMEIIRKGVESLKNMGFIS  324
Alvgi_297793383      KLNLGLGFISMEIIRKGVESLKNMGFIS  323
ColCCR2              KLNLGLGFISMEIIRKGVESLKNMGFIS  323
ColCCRgi_228480464    KLNLGLGFISMEIIRKGVESLKNMGFIS  323
                        *** ,** ,** ,* ,** ,** ,* ,** ,* ,** ,* ,** ,* ,** ,* ,** ,* ,**

```

FIGURA 9

Alineación de secuencia múltiple CLUSTAL 2.1

```

Col CCR3
Res CCR g1:255556687
Act CCR g1:15226935
-----
KRIQKREAVCVICANGPFGSRLIQTLQHGCTTIHASIYPASDPKHLPLISSESEGEI 60
---MKRETVCVICANGPFGSRLITLIEAGCTTIHASIYPGSDPTLLQLPQSDSK--- 53

Col CCR3
Res CCR g1:255556687
Act CCR g1:15226935
-----
INLKLVERALLVDIAICRAVEGQGVPHVASPCTLSEPKDPEKELVLPANQGTINYLEAR 130
--IKIFERALLDSRAISRAIDGCAGVPHVASPCTLSEFVLPKELVLPANVGTINYLEAR 111

Col CCR3
Res CCR g1:255556687
Act CCR g1:15226935
-----
--AIVKRNWNPQINMAFDETSWTDLEYCKSRQSWYFVVENTMAENIA 43
KRPKVERWVLTSEISIAVFNPNMP--AKVPTIESSWILLYKSRQSWYFVVKSLAEKDA 178
KRPVVERWVLTSEISIAVFNPNMP--EKVPTIESSWILLYKSRQSWYFVVKSLAEKDA 169
      *;*****      .**;***;*****;***;****;*

Col CCR3
Res CCR g1:255556687
Act CCR g1:15226935
-----
WEPAEKHMDVVAIHPSTICGLLQPELNASCAVLLQLLEGKDTQYEWLCAVHVKEVA 103
WEPAEKHMDVVAIHPSTICGLLQPELNASCAVLLQLLEGKDTQYEWLCAVHVKEVA 103
WEPAEKHMDVVAIHPSTICGLLQPELNASCAVLLQLLEGKDTQYEWLCAVHVKEVA 103
***;*** :;*:;*;***;*****;***;*** ***;***;***;*****;*****

Col CCR3
Res CCR g1:255556687
Act CCR g1:15226935
-----
KQILLFESPSASGRILDTNGIQFGTPTASTVHLPQYFVHRFTGDTQPGWVSCDKAAK 153
KAVLLFEAPASGRILDTNGIQFGTPTASTVHLPQYFVHRFTGDTQPGWVSCDKAAK 153
KAVLLFEAPASGRILDTNGIQFGTPTASTVHLPQYFVHRFTGDTQPGWVSCDKAAK 153
*_*;***;*.****;*****. ** **;***;*. ** *;*****;_*;****

Col CCR3
Res CCR g1:255556687
Act CCR g1:15226935
-----
RLIELGLVFTFVEEAVFETVESLQAKFLKQQQPSSE 203
RLIELGLVFTFVEEAVFETVESLQAKFLKHHH-SES 214
RLIELGLVFTFVEEAVFETVESLQAKFL----- 218
*****;*.**;*** *;_*; **

```

FIGURA 10

Alineación de secuencia múltiple CLUSTAL 2.1

```

Eg1 F5H g1:233970299 MDIFVYFSQQLSLVQQLQQLQSPWILLVQPLLLFLGLVARLARKEFFEPGPGSLFVIGM 60
Col F5H -----KALLPITVFPILLGLVERLRARKEFFEPGPKLPITIGM 36
Pto F5H g1:6688937 -----

Eg1 F5H g1:233970299 NLSMDELTHRCGLASIAKHYGQIFHRRMSEFLHNVAVESFEVARQVLQVHDMIFERRPATIA 120
Col F5H -----KSSSDQLTHRCGLAKLANKYQIFHRRMSEFLHNVAVESFEVARQVLQVQDMIFERRPATIA 96
Pto F5H g1:6688937 -----GLFHRMSEFLHNVAVESFEVARQVLQVQDMIPERRPATIA 40
                        *:***:***:***** *:***:*****:***:*****:**

Eg1 F5H g1:233970299 ISYLYDRADKAPAHYQCFPWQQRRLQNNKLFERRRAESKESVRIEVDIMKRTVAGSEGT 180
Col F5H ISYLYDRADKAPAHYQCFPWQQRRLQNNKLFERRRAESKESVRIEVDIMKRTVAGSEGT 156
Pto F5H g1:6688937 ISYLYDRADKAPAHYQCFPWQQRRLQNNKLFERRRAESKESVRIEVDIMKRTVAGSEGT 130
*****:***:***:*****:***:***:***:***:***:***:***:***:***:***:***:***

Eg1 F5H g1:233970299 AVNIGELVLFELTRDIIVRAAFQTSSTEQDEPISILQEPFKLPQAFNIADPFPYLSWIDP 240
Col F5H SINNGELIPKLTKNIIYRAAPQSSSQSQQEFICILQEPFKLPQAFNIADPFPYLSWIDP 216
Pto F5H g1:6688937 PVNNGELIPKLTKNIIYRAAPQ-ARNSQSQEFIKILQEPFKLPQAFNISDPFPYLSWIDP 188
.:***:***:***:***:***:***:***:***:***:***:***:***:***:***:***

Eg1 F5H g1:233970299 QGLTARLVKARQSLQGFIDHIIIDHMDKRRKSTSSQGGQDEVDIMWDDLAFYSIEAKV 300
Col F5H QGLNRRLKNAQCALDKFISTIIIDHMDKRRKSTSSQGGQDEVDIMWDDLAFYSIEAKV 276
Pto F5H g1:6688937 QGLTARLVKARQCALDKFIDHIIIDHMDKRRK-QNNYS---EERSTDWDDLAFYSIEAKV 248
***:***:***:***:***:***:***:***:***:***:***:***:***:***:***

Eg1 F5H g1:233970299 NEEDELQNSIELTRDNKAIIDVDMFGCTETVSAIEWAMELMSPEDEKRVQQLADV 360
Col F5H NEEDELQNAIELTRDNKAIIDVDMFGCTETVSAIEWAMELMSPEDEKRVQQLADV 336
Pto F5H g1:6688937 NEEDELQNAIELTRDNKAIIDVDMFGCTETVSAIEWAMELMSPEDEKRVQQLADV 308
***:***:***:***:***:***:***:***:***:***:***:***:***:***:***

Eg1 F5H g1:233970299 VGLDRVVEESDFDKLTPLKCTILKETERLHPPIPLLHETAEADAVGQYRIFAKE----- 420
Col F5H VGLDRVVEESDFDKLTPLKCTILKETERLHPPIPLLHETAEADAVGQYRIFAKE----- 396
Pto F5H g1:6688937 VGLDRVVEESDFDKLTPLKCTILKETERLHPPIPLLHETAEADAVGQYRIFAKE----- 368
***:***:***:***:***:***:***:***:***:***:***:***:***:***:***

Eg1 F5H g1:233970299 YAIQRDPGQSWTEPDKFKPFRFLSEGMSPYKGSNTEFIFPGSGRRSCPGMQLGLYALDMAY 480
Col F5H -----
Pto F5H g1:6688937 YAIQRDNKSWTEPDKFKPFRFLSEGMSPYKGSNTEFIFPGSGRRSCPGMQLGLYALDLAV 452

Eg1 F5H g1:233970299 AALLHSTWELFDGMKPSKMDAGDVGLTAPRSTRVAVPTPLVRLV 528
Col F5H -----
Pto F5H g1:6688937 AALLHSTWELFDGMKPSKMDAGDVGLTAPRSTRVAVPTPLVRLV 496

```

FIGURA 11

Alineación de secuencia múltiple CLUSTAL 2.1

```

Col COMT      MGSTG-ETQSTPTQVSDDEEANLFAMQLASASVLPWVLKSAIELLLEVMKAGPGAFLSP 59
Ghi gi:253509562 MGSTG-ETQMTPTQVSDDEEANLFAMQLTSASVLPWVLKSAIELLLEIMAKAGPGAFLSP 59
Eca gi:262474806 MGSTGSETQMTPTQVSDDEEANLFAMQLASASVLPWVLKSAIELLLEIMAKAGPGAFLSP 60
*****

Col COMT      TEVASQLPTKNPDAFVMLDRILRLLASYS ILTCSLRNLDPGKVERLYGLGPVCKVLVNE 119
Ghi gi:253509562 KELASQLPTSNPDAFVMLDRILRLLATYS ILTCSLRNLDPGKVERLYGLGPVCKFLTKNE 119
Eca gi:262474806 GEVAASQLPTQNPEAFVMLDRIFRLLASYSVLICTLRLELDPGKVERLYGLAPVCFTLVKNE 120
*:*:****,*:*****:***:*:*:*:*:*****,*:*:*

Col COMT      DGVALSALNLMNQEKVIMESWYYLKDAVLEGGIPFNKAYGMTAFEYHGTDPRFNKVFNRG 179
Ghi gi:253509562 DGVTLASALNLMNQEKVIMESWYYLKDAVLEGGIPFNKVYGMTAFEYHGTDPRFNKVFNRG 179
Eca gi:262474806 DGVSLAALNLMNQEKIIMESWYYLKDAVLEGGIPFNKAYGMTAFEYHGTDPRFNKIFNRG 180
***:;:*:*****:*****:*****:*****:*****:*****:*****

Col COMT      MSDHSTITMKMILETYDGFEGSLKILVWVGGSATLNMIVSKHPSIMG INFELPHVIE DA 239
Ghi gi:253509562 MSDHSTITMKMILETYDGFEGSLKILVWVGGSATLNMIVTKHPSIMG INFELPHVIE DA 239
Eca gi:262474806 MSDHSTITMKMILETYKGFEGLETVWVGGSATLNMIVARYPSMKG INFELPHVIE DA 240
*****:*****:*****:*****:*****:*****:*****:*****

Col COMT      PALPGVEHVGGDMFVSVEKGDALFMKWICHWSDEHCKFLKCYEALPDNGKVIVAE CI 299
Ghi gi:253509562 PAYPGVEHVGGDMFESVEKGDALFMKWICHWSDEHCKFLKCYEALPDNGKVIVAE CI 299
Eca gi:262474806 PPLPGVKHVGGDMFVSVEKGDALFMKWICHWSDDHCKFLKCYEALPDNGKVIVAE CV 300
*_*:*****:*****:*****:*****:*****:*****:*****

Col COMT      LPDYPDASLATKLVVHIDCIMLANNPGGKERTKEFEALANGAGFGQGVKCCAFGTYIM 359
Ghi gi:253509562 LPDYPDPSLATKLVVHIDCIMLANNPGGKERTKEFEALARSAGFGQGVKCCAFGTYIM 359
Eca gi:262474806 LPVYPDTSLATKLVVHIDCIMLANNPGGKERTKEFEETLANGAGFGQGVKCCAFGTHVM 360
**_*:*****:*****:*****:*****:*****:*****:*****

Col COMT      EFLKTV 365
Ghi gi:253509562 EFLKRV 365
Eca gi:262474806 EFLKIA 366
**:*

```

FIGURA 12

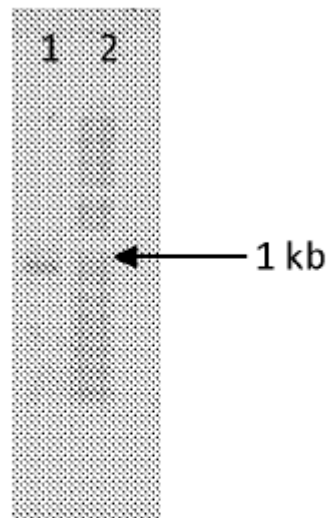


FIGURA 13

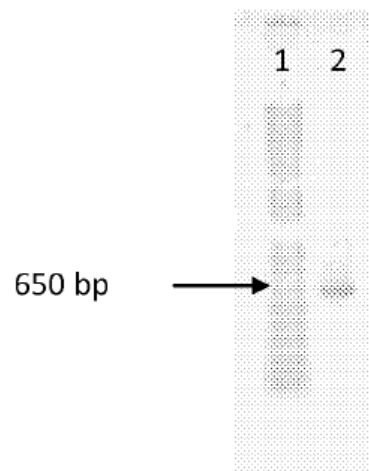


FIGURA 14

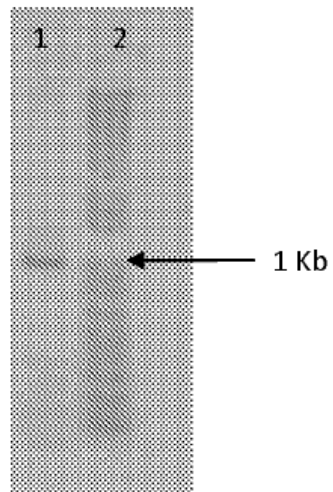


FIGURA 15

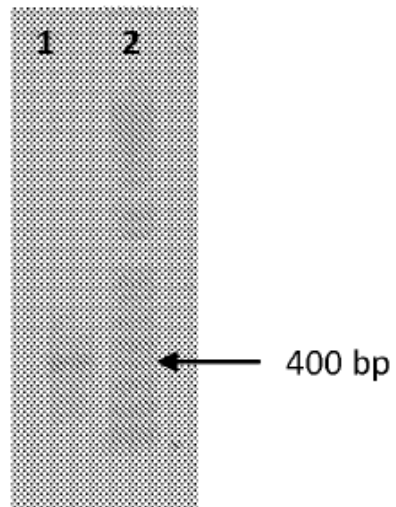


FIGURA 16

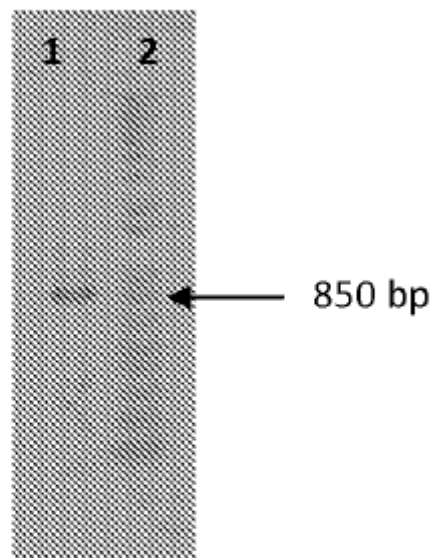


FIGURA 17