

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 685 605**

51 Int. Cl.:

C07K 14/005 (2006.01)

C12Q 1/70 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **24.06.2011 PCT/EP2011/003119**

87 Fecha y número de publicación internacional: **29.12.2011 WO11160848**

96 Fecha de presentación y número de la solicitud europea: **24.06.2011 E 11737889 (3)**

97 Fecha y número de publicación de la concesión europea: **30.05.2018 EP 2585477**

54 Título: **Moléculas de virus TT reorganizadas para su uso en diagnóstico, prevención y tratamiento del cáncer y autoinmunidad**

30 Prioridad:

23.11.2010 US 952300
23.06.2010 US 821634
23.11.2010 EP 10014907
23.06.2010 EP 10006541

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
10.10.2018

73 Titular/es:

DEUTSCHES KREBSFORSCHUNGSZENTRUM
(100.0%)
Im Neuenheimer Feld 280
69120 Heidelberg, DE

72 Inventor/es:

DE VILLIERS, ETHEL-MICHELE y
ZUR HAUSEN, HARALD

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 685 605 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

Moléculas de virus TT reorganizadas para su uso en diagnóstico, prevención y tratamiento del cáncer y autoinmunidad

5 La presente invención se refiere a moléculas reorganizadas de (a) una secuencia de virus TT específica y (b) a una secuencia de nucleótidos que codifica un polipéptido que muestra homología con proteínas de mamífero asociadas con el cáncer o una enfermedad autoinmunitaria que son capaces de replicarse de manera autónoma.

10 La familia *Anelloviridae* incluye torque teno virus (VTT), TT-midivirus (TTMDV) y TT-minivirus (TTMV), originándose la mayoría en muestras de origen humano (Nishizawa y col., 1997; Takahashi y col., 2000; Ninomiya y col., 2007; Okamoto, 2009; Biagini y de Micco, 2010). La pluralidad de esta familia de virus de ADNmc se refleja no solo en la secuencia de ADN, sino también en el tamaño y la organización del genoma.

15 Se han efectuado múltiples intentos para encontrar un sistema *in vitro* adecuado para la replicación y propagación de virus TT. Se han demostrado formas replicativas de su ADN en células de médula ósea y en el hígado (Kanda y col., 1999; Okamoto y col., 2000a, c, d). La sangre periférica actúa como reservorio para los virus TT (Okamoto y col., 2000b) y parece ser que la replicación *in vivo* se produce preferentemente en células mononucleares activadas (Maggi y col., 2001b; Mariscal y col., 2002; Maggi y col., 2010). Aunque se ha investigado la transcripción *in vitro* en una diversidad de líneas celulares (Kamahora y col., 2000; Kamada y col., 2004; Kakkola y col., 2007; 2009; Qiu y col., 2005; Müller y col., 2008), ha sido difícil lograr la replicación a largo plazo que dé lugar a la producción del virus (Leppik y col., 2007).

20 La presencia de una serie de moléculas subvéricas de TT reorganizadas intragenómicas en muestras de suero y la transcripción *in vitro* de una molécula subvérica que suponía únicamente el 10 % del genoma completo, dio lugar a la discusión de si los virus TT podrían compartir similitudes con la familia de plantivirus *Geminiviridae* (Leppik y col., 2007; de Villiers y col., 2009).

25 Los geminivirus tanto mono como bipartidos se asocian con satélites de ADN monocatenarios para formar complejos inductores de enfermedades (Saunders y col., 2000; Stanley, 2004; Nawaz-ul-Rehman y Fauquet, 2009; Jeske 2009; Paprotka y col., 2010; Patil y col., 2010).

30 Las infecciones se producen en los primeros días de vida, infectándose un 100 % de los bebés antes de cumplir un año de edad. Sin embargo, sigue sin estar clara la vía primaria de infección (Kazi y col., 2000; Peng y col., 2002; Ninomiya y col., 2008). La naturaleza ubicua de las infecciones por VTT ha dificultado los esfuerzos para asociarla con la patogénesis de enfermedades (Jelcic y col., 2004; Leppik y col., 2007; de Villiers y col., 2009; Okamoto, 2009). Se ha comunicado una posible asociación etiológica con enfermedades del hígado (revisado en Okamoto, 2009), el tracto respiratorio (Biagini y col., 2003; Maggi y col., 2003a,b; Pifferi y col., 2005), neoplasias malignas hematopoyéticas (Jelcic y col., 2004; Leppik y col., 2007; de Villiers y col., 2002; 2009; Shiramizu y col., 2002; Garbuglia y col., 2003; zur Hausen y de Villiers, 2005) y enfermedades autoinmunitarias (Sospedra y col., 2005; Maggi y col., 2001a; 2007; de Villiers y col., 2009). Durante los últimos años, se han compilado algunos datos que son indicativos de una asociación de la infección con virus TT con tumores malignos humanos. Una tasa elevada de carga de virus TT se ha observado en una biopsia de bazo de un paciente con linfoma de Hodgkin (24 genotipos individuales de VTT). Del mismo modo, otros informes describen una mayor tasa de prevalencia de VTT en el cáncer colorrectal y de esófago y en cánceres malignos hematopoyéticos, en comparación con tejido no tumoral del mismo paciente o de otros pacientes. Sin embargo, la ubicuidad de estas infecciones hizo que una interpretación de estos resultados fuera bastante difícil y no permitió relacionar estas observaciones con el desarrollo de tumores.

El documento WO 2008/138619 desvela moléculas subvéricas de virus TT reorganizadas y de replicación autónoma que se asocian con el cáncer.

Jelcic y col. (2004) desvela un poli(ácido nucleico) de virus TT reorganizado que comprende la secuencia detth25.

45 Sospedra (2005) desvela un genoma completo de un poli(ácido nucleico) de virus TT reorganizado y sugiere que el genoma completo del virus TT está asociado con enfermedades autoinmunitarias. El documento WO 00/46407 desvela un vector basado en VTT, en el que se inserta un gen de interés dentro del genoma completo de un virus TT.

50 Durante los experimentos que dieron como resultado la presente invención, se han aislado más de 200 genomas de virus TT. Los aislados que se agrupan en el género *Alphatorquevirus* (aproximadamente 3,8 kb de tamaño) comparten una homología de secuencia de ADN en la organización de su genoma. Una serie corta (71 pb) de la región intergénica está altamente conservada entre todos los aislados de VTT humanos (Peng y col., 2002) y se usa ampliamente para demostrar la infección por virus TT. Se analizaron muestras de un amplio espectro de enfermedades respecto de la presencia de ADN de torque teno virus aplicando amplificación por la PCR de esta región conservada (Jelcic y col., 2004; Leppik y col., 2007; de Villiers y col., 2009; Sospedra y col., 2005; de Villiers y Gunst, resultados no publicados). La identificación de tipos de virus TT individuales, sin embargo, requiere de la amplificación de genomas de longitud completa. Hasta ahora, se han aislado 93 genomas de longitud completa de VTT (aproximadamente 3,8 kb) de muestras humanas (Jelcic y col., 2004; Leppik y col., 2007; de Villiers y col., 2009;

presentes experimentos). Estos incluyeron muestras obtenidas de individuos sanos, pacientes con leucemia y linfoma, artritis reumatoide, esclerosis múltiple y enfermedad renal. La presente descripción describe la replicación y transcripción *in vitro* de 12 aislados tras la transfección inicial del ADN genómico y seguido de la propagación del virus usando células infectadas congeladas o partículas purificadas. Se clonaron y caracterizaron moléculas subvídicas reorganizadas de μ TTY (microTTY) que aparecieron en los países tempranos. Estas también se propagaron de manera independiente en cultivo celular, dando como resultado nuevas estructuras similares a partículas que son capaces de infectar a células 293TT sin virus.

La ubicuidad de los torques teno virus, junto con la ausencia de sistemas de cultivo *in vitro* adecuados, ha dificultado el progreso en la investigación de este grupo de virus. La multitud y heterogeneidad de tipos (Biagini y de Micco, 2010; Okamoto, 2009), así como su presencia ubicua en células hematopoyéticas (Takahashi y col., 2002; Kanda y col., 1999; Zhong y col., 2002), se han añadido al retraso en la recogida de información acerca de si estos virus están implicados en la patogénesis de cualquier enfermedad. Se aisló un espectro de tipos de VTT (Jelicic y col., 2004; Leppik y col., 2007; de Villiers y col., 2009; presente divulgación). Con frecuencia se han aislado genomas de longitud completa de una serie de tipos de VTT de una muestra individual, dependiendo de la composición de los cebadores usados para la amplificación por la PCR de larga distancia. La distribución dispersa de los nuevos aislados de la presente divulgación en un árbol filogenético del género *Alphatorquevirus* (figura 18) indica su heterogeneidad, independientemente del origen. Se ha observado con frecuencia la variación en la organización del genoma que resulta de diferencias en la identidad de secuencia por todo el genoma entre aislados del mismo tipo y ha suscitado preguntas acerca de la funcionalidad de estos genes modificados.

En el pasado, se ha intentado propagar genomas de VTT en una serie de líneas celulares y en monocitos de sangre periférica en diversas condiciones de cultivo *in vitro*. Se logró un éxito moderado con aislados individuales en estirpes celulares de linfoma de Hodgkin y en células 293T. Sin embargo, la replicación fue lenta y se produjo a bajos niveles (Leppik y col., 2007; Leppik y de Villiers, datos no publicados). Para los estudios de la presente divulgación, se modificó por ingeniería genética la estirpe de células de riñón embrionario humano 293TT para que expresase altos niveles del antígeno T grande de SV-40 (Buck y col., 2005). La transfección de genomas de VTT en estas células dio como resultado la replicación del ADN vírico y la producción de partículas viroides con un tamaño de aproximadamente 30 nm (figura 22). Las estructuras de estas partículas viroides difieren de aquellas publicadas previamente como partículas de VTT (Itoh y col., 2000). Esto es posiblemente a consecuencia del aislamiento de esta última de heces.

En la actualidad no pueden explicarse las diferencias en el nivel de replicación de ADN observadas entre aislados de VTT. La información filogenética no proporciona una respuesta. Cabe destacar que 6 aislados (VTT-HD14, VTT-HD15 y VTT-HD16) que se originaron a partir de biopsias cerebrales de pacientes con esclerosis múltiple se replicaron mucho menos en el sistema de la presente divulgación. La producción de virus (figura 22) o la propagación del virus (figuras 19 y 21) no parecieron estar influenciadas, a pesar de los diversos niveles de replicación de ADN o de las modificaciones en la organización del genoma, que incluyeron ORF1 modificadas. Sin embargo, los niveles de transcripción parecían estar influenciados y se detectó una menor cantidad de los transcritos comunes descritos para otros tipos de VTT en los cuatro aislados de VTT-14 que en los cultivos de VTT-HD15a y VTT-HD16a. Los transcritos comunicados previamente (Leppik y col., 2007; Kakkola y col., 2009) se aislaron de todos los cultivos infectados. De forma interesante, no se identificó ningún transcrito que codificase la proteína ORF1 de longitud completa (que se sospecha que desempeña un papel importante en la codificación de la cápside vírica, aunque aún no se ha demostrado) de cualquiera de los tipos de VTT-HD estudiados, a pesar del aislamiento de partículas viroides que portan el genoma de longitud completa de todos los cultivos infectados. Se identificó una serie de secuencias de supuestas proteínas que pueden ser el resultado de productos de fusión de cualesquiera dos o tres genes. Las estrategias de traducción que se sabe que usan los virus, tales como el rastreo ribosomal (*leaky scanning*), el reinicio y cambio de vía traduccional (*ribosomal shunting*) (Ryabova y col., 2006) podrían estar implicadas en este caso. La codificación dual en fases de lectura alternativas es un mecanismo adicional que podría estar implicado (Kovacs y col., 2010). De forma interesante, también se aislaron transcritos de la región de control. En este caso, se identificaron dos grupos de transcritos. Un grupo implicado en transcritos que abarcan al menos parte de la región intergénica y que se extienden al resto del genoma, abarcando los genes conocidos. El segundo grupo consistía en transcritos de longitudes diversas y sin capacidad de codificación reconocible. Se ha propuesto que la naturaleza de la región intergénica de VTT con su alto contenido en GC puede tener un papel en el bloqueo de la replicación dependiente de la transcripción (Belotserkovskii y col., 2010).

Una observación muy destacable en el presente estudio es la formación de moléculas subvídicas de manera temprana durante el ciclo de replicación de la mayoría de aislados obtenidos. Se distinguieron dos grupos de moléculas subvídicas. Con bastante frecuencia, se produjo la formación de múltiples moléculas de ADN subvídicas de diversos tamaños y de manera extensa en cultivos infectados por VTT-HD20a, VTT-HD3a y VTT-HD1a. Se ha demostrado la presencia de partículas subvídicas reorganizadas similares en muestras de suero (Leppik y col., 2007). La transfección en células L428 (estirpe celular de linfoma de Hodgkin) de un pequeño número de genomas subvídicos originarios de suero dio como resultado una replicación y transcripción limitada durante unos pocos días (de Villiers y col., 2009). Los datos mostrados en la presente divulgación indican un papel como partículas interferentes defectuosas durante la replicación *in vitro* del genoma de longitud completa. La replicación del genoma de longitud completa se reduce mientras que simultáneamente aumentan los niveles de moléculas subvídicas (figura 19b). Se han mostrado moléculas subvídicas similares de manera ocasional e inconsistente en cultivos de los otros 9 aislados, pero no tuvieron influencia en la replicación del genoma de longitud completa. Esta diferencia asimismo resalta no solo la diversidad entre tipos de VTT, sino también que este fenómeno no es el resultado de un artefacto

de la PCR. También se ha informado de moléculas interferentes defectuosas similares en Geminivirus, donde se acumulan durante una replicación inadecuada (Jeske, 2009).

El segundo grupo de moléculas subvíricas μ VTT se desprendió durante la replicación de los aislados de VTT VTT-HD14b, VTT-HD14c, VTT-HD14a y VTT-HD14e, VTT-HD15a, VTT-HD16a, VTT-HD1a, VTT-HD23b, VTT-HD23d y VTT-HD23a, y se mantuvo constante en cuanto a su tamaño y composición durante la propagación, según se evidenció tras la clonación y secuenciación. Su producción en el caso de los últimos 4 aislados parecía estar influenciada por las condiciones de cultivo. De forma interesante, la molécula subvímica μ VTT-HD1 en el cultivo infectado por VTT-HD1a fue detectable en el cultivo infectado incluso después de la pérdida del genoma de longitud completa parental detectable (figura 19c). Se aislaron dos moléculas, μ VTT-HD23.1 (409 bases) y μ VTT-HD23.2 (642 bases) de los 3 cultivos infectados por VTT-HD23. μ VTT-HD23.2 está formada por la molécula μ VTT-HD23.1 más una duplicación de 306 nt de la molécula más pequeña. Las moléculas subvíricas (μ VTT-HD14) que se aislaron de los 4 cultivos de VTT-HD14 fueron idénticas en cuanto a su secuencia y aparecieron de manera muy temprana tras la transfección inicial del genoma parental.

La producción de estas moléculas más pequeñas no parecía estar influenciada por la variación en la estructura del genoma entre aislados del mismo tipo de VTT. Todas las moléculas subvíricas estaban compuestas por partes del tipo de VTT parental, aunque las regiones genómicas implicadas, difirieron. Todas se amplificaron mediante PCR de larga distancia usando los mismos cebadores consecutivos que para la amplificación del genoma parental. Se ha observado previamente la replicación episómica de una molécula subvímica de VTT aislada de una muestra de suero a lo largo de un periodo de 23 días (de Villiers y col., 2009). Se ha demostrado la presencia de ARN subvímico multimérico durante este proceso. Las moléculas subvíricas comunicadas en la presente divulgación son capaces de replicarse de manera autónoma, pueden propagarse *in vitro* (figura 21) y parecen estar relacionadas con pequeñas estructuras proteicas observadas en estos cultivos por microscopía electrónica (figura 22). Se desconoce si se transmiten como parte de un virus TT infeccioso o si se inducen únicamente tras la infección por el virus parental y después se transmiten infectando de manera autónoma otras células. Se han asociado ADN subvímicos similares con el complejo de enfermedad por geminivirus (Stanley, 2004). Los satélites β potencian los fenotipos de síntomas en plantas. Comparten una red de interacciones proteicas con los geminivirus y dependen de ellos para la trans-replicación, la encapsidación y la transmisión del vector. La única secuencia compartida entre los satélites β y los geminivirus se encuentra en el origen de replicación corto (Nawaz-ul-Rehman y Fauquet, 2009; Patil y Fauquet, 2010; Paprotka y col., 2010). Esto contrasta con las moléculas subvíricas de VTT (μ VTT) que comparten secuencias prácticamente idénticas con el genoma parental. El efecto citopático observado durante la propagación *in vitro* de las moléculas subvíricas de VTT de la presente invención aporta indicios acerca de su posible papel como componente inductor de enfermedad de algunos torques virus. Se han identificado motivos de firma de proteínas implicadas en enfermedades autoinmunitarias mediante análisis *in silico* de supuestas proteínas expresadas por estas moléculas subvíricas, así como de transcritos de virus aislados de los cultivos infectados por VTT. La observación de un ADN que codifica una proteína que contiene un motivo de firma de una proteína de mamífero asociada con el cáncer o una enfermedad autoinmunitaria relacionada con la región altamente conservada de 71 pb del virus TT (HCR) es la base de la siguiente conclusión: Las fases abiertas de lectura reorganizadas de VTT y μ VTT codifican epítomos antigénicos que imitan a secuencias de proteínas celulares que son atacadas en el cáncer o en enfermedades autoinmunitarias. Su secuencia compartida pero no idéntica debería provocar una respuesta inmunitaria contra estos epítomos también presentes en tejido normal.

Un nuevo papel para los virus TT en el cáncer humano y la autoinmunidad

La sorprendente observación de ADN de células hospedadoras asociado a una forma aparentemente de cadena sencilla de la HCR del virus TT, es la base para la siguiente conclusión: todavía no se ha demostrado que secuencias víricas de TT estén integradas en el ADN celular de cadena doble, continuando dentro de los cromosomas de la célula hospedadora. Por lo tanto, el hallazgo opuesto de ADN de la célula hospedadora, relacionado en un estado de cadena sencilla a la HCR de VTT, debería tener una importancia biológica. Los datos actuales indican su persistencia a largo plazo como episomas en líneas celulares de cáncer humano, lo que apunta a un papel de esa persistencia en la proliferación celular. Dos aspectos parecen requerir una consideración específica: un posible papel de estos recombinantes en el cáncer y en la autoinmunidad.

Una posibilidad es la integración aleatoria de secuencias de la célula hospedadora en episomas de VTT. Esto puede suceder después de un desplazamiento de la cadena en el curso de una replicación de ADN aberrante o después de una transcripción inversa del ARN celular. En el caso de una integración aleatoria, una mayor cantidad de recombinantes debe ser inocua e inofensiva para las células que son portadoras de estos recombinantes. Una propiedad que favorece el crecimiento de los transcritos de la HCR de VTT, así como la integración y la transcripción de los genes de la célula hospedadora que estimulan el crecimiento, su modificación en el proceso de integración o su alteración por la HCR de VTT, darán lugar, sin embargo, a consecuencias proliferativas. Estos episomas deben adquirir propiedades inmortalizantes y, bajo ciertas condiciones, propiedades transformadoras. En combinación con modificaciones adicionales del genoma de la célula hospedadora, pueden dirigir el crecimiento maligno. Este modo de acción revela una semejanza distante con la inserción de oncogenes celulares en genomas retrovirales.

El concepto de oncogén de VTT

Las consideraciones anteriores se resumen en la Fig. 4. Obviamente, la recombinación entre la región reguladora de VTT y los ácidos nucleicos celulares debe ser un proceso relativamente frecuente, ya que tales recombinantes se

encuentran en la mayoría de las líneas celulares analizadas hasta el momento. También debe contribuir a la proliferación celular, de lo contrario la persistencia regular de tales moléculas, en parte durante décadas de proliferación continua, sería difícil de explicar. Se supone que este tipo de recombinación es un proceso aleatorio, que implica diferentes tipos de genes celulares. La función codificante de la HCR de VTT y/o la captación de genes que dirige la proliferación celular, o que bloquea la función de antagonistas de proliferación, o la inhibición de la diferenciación celular, deberían dar lugar a una acumulación de células que contienen estos tipos de recombinantes. Se prevé que esto, en combinación con eventos mutacionales o recombinantes adicionales de las células que albergan tales recombinantes de ácido nucleico de células hospedadoras de VTT, proporciona una ventaja selectiva para las células que son portadoras de este tipo de episomas. La presencia de estos últimos representaría un factor de riesgo principal para la conversión a cancerígeno. En este sentido, esas recombinaciones deben tener una importancia general para diferentes tipos de cánceres humanos, aunque se espera para tipos de cánceres individuales un cierto grado de especificidad para un conjunto limitado de genes.

Las implicaciones de este modelo son profundas. Alcanzan desde la prevención del cáncer, la detección temprana, hasta la terapia del cáncer. El papel importante de las infecciones con VTT y de la persistencia de HCR de VTT se destaca en la información disponible. La prevención de estas infecciones debe reducir el riesgo de desarrollar los recombinantes descritos. El diagnóstico de recombinantes específicos contribuirá probablemente a la evaluación del riesgo de cáncer. Se esperan profundas implicaciones en la terapia del cáncer: la HCR de VTT surge como el principal determinante de la persistencia y el mantenimiento de los episomas de una sola hebra. Puesto que esta región parece formar parte de un marco de lectura abierto, debería ser vulnerable frente a ARNs o ADN pequeños de interferencia. Por lo tanto, ofrece una diana adecuada para futuras deliberaciones terapéuticas.

Otros dos aspectos merecen un examen: ciertos paralelos que parecen existir en la carcinogénesis retroviral en roedor y pollo, y el uso de sistemas de vectores basados en VTT de replicación autónoma para la terapia génica. La mutagénesis por inserción, la captación y la modificación de genes que estimulan el crecimiento celular que los convierte en oncogenes, se han analizado frecuentemente en sistemas animales. Esto, por lo tanto, todavía no se ha descrito para los cánceres humanos. ¿Los virus TT reemplazan este nicho en las células de humanos y otros primates? ¿Compite el VTT con éxito con infecciones por retrovirus para hacerse cargo de su papel en especies específicas? La persistencia episómica del ADN de cadena sencilla surge, sin embargo, como una notable diferencia en la carcinogénesis inducida por retrovirus.

Las moléculas de ADN subvirales con replicación autónoma de aproximadamente 400 bases, originarias de VTT, se han descrito antes. Es tentador especular que ellas o recombinantes específicos de células hospedadoras de VTT puedan representar sistemas de vectores óptimos para futuros enfoques en la terapia génica y para la construcción de cromosomas artificiales.

El concepto de autoinmunidad del ADN de células hospedadoras de VTT recombinantes

La existencia de recombinantes de ácido nucleico de células hospedadoras de VTT también permite un enfoque novedoso de aspectos de enfermedades autoinmunes y otras enfermedades crónicas (potencialmente, incluso afecciones como la arteriosclerosis y la enfermedad de Alzheimer). La modificación o la alteración de proteínas celulares pueden proceder de eventos por inserción de genes celulares en el ADN de una sola cadena o sobre las diferentes HCRs, ejercida por elementos de VTT (Fig. 5). Podrían proporcionar una explicación conveniente para las reacciones autoinmunes, incluso para las locales, al igual que en la esclerosis múltiple (MS, del inglés "multiple sclerosis") o la enfermedad de Crohn. En los dos últimos casos, en particular, la reactivación de otras infecciones locales (potencialmente virus de tipo herpes) proporcionaría un estímulo para una amplificación local y la actividad génica de los respectivos recombinantes de ácido nucleico de células hospedadoras de VTT. En la MS, esto podría explicar episodios recurrentes de la progresión de la enfermedad. Un modelo del concepto de autoinmunidad se representa en la Figura 5.

Del mismo modo, se han identificado moléculas de virus TT reordenadas, de 719, 642, y 621 bases que se replican de manera autónoma después de la transfección de líneas celulares específicas. La composición de sus ADNs y el proceder de genotipos de VTT completos específicos, se muestra en la Figura 6. Aquí los reordenamientos dan como resultado nuevos marcos de lectura abiertos, en parte con epítomos relacionados con los de la diabetes juvenil y la artritis reumatoide.

Conclusión

Los modelos de la presente invención para un papel de los recombinantes de ácido nucleico de células hospedadoras de VTT, se basan en la demostración de las moléculas químicas de cadena sencilla entre la HCR de VTT y el ADN de la célula hospedadora y moléculas de VTT reordenadas con replicación autónoma de pesos moleculares sustancialmente reducidos. Ambos, el concepto de oncogén de VTT y el concepto de autoinmunidad de VTT, ofrecerán sin duda nuevos enfoques para la prevención, el diagnóstico y, en particular, para la terapia de estas afecciones y mejorarán el pronóstico de los pacientes respectivos.

Definiciones

A menos que se defina de otro modo, todos los términos técnicos y científicos usados en este documento tienen el

mismo significado que entienden comúnmente los expertos normales en la técnica a la que pertenece la invención. Aunque se puede emplear cualquier procedimiento y material, similar o equivalente a los descritos en el presente documento, en la práctica o al someter a ensayo la presente invención, se describen los procedimientos y materiales preferidos. Para los fines de la presente invención, se definen a continuación los siguientes términos.

5 Por "motivo de firma de una proteína de mamífero que esté asociada con una enfermedad autoinmunitaria" se entiende una secuencia de aminoácidos que muestre una identidad sorprendente con n motivo que pueda encontrarse en cualquiera de las proteínas listadas en la tabla 1. Preferentemente, la longitud del motivo de firma es de al menos 5 aa, preferentemente al menos 10 aa, más preferentemente al menos 20 aa y lo más preferentemente, al menos 30 aa y/o el grado de identidad de este motivo de firma respecto de un motivo correspondiente en una
10 proteína de mamífero es de al menos el 50 %, 60 %, 70 %, 80 %, 90 % o 95 %.

Por "*anticuerpo*" se entiende una proteína de la familia de las inmunoglobulinas que es capaz de combinarse, interactuar o de otro modo asociarse con un antígeno. El término "*antígeno*" se usa en el presente documento en su sentido más amplio para hacer referencia a una sustancia que es capaz de reaccionar y/o inducir una reacción inmunitaria. Típicamente, pero no necesariamente, los antígenos son extraños para el animal hospedador en el que
15 producen reacciones inmunitarias.

Por "*epítipo*" se entiende la parte de una molécula antigénica contra la que se dirige una respuesta inmunitaria particular. Típicamente, en un animal, los antígenos presentan varios o incluso muchos determinantes antigénicos simultáneamente. Por lo tanto, los términos "epítipo" y "determinante antigénico" significan una secuencia de aminoácidos que es inmunorreactiva. En general, un epítipo consiste en 4 y más normalmente 5, 6, 7, 8 o 9
20 aminoácidos contiguos. Sin embargo, ha de quedar claro que un epítipo no necesita estar compuesto por una secuencia de aminoácidos contigua. La secuencia inmunorreactiva puede estar separada por un enlazador, que no es una parte funcional del epítipo. El enlazador no es necesariamente una secuencia de aminoácidos, sino que puede ser cualquier molécula que permita la formación del epítipo deseado.

La expresión "*muestra biológica*" tal y como se emplea en esta memoria se refiere a una muestra que se puede extraer, sin tratar, tratada, diluida o concentrada a partir de un animal. Muestra biológica se refiere a cualquier muestra biológica (tejido o fluido) que contiene un poli(ácido nucleico) de VTT de la invención y se refiere más particularmente a muestras de suero sanguíneo, muestras de plasma, muestras de biopsias, muestras de líquido cefalorraquídeo, etc..
25

Por "*vehículo*" se entiende cualquier sustancia de peso molecular normalmente alto a la que una sustancia no inmunogénica o poco inmunogénica (por ejemplo, un hapteno) está ligada de forma natural o artificial para mejorar su inmunogenicidad.
30

El término "*diagnosis*" se emplea en esta memoria en su sentido más amplio para incluir la detección de un antígeno reactivo con una molécula que se une a un antígeno de una subinmunoglobulina. También se incluye dentro de su alcance, el análisis de mecanismos de trastornos. De acuerdo con ello, el término "diagnosis" incluye el uso de anticuerpos monoclonales para fines de investigación, como herramientas para detectar y comprender los mecanismos asociados con una enfermedad o afección de interés. También incluye el uso diagnóstico del poli(ácido nucleico) de VTT de la invención para la detección de ARN homólogo o complementario transcrito a partir de tales moléculas. El término "*inmunogenicidad*" se emplea en esta memoria en su sentido más amplio para incluir la propiedad de provocar una respuesta inmune dentro de un organismo. La inmunogenicidad depende normalmente en parte del tamaño de la sustancia en cuestión, y en parte de en qué grado es no similar a las moléculas hospedadoras. Se considera generalmente que las proteínas altamente conservadas tienden a tener una inmunogenicidad más bien baja.
35
40

El término "*paciente*" se refiere a pacientes de origen humano o de otro mamífero, e incluye a cualquier individuo que se desea examinar o tratar usando los procedimientos divulgados en la presente memoria. Sin embargo, se entenderá que "*paciente*" no implica que los síntomas estén presentes. Mamíferos adecuados que están comprendidos dentro del alcance de la invención incluyen, pero no se limitan a, primates, animales de granja (por ejemplo, ovejas, vacas, caballos, burros, cerdos), animales de ensayo de laboratorio (por ejemplo, conejos, ratones, ratas, cobayas, hámsteres), animales de compañía (por ejemplo, gatos, perros) y animales silvestres cautivos (por ejemplo, zorros, ciervos, dingos).
45

Por "*vehículo farmacéuticamente aceptable*" se entiende una carga sólida o líquida, diluyente o sustancia encapsulante que se puede utilizar de manera segura en cualquier tipo de administración.
50

La expresión "*enfermedad o afección relacionada*" se emplea en esta memoria para referirse a una enfermedad o una afección que se relaciona anatómica, fisiológica, patológica y/o sintomáticamente con una enfermedad o afección de referencia. Por ejemplo, las enfermedades o afecciones pueden estar relacionadas entre sí mediante una afección a localizaciones anatómicas similares (por ejemplo, que afectan al mismo órgano o parte del cuerpo), que afectan a diferentes órganos o partes del cuerpo con función fisiológica similar (por ejemplo, el esófago, el duodeno y el colon que se basan en el peristaltismo para mover los alimentos desde un extremo del canal alimentario a otro), por tener patologías similares o solapantes (por ejemplo, lesión o ruptura de un tejido, apoptosis, necrosis) o por tener síntomas similares o solapantes (es decir, respuesta alérgica, inflamación, linfocitosis). Por lo tanto, por ejemplo, un antígeno asociado con una colitis ulcerosa también se puede asociar con una perforación del colon, ya que esa enfermedad afecta al mismo órgano (es decir, el colon).
55
60

El término "tratar" se emplea en esta memoria en su sentido más amplio para incluir tanto un tratamiento terapéutico como profiláctico (es decir, preventivo), diseñado para mejorar la enfermedad o la afección.

El término "episoma" se emplea en esta memoria para referirse a una porción de material genético que puede existir de forma independiente del cuerpo principal del material genético (cromosoma) algunas veces o de manera continua y replicarse de manera autónoma, mientras que otras veces es capaz de integrarse en el cromosoma. Ejemplos de episomas incluyen secuencias de inserción, transposones y el VTT de la invención.

Leyendas de las figuras

Figura 1: amplificación con PCR de un fragmento de 71 bases que contiene la región de VTT altamente conservada (HCR) en 4 líneas celulares diferentes. L1236 (línea de linfoma de Hodgkin negativa para EBV), HSB-2 (línea de leucemia linfoblástica aguda), KR e IGL (líneas celulares de melanoma) y ADN de placenta

Figura 2: ADN enrollado que permanece en el material sobrenadante de las células L1236 después de una precipitación y recogida del ADN de peso molecular elevado y digestión con ARNasa
 Dos bandas son visibles en la región entre las bandas de 4,3 y 6,6 bases.

Figura 3: PCR larga dirigida hacia afuera, usando cebadores de la región HCR de VTT de 71 bases en ADN de HSB-2
 Dos bandas son visibles en regiones correspondientes a 4,5 a 7 kb. Además, las bandas surgen en la región correspondiente a 0,4 a 0,7 kb.

Figura 4: Representación esquemática del concepto de oncogén de VTT
 La parte izquierda representa la organización genómica de los genomas de VTT de tipo silvestre. La parte derecha prevé la integración del ADN de la célula hospedadora en los plásmidos de cadena sencilla.

Figura 5: Representación esquemática del concepto de autoinmunidad del ADN de células hospedadoras de VTT. Los genes modificados de las células hospedadoras deben codificar epítopos antigénicos inmunorreactivos.

Figura 6: Amplificación con PCR de la región altamente conservada (HCR) de 71 bases a partir del ADN de 4 líneas celulares diferentes. Las flechas apuntan a los dos sitios con variaciones en las secuencias de nucleótidos.

Figura 7:

(A) El ADN con replicación autónoma de 719 bases de VTT (derecha) y la secuencia de VTT completa a partir de la cual se obtiene. La composición de nucleótidos de ambas moléculas se encuentra en la Figura 11A + B.

(B) El ADN con replicación autónoma de 621 bases de VTT (derecha) y la secuencia de ADN completa a partir de la cual se obtiene. La composición de nucleótidos de ambas moléculas se encuentra en la Figura 12A + B.

(C) El ADN con replicación autónoma de 642 bases de VTT (derecha) y la secuencia de ADN completa a partir de la cual se obtiene. La composición de nucleótidos de ambas moléculas se encuentra en la Figura 13A + B.

Figura 8: Tres secuencias ejemplares de ADN quimérico de VTT / célula hospedadora truncada procedentes de biopsias cerebrales de pacientes con esclerosis múltiple

(A) Secuencias celulares quiméricas obtenidas a partir del cromosoma 1 con algunas homologías con secuencias de priones y de tumor de Wilms y el extremo 3' del seudogén de leucemia linfocítica mieloide 3 (MLL3). Secuencia de ADN humano del clon RP11-14N7 en el cromosoma 1. Contiene un extremo 3' de un seudogén de leucemia linfocítica mieloide 3 o un linaje mixto (MLL3), un seudogén del receptor de siete hélices transmembranales, el extremo 5' de un gen novedoso.

(B) Secuencias celulares quiméricas obtenidas a partir del cromosoma 16. Homologías con el factor de transcripción 3 (TF 3C), firmas proteicas para receptores de quimiocinas y receptor de leucotrienos B4.

(C) Secuencias celulares quiméricas obtenidas a partir del cromosoma 10, secuencia truncada de la miosina, reactividad descrita en pacientes con esclerosis múltiple y aquellos con artritis reumatoide (la secuencia contiene ambos cebadores completos delantero y posterior).

Figura 9: Tres secuencias ejemplares de ADN quimérico de VTT / célula hospedadora truncada procedentes de líneas celulares obtenidas a partir de pacientes con la enfermedad de Hodgkin o leucemia

(A) Secuencias del cromosoma 1 con parte de transgelina 2, el gen IGSF9 para miembros de la superfamilia de inmunoglobulinas 9, el gen SLAMF9.

(B) Secuencias de proteínas traducidas con homología sustancial con los oncogenes v-myb (oncogén vírico de la mieloblastosis aviar), pero también con c-myb. Esta secuencia se amplificó con el cebador directo en ambos extremos.

(C) Obtenida a partir del cromosoma 10. Homología elevada con la proteína "Deleted in malignant 1 Protein" (DMBT), un gen supresor tumoral identificado. Esta secuencia se amplificó con el cebador directo en ambos extremos.

5 Figura 10: Secuencias de cebadores utilizadas en las reacciones descritas en los Ejemplos, obtenidas a partir de la HCR de 71 bases.

Figura 11:

(A) Secuencia completa de VTT a partir de la cual se ha obtenido el ADN de 719 bases que se replica autónomamente.

(B) Secuencia completa del ADN de TTV de 719 bases que se replica autónomamente.

10 Figura 12:

(A) Secuencia completa de VTT (tth25) a partir de la cual se ha obtenido el ADN de 621 bases que se replica autónomamente.

(B) Secuencia completa del ADN de TTV de 621 bases que se replica autónomamente.

Figura 13:

15 (A) Secuencia completa de VTT (ttrh215) a partir de la cual se ha obtenido el ADN de 642 bases que se replica autónomamente.

(B) Secuencia completa del ADN de TTV de 642 bases que se replica autónomamente.

20 Figura 14: Marcos de lectura abierta (ORFs) que se encuentran dentro de la secuencia de nucleótidos de 71 nt zyb2.1.pep, zyb9.1.pep y zkb69.1.pep están empezando en el primer triplete, zyb2.3.pep, zyb9.3.pep, zkb5.3.pep y zkb69.3.pep están empezando a partir del tercer triplete. Esta región se transcribe activamente.

Figura 15: Digestión de ADN monocatenario mediante la nucleasa de judía de mungo (MBN). Los carriles 2 y 3 muestran que el ADN amplificado puede ser digerido mediante un pretratamiento con MBN. Los carriles 5 y 6 muestran que ADN plasmídico tratado previamente de la misma manera, no es digerido por MBN.

25 Figura 16: Presentación esquemática del ORF1 de una serie de aislados de VTT-HD

El ORF1 se dividió en uno a varios ORF más pequeños o se fusionó a otros ORF.

Figura 17: Transcritos aislados durante la replicación *in vitro* de aislados de VTT-HD

El marcaje de transcritos individuales indica "aislado 5'- o 3'-race (s - monocatenario).no". Los números de aislado de VTT (1-12) indicados con el genoma esquemático respectivo y el número de VTT-HD. * - transcritos que se aislaron con más frecuencia.

30 Figura 18: Árbol filogenético que muestra las especies de VTT y aislados del género Alphatoquevirus, así como todos los tipos de VTT-HD, los tipos de VTT-HD propagados en cultivos celulares *in vitro* se encuentran rodeados con un círculo.

Figura 19: Propagación de genomas de VTT-HD de longitud completa en células 293TT Ejemplos de propagación de

35 (A) VTT-HD14b, VTT-HD14c, VTT-HD14a y VTT-HD14e (carriles 1-4), VTT-HD15a (carril 5) y VTT-HD16a (carril 16) tras la amplificación por PCR anidada;

40 (B) VTT-HD20a (carril 7), VTT-HD3a (carril 8), VTT-HD1a (carril 9), VTT-HD23b, VTT-HD23d y VTT-HD23a (carriles 10-12) tras una sola amplificación por PCR. a, b y c - ejemplos de propagaciones, aproximadamente 7 días después de la infección. b-1, b-2 y b-3 indican la variabilidad observada cuando se propaga el mismo pase.

(C) Muestreo diario de cultivos de CTT-HD14e (PCR anidada) y VTT-HD23b. M - marcador de tamaño del ADN; * - indica moléculas subvircas de diferentes cultivos.

Figura 20: Presentación esquemática de VTT-HD de longitud completa con sus respectivas moléculas de μ VTT-HD Los números indican los ORF en el genoma del ADN.

45 Figura 21: Propagación independiente de μ VTT-HD μ VTT-HD15 se replicó con más fuerza tras la transfección inicial, pero se redujo con el paso del tiempo (*-indica amplificación por PCR anidada). μ VTT-HD1 y μ VTT-HD23.2 se replicaron de manera creciente tras las etapas iniciales de propagación. Moléculas de μ VTT-HD23.2 formadas durante la replicación de μ VTT-HD23.1.

Figura 22:

50 (A) Partículas viroides parcialmente purificadas. Las partículas se lisaron y el contenido se separó en gel de agarosa.

(B) Partículas de mVTT parcialmente purificadas. Las partículas se lisaron y el contenido de ADN se separó en gel de agarosa.

3-VTT-HD14a, 5 - μ VTT-14, 6 - VTT-HD16a, 8 - VTT-HD3a, 9 - μ VTT-HD1, 12 - VTT-HD23a, 12a - μ VTT-HD12.1, 12b - μ VTT-HD12.2

5 En el presente documento se divulga un poli(ácido nucleico) de virus TT que comprende (o consiste en)

(a) una secuencia de nucleótidos mostrada en la figura 6;

(b) una secuencia de nucleótidos que muestra al menos un 70 %, 80 %, 90 %, 95 % o al menos un 98 % de identidad respecto de una secuencia de nucleótidos de (a) y es capaz de replicarse de manera autónoma;

(c) una secuencia de nucleótidos que es el complemento de la secuencia de nucleótidos de (a) o (b); o

10 (d) una secuencia de nucleótidos que codifica una secuencia de aminoácidos seleccionada entre las secuencias de aminoácidos zyb2.1.pep, zyb9.1.pep, zkb69.1.pep, zyb2.3.pep, zyb9.3.pep, zkb5.3.pep y zkb69.3.pep mostradas en la figura 14 y dicho poli(ácido nucleico) de virus TT es capaz de replicarse de manera autónoma tras su transfección en células 293TT,

15 en la que dicha secuencia de nucleótidos de (a), (b), (c) o (d) está unida a un poli(ácido nucleico) que codifica una proteína que contiene un motivo de firma de una proteína que está asociada con el cáncer o una enfermedad autoinmunitaria mediante un enlace fosfodiéster.

La proteína es una proteína de mamífero. Se prefiere particularmente que la proteína de mamífero sea una proteína humana. También se divulgan en el presente documento fragmentos de las secuencias de nucleótidos descritas anteriormente que son capaces de replicarse de manera autónoma. El experto en la técnica puede obtener fragmentos que sigan teniendo la actividad biológica de la molécula de longitud completa sin experimentación innecesaria.

El experto en la técnica puede determinar fácilmente qué secuencias de ácidos nucleicos están relacionadas con la secuencia de nucleótidos de la Figura 6 o qué fragmentos todavía son capaces de replicarse de forma autónoma mediante el uso de ensayos convencionales o los ensayos descritos en los ejemplos, a continuación.

25 El término "poli(ácido nucleico)" se refiere a una secuencia de ácido nucleico de cadena sencilla o de cadena doble. Un poli(ácido nucleico) puede consistir en desoxirribonucleótidos o ribonucleótidos, análogos de nucleótidos o nucleótidos modificados, o se puede haber adaptado para fines terapéuticos. Preferentemente, el poli(ácido nucleico) del virus TT es un ADN de cadena sencilla.

30 Preferentemente, el poli(ácido nucleico) del virus TT de la invención está presente en forma de un episoma extracromosómico.

La proteína de mamífero asociada con el cáncer o una enfermedad autoinmunitaria o el alérgeno asociado con una enfermedad autoinmunitaria es una proteína mostrada en la tabla 1.

Tabla 1

35 **(A) Ejemplos de motivos de firma identificados en supuestas proteínas resultantes de transcritos y genomas de longitud completa de VTT-HD**

Protamina 1+2

Receptor B4 de leucotrienos

AIRE (Regulador autoinmunitario)

Gliadina

40 Neuropéptido Y

CHLAMIDIAOM3 - Chlamidia mol. mimicry - enfermedad cardíaca

Rica en arginina (re. Sospedra y col., 2005 - mimicry molecular en MS)

Opsina

Ciclina cinasa

45 Proxisoma (diabetes, receptor de esteroides)

Vasopresina

Factor BDNF (factor neurotrófico derivado de cerebro)

prepro-orexina

Repetición de hélice de colágeno

50 Receptor GIP

Neurotensina Prion

antígeno CD36 (deficiencia de resistencia a insulina, arterioesclerosis)

Calcitonina

Prostanoide

55 Prostanoide

Receptor de GABA (principal neurotransmisor inhibitor en el cerebro)

Arginina desaminasa
Opioide, receptor de factor de crecimiento

Galanina
Plexina/semaforina

5 NURR (receptor de hormona nuclear huérfano de rata)
Factor neurotrófico derivado de cerebro (BDN)
Colagenasa + endostatina
Aerolisina

10 Proteolípido de mielina
Serotonina
Receptor muscarínico

Receptor de hormona concentradora de melanina

15 Autoantígeno p27 de síndrome de Sjögren/esclerodermia
Firma de repetición de tipo plexina/semaforina/integrina
Proteína específica masculina
Gastrina
Colágeno
Colagenasa metaloproteasa

(B) alineamientos de secuencias de aa

20 **DomainSweep emplea una serie de procedimientos de búsqueda para efectuar búsquedas en las siguientes bases de datos de familias de proteínas:**

BLOCKS
PFAMA
PRINTS
25 **PRODOM**
PROSITE
SMART
SUPERFAMILY
TIGRFAMS

30 **OPSINA**
gbCsCt38.4ikn.2.154

OPSINRH3RH4_3: dominio 1 de 1, de 46 a 56: puntuación 8,4, E = 5,1
->iynsFhrGfAlg<-
y sFhrG+A
gbCsCt38.4 46 -YESFHRGHAAF 56

zc55s.B4.l8dek.281

35 **OPSINRH3RH4_3:** dominio 1 de 1, de 19 a 29: puntuación 8,4, E = 5,1
->iynsFhrGfAlg<-
y sFhrG+A
zc55s.B4.1 19 -YESFHRGHAAF 29

rheu.cd.215rev.1.736

OPSINRH3RH4_7: dominio 1 de 1, de 665 a 683: puntuación 7,8, E = 5,3
->R1ELqKR1PWLeInEKave<-
R+ +q+R1PW+ + + +
rheu.cd.21 665 RFGVQQRLPWVHSSQETQS 683

40 **OPSINRH3RH4_7:** dominio 1 de 1, de 23 a 41: puntuación 8,2, E = 4,4
->R1ELqKR1PWLeInEKave<-
R+ +q+R1PW+ + + +
zc3r11.B4. 23 RFRVQQRLPWVHSSQETQS 41

gc; **OPSINRH3RH4**
 gx; PR00577
 gn; COMPUESTO(7)
 ga; II-SEP-1996; ACTUALIZADO 07-JUN-1999
 gt; Firma de Opsina RH3/RH4
 gp; PRINTS; [PR00237](#) GPCRRHODOPSN; [PR00247](#) GPCRCAMP; [PR00248](#) GPCRMGR
 gp; PRINTS; [PR00249](#) GPCRSECRETIN; [PR00250](#) GPCRSTE2; [PR00899](#) GPCRSTE3
 gp; PRINTS; [PR00251](#) BACTRLOPSIN
 gp; PRINTS; [PR00238](#) OPSIN; [PR00574](#) OPSINBLUE; [PR00575](#) OPSINREDGRN
 gp; PRINTS; [PR00576](#) OPSINRH1RH2; [PR00578](#) OPSINLTRLEYE; [PR01244](#) PEROPSIN
 gp; PRINTS; [PR00666](#) PINOPSIN; [PR00579](#) RHODOPSIN; [PR00239](#) RHODOPSNTAIL
 gp; PRINTS; [PR00667](#) RPERETINALR
 gp; INTERPRO; [IPR000856](#)
 gr; 1. APPLEBURY, M.L. Y HARGRAVE, P.A.
 gr; Molecular biology of the visual pigments.
 gr; VISION RES. 26(12) 1881-1895 (1986).
 gr; 2. FRYXELL, K.J. Y MEYEROWITZ, E.M.
 gr; The evolution of rhodopsins and neurotransmitter receptors.
 gr; J.MOL.EVOL. 33(4) 367-378 (1991).
 gr; 3. ATTWOOD, T.K. Y FINDLAY, J.B.C.
 gr; Design of a discriminating fingerprint for G protein-coupled receptors.
 gr; PROTEIN ENG. 6(2) 167-176 (1993).
 gr; 4. ATTWOOD, T.K. Y FINDLAY, J.B.C.
 gr; Fingerprinting G protein-coupled receptors.
 gr; PROTEIN ENG. 7(2) 195-203 (1994).
 gr; 5. FRYXELL, K.J. Y MEYEROWITZ, E.M.
 gr; An opsin gene that is expressed only in the R7 photoreceptor cell of
 gr; Drosophila.
 gr; EMBO J. 6(2) 443-451 (1987).
 gr; 6. ZUKER, C.S., MONTELL, C., JONES, K., LAVERTY, T. Y RUBIN, G.M.J.
 gr; A rhodopsin gene expressed in photoreceptor cell R7 of the Drosophila
 gr; eye - homologies with other signal-transducing molecules.
 gr; NEUROSCIENCE 7(5) 1550-1557 (1987).
 gr; 7. MONTELL, C., JONES, K., ZUKER, C.S. Y RUBIN, G.M.J.
 gr; A second opsin gene expressed in the ultraviolet-sensitive R7
 gr; photoreceptor cells of Drosophila melanogaster.
 gr; NEUROSCIENCE 7(5) 1558-1566 (1987).
 gd; **Las opsinas**, las moléculas absorbedoras de luz que median la visión [1,2], son
 gd; proteínas integrales de membrana que pertenecen a una superfamilia de receptores
 gd; acoplados a proteína G (GPCR). Los ligandos activadores de los diferentes miembros
 gd; de la superfamilia varían ampliamente en cuanto a su estructura y carácter, aunque
 gd; las proteínas parecen tener fielmente conservado un armazón estructural básico, que
 gd; se cree que consiste en 7 hélices transmembrana (TM). Aunque las secuencias de
 gd; estas proteínas son muy diversas, reflejando hasta cierto punto esta gran variedad de
 gd; ligandos activadores, sin embargo, se han identificado motivos en las regiones TM que
 gd; son característicos de prácticamente toda la superfamilia [3,4].
 gd; Entre las excepciones se encuentran los receptores olfatorios, que se agrupan en una
 gd; superfamilia, que carece de coincidencias significativas con los dominios 2, 4 y 6. De
 gd; forma interesante, las opsinas también parecen emerger como crecientemente
 gd; atípicas de la superfamilia, agrupándose con más fuerza, en los análisis filogenéticos,
 gd; con los receptores olfatorios [4].
 gd; Los pigmentos visuales comprenden una apoproteína (opsina), unida covalentemente
 gd; al cromóforo cis-terminal 11. El enlace covalente es en forma de una base de Schiff
 gd; protonada entre el resto retinal y uno de lisina ubicado en el dominio 7 TM. La visión
 gd; se efectúa mediante la absorción de un fotón por el cromóforo, que se isomeriza en la
 gd; forma todo trans, que promueve un cambio conformacional en la proteína.
 gd; A diferencia de la rodopsina de vertebrados, que se encuentra en los bastones, los
 gd; fotorreceptores de insectos se encuentran en los omatidios, que comprende los ojos
 gd; compuestos. Cada ojo de drosófila tiene 800 omatidios, cada uno de los cuales
 gd; contiene 8 células fotorreceptoras (denominadas R1-R8): R1-R6 son células externas,
 gd; mientras que R7 y R8 son células internas. Las opsinas RH3 y RH4 son sensibles a la
 gd; luz UV [5-7]. OPSINRH3RH4 es una huella de 7 elementos que proporciona una firma
 gd; para las opsinas RH3 y RH4. La huella se obtuvo de un alineamiento inicial de 5
 gd; secuencias: los motivos se representaron a partir de secciones conservadas dentro de
 gd; cada bucle o las regiones N y C-terminales, centrándose en aquellas áreas del
 gd; alineamiento que caracterizan a las opsinas RH3/RH4 pero que las distinguen del

gd; resto de la superfamilia similar a rodopsina - los motivos 1 y 2 se encuentran en el
 gd; extremo N-terminal; el motivo 3 abarca el primer bucle externo; el motivo 4 se
 gd; encuentra en el segundo bucle externo; el motivo 5 abarca la mitad C-terminal del
 gd; dominio 5 TM; el motivo 6 se encuentra en el tercer bucle citoplásmico; y el motivo 7
 gd; se encuentra en el extremo C-terminal. Para alcanzar la convergencia se necesitó una
 gd; sola interacción en OWL28.1, sin identificar secuencias adicionales más allá del
 gd; conjunto de inicio.

c; OPSINRH3RH43

il; 12

it; Motivo III de opsina RH3/RH4 - 1

id;	IFNSFHGRGFAIY	OPS4_DROME	109	52
id;	IYNSFHGRGFALG	OPS4_DROPS	112	54
id;	IYNSFHGRGFALG	OPS4_DROVI	115	54
id;	IYNSFHQGYALG	OPS3_DROME	115	54
id;	IYNSFHQGYALG	OPS3_DROPS	114	54

bb; :

fc; : OPSINRH3RH43

fl; : 12

ft; Motivo III de opsina RH3/RH4 - 2

fd;	IYNSFHGRGFALG	OPS4_DROVI	115	54
fd;	IYNSFHQGYALG	OPS3_DROME	115	54
fd;	IYNSFHGRGFALG	OPS4_DROPS	112	54
fd;	IYNSFHQGYALG	OPS3_DROPS	114	54
fd;	IFNSFHGRGFAIY	OPS4_DROME	109	52
fd;	IYNSFHTGFATG	O61474	105	54
fd;	IYNSFNTGFATG	O61473	106	54
fd;	IYNSFNTGFALG	OPSV_APIME	105	54

fc; : OPSINRH3RH47

fl; : 19

ft; Motivo VII de opsina RH3/RH4 - 2

fd;	RMELQKRCPWLAIDEKAPE	OPS4_DROVI	346	62
fd;	RMELQKRCPWLALNEKAPE	OPS3_DROME	346	62
fd;	RMELQKRCPWLGVNEKSGE	OPS4_DROPS	343	62
fd;	RMELQKRCPWLAISEKAPE	OPS3_DROPS	345	62
fd;	RLELQKRCPWLGVNEKSGE	OPS4_DROME	342	62
fd;	RLELQKRLPWLELQEKPV	O61474	336	62
fd;	RLELQKRLPWLELQEKPIE	O61473	337	62
fd;	RLELQKRLPWLELQEKPI	OPSV_APIME	336	62

RICA EN ARG

PERFILES DE PROSITE

RICA EN ARG

Región rica en arginina

NLS BP

Lo nuclear bipartito

5 PFSCAN usando la secuencia **gbCsCt38.2ikn.1.726**

y perfil(es) PRFDIR:prosite.prf,

Parámetros de línea de comandos usados:

-CUTLEV=-1

..

Puntuación	Raw	seq-f	seq-t	prf-f	prf-t	Nombre	Descripción
30,1607	170	4 -	67	1-	2	RICA EN ARG	Región rica en arginina
4,0000	4	10 -	26	1-	17	NLS_BP	Lo nuclear bipartito
4,0000	4	32 -	46	1-	17	NLS_BP	Lo nuclear bipartito
5,0000	5	52 -	66	1-	17	NLS_BP	Lo nuclear bipartito

10 PFSCAN usando la secuencia **gbDhD143.4rp.1.765**

y perfil(es) PRFDIR:prosite.prf, 15 de octubre de 2010 15:31

Parámetros de línea de comandos usados:

-CUTLEV=-1

..

Puntuación	Raw	seq-f	seq-t	prf-f	prf-t	Nombre	Descripción
33,0880	187	9 -	- 73	1 -	2	ARG_RICH	Región rica en arginina

PFSCAN usando la secuencia **zpr5.B4.12dk.209**

Parámetros de línea de comandos usados:

-CUTLEV=-1

Puntuación	Raw	seq-f	seq-t	prf-f	prf-t	Nombre	Descripción
30,1607	170	4 -	67	1 -	2	RICA EN ARG	Región rica en arginina

5 PFSCAN usando la secuencia **zc55s.B4.18dek.117**
y perfil(es) PRFDIR:prosite.prf,

Parámetros de línea de comandos usados:

-CUTLEV=-1

Puntuación	Raw	seq-f	seq-t	prf-f	prf-t	Nombre	Descripción
18,7959	104	4 -	85	1 -	2	RICA EN ARG	Región rica en arginina

10 PFSCAN usando la secuencia **zc37.B9.2de.p1**

Parámetros de línea de comandos usados:

-CUTLEV=-1

Puntuación	Raw	seq-f	seq-t	prf-f	prf-t	Nombre	Descripción
24,3061	136	7 -	86	1 -	2	ARG_RICH	Región rica en arginina

PROTAMINA 1 Y PROTAMINA 2

15 BLKPROB Versión 5/21/00.1

Base de datos=*gcg/hsar/gcgdata/gcgblimps/blocksplus.dat*

Búsqueda=gbCsCt38.2ikn.1.726 Longitud: Tamaño=726 aminoácidos

Familia

	Hebra	Bloques	Valor de E combinado
IPB000221 Protamina P1	1	1 de 1	<u>1,3e-09</u>
HSP1_CHICK P15340 1	ARYRRSRTSRSPRSRRRRRRSGRRRSPRRRRY		
IPB000492 Protamina 2, PRM2	1	1 de 2	<u>2,2e-09</u>
HSP2_PIG P19757 55	HTRRRRSCRRRRRACRHRHRRRGCRRI RRRRRCR		

Búsqueda=gbDhDi43.xp.1.765 Longitud: 7 65

Familia	Hebra	Bloques	Valor de E combinado
IPB000221 Protamina P1	1	1 de 1	<u>1,2e-11</u>
HSP1 DIDMA P35305 1	ARYRRRSRSRSRSRYGRRRRRSR RRRRSRRRRR		
IPB000492 Protamina 2, PRM2 1	1	1 de 2	<u>2,8e-10</u>
HSP2_CALJA Q28337 69	RRRSRSCRRRRRRSCRYRRRPRRGCRSRRRRRCRR		

Búsqueda=rheu.ef.242.746 Longitud: 746

Familia	Hebra	Bloques	Valor de E combinado
IPB000492 Protamina 2, PRM2	1	1 de 2	<u>1,4e-08</u>
HSP2 CALJA Q28337 69	RRRSRSCRRRRRRSCRYRRRPRRGCRSRRRRRCRR		
IPB000221 Protamina P1	1	1 de 1	<u>1,5e-07</u>
HSP1 DIDMA P35305 1	ARYRRRSRSRSRSRYGRRRRRSR RRRRSRRRRR		

Búsqueda=uro705rev.1a.74 Longitud: 74

IPB000221 Protamina P1	1/1 bloques	Valor de E combinado= 2,8e-12
HSP1 DIDMA P35305 1	ARYRRRSRSRSRSRYGRRRRRSR RRRRSRRRRR	
IPB000492 Protamina 2, PRM2	1/2 bloques	Valor de E combinado= 2,3e-10
HSP2_CALJA Q28337 69	RRRSRSCRRRRRRSCRYRRRPRRGCRSRRRRRCRR	

Búsqueda=zpr5.B4.12dk Longitud: 209

IPB000221 Protamina P1	1	1 de 1	<u>4,1e-10</u>
HSP1 CHICK P15340 1	ARYRRSRTSRSPRSRRRRRRSGRRRSPRRRRY		
IPB000492 Protamina 2, PRM2	1	1 de 2	<u>7,1e-10</u>
HSP2 PIG P19757 55	HTRRRRSCRRRRRACRHRHRRRGCRRI RRRRRCR		

Búsqueda=zc55s.B4.18dek.117 Longitud: 117

Familia	Hebra	Bloques	Valor de E combinado
IPB000492 Protamina 2, Q91V94 Q91V94_MESAU63	PRM2 1	1 de 2	3,4e-05
IPB000221 Protamina P1 HSP1_MOUSE P02319 1	1	1 de 1	0,0013

Búsqueda=zc37.B9.2de.p1 Longitud: 918

Familia	Hebra	Bloques	Valor de E combinado
IPB000492 Protamina 2, HSP2 ERYPA Q9GKM0 69	PRM2 1	1 de 2	2,8e-05
IPB000221 Protamina P1 HSPL_CAVPO P35304 1	1	1 de 1	0,0001

Secuencias presentadas como ejemplos:

Genomas de longitud completa (VTT) de:

- 5 gbCsCt38.2ikn.1.726 (VTT-HD15, ORF1=726aa)
- gbDhDi43.4rp.1.765 (VTT-HD16, ORF1=765aa)
- rheu.ef.242.746 (VTT-HD19, ORF1=746aa)
- uro705rev.1a.74 (VTT-HD18, ORF1a=74aa)

Genoma de longitud completa (µVTT) de:

zpr5.B4.12dk (µVTT-HD15, ORF=208aa)

Transcritos (de -):

- 10 zc55s.B4.18dek.117 (VTT-HD15, ORF=117aa)
- zc37.B9.2de.p1 (VTT-HD20, ORF=109aa)

GALANINA:

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine Distribuido libremente bajo la licencia pública general GNU (GPL)

Archivo HMM: smart.hmm

Archivo de secuencia: **gbDhDi33.33ik.1c.417**

galanina: dominio 1 de 1, de 264 a 367: puntuación -22,9, E = 6,5

```

*->atlGLgsPvkekrGwTLnsAGYLLGPHaidnHRsFsdKhGLtgKREL
      t L P + r + s LGP ++ ++G+ +KR +
gbDhDi33.3 264 STHelpDPDRHPRMLQV-SDPTKLGPKT--AFHKWDWRRGMLSKRSI 307

e..pEdearpGsfdprlpsenivrtiiefLsfLhLkeaGalDrLpplPa
++ Ed +++pl+ ++n t + L+ L +
gbDhDi33.3 308 KrvQEDSTDDEYVAGPLPRkrNKFDTRVQGPPTPEKESYTLQALQESGQ 357

aasseDlers<-*
      sseD e++
gbDhDi33.3 358 ESSSEDQEQA 367
    
```

gbDfDg33.48ikn.1b.179

galanina: dominio 1 de 1, de 26 a 129: puntuación -21,0, E = 3,9

```

*->atlGLGsPvkekrGwTlnsAGYLLGPHaidnHRsFsdKhGLtgKREL
t L P + r + s LGP + ++ ++G+ +KR +
gbDfDg33.4 26 STHELDPDRHPRMLQV-SDPTKLGPKTV--FHKWDWRRGMLSKRSI 69

e..pEdearpGsfdprlses.nivrtiiefLsfLhLkeaGaLdrLpplPa
++ Ed +++pl+ ++n t + L+ L +
gbDfDg33.4 70 KrvQEDSTDDEYVAGPLPRKrNKFDTRVQGPPTPEKESYTLQALQESGQ 119

aasseDlers<-*
sseD e++
gbDfDg33.4 120 ESSSEDQEQA 129
    
```

Archivo HMM: smart.hmm
 Archivo de secuencia: **gbDhDi33.32ikn.1.648**
galanina: dominio 1 de 1, de 495 a 598: puntuación -24,5, E = 9,7

```

*->atlGLGsPvkekrGwTlnsAGYLLGPHaidnHRsFsdKhGLtgKREL
t L P + r + s LGP + ++ ++G+ +KR +
gbDhDi33.3 495 STHELDPDRHPRMLQV-SDPTKLGPKTV--FHKWDWRRGMLSKRSI 538

.epEdearpGs.fdrplses.nivrtiiefLsfLhLkeaGaLdrLpplPa
++ + G +++pl+ ++n t + L+ L +
gbDhDi33.3 539 kRVQGDSTDGEyVAGPLPRKrNKFDTRVQGPPTPEKESYTLQALQESGQ 588

aasseDlers<-*
sseD e++
gbDhDi33.3 589 ESSSEDQEQA 598
    
```

gbDfDg33.45ikn.1b.210
galanina: dominio 1 de 1, de 57 a 160: puntuación -23,1, E = 6,8

```

*->atlGLGsPvkekrGwTlnsAGYLLGPHaidnHRsFsdKhGLtgKREL
t L P + r + s LGP + ++ +G+ +KR +
gbDfDg33.4 57 STHELDPDRHPRMLQV-SDPTKLGPKTV--FHKWDWRRGMLSKRSI 100

e..pEdearpGsfdprlses.nivrtiiefLsfLhLkeaGaLdrLpplPa
++ Ed +++pl+ ++n t + L+ L +
gbDfDg33.4 101 KrvQEDSTDDEYVAGPLPRKrNKFDTRVQGPPTPEKESYTLQALQESGQ 150

aasseDlers<-*
sseD e++
gbDfDg33.4 151 ESSSEDQEQA 160
    
```

FIRMAS DE REPETICIÓN DE TIPO PLEXINA/SEMAFORINA/INTEGRINA

HMMER 2.3.2 (Oct 2003)
 Copyright (C) 1992-2003 HHMI/Washington University School of Medicine
 Distribuido libremente bajo la licencia pública general GNU (GPL)

5

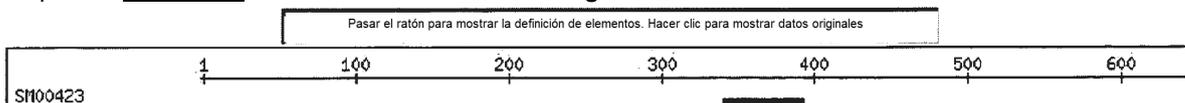
```

-----
Archivo HMM: smart.hmm
Archivo de secuencia: gbDhDi33.32ikn.1.648
psinew7: dominio 1 de 1, de 341 a 394: puntuación -16,8, E = 3,9
*->rCsqygv...tsCseCllardpyg.....CgWCssegrCtrg.erC
Cs +++ ++ s C+l++ p + C W + +Ct +++++
gbDhDi33.3 341 WCSEKSSkldTTKSKCILRDFPLWamaygyCDWVV---KCTGVsSAW 384

derrgerqnwasgppssqCp<-*
+ +r+ + Cp
gbDhDi33.3 385 TDMRI----AI-----ICP 394
    
```

10

Interpro: [IPR003659](#) **Plexina/semaforina/integrina**



IPR003659; IPR016201; (Similar a plexina) coincidencias 1383 proteínas NombreCorto Plexina-ilice
 NombreCompleto firmas de repetición de tipo plexina/semaforina/integrina SMART: [SM00423](#) PSI/
 ▼ Children [IPR016201](#) Pliegue de tipo plexina (coincidencias 619 proteínas) ▲ Encontrada en
[IPR012013](#) subunidad beta-4 de integrina (coincidencias 9 proteínas)
 5 [IPR020707](#) Tirosina-proteína cinasa, receptor de factor de crecimiento de hepatocitos (coincidencias 82
 proteínas)
[IPR020739](#) Tirosina-proteína cinasa, receptor MSP (coincidencias 18 proteínas) Resumen Este es un dominio
 que se ha encontrado en plexinas, semaforinas e integrinas. La plexina está implicada en el desarrollo de
 tejidos neurales y epiteliales; las semaforinas inducen el colapso y la parálisis de los conos de crecimiento
 10 neuronal; y las integrinas pueden mediar las funciones adhesivas o migratorias de las células epiteliales.
 Ejemplos

 Archivo HMM: smart.hmm
 Archivo de secuencia: **gbDhDi33.31ikn.1.712**
 15 **psinew7**: dominio 1 de 1, de 341 a 378: puntuación -14,4, E = 2,3

```

          *->rCsqqygv...tsCseCllardpygCgWCssegrCtrgerCderrgsr
          Cs  +++ +t+ s C+l++ p  W  +++++Cd
gbDhDi33.3  341  WCSEKSSklDttKSKCILRDFP---LWA-----MAYGHCD----- 372

          qnwssgpssqCp<-*
          w+      +C+
gbDhDi33.3  373  --wVv----KCT      378
  
```

GASTRINA

HMMER 2.3.2 (Oct 2003)
 Copyright (C) 1992-2003 HHMI/Washington University School of Medicine
 20 Distribuido libremente bajo la licencia pública general GNU (GPL)

 Archivo HMM: prints.hmm
 Archivo de secuencia: **gbDhDi33.32ikn.1.648**
GASTRINR_8: dominio 1 de 1, de 541 a 559:

```

          *->vaGEDsDGCyvq..LPRsR<-*
          v G+  DG yv ++LPR R
gbDhDi33.3  541  VQGDSTDGEYVAgpLPRKR      559
  
```

gc; **GASTRINR**
 gx; PRO0527 gn; COMPUESTO(9)
 qa; 03-JUN-1996; ACTUALIZADO 10-JUN-1999

gt; **Firma de receptor de gastrina**
 gp; PRINTS; [PR00237](#) GPCRRHODOPSN; [PR00247](#) GPCRCAMP; [PR00248](#) GPCRMGR
 gp; PRINTS; [PR00249](#) GPCRSECRETIN; [PR00250](#) GPCRSTE2; [PR00899](#) GPCRSTE3
 gp; PRINTS; [PR00251](#) BACTRLOPSIN
 gp; PRINTS; [PRO1822](#) CCYSTOKININR; [PR00524](#) CCYSTOKININAR
 gp; INTERPRO; [IPR000314](#)
 gr; 1. ATTWOOD, T.K. Y FINDLAY, J.B.C.
 gr; Fingerprinting G protein-coupled receptors.
 gr; PROTEIN ENG. 7(2) 195-203 (1994).

Las gastrinas y las colecistocininas (CCK) son péptidos de origen natural que comparten una
 gd; secuencia C-terminal común, GWMDF; toda la actividad biológica reside en esta región
 gd; [6]. El principal papel fisiológico de la gastrina es estimular la secreción de ácido en el
 gd; estómago; también tiene efectos tróficos en la mucosa gástrica [6]. La gastrina se
 gd; produce a partir de un solo transcrito génico y se encuentra predominantemente en el
 gd; estómago e intestino, pero también en los nervios vagales. El receptor CCKB tiene una
 gd; amplia distribución en el SNC y se ha relacionado con la patogénesis de los ataques de
 gd; pánico-ansiedad causados por péptidos relacionados con CCK [6]. Tiene una distribución
 gd; más limitada en la periferia, donde se encuentra en el músculo liso y las glándulas
 gd; secretorias.
 gd; GASTRINR es una huella de 9 elementos que proporciona una firma para los receptores
 gd; de gastrina (CCKB). La huella se obtuvo de un alineamiento inicial de 5 secuencias: los
 gd; motivos se representaron a partir de secciones conservadas dentro de cada bucle o las
 gd; regiones N y C-terminales, centrándose en aquellas áreas del alineamiento que
 gd; caracterizan a los receptores de gastrina, pero que los distinguen del resto de la
 gd; superfamilia similar a rodopsina - los motivos 1 y 2 se encuentran en el extremo N-
 gd; terminal; el motivo 3 abarca el primer bucle externo; el motivo 4 abarca el segundo bucle
 gd; citoplásmico; los motivos 5 y 6 abarcan el segundo bucle externo; los motivos 7 y 8
 gd; abarcan el tercer bucle citoplásmico; y el motivo 9 se encuentra en el extremo C-terminal.
 gd; Para alcanzar la convergencia, se necesitaron dos interacciones en OWL28.0, punto en
 gd; el cual se identificó un conjunto auténtico que comprendía 7 secuencias. También se
 gd; encontraron varias coincidencias parciales, todas las cuales eran fragmentos de gastrina
 gd; o miembros de la familia de receptor de tipo A de colecistocinina.

fc; GASTRINR8
 fl; 17

ft; Motivo VIII de receptor de gastrina - 2

fd; LAGEDGDGCYVQLPRSR	GASR_RABIT	288	31
fd; VAGEDNDGCYVQLPRSR	GASR_PRANA	289	30
fd; LAGEDGDGCYVQLPRSR	GASR_BOVIN	290	31
fd; AVGEDSDGCYVQLPRSR	GASR_HUMAN	285	26
fd; LAGEDGDGCYVQLPRSR	GASR_CANFA	289	29
fd; LTGEDSDGCYVQLPRSR	GASR_MOUSE	291	32
fd; VAGEDSDGCCVQLPRSR	GASR_RAT	290	31

COLAGENASA

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine

Distribuido libremente bajo la licencia pública general GNU (GPL)

5

 Archivo HMM: pfam.hmm
 Archivo de secuencia: **rheu.ef.241.736**

Peptidase_M9: dominio 1 de 1, de 125 a 412: puntuación -152,5, E = 7,5

ES 2 685 605 T3

```

*->msrlaelyllGdsiKgrhDnlWLaaaemlsYyApegkselgidicqa
  l ly      r n W + +e l+      g+      +
rheu.ef.24  125  --TLRILYDEF----TRFMNFWTVSNEDLDLCRYVVGCKLIF--FKHP 163

klelaakVlPy..lyeCsgpaa.irsqldtdgqaAsaCdilrnkekdfhq
+ +      + +++++      +aa+i      + ++      +l      h+
rheu.ef.24  164 TVDFIVQINTQppFLDTHLTAAAsIHPGIMMISKRRILIPSLKTRPSRKHR 213

vkytGktPVaDDgntrveVgvfvseedykrYSafaSKEVkaqFgrvtdNG
v+      V      ++ +      d + +S fa      t +
rheu.ef.24  214 VVVR----VGAPRLFQDKWYPQSDLCDTVLLSIFA-----TACD 248

GmYLEGNPsdagNqvrF..iAYEeaklnadlsigNlehEYthY...LDgR
+Y G P      + v+F+      + + k      ++s N+e + thY+++L +
rheu.ef.24  249 LQYFPGSPLTENPCVNFqiLGPYKHL-SISSTNDETnkThYesnLFNK 297

fdtYgtFsrnleeshivWweEGfAEYvhYkqgGvPyqaApeligggskly
+Y tF      ++      + e G+      v      v ++      + ++g      +
rheu.ef.24  298 TELYNTFQTIAQ-----LKETGRtSGVNPnWtSVQNTTPLNQAGNN---A 339

lsdvftTTeeGyAEI FAGShDtdRIyRWGYLA.vrf.....mletnHnr
++ + t++      G + d I      +++++rf++ + ++l n +
rheu.ef.24  340 QNSRDTWY---K-----GNTYNDNISKLAEITrQRfksatisALP-NYPT 380

dvesllvhsRyGnsfafyaylvkllgymYnnefgiw<-*
+ ++l ++ +G      y+ ++ +g Y      g++
rheu.ef.24  381 IMSTDLYEYHSG----IYSSIFLSAGRSYFETTGAY 412

```

rheu.ef.241.736

Peptidase_M9: dominio 1 de 1, de 125 a 412: puntuación -152,5, E = 7,5

```

*->msrlaelyllGdsiKgrhDnlWLaaaemlsYyApegkselgidicqa
  l ly      r n W + +e l+      g+      +
rheu.ef.24  125  --TLRILYDEF----TRFMNFWTVSNEDLDLCRYVVGCKLIF--FKHP 163

klelaakVlPy..lyeCsgpaa.irsqldtdgqaAsaCdilrnkekdfhq
+ +      + +++++      +aa+i      + ++      +l      h+
rheu.ef.24  164 TVDFIVQINTQppFLDTHLTAAAsIHPGIMMISKRRILIPSLKTRPSRKHR 213

vkytGktPVaDDgntrveVgvfvseedykrYSafaSKEVkaqFgrvtdNG
v+      V      ++ +      d + +S fa      t +
rheu.ef.24  214 VVVR----VGAPRLFQDKWYPQSDLCDTVLLSIFA-----TACD 248

GmYLEGNPsdagNqvrF..iAYEeaklnadlsigNlehEYthY...LDgR
+Y G P      + v+F+      + + k      ++s N+e + thY+++L +
rheu.ef.24  249 LQYFPGSPLTENPCVNFqiLGPYKHL-SISSTNDETnkThYesnLFNK 297

fdtYgtFsrnleeshivWweEGfAEYvhYkqgGvPyqaApeligggskly
+Y tF      ++      + e G+      v      v ++      + ++g      +
rheu.ef.24  298 TELYNTFQTIAQ-----LKETGRtSGVNPnWtSVQNTTPLNQAGNN---A 339

lsdvftTTeeGyAEI FAGShDtdRIyRWGYLA.vrf.....mletnHnr
++ + t++      G + d I      +++++rf++ + ++l n +
rheu.ef.24  340 QNSRDTWY---K-----GNTYNDNISKLAEITrQRfksatisALP-NYPT 380

dvesllvhsRyGnsfafyaylvkllgymYnnefgiw<-*
+ ++l ++ +G      y+ ++ +g Y      g++
rheu.ef.24  381 IMSTDLYEYHSG----IYSSIFLSAGRSYFETTGAY 412

```

```

#=GF ID      Peptidasa M9
#=GF AC      PF01752.9
#=GF DE      Colagenasa
#=GF AU      Bateman A
#=GF SE      SWISS-PROT
#=GF RM      7582017
#=GF RT      Molecular analysis of an extracellular protease gene from Vibrio
#=GF RT      parahaemolyticus.
#=GF RA      Lee CY, Su SC, Liaw RB;
#=GF RL      Microbiology 1995;141:2569-2576.
#=GF RM      8282691
#=GF RT      Purification and characterization of Clostridium perfringens
#=GF RT      120-kilodalton collagenase and nucleotide sequence of the
#=GF RT      corresponding gene.
#=GF RA      Matsushita O, Yoshihara K, Katayama S, Minami J, Okabe A;
#=GF RL      J Bacteriol 1994;176:149-156.
#=GF DR      INTERPRO; IPR013510;
#=GF DR      MEROPS; M9;
#=GF CC      Esta familia de enzimas degradan los colágenos.
    
```

REPETICIÓN DE HÉLICE DE COLÁGENO

BLKPROB Versión 5/21/00.1

Base de datos=/gcg/hsar/gcgdata/gcgblimps/blocksplus.dat

5

```

=====
Copyright (c) 1992-6 por el Fred Hutchinson Cancer Research Center
En caso de que use BLOCKS en su investigación, sírvase de citar:
Steven Henikoff y Jorja G. Henikoff, Protein Family Classification Based on Searching a
Database of Blocks, Genomics 19:97-107 (1994).
    
```

10

Cada resultado numerado consiste en uno o más bloques de PROSITE o PRINTS

gbDhDi33.35ikn.2.128.pep

Familia		Hebra	Bloques	Valor de E combinado
<u>IPB008161</u>	Repetición de hélice de colágeno	1	1 de 1	<u>0,0077</u>

>IPB008161 1/1 bloques Valor de E combinado= 0,0077: Repetición de hélice de colágeno

Bloque	Marco	Ubicación (aa)	Valor de E de bloque
IPB008161	0	49-91	0,007

Otros alineamientos comunicados:

```

IPB008161 <-> GAPGPPFLPGPKGPRGPAGIEGKPGRLGEDNRPGLPPGPPGVRG
016787|016787_CAEEL143 | | | | | | | | | | | | | | | |
GPPRPPPGLDQLNPEGPPAGPGGPPAILPALPAPADPEPAPRRG
    
```

gbDhDi33.35ikn.2.1 49

Búsqueda=**rheu.ef.241,148** Longitud: 148 Tipo: P

>IPB008161 1/1 bloques Valor de E combinado= 0,0075: Repetición de hélice de colágeno

Bloque	Marco	Ubicación (aa)	Valor de E de bloque
IPB008161	0	67-109	<u>0,0068</u>

Otros alineamientos comunicados:

```

IPB008161 <->
Q9L470_STAEP 1076 GKPAEPGKPAEPGKPAEPGTPAEPGKPAEPGTPAEPGKPAEPG
| | | | | | | | | | | | | | | |
rheu.ef.241.148 67 HLATTLGRPPRPGLPPGPRTPQIRnLPALPAPQGEPPGDRATWR
    
```

Búsqueda=**rheu.ef.238rev.148** 2774.sreformat Longitud: 148

>IPB008161 1/1 bloques Valor de E combinado= 0,0075: **Repetición de hélice de colágeno**

Bloque	Marco	Ubicación (aa)	Valor de E de bloque
IPB008161	0	67-109	<u>0,0068</u>

Otros alineamientos comunicados:

```

IPB008161          <->
Q9L470_STAEP      1076  GKPAEPGKPAEPGKPAEPGTPAEPGKPAEPGTPAEPGKPAEPG
                   | | | | | | | | | | | | | | | |
rheu.ef.238rev.148 67   HLATTLGRPPRPGPPGGPRTPQIRnLPALPAPQGEPEGDRATWR
=
    
```

HMMER 2.3.2 (Oct 2003)
 Copyright (C) 1992-2003 HHMI/Washington University School of Medicine
 Distribuido libremente bajo la licencia pública general GNU (GPL)

```

Archivo HMM:          pfam.hmm
Archivo de secuencia: rheu.ef.241.148
Colágeno: dominio 1 de 1, de 73 a 133: puntuación -74,8, E = 3,5
                   *->GppGppGppGppGppGppGppGppGpaGapGppGppGe . pGpPGppGppG
                   G+p +pGppG p p + p + ++G+pG++ +G+ G++ + G
rheu.ef.24          73   GRFPRPGPPGGPRTPQIRNLPALPAPQGEPEGDRATWRGASGADAAGG 119

                   ppGppGapGapGpp<-*
                   G++Ga+G
rheu.ef.24          120 DGGERGADGGDPGD          133
    
```

rheu.ef.238rev.148
Colágeno Repetición de tripe hélice de colágeno (20 copias)
Colágeno: dominio 1 de 1, de 73 a 133: puntuación -74,8, E = 3,5

```

                   *->GppGppGppGppGppGppGppGppGpaGapGppGppGe . pGpPGppGppG
                   G+p +pGppG p p + p + ++G+pG++ +G+ G++ + G
rheu.ef.23          73   GRFPRPGPPGGPRTPQIRNLPALPAPQGEPEGDRATWRGASGADAAGG 119

                   ppGppGapGapGpp<-*
                   G++Ga+G
rheu.ef.23          120 DGGERGADGGDPGD          133
    
```

```

#=GF ID Colágeno
#=GF AC PF01391.10
#=GF DE Repetición de tripe hélice de colágeno (20 copias)
#=GF AU Bateman A, Eddy SR
#=GF SE Swissprot
#=GF TP Repetición
#=GF BM hmmbuild -F --prior PRIORHMM ls.ann SEED.ann
#=GF BM hmmbuild -f -F --prior PRIORHMM fs.ann SEED.ann
#=GF BM hmmbuild -f -F --prior PRIORHMM fs.ann SEED.ann
#=GF BM hmmbuild -f -F --prior PRIORHMM fs.ann SEED.ann
#=GF AM byscore
#=GF RM 8240831
    
```

#=GF RT New members of the collagen superfamily
 #=GF RA Mayne R, Brewton RG;
 #=GF RL Curr Opin Cell Biol 1993;5:883-890.
 #=GF DR INTERPRO; [IPR008160](#);
 #=GF DR SCOP; [1a9a](#); fa;
 #=GF DR MIM; [240400](#);
 #=GF DC El escorbuto se asocia con los colágenos.
 #=GF CC Los miembros de esta familia pertenecen a la superfamilia de colágeno
 #=GF CC [1]. En general, los colágenos son proteínas estructurales extracelulares
 #=GF CC implicadas en la formación de la estructura del tejido conectivo. El
 #=GF CC alineamiento contiene 20 copias de la repetición de G-X-Y que forma una
 #=GF CC triple hélice. La primera posición de la repetición es glicina, las posiciones
 #=GF CC segunda y tercera pueden ser cualquier resto, pero frecuentemente son
 #=GF CC prolina e hidroxiprolina. Los colágenos se modifican
 #=GF CC postraduccionalmente mediante prolina hidroxilasa para formar los restos
 #=GF CC de hidroxiprolina. La hidroxilación defectuosa es la causa del escorbuto.
 #=GF CC Algunos miembros de la superfamilia del colágeno no están implicados en
 #=GF CC la estructura del tejido conectivo, sino que comparte la misma estructura
 #=GF CC de triple hélice.

PROTEÍNA ESPECÍFICA MASCULINA DE ESPERMA

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine

Distribuido libremente bajo la licencia pública general GNU (GPL)

5

 Archivo HMM: pfam.hmm
 Archivo de secuencia: **gbDhDi33.34ik.2.128**

MSSP: dominio 1 de 1, de 59 a 116: puntuación -9,5, E = 8,9

```

          *->vgGPCgpCGPCggpcCGsccsPCg.gpCgPCgpCGpCGPccggCGPC
          P gp GP g+p+ P ++p P p CG ++ g
gbDhDi33.3 59 QLNPEGPAGPGGPPAIL----PALpAPADPE-PAPRCGGRADGGAAA 100

          GpCGPCCGttekycGl<-*
          G t + l
gbDhDi33.3 101 GAAADADHTGYEEGDL 116
    
```

10

#=GF ID **MSSP**
 #=GF AC PF03940.5
 #=GF DE **Proteína específica masculina de esperma**

15

Esta familia de proteínas de Drosophila está tipificada por el motivo repetitivo C-G-P.

FIRMA DE METALOPROTEASA COLAGENASA MICROBIANA (M9)

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine

Distribuido libremente bajo la licencia pública general GNU (GPL)

20

 Archivo HMM: prints.hmm
 Archivo de secuencia: **gbDhDi43.4rp.1.765**

MICOLLPTASA_1: dominio 1 de 1, de 311 a 328: puntuación 5,3, E = 5,7

```

          *->gletLveflRAGYYvrfyn<-*
          le+ +++ RA Y f++
gbDhDi43.4 311 TLEN-ILYTRASYWNSFHA 328
    
```

MICOLLPTASA

gx; PR00931
 gn; COMPUESTO(5)
 ga; 09-SEP-1998; ACTUALIZADO 07-JUN-1999

gt; **Firma de metaloproteasa colagenasa microbiana (M9)**
gp; PRINTS; [PR00756](#) ALADIPTASE; [PR00791](#) PEPDIPTASEA; [PR00730](#) THERMOLYSIN
gp; PRINTS; [PR00787](#) NEUTRALPTASE; [PR00782](#) LSHMANOLYSIN; [PR00997](#) FRAGILYSIN
gp; PRINTS; [PR00786](#) NEPRILYSIN; [PR00765](#) CRBOXYPTASEA; [PR00932](#) AMINOIPTASE
gp; PRINTS; [PR00789](#) OSIALOPTASE; [PR00933](#) BLYTICPTASE; [PR00934](#) XHISDIPTASE
gp; PRINTS; [PR00919](#) THERMOPTASE; [PR00998](#) CRBOXYPTASET; [PR00768](#) DEUTEROLYSIN
gp; PRINTS; [PRO0999](#) FUNGALYSIN; [PR01000](#) SREBPS2PTASE
gp; INTERPRO; [IPR002169](#)
gp; PROSITE; [PS00142](#) PROTEASA DE CINC
gp; PFAM; [PF00099](#)
gr; 1. RAWLINGS, N.D. Y BARRETT, A.J.
gr; Evolutionary families of metallopeptidases.
gr; METHODS ENZYMOL. 248 183-228 (1995).
gr; 2. RAWLINGS, N.D. Y BARRETT, A.J.
gr; MEROPS - Base de datos de peptidasas
gr; <http://www.bi.bbsrc.ac.uk/merops/merops.htm>
gr; 3. RAWLINGS, N.D. Y BARRETT, A.J.
gr; Familia M9 - Clan MA - Colagenasa microbiana
gr; <http://www.bi.bbsrc.ac.uk/merops/famcards/m9.htm>
gr; 4. BARRETT, A.J., RAWLINGS, N.D. ANDWOESSNER, J.F.
gr; Vibrio collagenase.
gr; EN HANDBOOK OF PROTEOLYTIC ENZYMES, ACADEMIC PRESS, 1998, PP.1096-1098.
gr; 5. BARRETT, A.J., RAWLINGS, N.D. Y WOESSNER, J.F.
gr; Clostridium collagenases.
gr; EN HANDBOOK OF PROTEOLYTIC ENZYMES, ACADEMIC PRESS, 1998, PP.1098-1102.
gr; 6. MATSUSHITA, O., YOSHIHARA, K., KATAYAMA, S., MINAMI, J. Y OKABE, A.
gr; Purification and characterization of Clostridium perfringens 120-kilodalton collagenase and nucleotide
gr; sequence of the corresponding gene.
gr; J.BACTERIOL. 176 149-156 (1994).
gd; Las metaloproteasas son las más diversas de los cuatro tipos principales de
gd; proteasas, habiéndose identificado hasta ahora más de 30 familias [1]. De
gd; estas, aproximadamente la mitad contienen el motivo HEXXH, que se ha
gd; demostrado en estudios cristalográficos que forma parte del sitio de unión a
gd; metales [1]. El motivo HEXXH es relativamente común, pero puede definirse de
gd; manera más rigurosa para metaloproteasas como abXHEbbHbc, donde a suele
gd; ser valina o treonina y forma parte del sustituto S1' en termolisina y neprilisina, b
gd; es un resto no cargado y c un resto hidrófobo. En este sitio nunca hay prolina,
gd; posiblemente debido a que podría romper la estructura helicoidal adoptada por
gd; este motivo en las metaloproteasas [1].
gd; Las metaloproteasas pueden dividirse en cinco grupos basándose en sus restos
gd; de unión a metales: los tres primeros contienen el motivo HEXXH, los otros dos,
gd; no [1]. En el primer grupo, un ácido glutámico completa el sitio activo - estos se
gd; denominan HEXXH+E: todas las familias en este grupo muestran cierta relación
gd; de secuencia y se han asignado al clan MA [1]. El segundo grupo, que tiene una
gd; tercera histidina como resto de unión a metal adicional, se denomina HEXXH+H
gd; y se agrupa en el clan MB basándose en su interrelación [1]. En el tercer grupo,
gd; los restos adicionales de unión a metal no están identificados. El cuarto grupo
gd; es diverso - se conocen los restos de unión a metales pero no forman el motivo
gd; HEXXH. Finalmente, el quinto grupo comprende las demás familias en las que
gd; los restos de unión a metales aún se desconocen [1,2]. Se han identificado
gd; colagenasas microbianas de bacterias de los géneros tanto Vibrio como
gd; Clostridium. Son metalopeptidasas que contienen cinc que pertenecen a la
gd; familia de proteasa M25, que forman parte del clan MA [1,3]. La colagenasa se
gd; usa durante el ataque bacteriano para degradar la barrera de

gd; colágeno del hospedador durante la invasión. Las bacterias del género Vibrio no son patógenas y
 gd; en ocasiones se usan en hospitales para eliminar tejido muerto de heridas y úlceras [4]. Clostridium
 gd; histolyticum es un patógeno que provoca la gangrena gaseosa; sin embargo, la colagenasa
 gd; aislada se ha usado para tratar las úlceras por presión [5]. La escisión del colágeno se produce
 gd; en los enlaces Xaa+Gly en las colagenasas de bacterias del género Vibrio y en Yaa+Gly en las
 gd; de Clostridium [4,5].
 gd; El análisis de la estructura primaria del producto génico de Clostridium perfringens ha revelado
 gd; que la enzima se produce con una serie de 86 restos que contienen una supuesta secuencia de
 gd; señal [6]. Dentro de esta serie se encuentra PLGP, una secuencia de aminoácidos típica de
 gd; sustratos de colagenasa.
 gd; Por lo tanto, esta secuencia puede estar implicada en el auto-procesamiento de la
 gd; colagenasa [6].
 gd; MICOLLPTASA es una huella de 5 elementos que proporciona una firma para metalopeptidasas
 gd; de cinc de colagenasa microbianas (M9). La huella se obtuvo de un alineamiento inicial de 4
 gd; secuencias: los motivos se representaron a partir de regiones conservadas que abarcan
 gd; virtualmente la longitud de alineamiento completa - el motivo 4 incluye la región codificada por el
 gd; patrón PROSITE de PROTEASA DE CINCO (PS00142), que describe el sitio activo de HEXXH; y
 gd; el motivo 5 contiene el glutamato del sitio activo. Para alcanzar la convergencia, se necesitaron
 gd; dos interacciones en OWL31.1, punto en el cual se identificó un conjunto auténtico que
 gd; comprendía 8 secuencias.
 tp; COLA CLOPE 054108 COLA VIBAL Q46085
 tp; COLA_VIBPA
 sn; Códigos que implican 4 elementos
 st; 086030
 tt; COLA CLOPE PRECURSOR DE COLAGENASA MICROBIANA (EC 3.4.24.3) (120 KD
 COLAGENASA) - CLOSTRIDIUM
 tt; 054108 SUPUESTA PROTEASA SECRETADA - STREPTOMYCES COELICOLOR.
 tt; COLA VIBAL PRECURSOR DE COLAGENASA MICROBIANA (EC 3.4.24.3) - VIBRIO
 ALGINOLYTICUS.
 tt; Q46085 PRECURSOR DE COLAGENASA - CLOSTRIDIUM HISTOLYTICUM.
 tt; COLA VIBPA PRECURSOR DE COLAGENASA MICROBIANA (EC 3.4.24.3) - VIBRIO
 PARAHAEMOLYTICUS.
 tt; 086030 COLAGENASA - VIBRIO CHOLERAE.
 ic; MICOLLPTASE1
 il; 19
 it; Motivo I de colagenasa microbiana - 1
 id; GIPTLVEFLRAGYYLGFYN COLA CLOPE 159 159
 id; ELETFLYLRAAGYYAEFYN COLA_VIBAL 144 144
 id; VLENLGEFVRAAYVRYNA COLA_VIBPA 97 97
 id; RLENYGEFIRAAYVRYNA AF080248 97 97
 bb;

Firma de proteína MIC1 de micronema

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine

Distribuido libremente bajo la licencia pública general GNU (GPL)

5

Archivo HMM: prints.hmm
 Archivo de secuencia: rheu.ef.242.746
MIC1MICRNEME_5: dominio 1 de 1, de 448 a 463: puntuación **6,6**, E = 4,4
 ->TyiStkLdVaVGsChk<-
 T t+L Va GSC
 rheu.ef.24 448 TKADTQLIVAGGSCKA 463

gc; MIC1MICRNEME
 gx; PR01744
 gn; COMPUESTO(7)
 ga; 03-JUL-2002

gt; **Firma de proteína MIC1 de micronema**
 gr; 1. SIBLEY, L.D., MORDUE, D. Y HOWE, K.
 gr; Experimental approaches to understanding virulence in toxoplasmosis.
 gr; IMMUNOBIOLOG. 201 210-224 (1999).
 gr; 2. CARRUTHERS, V.B.
 gr; Armed and dangerous: Toxoplasma gondii uses an arsenal of secretory proteins
 gr; to infect host cells.
 gr; PARASITOL.INT. 48 1-10 (1999).
 gr; 3. FOURMAUX, M.N., ACHBAROU, A., MERCEREAU-PDIJALON, O., BIDERRE, C.,
 gr; BRICHE, I., LOYENS, A., ODBERG-FERRAGUT, C., CAMUS, D. Y DUBREMETZ, J.F.
 gr; The MIC1 microneme protein of Toxoplasma gondii contains a duplicated
 gr; receptor-like domain and binds to host cell surface.
 gr; MOL.BIOCHEM.PARASITOL. 20 201-210 (1996).
 gr; 4. LOURENCO, E.V., PEREIRA, S.R., FACA, V.M., COELHO-CASTELO, A.A., MINEO, J.R., ROQUE-
 gr; BARREIRA, M.C., GREENE, L.J. Y PANUNTO-CASTELO, A.
 gr; Toxoplasma gondii micronemal protein MIC1 is a lactose-binding lectin.
 gr; GLYCOBIOLOG. 11 541-547 (2001).
 gr; 5. KELLER, N., NAGULESWARAN, A., CANNAS, A., VONLAUFEN, N., BIENZ, M., BJORKMAN, C.,
 gr; BOHNE, W. Y HEMPHILL, A.
 gr; Identification of a Neospora caninum microneme protein (NcMIC1) which interacts with sulphated host cell
 gr; surface glycosaminoglycans.
 gr; INFECT.IMMUN. 70 187-198 (2002).
 gd; Toxoplasma gondii es un protozoo parásito intracelular obligado del género
 gd; Apicomplexa, con un modo de vida complejo que implica diversos hospedadores
 gd; [1]. Tiene dos fases de crecimiento: una fase intestinal en hospedadores felinos y
 gd; una fase extraintestinal en otros mamíferos. Los ovocistos de gatos infectados se
 gd; desarrollan en taquizoitos y en última instancia, bradizoitos y zoocistos en el
 gd; hospedador extraintestinal [1]. La transmisión del parásito se produce por el
 gd; contacto con gatos infectados o carne cruda/poco cocinada; en individuos
 gd; inmunocomprometidos, puede provocar toxoplasmosis grave y en ocasiones letal.
 gd; La infección aguda en humanos sanos también puede causar en ocasiones daño
 gd; tisular [1].
 gd; El protozoo utiliza una serie de proteínas secretoras y antigénicas para invadir
 gd; un hospedador y obtener acceso al ambiente intracelular [2]. Estas se originan
 gd; en distintos orgánulos en la célula de T. gondii denominados micronemas,
 gd; roptrias y gránulos densos. También se liberan en momentos específicos
 gd; durante la invasión para asegurar que las proteínas se sitúan en sus destinos
 gd; diana correctos [2].
 gd; MIC1, una proteína secretada del micronema, es un residuo de 456 restos
 gd; implicado en el reconocimiento de la célula hospedadora por el parásito [3]. La
 gd; proteína se libera desde el polo apical de T. gondii durante la infección y se une a
 gd; receptores específicos del hospedador [4]. Estudios recientes han demostrado que
 gd; Mic1 es una lectina de unión a lactosa y la utiliza para mejorar su unión a células
 gd; endoteliales del hospedador [4]. Un homólogo de Mic1 encontrado en Neospora
 gd; caninum interactúa con glucosaminoglucanos sulfatados de la superficie de la
 gd; célula hospedadora [5]. MIC1MICRNEMA es una huella de 7 elementos que
 gd; proporciona una firma para las proteínas micronema MIC1. La huella se obtuvo de
 gd; un alineamiento inicial de 2 secuencias: los motivos se representaron a partir de
 gd; regiones conservadas que abarcan la parte C-terminal del alineamiento (~380
 gd; aminoácidos). Para alcanzar la convergencia se necesitó una sola interacción en
 bb; SPTR40_20f, sin identificar secuencias adicionales más allá del conjunto de inicio.
 ic; **MIC1MICRNEMA5**
 il; 16
 it; **motivo V de proteína MIC1 de micronema - 1**
 id; TFISTKLDVAVGSCHS 000834 341 133
 id; TYSSPQLHVSVGSCHK 08WRS0 344 138

FIRMA DE REGULADOR AUTOINMUNITARIO (AIRE)

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine

Distribuido libremente bajo la licencia pública general GNU (GPL)

Archivo HMM: prints.hmm

Archivo de secuencia: rheu.ef.241.736

AIREGULATOR_4: dominio 1 de 1, de 138 a 152: puntuación 6,4, E = 9,2

5

```

          *->DFWRvLFKDYnLERY<-*
          FW v   D L RY
rheu.ef.24   138   NFWTVSNEDLDLCRY   152

```

rheu.ef.234rev.628

AIREGULATOR_4: dominio 1 de 1, de 30 a 44: puntuación 6,4, E = 9,2

```

          *->DFWRvLFKDYnLERY<-*
          FW v   D L RY
rheu.ef.23   30   NFWTVSNEDLDLCRY   44

```

5

rheu.cd.215rev.1.736

AIREGULATOR_4: dominio 1 de 1, de 138 a 152: puntuación 6,4. E = 9,2

```

          *->DFWRvLFKDYnLERY<-*
          FW v   D L RY
rheu.cd.21   138   NFWTVSNEDLDLCRY   152

```

gc; **AIREGULATOR**

gx; PR01711

gn; COMPUESTO(8)

ga; 13-MAR-2002

gt; **Firma de regulador autoinmunitario (AIRE)**

gr; 1. The Finnish-German APECED Consortium.

gr; An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains.

gr; NAT.GENET. 17 399-403 (1997).

gr; 2. MITTAZ, L., ROSSIER, C., HEINO, M., PETERSON, P., KROHN, K.J.E., GOS, A., MORRIS, M.A.,

gr; KUDOH, J., SHIMIZU, N., ANTONARAKIS, S.E. Y SCOT, H.S.

gr; Isolation and characterization of the mouse Aire gene.

gr; BIOCHEM.BIOPHYS.RES.COMMUN. 255 483-490 (1999).

gr; 3. PETERSON, H.M., KUDOH, J., NAGAMINE, K., LAGERSTEDT, A., OVOD, V.,

gr; RANKI, A., RANTALA, I., NIEMINEN, M., TUUKKANEN, J., SCOTT, H.S., ANTONARAKIS, S.E., SHIMIZU,

gr; N. Y KROHN, K.

gr; Autoimmune regulator is expressed in the cells regulating immune

gr; tolerance in thymous medulla.

gr; BIOCHEM.BIOPHYS.RES.COMMUN. 257 821-825 (1999).

gr; 4. KUMAR, P.G., LALORAYA, M., WANG, C.Y., RUAN, Q.G., SEMIROMI, A.D., KAO, K.J. Y SHE, J.X.

gr; The autoimmune regulator (AIRE) is a DNA-binding protein. J.BIOL.CHEM. 276 41357-41364 (2001).

gr; **AIRE (REgulador AutoInmunitario)** es la proteína predicha responsable

gr; de una enfermedad de herencia autosómica recesiva denominada

gd; APECED. APECED, también denominado síndrome poliglandular

gd; autoinmunitario de tipo I (APS 1) es la única enfermedad autoinmunitaria

gd; descrita con un origen monogénico demostrado, localizándose fuera de la

gd; región del complejo mayor de histocompatibilidad. Se caracteriza por la

gd; presencia de dos de las tres entidades clínicas principales, candidiasis

gd; mucocutánea crónica, hipoparatiroidismo y enfermedad de Addison. Otros

gd; fenotipos mediados inmunológicamente, incluyendo diabetes mellitus

gd; insulino dependiente (IDDM), insuficiencia gonadal, gastritis crónica,

gd; vitíligo, enfermedad tiroidea autoinmunitaria, hipoplasia del esmalte y

gd; alopecia también pueden estar presentes. Inmunológicamente, los

gd; pacientes con APECED tienen respuestas

gd; de linfocitos T defectuosas contra antígenos de Candida y síntomas clínicos tanto dentro
gd; como fuera del sistema endocrino, principalmente a consecuencia de la autoinmunidad
gd; contra autoantígenos específicos de órganos [1,2].
gd; AIRE tiene motivos que sugieren una proteína reguladora transcripcional. Porta dos dedos
gd; de cinc de tipo homodominio de planta (PHD). Un supuesto dominio de unión a ADN,
denominado SAND, así como cuatro motivos LXXLL de unión a receptor nuclear, un
gd; dominio LXXLL invertido y una variante de este último (FXXLL), sugieren que esta
gd; proteína funciona como coactivador de la transcripción. Además, un dominio N-terminal de
gd; 100 aminoácidos altamente conservado en AIRE muestra una similitud significativa con el
gd; dominio de tinción homogénea (HSR) de las proteínas Sp100 y Spl40, que se ha
gd; demostrado que funciona como dominio de dimerización en varias proteínas relacionadas
gd; con Sp-100 [2-4].
gd; AIRE tiene una ubicación subcelular dual. No solo se expresa en múltiples tejidos
gd; inmunológicamente relevantes, tales como el timo, bazo, ganglios linfáticos y médula
gd; ósea, sino que también se ha detectado en varios tejidos distintos, tales como riñón,
gd; testículos, glándulas suprarrenales, hígado y ovario, lo que sugiere que las proteínas
gd; APECED también podrían tener una función fuera del sistema inmunitario.
gd; Sin embargo, AIRE no se expresa en los órganos diana de la destrucción autoinmunitaria.
gd; A nivel subcelular, puede encontrarse AIRE en el núcleo celular con un patrón moteado en
gd; dominios que se asemejan a cuerpos nucleares de leucemia promielocítica, también
gd; conocidos como ND10, puntos nucleares o dominios oncogénicos potenciales asociados
gd; con las proteínas nucleares homólogas de AIRE, Sp100, Spl40 y Lysp100. La localización
gd; nuclear de AIRE, en concordancia con sus dominios de proteína predichos, sugiere que
gd; podría regular los mecanismos implicados en la inducción y el mantenimiento de la
gd; tolerancia inmunitaria [3,4].
gd; AIREGULATOR es una huella de 8 elementos que proporciona una firma para los
gd; reguladores autoinmunitarios AIRE. La huella se obtuvo de un alineamiento inicial de 6
gd; secuencias: los motivos se representaron a partir de regiones conservadas que abarcan
gd; en gran medida las porciones N-terminales y centrales del alineamiento, centrándose en
gd; aquellas secciones que caracterizan a los autorreguladores pero que los distinguen de los
gd; que poseen dominios SAND y PHD. Para alcanzar la convergencia, se necesitaron dos
gd; interacciones en SPTR39_17f, punto en el cual se identificó un conjunto auténtico que
gd; comprendía 14 secuencias.

fc; **AIREGULATOR4**

fl; 15

ft; **Motivo IV de regulador autoinmunitario (AIRE) - 1**

fd;	DFWRILFKDYNLERY	Q9JLW0	77	18
fd;	DFWRILFKDYNLERY	Q9Z0E3	77	18
fd;	DFWRILFKDYNLERY	Q9JLX0	77	18
fd;	DFWRILFKDYNLERY	Q9JLW9	77	18
fd;	DFWRILFKDYNLERY	Q9JLW8	77	18
fd;	DFWRILFKDYNLERY	Q9JLW7	77	18
fd;	DFWRILFKDYNLERY	Q9JLW6	77	18
fd;	DFWRILFKDYNLERY	Q9JLW5	77	18
fd;	DFWRILFKDYNLERY	Q9JLW4	77	18
fd;	DFWRILFKDYNLERY	Q9JLW3	77	18
fd;	DFWRILFKDYNLERY	Q9JLW2	77	18
fd;	DFWRILFKDYNLERY	Q9JLW1	77	18
fd;	DFWRVLFKDYNLERY	AIRE HUMANA	76	18
fd;	DFWRVLFKDYNLERY	075745	76	18

GLIADINA

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine Distribuido libremente bajo la licencia pública general GNU (GPL)

5 Archivo HMM: prints.hmm
 Archivo de secuencia: **rheu.ef.241.736**

GLIADIN_7: dominio 1 de 1, de 688 a 708: puntuación 17,7, E = 0,056

->PqaaGsvqPqqLPqFeEiRnL<-
 qaaGsvq q L q E R L

rheu.ef.24 688 TQAQGSVQEQLLQLREQRVL 708

rheu.ef.234rev.628

10 **GLIADIN_7:** dominio 1 de 1, de 580 a 600: puntuación 17,7, E = 0,056

```

      *->PqaqGsvqPqqLPqFeEiRnL<-*
      qaqGsvq q L q E R L
rheu.ef.23 580 TQAQGSVQEQLLLQLREQRVL 600

```

rheu.cd.215rev.1.736

GLIADINA_7: dominio 1 de 1, de 688 a 708: puntuación 18,3, E = 0,037

```

      *->PqaqGsvqPqqLPqFeEiRnL<-*
      qaqGsvq q L q E R L
rheu.cd.21 688 TQAQGSVQDQLLLQLREQRVL 708

```

5 **GLIADIN_7:** dominio 1 de 1, de 46 a 66: puntuación 18,3, E = 0,037

```

      *->PqaqGsvqPqqLPqFeEiRnL<-*
      qaqGsvq q L q E R L
zc3r11.B4. 46 TQAQGSVQDQLLLQLREQRVL 66

```

gc; **GLIADINA**

gx; PR00209

gn; COMPUESTO(9)

ga; 21-OCT-1992; ACTUALIZADO 19-JUN-1999

gt; **Firma de la familia de alfa/beta gliadina**

gp; PRINTS; [PR00208](#) GLIADGLUTEN; [PR00211](#) GLUTELIN; [PR00210](#) GLUTENIN

gp; INTERPRO; [IPR001376](#)

gr; 1. SHEWRY, P. Y MORGAN, M.

gr; Gluten - proteins that put the springiness into bread and are implicated

gr; in food intolerance syndromes such as coeliac disease.

gr; EN PROTEIN POWER AFRC NEWS SUPPLEMENT (1992).

gr; 2. OKITA T.W., CHEESBROUGH V. Y REEVES C.D.

gr; Evolution and heterogeneity of the alpha-type, beta-type, and gamma-type

gr; gliadin DNA sequences.

gr; J.BIOL.CHEM. 260(13) 8203-8213 (1985).

gr; 3. RAFALSKI J.A.

gr; Structure of wheat gamma-gliadin genes.

gr; GENE 43(3) 221-229 (1986).

gd; El gluten es la proteína componente de la harina de trigo. Consta de numerosas

gd; proteínas, que son de 2 tipos diferentes, responsables de diferentes propiedades físicas

gd; de la masa [1]: las glutelinas, que son responsables principalmente de la elasticidad y las

gd; gliadinas, que contribuyen a la extensibilidad.

gd; Las gliadinas en sí son de diferentes tipos (por ejemplo, alfa/beta o gamma) y, al igual

gd; que las glutelinas, contienen secuencias repetitivas [2] que forman estructuras

gd; helicoidales sueltas, pero normalmente se asocian con regiones no repetitivas más

gd; extensas, que son compactas y globulares [3].

gd; GLIADINA es una huella de 9 elementos que proporciona una firma para las alfa/beta

gd; gliadinas. La huella se obtuvo de un alineamiento inicial de 5 secuencias: los motivos 2 y

gd; 3 codifican las repeticiones en tándem ricas en Gln/Pro. Para alcanzar la convergencia, se

gd; necesitaron dos interacciones en OWL18.0, punto en el cual se identificó un conjunto

gd; auténtico que comprendía 14 secuencias. También se encontraron varias coincidencias

gd; parciales: 3 de estas son fragmentos de alfa/beta gliadina:

gd; tanto GDA1 WHEAT como B22364 carecen de la parte C-terminal de la secuencia que

gd; porta los 2 últimos motivos y GDA8 WHEAT carece de la parte N-terminal de la secuencia

gd; que porta los 3 primeros motivos.

gd; Además de los fragmentos de alfa/beta gliadina, se identificó una serie de otras

gd; coincidencias parciales: estas incluyen gamma-gliadinas, gluteninas de bajo

gd; peso molecular, aveninas, secalinas y así sucesivamente. La mayoría de estas

gd; no logran emparejarse o al menos se emparejan mal, con los motivos que

gd; codifican las repeticiones en tándem - evidentemente, se caracterizan por sus

gd; propias firmas distintivas en esta región. La huella proporciona, por lo tanto, una

gd; discriminación razonable entre las gliadinas de tipo alfa/beta y las de tipo gamma

gd; y proteínas relacionadas.

c; **GLIADIN7**

fl; 21

ft; Motivo VII de gliadina - 2

fd;	PQAQGSVQPQQLPQFEEIRNL	GDA9_WHEAT	259	6
fd;	PQAQGSVQPQQLPQFEEIRNL	GDA6_WHEAT	246	6
fd;	PQAQGSVQPQQLPQFEEIRNL	Q41509	239	6
fd;	PQAQGSVQPQQLPQFEEIRNL	Q41531	241	6
fd;	PQAQGSVQPQQLPQFEEIRNL	GDA0_WHEAT	238	6
fd;	PQAQGSVQPQQLPQFAEIRNL	GDA7_WHEAT	263	6
fd;	PQAQGSVQPQQLPQFAEIRNL	Q41546	263	6
fd;	PQAQGSFQPQQLPQFEEIRNL	GDA2_WHEAT	243	6
fd;	PQAQGSVQPQQLPQFEEIRNL	Q41632	246	6
fd;	PQAQGSVQPQQLPQFEEIRNL	Q41530	240	6
fd;	PQAQGSVQPQQLPQFAEIRNL	Q41529	263	6
fd;	PQAQGSVQPQQLPQFAEIRNL	GDA5_WHEAT	269	6
fd;	PQAQGSVQPQQLPQFAEIRNL	Q41545	268	6
fd;	PQAQGSVQPQQLPQFEEIRNL	Q41528	239	6
fd;	PQAQGSVQPQQLPQFEEIRNL	GDA4_WHEAT	249	6
fd;	PQAQGSVQPQQLPQFQEIIRNL	GDA3_WHEAT	232	6

FIRMA DE RECEPTOR Y2 DE NEUROPEPTIDO

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine Distribuido libremente bajo la licencia pública general GNU (GPL)

5 Archivo HMM: prints.hmm

Archivo de secuencia: **rheu.ef.241.736**

NRPEPTIDEY2R_9: dominio 1 de 1, de 664 a 677: puntuación 8,9, E = 3,1

```

*->AFLsAFRCEqRLDAiHs<-*
      sAFR qR+ +Hs
rheu.ef.24 664 ---SAFRVQQRVPWVHS 677

```

rheu.ef.234rev.628

10 **NRPEPTIDEY2R_9:** dominio 1 de 1, de 556 a 569: puntuación 8,9, E = 3,1

```

*->AFLsAFRCEqRLDAiHs<-*
      sAFR qR+ +Hs
rheu.ef.23 556 ---SAFRVQQRVPWVHS 569

```

NRPEPTIDEY2R_9: dominio 1 de 1, de 22 a 35: puntuación 7,2, E = 6,3

```

*->AFLsAFRCEqRLDAiHs<-*
      s FR qRL +Hs
zc3r11.34. 22 ---SRFRVQQRLPWVHS 35

```

gc; **NRPEPTIDEY2R**

gx; PRO1014 gn;

gn; COMPUESTO(11)

ga; 30-NOV-1998; ACTUALIZADO 07-JUN-1999

gt; **Firma de receptor Y2 de neuropéptido**

gp; PRINTS; [PR00237](#) GPCRRHODOPSN; [PR00247](#) GPCRCAMP; [PR00248](#) GPCRMGR

gp; PRINTS; [PR00249](#) GPCRSECRETIN; [PR00250](#) GPCRSTE2; [PR00899](#) GPCRSTE3

gp; PRINTS; [PR00251](#) BACTRLOPSIN

gp; PRINTS; [PR01012](#) NRPEPTIDEYR; [PR01013](#) NRPEPTIDEY1R; [PR01015](#) NRPEPTIDEY4R

gp; PRINTS; [PR01016](#) NRPEPTIDEY5R; [PR01017](#) NRPEPTIDEY6R

gp; INTERPRO; [IPR001358](#)

gr; 1. ATTWOOD, T.K. Y FINDLAY, J.B.C.

gr; Fingerprinting G protein-coupled receptors.

gr; PROTEIN ENG. 7(2) 195-203 (1994).

gr; 2. ATTWOOD, T.K. Y FINDLAY, J.B.C.

gr; G protein-coupled receptor fingerprints.

gr; 7TM, VOLUMEN 2, EDS. G.VRIEND Y B.BYWATER (1993).
 gr; 3. BIRNBAUMER, L.
 gr; G proteins in signal transduction.
 gr; ANNU.REV.PHARMACOL.TOXICOL. 30 675-705 (1990).
 gr; 4. CASEY, P.J. Y GILMAN, A.G.
 gr; G protein involvement in receptor-effector coupling.
 gr; J.BIOL.CHEM. 263(6) 2577-2580 (1988).
 gr; 5. ATTWOOD, T.K. Y FINDLAY, J.B.C.
 gr; Design of a discriminating fingerprint for G protein-coupled receptors.
 gr; PROTEIN ENG. 6(2) 167-176 (1993).
 gr; 6. WATSON, S. Y ARKINSTALL, S.
 gr; Neuropeptide Y.
 gr; EN THE G PROTEIN-LINKED RECEPTOR FACTSBOOK, ACADEMIC PRESS, 1994,
 gd; PP.194-198 Los receptores acoplados a proteína G (GPCR) constituyen una vasta familia
 gd; de proteínas que abarca una gran variedad de funciones (incluyendo diversos procesos
 gd; autocrinos, paracrinos y endocrinos). Muestran una considerable diversidad a nivel de
 gd; secuencia, que puede usarse como base para separarlas en distintos grupos. Se usa el
 gd; término clan para describir las GPCR, ya que abarcan un grupo de familias para las que
 gd; hay indicaciones de relación evolutiva, pero entre las cuales no hay una similitud
 gd; estadísticamente significativa en su secuencia [1,2]. Los miembros del clan conocidos en
 gd; la actualidad incluyen las GPCR similares a rodopsina, las GPCR similares a secretina,
 gd; los receptores de AMPc, los receptores de feromonas de emparejamiento fúngico y la
 gd; familia de receptor de glutamato metabotrópico.
 gd; Las GPCR similares a rodopsina representan por sí mismas una familia de proteínas
 gd; ampliamente diseminada que incluye receptores de hormonas, neurotransmisores y de
 gd; luz, todos los cuales transducen señales extracelulares mediante su interacción con
 gd; proteínas de unión a nucleótidos guanina (G). Aunque sus ligandos activadores varían
 gd; ampliamente en su estructura y carácter, las secuencias de aminoácidos de los
 gd; receptores son muy similares y se cree que adoptan un armazón estructural común que
 gd; comprende 7 hélices transmembrana (TM) [3-5].
 gd; El neuropéptido Y [NPY] es uno de los péptidos más abundantes en el cerebro de
 gd; mamíferos, e induce una serie de efectos en el comportamiento (por ejemplo,
 gd; estimulación de la ingesta de comida, ansiedad, facilitación del aprendizaje y la memoria
 gd; y regulación de los sistemas cardiovascular y neuroendocrino) [6]. En la periferia, NPY
 gd; estimula la contracción del músculo liso vascular y modula la secreción de hormonas. Se
 gd; ha relacionado a NPY con la fisiopatología de la hipertensión, la insuficiencia cardíaca
 gd; congestiva, los trastornos afectivos y la regulación del apetito [6]. Se han caracterizado
 gd; varios receptores Y de neuropéptido Y farmacológicamente distintos, denominados NPY
 gd; Y1-Y6. Hay presencia de altas densidades de receptores Y2 en el hipocampo de ratas y
 gd; también se encuentran a altos niveles en las capas superficiales del córtex, ciertos
 gd; núcleos talámicos, el septo lateral y el núcleo olfatorio anterior; se encuentran niveles
 gd; menores en el striatum [6]. Los receptores se encuentran a altos niveles en el músculo
 gd; liso (por ejemplo, conductos deferentes e intestino), los túbulos proximales del riñón y en
 gd; estirpes celulares [6]. Se cree que tienen una ubicación predominantemente presináptica
 gd; y están implicados en la inhibición de la adenililo ciclasa y de los canales de calcio
 gd; dependientes de voltaje mediante una proteína G sensible a la toxina tósferínica,
 gd; probablemente de la clase G_O/G_i [6]. NRPEPTIDEY2R es una huella de 11 elementos
 gd; que proporciona una firma para receptores de neuropéptido Y2. La huella se obtuvo de
 gd; un alineamiento inicial de 2 secuencias: los motivos se representaron a partir de
 gd; secciones conservadas dentro de cada bucle o las regiones TM, centrándose en aquellas
 gd; áreas del alineamiento que caracterizan los receptores Y2, pero que lo distinguen del
 gd; resto de la familia de neuropéptido Y - los motivos 1-3 abarcan el extremo N-terminal,
 gd; desembocando en el dominio 1 TM; los motivos 4 y 5 abarcan el extremo C-terminal del
 gd; dominio 4 TM y el segundo bucle externo; los motivos 6 y 7 abarcan el extremo C-
 gd; terminal del dominio 5 TM y el tercer bucle citoplásmico; el motivo 8 abarca el extremo C-
 gd; terminal del dominio 6 TM y el tercer bucle externo; y los motivos 9-11 se encuentran en
 gd; el extremo C-terminal. Para alcanzar la convergencia, se necesitaron dos interacciones
 gd; en OWL30.2, punto en el cual se identificó un conjunto auténtico que comprendía 5
 gd; secuencias. También se encontraron dos coincidencias parciales: OAU83458 es un
 gd; fragmento de receptor de neuropéptido Y2 ovino que coincide con los motivos 4-6;
 gd; yAF054870 es un fragmento de receptor de neuropéptido Y2 de rata que coincide con los
 gd; motivos 5 y 6.
 fc; NRPEPTIDEY2R9
 fl; 17
 ft; Motivo IX del receptor de neuropéptido Y2 - 2

fd;	AFLSAFRCEQRLDAIHS	NY2R HUMAN	335	29
fd;	AFLSAFRCEQRLDAIHS	NY2R_BOVIN	338	29
fd;	AFLSAFRCEQRLDAIHS	NY2R_MOUSE	339	29
fd;	AFLSAFRCEQRLDAIHS	NY2R_PIG	337	29

AEROLISINA

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine

Distribuido libremente bajo la licencia pública general GNU (GPL)

5

Archivo HMM: prints.hmm
 Archivo de secuencia: **rheu.ef.241.736**

AEROLYSINA_7: dominio 1 de 1, de 602 a 621: puntuación 3,4, E = 9,3

->wDKRYiPGEvKWWDWnWtiq<-
 +D +Y+ Ev W W

rheu.ef.24 602 VDPKYVTPEVTWHSWDIRRG 621

10

rheu.ef.234rev.628

AEROLYSINA 7: dominio 1 de 1, de 494 a 513: puntuación 3,4, E = 9,3

->wDKRYiPGEvKWWDWnWtiq<-
 +D +Y+ Ev W W

rheu.ef.23 494 VDPKYVTPEVTWHSWDIRRG 513

Archivo HMM: prints.hmm

Archivo de secuencia: **uro742rev.109r**

15

AEROLYSINA_7: dominio 1 de 1, de 65 a 84: puntuación 3,6, E = 8,6

->wDKRYiPGEvKWWDWnWtiq<-
 + G K W WnW+ +

uro742rev. 65 FAWVLASGTAKCWSWNWSAR 84

AEROLYSIN_7: dominio 1 de 1, de 65 a 84: puntuación 3,6, E = 8,6

->wDKRYiPGEvKWWDWnWtiq<-
 + G K W WnW+ +

zc37.B9.2d 65 FAWVLASGTAKCWSWNWSAR 84

gc; **AEROLISINA**

gx; PR00754

gn; COMPUESTO(9)

ga; 25-AUG-1997; ACTUALIZADO 06-JUN-1999

gt; **Firma de aerolisina**

gp; INTERPRO; IPR001776

gp; PROSITE; PS00274

gp; AEROLYSIN PFAM; PF01117 Aerolysin

gr; 1. PARKER, M.W., BUCKLEY, J.T., POSTMA, J.P., TUCKER, A.D., LEONARD, K.,

gr; PATTUS, F. Y TSENOGLOU, D.

gr; Structure of the aeromonas toxin proaerolysin in its water-soluble and membrane-

gr; channel states.

gr; NATURE 367 292-295 (1994).

gd; La aerolisina es responsable de la patogenicidad de Aeromonas hydrophila, una

gd; bacteria asociada con enfermedades diarreicas e infecciones de heridas profundas

gd; [1]. En común con otras toxinas microbianas, la proteína cambia en un proceso

gd; multi-etapa desde una forma hidrosoluble hasta producir un canal transmembrana

gd; que destruye a las células sensibles, rompiendo sus barreras de permeabilidad [1].

gd; Se ha determinado la estructura de la proaerolisina con una resolución de 2,8Å y

gd; demuestra que la protoxina adopta un nuevo plegamiento [1]. Las imágenes del

gd; oligómero de aerolisina obtenidas por microscopía electrónica han ayudado a

gd; construir un modelo de la proteína y a esbozar un mecanismo mediante el cual

gd; podría insertarse en las bicapas lipídicas para formar canales de iones [1].

gd; AEROLYSINA es una huella de 9 elementos que proporciona una firma para las

gd; aerolisinas. La huella se obtuvo de un alineamiento inicial de 10 secuencias: los

gd; motivos se representaron a partir de regiones conservadas que abarcan

gd; prácticamente la longitud completa del alineamiento. Para alcanzar la convergencia

gd; se necesitó una sola interacción en OWL29.4, sin identificar secuencias adicionales

gd; más allá del conjunto de inicio. Se encontró una sola coincidencia parcial,

gd; CLOALPTOX, una alfa-toxina relacionada de Clostridium septicum que coincide

gd; con los motivos 4 y 6.

```

fc; AEROLYSIN7
fl; 20
ft; Motivo VII de aerolisina - 2
fd; WDKRYIPGEVKWWDWNWTIQ      AERA_AERHY      382      21
fd; WDKRYIPGEVKWWDWNWTIQ      Q44063          382      21
fd; WDKRYIPGEVKWWDWNWTIQ      AER3_AERHY      382      21
fd; WDKRYIPGEVKWWDWNWTIQ      AER5_AERHY      382      21
fd; WDKRYIPGEVKWWDWNWTIQ      AER4_AERHY      382      21
fd; WDKRYIPGEVKWWDWNWTIQ      P94128          382      21
fd; WDKRYLPGEMKWWWDWNWAIQ     AERA_AERTR      382      21
fd; WDKRYLPGEMKWWWDWNWAIQ     085370          382      21
fd; VDKRYIPGEVKWWDWNWTIS      AERA_AERSA      383      21
fd; VDKRYIPGEVKWWDWNWTIS      AERA_AERSO      382

```

OREXINA:

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine

Distribuido libremente bajo la licencia pública general GNU (GPL)

5

```

-----
Archivo HMM:                pfam.hmm
Archivo de secuencia:       rheu.ef.241.148
Orexina: dominio 1 de 1, de 10 a 122: puntuación -38,9, E = 4,1
      *->mnlPsaKvsWAavtL LLLLLLLP PAlLslGvdAqPLPDCCRqKtCsC
      +   v  A   LL +  PP   +G++   C R   C
rheu.ef.24   10   RKVLLQTVRAAKKARRLLGMWQPPVHNVPGIERNWYESCFRSHA AVC 56

      RLYELLHGAGnHAAGiLtLGK.RRPGPPGLqGRLqRLLqAsGnHAAGiLt
      + + G  nH A   tLG++  RPGPPG           G   i
rheu.ef.24   57   GCGDFV-GHINHLAT--TLGRpPRPGPPG-----GPRTPOI-- 89

      mGRRAGAElePrLCPGRRClAaAAsalAPrGrsvr<-*
      R   A ++P+  PG R   As   G+ +
rheu.ef.24   90   --RNLPALPAPQGE PGDRATWRGASGADAAGGDGG      122

```

10

rheu.ef.238rev.148

Orexina: dominio 1 de 1, de 10 a 122: puntuación -38,9, E = 4,1

```

      *->mnlPsaKvsWAavtL LLLLLLLP PAlLslGvdAqPLPDCCRqKtCsC
      +   v  A   LL +  PP   +G++   C R   C
rheu.ef.23   10   RKVLLQTVRAAKKARRLLGMWQPPVHNVPGIERNWYESCFRSHA AVC 56

      RLYELLHGAGnHAAGiLtLGK.RRPGPPGLqGRLqRLLqAsGnHAAGiLt
      + + G  nH A   tLG++  RPGPPG           G   i
rheu.ef.23   57   GCGDFV-GHINHLAT--TLGRpPRPGPPG-----GPRTPOI-- 89

      mGRRAGAElePrLCPGRRClAaAAsalAPrGrsvr<-*
      R   A ++P+  PG R   As   G+ +
rheu.ef.23   90   --RNLPALPAPQGE PGDRATWRGASGADAAGGDGG      122

```

15

```

#=GF ID      Orexina
#=GF AC      PF02072.7
#=GF DE      Prepro-orexina
#=GF AU      Mian N, Bateman A
#=GF SE      IPR001704

```

20

```

#=GF TP      Familia
OREX HUMAN/1-131
MNL PSTKVSWA AVT L L L L L L L L P P A L L S S G A A A Q P L P D C C R Q K T C S C R L Y E L L H G A G N H A A G I L T L G K R R S G P P G L Q G R L Q R L L
A S G N H A A G I L T M G R R A G A E P A P R P C L G R R C S A P A A A S V A P G G Q S G I

```

RECEPTOR GIP

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine

Distribuido libremente bajo la licencia pública general GNU (GPL)

25

```

-----
Archivo HMM:                prints.hmm

```

Archivo de secuencia: **rheu.ef.241.148**
GIPRECEPTOR_7: dominio 1 de 1, de 76 a 97: puntuación 7,9, E = 3,7
 ->Pr1GPy1GdqtlLwnq.ALAA<-

rheu.ef.24 76 PRPGPPGGPRTpQIRNLpALPA 97
 Pr+GP G +t+ ++n +AL A

rheu.ef.238rev

5 **GIPRECEPTOR_7**: dominio 1 de 1, de 76 a 97: puntuación 7,9, E = 3,7
 ->Pr1GPy1GdqtlLwnq.ALAA<-

rheu.ef.23 76 PRPGPPGGPRTpQIRNLpALPA 97
 Pr+GP G +t+ ++n +AL A

GIPRECEPTOR

- gx; PR01129
- gn; COMPUESTO(11)
- ga; 22-MAY-1999
- gt; **Firma de precursor de receptor de polipéptido inhibidor gástrico**
- gp; PRINTS; [PR00237](#) GPCRRHODOPSN; [PR00247](#) GPCRCAMP; [PR00248](#) GPCRMGR
- gp; PRINTS; [PR00249](#) GPCRSECRETIN; [PR00250](#) GPCRSTE2; [PR00899](#) GPCRSTE3
- gp; PRINTS; [PR00251](#) BACTRLOPSIN
- gp; INTERPRO; [IPR001749](#)
- gr; 1. ATTWOOD, T.K. Y FINDLAY, J.B.C.
- gr; Fingerprinting G protein-coupled receptors.
- gr; PROTEIN ENG. 7(2) 195-203 (1994).
- gr; 2. ISHIHARA T., NAKAMURA S., KAZIRO, Y., TAKAHASHI, T., TAKAHASHI, K. Y
- gr; NAGATA, S.
- gr; Molecular cloning and expression of a cDNA encoding the secretin receptor
- gr; EMBO J. 10 1635-1641 (1991).
- gr; 3. LIN, H.Y., HARRIS, T.L., FLANNERY, M.S., ARUFFO, A., KAJI, E.H.,
- gr; GORN, A., KOLAKOWSKI, L.F., LODISH, H.F. Y GOLDRING, S.R.
- gr; Expression cloning of adenylate cyclase-coupled calcitonin receptor
- gr; SCIENCE 254 1022-1024 (1991).
- gr; 4. JUEPPNER, H., ABOU-SAMRA, A.-B., FREEMAN, M., KONG, X.F.,

gr; SCHIPANI, E., RICHARDS, J., KOLALOWSKI, L.F., HOCK, J., POTTS, J.T., KRONENBERG, H.M. Y
gr; SEGRE, G.E.
gr; A G protein linked receptor for parathyroid hormone and parathyroid
gr; hormone-related peptide.
gr; SCIENCE 254 1024-1026 (1991).
gr; 5. ISHIHARA, T., SHIGEMOTO, R., MORI, K., TAKAHASHI, K. Y NAGATA, S.
gr; Functional expression and tissue distribution of a novel receptor for vasoactive intestinal polypeptide.
gr; NEURON 8(4) 811-819 (1992).
gr; 6. VOLZ, A., GOKE, R., LANKAT-BUTTGEREIT, B., FEHMANN, H.C., BODE, H.P. Y
gr; GOKE, B.
gr; Molecular cloning, functional expression, and signal transduction of the GIP-receptor cloned from a human
gr; insulinoma.
gr; FEBS LETT. 373(1) 23-9 (1995).
gr; Los receptores acoplados a proteína G (GPCR) constituyen una vasta familia de
gd; proteínas que abarca una gran variedad de funciones (incluyendo diversos
gd; procesos autocrinos, paracrinos y endocrinos). Muestran una considerable
gd; diversidad a nivel de secuencia, que puede usarse como base para separarlas en
gd; distintos grupos. Se usa el término clan para describir las GPCR, ya que abarcan
gd; un grupo de familias para las que hay indicaciones de relación evolutiva, pero entre
gd; las cuales no hay una similitud estadísticamente significativa en su secuencia [1].
gd; Los miembros del clan conocidos en la actualidad incluyen las GPCR similares a
gd; rodopsina, las GPCR similares a secretina, los receptores de AMPc, los receptores
gd; de feromonas de emparejamiento fúngico y la familia de receptor de glutamato
gd; metabotrópico.
gd; Los GPCR similares a secretina incluyen secretina [2], calcitonina [3], hormona
gd; paratiroidea/péptidos relacionados con la hormona paratiroidea [4] y péptido
gd; intestinal vasoactivo [5], todos los cuales activan la adenililo ciclase y la vía de
gd; fosfatidilo-inositol-calcio. Las secuencias de aminoácidos de los receptores
gd; contienen altas proporciones de restos hidrófobos agrupados en 7 dominios, de un
gd; modo que recuerda a las rodopsinas y otros receptores que se cree que interactúan
gd; con las proteínas G. Sin embargo, aunque se ha propuesto que esto se debe a una
gd; estructura en 3D similar, no hay una similitud de secuencia significativa entre estas
gd; familias: por lo tanto, los receptores similares a secretina portan su firma de "7TM"
gd; única.
gd; El polipéptido insulínico dependiente de glucosa (GIP) desempeñan un papel
gd; importante en la regulación de la secreción de insulina postprandial y la expresión del
gd; gen de proinsulina de las células beta pancreáticas [6]. El receptor de GIP humano
gd; codifica una proteína de 7TM que es similar al receptor del péptido 1 similar a glucagón
gd; humano (GLP-1). Se espera que la comprensión de la regulación del receptor de GIP y
gd; de la transducción de señales arroje luz sobre la incapacidad de la hormona para
gd; ejercer su acción biológica en las células B pancreáticas en la diabetes mellitus de tipo
gd; II. GIPRECEPTOR es una huella de 11 elementos que proporciona una firma para
gd; receptores de polipéptido inhibidor gástrico. La huella se obtuvo de un alineamiento
gd; inicial de 3 secuencias: los motivos se representaron a partir de regiones conservadas
gd; que abarcan la longitud completa del alineamiento, centrándose en aquellas secciones
gd; que caracterizan los receptores de polipéptido inhibidor gástrico, pero lo distinguen del
gd; resto de la superfamilia similar a secretina - los motivos 1-6 abarcan el dominio N-
gd; terminal; el motivo 7 se encuentra en el bucle entre los dominios 2 y 3 TM; el motivo 8
gd; abarca el bucle entre los dominios 3 y 4 TM; el motivo 9 abarca la parte C-terminal del
gd; dominio 6 TM y
gd; el bucle entre los dominios 4 y 5 TM; y los motivos 10 y 11 se encuentran en el extremo C-
gd; terminal. Para alcanzar la convergencia se necesitó una sola interacción en SPTR37 9f, sin
gd; identificar secuencias adicionales más allá del conjunto de inicio. También se encontraron
gd; dos coincidencias parciales, receptores de secretina y glucagón, que coinciden con los
gd; motivos 1, 8 y 9.
bb;
fc; GIPRECEPTOR7
fl; 21
ft; Motivo VII de precursor de receptor de polipéptido inhibidor gástrico - 1
fd; PTLGPYPGDRTLTLRNQALAA GIPR_MESAU 192 56
fd; PPLGPYTGNTPTLWNQALAA GIPR_RAT 192 56 56
fd; PRPGPYLGDQALALWNQALAA GIPR_HUMAN 195

PRION

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine

Distribuido libremente bajo la licencia pública general GNU (GPL)

Archivo HMM: prints.hmm
 Archivo de secuencia: rheu.ef.241.148

PRION_2: dominio 1 de 1, de 68 a 89: puntuación 5,4, E = 8,6

->snggggsrypgqGSPGGNRYPpq<-
 + r+p +G PGG R P

rheu.ef.24 68 LATTLGRPPRPFGPPGGPRT PQI 89

5

rheu.ef.238rev.148

PRION_2: dominio 1 de 1, de 68 a 89: puntuación 5,4, E = 8,6

->snggggsrypgqGSPGGNRYPpq<-
 + r+p +G PGG R P

rheu.ef.23 68 LATTLGRPPRPFGPPGGPRT PQI 89

gc; **PRION**

gx; PR00341

gn; COMPUESTO(8)

ga; 19-OCT-1992; ACTUALIZADO 07-JUN-1999

gt; **Firma de proteína prion**

gp; INTERPRO; IPR000817

gp; PROSITE; PS00291 PRION_1; PS00706 PRION_2

gp; PFAM; PF00377 prion

gr; 1. STAHL, N. Y PRUSINER, S.B.

gr; Prions and prion proteins.

gr; FASEB J. 5 2799-2807 (1991).

gr; 2. BRUNORI, M., CHIARA SILVESTRINI, M. Y POCCHIARI, M.

gr; The scrapie agent and the prion hypothesis.

gr; TRENDS BIOCHEM.SCI. 13 309-313 (1988).

gr; 3. PRUSINER, S.B.

gr; Scrapie prions.

gr; ANNU.REV.MICROBIOL. 43 345-374 (1989).

gd; La proteína prion (PrP) es una glucoproteína pequeña encontrada en alta cantidad en los cerebros de animales infectados con ciertas enfermedades neurológicas degenerativas, tales como el tembleque de las ovejas y la encefalopatía espongiforme bovina (BSE) y las demencias humanas, la enfermedad de Creutzfeldt-Jacob (CJD) y el síndrome de Gerstmann-Straussler (GSS). PrP se codifica en el genoma del hospedador y se expresa en células tanto normales como infectadas. Durante la infección, sin embargo, las moléculas de PrP se alteran y polimerizan, produciendo fibrillas de proteína PrP modificada. Se han encontrado moléculas de PrP en la superficie externa de las membranas plasmáticas de células nerviosas, a las que se anclan a través de un glucolípido unido covalentemente, lo que sugiere un papel como receptor de membrana. PrP también se expresa en otros tejidos, lo que indica que puede tener diferentes funciones, dependiendo de su ubicación.

gd; Las secuencias primarias de las PrP de diferentes fuentes son altamente similares: todas portan un dominio N-terminal que contiene múltiples repeticiones en tándem de un octapéptido rico en Pro/Gly; sitios de glucosilación unidos a Asn; un enlace disulfuro esencial; y 3 segmentos hidrófobos. Estas secuencias muestran cierta similitud con una glucoproteína de pollo, que se cree que es una molécula inductora de actividad de receptor de acetilcolina (ARIA). Se ha sugerido que los cambios en la región de repetición del octapéptido pueden indicar una predisposición a enfermedades, pero no se sabe a ciencia cierta si la repetición puede usarse de manera significativa como huella para indicar susceptibilidad.

gd; PRION es una huella de 8 elementos que proporciona una firma para las proteínas prion. La huella se obtuvo de un alineamiento inicial de 5 secuencias: los motivos se representaron a partir de regiones conservadas que abarcan prácticamente la longitud completa del alineamiento, incluyendo los 3 dominios hidrófobos y las repeticiones de octapéptido (WGQPHGGG). Para alcanzar la convergencia, se necesitaron dos interacciones en OWL18.0, punto en el cual se identificó un conjunto auténtico que comprendía 9 secuencias. También se encontraron varias coincidencias parciales: estas incluyen un fragmento (PRIO_RAT) que carece de parte de la secuencia que porta el primer motivo y el homólogo de PrP encontrado en pollos - este se empareja bien con solo 2 de los 3 motivos hidrófobos (1 y 5) y una de las otras regiones conservadas (6), pero tiene una firma N-terminal basada en una repetición de sextapéptido (YPHNPG) en lugar del octapéptido característico de PrP.

c;	PRION2			
fl;	22			
ft;	Motivo II de proteína prion - 2			
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_COLGU	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_MACFA	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_CEREL	34	9
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_ODOHE	34	9
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_GORGO	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_PANTR	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_HUMAN	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	046648	34	9
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_SHEEP	34	9
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_CALJA	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_BOVIN	34	9
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRP2_BOVIN	34	9
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_ATEPA	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_SAISC	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_PREFR	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_PONPY	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	075942	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_CAPHI	34	9
fd;	WNTGGSRYPGQGSPGGNLYPPQ	PRIO_CEBAP	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_CAMDR	34	9
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_FELCA	34	9
fd;	WNTGGSRYPGQGSPGGNRYPSQ	PRP1_TRAST	34	9
fd;	WNTGGSRYPGQSSPGGNRYPPQ	PRIO_RABIT	32	9
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRP2_TRAST	34	9
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_PIG	34	9
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_CANFA	34	9
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_CRIGR	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_CRIMI	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	Q15216	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_RAT	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_CERAE	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_MUSPF	34	9
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_MUSVI	34	9
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_MESAU	31	8
fd;	WNTGGSRYPGQGSPGGNRYPPQ	PRIO_MOUSE	31	8
fd;	NTGGSRYPGQGSPGGNRYPPQ	046593	34	9
fd;	SGGSNRYPGQPGSPGGNRYPGW	PRIO_TRIVU	37	12
bb;				

NEUROTENSINA

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine

Distribuido libremente bajo la licencia pública general GNU (GPL)

5 Archivo HMM: prints.hmm

Archivo de secuencia: **rheu.ef.241.148**

NEUROTENSIN2R_1: dominio 1 de 1, de 68 a 80: puntuación 6,8, E = 8,7

->mEtsspWPPRPsp<-

+ t +PPRP p

rheu.ef.24 68 LATTTLGRPPRPGP 80

rheu.ef.238rev.148

10 **NEUROTENSIN2R_1**: dominio 1 de 1, de 68 a 80: puntuación 6,8, E = 8,7

->mEtsspWPPRPsp<-

+ t +PPRP p

rheu.ef.23 68 LATTTLGRPPRPGP 80

c; **NEUROTENSIN2R**
 gx; PR01481
 gn; COMPUESTO(6)
 ga; 12-MAR-2001
 gt; **Firma de receptor de tipo 2 de neurotensina**
 gp; PRINTS; PR00237 GPCRRHODOPSN; PR00247 GPCRCAMP; PR00248 GPCEIMGR
 gp; PRINTS; PR00249 GPCRSECRETIN; PR00250 GPCRSTE2; PR00899 GPCRSTE3
 gp; PRINTS; PR00251 BACTRLOPSIN
 gp; PRINTS; PR01479 NEUROTENSINR; PR01480 NEUROTENSIN1R
 gr; 1. ATTWOOD, T.K. Y FINDLAY, J.B.C.
 gr; Fingerprinting G protein-coupled receptors.
 gr; PROTEIN ENG. 7(2) 195-203 (1994).
 gr; 2. ATTWOOD, T.K. Y FINDLAY, J.B.C.
 gr; G protein-coupled receptor fingerprints.
 gr; 7TM, VOLUMEN 2, EDS. G.VRIEND Y B.BYWATER (1993).
 gr; 3. BIRNBAUMER, L.
 gr; G proteins in signal transduction.
 gr; ANNU.REV.PHARMACOL.TOXICOL. 30 675-705 (1990).
 gr; 4. CASEY, P.J. Y GILMAN, A.G.
 gr; G protein involvement in receptor-effector coupling.
 gr; J.BIOL.CHEM. 263(6) 2577-2580 (1988).
 gr; 5. ATTWOOD, T.K. Y FINDLAY, J.B.C.
 gr; Design of a discriminating fingerprint for G protein-coupled receptors.
 gr; PROTEIN ENG. 6(2) 167-176 (1993).
 gr; 6. WATSON, S. Y ARKINSTALL, S.
 gr; Neurotensin.
 gr; EN THE G PROTEIN-LINKED RECEPTOR FACTSBOOK, ACADEMIC PRESS, 1994, PP.199-201.
 gr; 7. VINCENT, J-P., MAZELLA, J. Y KITABGI, P.
 gr; Neurotensin and neurotensin receptors.
 gr; TRENDS PHARMACOL.SCI. 20(7) 302-309 (1999).
 gr; 8. VITA, N., OURY-DONAT, F., CHALON, P., GUILLEMOT, M., KAGHAD, M., BACHY,
 gr; A., THURNEYSSSEN, O., GARCIA, S., POINOT-CHAZEL, C., CASELLAS, P., KEANE,
 gr; P., LE FUR, G., MAFFRAND, J.P., SOUBRIE, P., CAPUT, D. Y FERRARA, P.
 gr; Neurotensin is an antagonist of the human neurotensin NT2 receptor expressed in Chinese
 gr; hamster ovary cells.
 gr; EUR.J.PHARMACOL. 360(2-3) 265-272 (1998).
 gr; 9. YAMADA, M., YAMADA, M., LOMBET, A., FORGEZ, P. Y ROSTENE, W.
 gr; Distinct functional characteristics of levocabastine sensitive rat neurotensin NT2 receptor expressed in
 gr; Chinese hamster ovary cells.
 gr; LIFE SCI. 62(23) PL 375-380 (1998).

gd; Los receptores acoplados a proteína G (GPCR) constituyen una vasta familia de
 gd; proteínas que abarca una gran variedad de funciones (incluyendo diversos
 gd; procesos autocrinos, paracrinos y endocrinos). Muestran una considerable
 gd; diversidad a nivel de secuencia, que puede usarse como base para separarlas en
 gd; distintos grupos. Se usa el término clan para describir las GPCR, ya que abarcan
 gd; un grupo de familias para las que hay indicaciones de relación evolutiva, pero
 gd; entre las cuales no hay una similitud estadísticamente significativa en su secuencia
 gd; [1,2]. Los miembros del clan conocidos en la actualidad incluyen las GPCR
 gd; similares a rodopsina, las GPCR similares a secretina, los receptores de AMPc, los
 gd; receptores de feromonas de emparejamiento fúngico y la familia de receptor de
 gd; glutamato metabotrópico.

gd; Las GPCR similares a rodopsina representan por sí mismas una familia de
 gd; proteínas ampliamente diseminada que incluye receptores de hormonas,
 gd; neurotransmisores y de luz, todos los cuales transducen señales extracelulares
 gd; mediante su interacción con proteínas de unión a nucleótidos guanina (G). Aunque
 gd; sus ligandos activadores varían ampliamente en su estructura y carácter, las
 gd; secuencias de aminoácidos de los receptores son muy similares y se cree que
 gd; adoptan un armazón estructural común que comprende 7 hélices transmembrana
 gd; (TM) [3-5].

gd; La neurotensina es un transmisor peptídico de 13 restos, que comparte una
 gd; similitud significativa en sus 6 aminoácidos C-terminales con varios neuropéptidos
 gd; diferentes, incluyendo la neuromedina N. Esta región es responsable de la
 gd; actividad biológica, teniendo la porción N-terminal un papel modulador. La
 gd; neurotensina se distribuye por el sistema nervioso central, con niveles máximos en
 gd; el hipotálamo, la amígdala y el núcleo accumbens. Induce una serie de efectos,
 gd; incluyendo: analgesia, hipotermia y aumento de la actividad locomotora. También
 gd; está implicada en la regulación de las vías de dopamina. En la periferia, la
 gd; neurotensina se encuentra en células endocrinas del intestino delgado, donde
 gd; ocasiona secreción y contracción del músculo liso [6].

gd; La existencia de 2 subtipos de receptor de neurotensina, con diferentes afinidades
 gd; por la neurotensina y diferentes sensibilidades al antihistamínico levocabastina, se
 gd; demostró originariamente mediante estudios de unión en el cerebro de roedores.
 gd; Desde entonces, se han clonado dos receptores de neurotensina (NT1 y NT2) con
 gd; dichas propiedades y se ha descubierto que son miembros de la familia de
 gd; receptores acoplados a proteína G [7].

gd; El receptor NT2 se clonó de cerebros de rata, ratón y ser humano basándose en
 gd; su similitud con el receptor NT1. Se observó que el receptor es un receptor de baja
 gd; afinidad sensible a la levocabastina para neurotensina. A diferencia del receptor de
 gd; alta afinidad, NT1, NT2 es insensible al trifosfato de guanosa y tiene baja
 gd; sensibilidad por los iones sodio [7]. Los mayores niveles de expresión del receptor
 gd; se encuentran en el cerebro, en regiones que incluyen: el sistema olfatorio, las
 gd; cortezas cerebral y cerebelar, el hipocampo y el núcleo hipotalámico. La
 gd; distribución es distinta a la del receptor NT1, expresándose ambos subtipos solo
 gd; en unas pocas áreas (banda diagonal de Broca, núcleo septal medial y núcleo
 gd; supraquiasmático) [7]. También se ha encontrado el receptor a niveles menores en
 gd; el riñón, útero, corazón y pulmón [8]. La activación del receptor NT2 por agonistas
 gd; no peptídicos sugiere que el receptor puede acoplarse a fosfolipasa C, fosfolipasa
 gd; A2 y MAP cinasa. Sin embargo, una respuesta funcional a la neurotensina es débil
 gd; [9] o está ausente y parece ser que la neurotensina actúa como antagonista del
 gd; receptor [8]. También se ha sugerido que una sustancia distinta de la neurotensina
 gd; puede actuar como ligando natural para este receptor [8].

gd; NEUROTENSIN2R es una huella de 6 elementos que proporciona una firma para
 gd; los receptores de tipo 2 de neurotensina. La huella se obtuvo de un alineamiento
 gd; inicial de 3 secuencias: los motivos se representaron a partir de secciones
 gd; conservadas dentro del extremo N-terminal y las regiones de bucle, centrándose
 gd; en aquellas áreas del alineamiento que caracterizan los receptores de tipo 2 de
 gd; neurotensina, pero que los distinguen del resto de la familia de receptores de
 gd; neurotensina - los motivos 1 y 2 abarcan el extremo N-terminal; los motivos 3 y 4
 gd; abarcan el segundo bucle externo; y los motivos 5 y 6 abarcan el tercer bucle
 gd; citoplásmico. Para alcanzar la convergencia se necesitó una sola interacción en
 bb; SPTR39 15f, sin identificar secuencias adicionales más allá del conjunto de inicio.

fc; NEUROTENSIN2R1

fl; 13

ft; Motivo I de receptor de tipo 2 de neurotensina - 1

fd; METSSPWPPRPSP	NTR2_RAT	1	1
fd; METSSLWPPRPSP	NTR2_MOUSE	1	1
fd; METSSPRPPRPSS	NTR2_HUMAN	1	1

FIRMA DE LA FAMILIA DE RECEPTOR NUCLEAR HUÉRFANO (RECEPTOR NUCLEAR A4)

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine Distribuido libremente bajo la licencia pública general GNU (GPL)

5

Archivo HMM: prints.hmm

Archivo de secuencia: uro742rev.1,780

NUCLEARECPTR_5: dominio 1 de 1, de 326 a 341: puntuación 7,2, E = 5

->PvnLlnaLVRAhvDStP<-

· + + n++VRAh+D+

uro742rev. 326 -TFITNSMVRAHIDADK 341

gc; NUCLEARECPTR

gx; PRO 128 4

gn; COMPUESTO(11)

ga; 16-FEB-2000

gt; Firma de la familia de receptor nuclear huérfano (receptor nuclear A4)

gp; PRINTS; [PR00398](#) STRDHORMONER; [PR00047](#) STROIDFINGERgp; PRINTS; [PR01285](#) HMRNUCRECPTR; [PR01286](#) NORNUCRECPTR; [PR01287](#) NURRNUCRCPTR

gr; 1. NUCLEAR RECEPTORS NOMENCLATURE COMMITTEE

gr; A unified nomenclature system for the nuclear receptor superfamily.

gr; CELL 97 161-163 (1999).

gr; 2. NISHIKAWA, J-I., KITAURA, M., IMAGAWA, M. Y NISHIHARA, T.

gr; Vitamin D receptor contains multiple dimerisation interfaces that are functionally different.

gr; NUCLEIC ACIDS RES. 23(4) 606-611 (1995).

gr; 3. DE VOS, P., SCHMITT, J., VERHOEVEN, G. Y STUNNENBERG, G.

gr; Human androgen receptor expressed in HeLa cells activates transcription *in vitro*.

gr; NUCLEIC ACIDS RES. 22(7) 1161-1166 (1994).

gr; 4. OHKURA, N., HIJIKURO, M., YAMAMOTO, A. Y MIKI, K.

gr; Molecular cloning of a novel thyroid/steroid receptor superfamily gene from cultured rat neuronal cells.

gr; BIOCHEM.BIOPHYS.RES.COMMUN. 205 1959-1965 (1994).

gr; 5. LAW, S.W., CONNEELY, O.M., DEMAYO, F.J. Y O'MALLEY, B.W.

gr; Identification of a **new brain-specific transcription factor, NURR1**.

gr; MOL.ENDOCRINOL. 2129-2135 (1992).

gr; 6. WILSON, T.E., PAULSEN, R.E., PADGETT, K.A. Y MILBRANDT, J.

gr; Participation of non-zinc finger residues in DNA binding by two nuclear orphan receptors.

gr; SCIENCE 256 107-110 (1992).

gr; 7. CLARK, J., BENJAMIN, H., GILL, S., SIDHAR, S., GOODWIN, G., CREW, J.,

gr; GUSTERSON, B.A., SHIPLEY, J. Y COOPER, C.S.

gr; Fusion of the EWS gene to CHN, a member of the steroid/thyroid receptor

gr; gene superfamily, in a human myxoid chondrosarcoma.

gr; ONCOGENE 12 229-235 (1996).

gd; Los receptores de esteroides o de hormonas nucleares (NR) forman una superfamilia

gd; importante de reguladores de la transcripción que están implicados en funciones fisiológicas

gd; ampliamente diversas, incluyendo el control del desarrollo embrionario, la diferenciación celular

gd; y la homeostasia [1]. Los miembros de la superfamilia incluyen los receptores de hormonas

gd; esteroideas y los receptores para hormonas tiroideas, retinoides,

gd; 1,25-dihidroxi-vitamina D3 y varios ligandos diferentes. Las proteínas funcionan como

gd; moléculas diméricas en los núcleos para regular la transcripción de genes diana de una

gd; manera sensible al ligando [2,3]. Además de los dominios de unión a ligando C-terminales,

gd; estos receptores nucleares contienen un dedo de cinc N-terminal altamente conservado que

gd; media la unión específica a secuencias de ADN diana, denominados elementos sensibles a

gd; ligando. En ausencia de ligando, se cree que los recetores de hormonas esteroideas se

gd; asocian débilmente con componentes nucleares; la unión a hormonas aumenta en gran medida

gd; la afinidad del receptor.

gd; Los NR son extremadamente importantes en investigación médica, estando implicado un gran

gd; número de estos en enfermedades tales como cáncer, diabetes, síndromes de resistencia a

gd; hormonas, etc. [1]. Aunque varios NR actúan como factores de transcripción inducibles por

gd; ligando, muchos aún no tienen un ligando definido y por consiguiente, se les denomina
 gd; receptores "huérfanos". Durante la última década, se han descrito más de 300 NR, muchos de
 gd; los cuales se encuentran huérfanos, que no pueden nombrarse fácilmente debido a las
 gd; confusiones de nomenclatura actual en la bibliografía. Sin embargo, recientemente se ha
 gd; introducido un nuevo sistema en un intento de racionalizar el conjunto de nombres cada vez
 gd; más complejo usado para describir a miembros de superfamilias [1].
 gd; **Los nuevos miembros de la superfamilia de receptores de esteroides, denominados**
 gd; **NOR-1 (receptor huérfano derivado de neuronas) [4], Nurr1 (factor 1 relacionado con Nur)**
 gd; [5] y NGFI-B [6] se han identificado a partir de células neuronales del cerebro anterior que
 gd; sufren apoptosis, de la corteza cerebral y de pulmón, ganglios cervicales superiores y tejido
 gd; suprarrenal, respectivamente. La proteína NOR-1 se une al elemento de respuesta a B1a, que
 gd; se ha identificado como la secuencia diana de la familia Nur77, lo que sugiere que tres
 gd; miembros de la familia Nur77 pueden transactivar genes diana comunes en diferentes
 gd; situaciones [4]. El sarcoma de Ewing se caracteriza por traslocaciones cromosómicas que
 gd; implican a la proteína NOR [7].
 gd; **NUCLEARECPTR es una huella de 11 elementos que proporciona una firma para la**
 gd; **familia de receptores nucleares huérfanos.** La huella se obtuvo de un alineamiento inicial de
 gd; 11 secuencias: los motivos se representaron a partir de regiones conservadas que abarcan
 gd; prácticamente la longitud completa del alineamiento, centrándose en aquellas secciones que
 gd; caracterizan a miembros de la familia de receptores nucleares, pero que los distingue del resto
 gd; de la superfamilia de receptores de hormonas esteroideas - los motivos 1-3 se encuentran en
 gd; N-terminal respecto del dominio de dedo de cinc; los motivos 4 y 5 se encuentran entre los
 gd; dedos de cinc y el supuesto dominio de unión a ligando; los motivos 6 y 7 codifican los
 gd; extremos N y C-terminales del dominio de unión a ligando; y los motivos 8-11 se encuentran en
 gd; el extremo C-terminal. Para alcanzar la convergencia se necesitó una sola interacción en
 gd; SPTR37_10f, sin identificar secuencias adicionales más allá del conjunto de inicio. Se
 gd; encontraron varias coincidencias parciales, pareciendo ser todas ellas homólogos truncados en
 gd; N o C-terminal.
 fc; NUCLEARECPTR5
 fl; 17
 ft; Motivo V de la familia de receptor nuclear huérfano - 1

fd;	PANLLTSLVRAHLDSGP	NR41_HUMAN	361	6
fd;	PANLLTSLVRAHLDSGP	NR41_CANFA	361	6
fd;	PVSLISALVRAHVDSNP	NR42_RAT	361	10
fd;	PVSLISALVRAHVDSNP	R4_2_MOUSE	361	10
fd;	PVSLISALVRAHVDSNP	NR4_2-HUMAN	361	10
fd;	PTNLLTSLIRAHLDSPG	NR41_RAT	360	6
fd;	PTNLLTSLIRAHLDSPG	NR41_MOUSE	364	6
fd;	PVDLINSLVRAHIDSIP	NR42_XENLA	340	6
fd;	PVCMMNALVRALTDSTP	097726	412	15
fd;	PICMMNALVRALTDSTP	NR43_HUMAN	395	15
fd;	PICMMNALVRALTDATP	NR4_3_RAT	397	15

FACTOR NEUROTRÓFICO DERIVADO DE CEREBRO (BDN)

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine Distribuido libremente bajo la licencia pública general GNU (GPL)

5

Archivo HMM: prints.hmm

Archivo de secuencia: **uro742rev.1.780**

BDNFACTOR_3: dominio 1 de 2, de 496 a 512: puntuación 3,1, E = 42

```
*->PLLFLLEEYKnYLDAAAn<-*
      PL LL Y YL+
```

```
uro742rev. 496 PLWALLNGYVDYLETQI 512
```

BDNFACTOR_3: dominio 2 de 2, de 690 a 706: puntuación 7,7, E = 5,7

```
*->PLLFLLEEYKnYLDAAAn<-*
      PLLFL EY+ AA
```

```
uro742rev. 690 PLLFLPSEYQREDGAAE 706
```

10

gc; **BDNFACTOR**
 gx; PR01912
 gn; COMPUESTO(5)
 ga; 29-AUG-2008
 gt; **Firma de factor neurotrófico derivado de cerebro**
 gp; PRINTS; [PR00268](#) NGF; [PR01913](#) NGFBETA; [PR01914](#) NEUROTROPHN3
 gp; PRINTS; [PR01915](#) NEUROTROPHN4; [PR01916](#) NEUROTROPHN6
 gp; PDB; [1BND](#); [1B8M](#)
 gp; SCOP; [1BND](#); [1B8M](#)
 gp; CATH; [1BND](#); [1B8M](#)
 gp; MIM; 113505
 gr; 1. HOFER, M., PAGLIUSI, S.R., HOHN, A., LEIBROCK, J. Y BARDE, Y.A.
 gr; Regional distribution of brain-derived neurotrophic factor messenger RNA
 gr; in the adult mouse brain.
 gr; EMBO J. 9(8) 2459-2464 (1990).
 gr; 2. KOYAMA, J.I., INOUE, S., IKEDA, K. Y HAYASHI, K.
 gr; Purification and amino acid sequence of a nerve growth factor from the venom
 gr; of *Vipera russelli russelli*.
 gr; BIOCHIM.BIOPHYS.ACTA 1160 287-292 (1992).
 gr; 3. INOUE, S., ODA, T., KOYAMA, J., IKEDA, K. Y HAYASHI, K.
 gr; Amino acid sequences of nerve growth factors derived from cobra venoms.
 gr; FEBS LETT. 279(1) 38-40 (1991).
 gr; 4. BARDE, Y., EDGAR, D. Y THOENEN, H.
 gr; Purification of a new neurotrophic factor from mammalian brain.
 gr; EMBO J. 1 549-553 (1982).
 gr; 5. HIBBERT, A., KRAMER, B., MILLER, F. Y KAPLAN, D.
 gr; The localization, trafficking and retrograde transport of BDNF bound to p75NTR
 gr; in sympathetic neurons.
 gr; MOL.CELL.NEUROSCI. 32 387-402 (2006).
 gr; 6. LINNARSSON, S., BJORKLUND, A. Y ERNFORS, P.
 gr; Learning deficit in BDNF mutant mice.
 gr; EUR.J.NEUROSCI. 9 2581-2587 (1997).
 gr; 7. LEBRUN, B., BARIOHAY, B., MOYSE, E. Y JEAN, A.
 gr; Brain-derived neurotrophic factor (BDNF) and food intake
 gr; regulation: a minireview.
 gr; AUTON.NEUROSCI. 126-127 30-38 (2006).
 gr; 8. KOZISEK, M., MIDDLEMAS, D. Y BYLUND, D.

gr; Brain-derived neurotrophic factor and its receptor tropomyosin-related kinase B in the mechanism of
 gr; action of antidepressant therapies.
 gr; PHARMACOL.THER. 117 30-51 (2008).
 gd; Durante el desarrollo del sistema nervioso de los vertebrados, muchas neuronas se vuelven
 gd; redundantes (debido a que han muerto, no han logrado conectarse a células diana, etc.) y se
 gd; eliminan. Al mismo tiempo, las neuronas en desarrollo proyectan recrecimientos axónicos que
 gd; entran en contacto con sus células diana [1]. Dichas células controlan su grado de innervación
 gd; (el número de conexiones axónicas) mediante la secreción de diversos factores neurotróficos
 gd; que son esenciales para la supervivencia neuronal. Uno de estos es el factor de crecimiento
 gd; nervioso (NGF), que está implicado en la supervivencia de ciertas clases de neuronas
 gd; embrionarias (por ejemplo, neuronas simpáticas periféricas) [1]. NGF se encuentra
 gd; principalmente fuera del sistema nervioso central (SNC), pero pueden encontrarse pequeñas
 gd; cantidades residuales en tejidos del SNC de adultos, aunque se desconoce el papel fisiológico
 gd; de esto [1]; también se ha encontrado en varios venenos de serpiente [2,3]. Las proteínas
 gd; similares a NGF incluyen el factor neurotrófico derivado de cerebro (BDNF) y las neurotrofinas
 gd; 3 a 7, todas las cuales han demostrado actividades de supervivencia y recrecimiento
 gd; neuronal.
 gd; Originariamente purificado a partir de cerebro de cerdo [4], la neurotrofina BDNF se expresa
 gd; en una serie de tejidos y tipos celulares en el SNC y la periferia. Ejerce sus efectos uniéndose
 gd; al receptor de tirosina cinasa neurotrófico de tipo 2 (NTRK2; también denominado TrkB) y al
 gd; receptor de factor de crecimiento nervioso de baja afinidad, p75NTR. Aunque este último
 gd; receptor media las funciones promotoras de la supervivencia de las neurotrofinas, se ha
 gd; demostrado que la activación de p75NTR por BDNF promueve la apoptosis e inhibe el
 gd; crecimiento axonal [5].
 gd; El BDNF es un regulador clave de la plasticidad sináptica y desempeña un papel importante
 gd; en el aprendizaje y la memoria [6]. Diversos indicios sugieren que también está implicado en
 gd; el control de la ingesta de alimento y el peso corporal [7]. Una serie
 gd; de estudios clínicos ha demostrado una asociación entre los niveles aberrantes de BDNF y
 gd; trastornos y patologías, tales como depresión, epilepsia, trastorno bipolar, enfermedad de
 gd; Parkinson y enfermedad de Alzheimer [8].
 gd; BDNFACTOR es una huella de 5 elementos que proporciona una firma para el factor
 gd; neurotrófico derivado de cerebro. La huella se obtuvo de un alineamiento inicial de 33
 gd; secuencias: los motivos se representaron a partir de regiones conservadas que abarcan
 gd; prácticamente la longitud completa del alineamiento - el motivo 1 incluye parte de la secuencia
 gd; de señal. Para alcanzar la convergencia, se necesitaron tres interacciones en SPTR55 38f,
 gd; punto en el cual se identificó un conjunto auténtico que comprendía 47 secuencias. También
 gd; se encontró una sola coincidencia parcial de humano, Q6YNR1 HUMAN, una variante de
 gd; corte y empalme de BDNF humano que no coincide con los motivos 4 y 5.
 fc; BDNFACTOR3
 fl; 17
 ft; Motivo III de factor de crecimiento neurotrófico derivado de cerebro - 3

fd; PLLFLLEEYKNYLDAAN	A2AII2_MOUSE	115	31
fd; PLLFLLEEYKNYLDAAN	Q8CCH9_MOUSE	107	31
fd; PLLFLLEEYKNYLDAAN	Q6YNR3_HUMAN	113	31
fd; PLLFLLEEYKNYLDAAN	Q6YNR2_HUMAN	120	31
fd; PLLFLLEEYKNYLDAAN	Q598Q1_HUMAN	105	31
fd; PLLFLLEEYKNYLDAAN	Q541P3_MOUSE	107	31
fd; PLLFLLEEYKNYLDAAN	BDNF_URSML	105	31
fd; PLLFLLEEYKNYLDAAN	BDNF_URSAR	105	31
fd; PLLFLLEEYKNYLDAAN	BDNF_SPECI	105	31
fd; PLLFLLEEYKNYLDAAN	BDNF_SELTH	105	31
fd; PLLFLLEEYKNYLDAAN	BDNF_RAT	107	31
fd; PLLFLLEEYKNYLDAAN	BDNF_PROLO	105	31
fd; PLLFLLEEYKNYLDAAN	BDNF_PIG	110	31
fd; PLLFLLEEYKNYLDAAN	BDNF_PANTR	105	31
fd; PLLFLLEEYKNYLDAAN	BDNF_MOUSE	107	31
fd; PLLFLLEEYKNYLDAAN	BDNF_HUMAN	105	31
fd; PLLFLLEEYKNYLDAAN	BDNF_FELCA	105	31
fd; PLLFLLEEYKNYLDAAN	BDNF_CANFA	105	31
fd; PLLFLLEEYKNYLDAAN	BDNF_BOVIN	108	31
fd; PLLFLLEEYKNYLDAAN	BDNF_AILME	105	31
fd; PLLFLLEEYKNYLDAAN	BDNF_AILFU	105	31
fd; PLLFLLEEYKNYLDAAN	A7LA92_HUMAN	187	31
fd; PLLFLLEEYKNYLDAAN	A7LA85_HUMAN	134	31
fd; PLLFLLEEYKNYLDAAN	BDNF_CAVPO	113	31
fd; PLLFLLEEYKNYLDAAN	BDNF_HORSE	105	31
fd; PLLFLLEEYKNYLDAAN	Q8VHH4_MOUSE	107	31
fd; PLLFLLEEYKNYLDAAN	Q6DN19_HUMAN	105	31
fd; PLLFLLEEYKNYLDAAN	BDNF_LIPVE	106	31
fd; PLLFLLEEYKNYLDAAN	BDNF_CHICK	104	30
fd; PLLFLLEEYKNYLDAAN	Q8AV78_NIPNI	104	30
fd; PLLFLLEEYKNYLDAAN	Q4JHT7_POEGU	104	30
fd; PLLFLLEEYKNYLDAAN	A4L7M3_BOMOR	105	30
fd; PLLFLLEEYKNYLDAAN	Q63ZM5_XENLA	105	30
fd; PLLFLLEEYKNYLDAAN	A3FPG9_XENTR	105	30
fd; PLLFLLEEYKNYLDAAN	Q8QG75_9SAUR	104	30
fd; PLLFLLEEYKNYLDAAN	Q8QG76_9SAUR	104	30
fd; PLLFLLEEYKNYLDAAN	A4L7M4_9SALA	105	30
fd; PLLFLLEEYKNYLDAAN	A4L7M5_SALSL	105	30
fd; PLLFLLEEYKNYLDAAN	A2ICR4_AMBME	105	30
fd; PLLFLLEEYKNYLDAAN	Q8QG77_9SALA	105	30
fd; PLLFLLEEYKNYLDAAN	Q6NZ01_DANRE	128	47
fd; PLLFLLEEYKNYLDAAN	Q9YH42_DANRE	128	47
fd; PLLFLLEEYKNYLDAAN	Q8JGW4_PAROL	127	48
fd; PLLFLLEEYKNYLDAAN	Q06B76_DICLA	127	48
fd; PLLFLLEEYKNYLDAAN	BDNF_CYPCA	128	47
fd; PLLFLLEEYKNYLDAAN	Q8QG74_9SAUR	104	30
fd; PLLFLLEEYKNYLDAAN	BDNF_XIPMA	127	48

CALCITONINA

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine

Distribuido libremente bajo la licencia pública general GNU (GPL)

5

Archivo HMM: prints.hmm
 Archivo de secuencia: **uro742rev.154**

CALCITONINR_2: dominio 1 de 1, de 91 a 108: puntuación 6,0, E = 9,4
 ->kCYDRmqqLPpYeGEGpY<-

R+ LP+Y GEGp
 uro742rev. 91 TPVRRLLPLPSYPGEGPQ 108

CALCITONINR_2: dominio 1 de 1, de 72 a 89: puntuación 6,0, E = 9,4
 ->kCYDRmqqLPpYeGEGpY<-

R+ LP+Y GEGp
 zc37.B9.2d 72 TPVRRLLPLPSYPGEGPQ 89

10

gc; **CALCITONINR**
 gx; PR00361
 gn; COMPUESTO(6)
 ga; 15-ABR-1995; ACTUALIZADO 06-JUN-1999
 gt; **Firma de receptor de calcitonina**
 gp; PRINTS; [PR00237](#) GPCRRHODOPSN; [PR00247](#) GPCRCAMP; [PR00248](#) GPCRMGR
 gp; PRINTS; [PR00249](#) GPCRSECRETIN; [PR00250](#) GPCRSTE2; [PR00899](#) GPCRSTE3
 gp; PRINTS; [PR00251](#) BACTRLOPSIN
 gp; PRINTS; [PR01350](#) CTRFAMILY; [PR01351](#) CGRPRECEPTOR
 gp; INTERPRO; [IPR001688](#)
 gr; 1. ATTWOOD, T.K. Y FINDLAY, J.B.C.
 gr; Fingerprinting G protein-coupled receptors.
 gr; PROTEIN ENG. 7(2) 195-203 (1994).
 gr; 2. ISHIHARA T., NAKAMURA S., KAZIRO, Y., TAKAHASHI, T., TAKAHASHI,
 gr; K. Y NAGATA, S.
 gr; Molecular cloning and expression of a cDNA encoding the secretin receptor.
 gr; EMBO J. 10 1635-1641 (1991).
 gr; 3. LIN, H.Y., HARRIS, T.L., FLANNERY, M.S., ARUFFO, A., KAJI, E.H., GORN, A.,
 gr; KOLAKOWSKI, L.F., LODISH, H.F. Y GOLDRING, S.R.
 gr; Expression cloning of adenylate cyclase-coupled calcitonin receptor.
 gr; SCIENCE 254 1022-1024 (1991).
 gr; 4. JUEPPNER, H., ABOU-SAMRA, A.-B., FREEMAN, M., KONG, X.F., SCHIPANI, E., RICHARDS,
 gr; J., KOLALOWSKI, L.F., HOCK, J., POTTS, J.T., KRONENBERG, H.M. Y SEGRE, G.E.
 gr; A G protein linked receptor for parathyroid hormone and parathyroid
 gr; hormone-related peptide.
 gr; SCIENCE 254 1024-1026 (1991).
 gr; 5. ISHIHARA, T., SHIGEMOTO, R., MORI, K., TAKAHASHI, K. Y NAGATA, S.
 gr; Functional expression and tissue distribution of a novel receptor for
 gr; vasoactive intestinal polypeptide.
 gr; NEURON 8(4) 811-819 (1992).
 gr; 6. WATSON, S. Y ARKINSTALL, S.
 gr; Calcitonin.
 gr; EN THE G PROTEIN-LINKED RECEPTOR FACTSBOOK, ACADEMIC PRESS, 1994, PP.74-76.
 gr; 7. NJUKI, F., NICHOLL, C.G., HOWARD, A., MAK, J.C., BARNES, P.J.,
 gr; GIRGIS, S.I. Y LEGON, S.A.
 gr;

gr; A new calcitonin-receptor-like sequence in rat pulmonary blood vessels. CLIN.SCI. 85(4) 385-388
 gr; (1993).
 gd; Los receptores acoplados a proteína G (GPCR) constituyen una vasta familia de
 gd; proteínas que abarca una gran variedad de funciones (incluyendo diversos procesos
 gd; autocrinos, paracrinos y endocrinos). Muestran una considerable diversidad a nivel de
 gd; secuencia, que puede usarse como base para separarlas en distintos grupos. Se usa
 gd; el término clan para describir las GPCR, ya que abarcan un grupo de familias para las
 gd; que hay indicaciones de relación evolutiva, pero entre las cuales no hay una similitud
 gd; estadísticamente significativa en su secuencia [1]. Los miembros del clan conocidos
 gd; en la actualidad incluyen las GPCR similares a rodopsina, las GPCR similares a
 gd; secretina, los receptores de AMPc, los receptores de feromonas de emparejamiento
 gd; fúngico y la familia de receptor de glutamato metabotrópico.
 gd; Los GPCR similares a secretina incluyen secretina [2], calcitonina [3], hormona
 gd; paratiroidea/péptidos relacionados con la hormona paratiroidea [4] y péptido intestinal
 gd; vasoactivo [5], todos los cuales activan la adenililo ciclase y la vía de fosfatidilo-
 gd; inositol-calcio. Las secuencias de aminoácidos de los receptores contienen altas
 gd; proporciones de restos hidrófobos agrupados en 7 dominios, de un modo que
 gd; recuerda a las rodopsinas y otros receptores que se cree que interactúan con las
 gd; proteínas G. Sin embargo, aunque se ha propuesto que esto se debe a una estructura
 gd; en 3D similar, no hay una identidad de secuencia significativa entre estas familias: por
 gd; lo tanto, los receptores similares a secretina portan su firma de "7TM" única.
 gd; El papel fisiológico principal de la **calcitonina** es inhibir la resorción ósea,
 gd; ocasionando por tanto una reducción en el Ca⁺⁺ plasmático [6]. Además, mejora la
 gd; excreción de iones en el riñón, previene la absorción de iones en el intestino e inhibe
 gd; la secreción en células endocrinas (por ejemplo, páncreas y pituitaria). En el SNC, se
 gd; ha comunicado que la calcitonina es analgésica y que suprime la alimentación y la
 gd; secreción de ácidos gástricos. Se usa para tratar la enfermedad de Paget del hueso.
 gd; Los receptores de calcitonina se encuentran principalmente en los osteoclastos o en
 gd; estirpes celulares inmortales procedentes de estas células. Se encuentra en
 gd; cantidades menores en el cerebro (por ejemplo, en los tejidos hipotalámicos y
 gd; pituitarios) y en tejidos periféricos (por ejemplo, testículos, riñón, hígado y linfocitos).
 gd; También se ha descrito en estirpes celulares de cáncer de pulmón y de mama. La vía
 gd; de señalización predominante es la activación de la adenililo ciclase mediante G, pero
 gd; también se ha descrito que la calcitonina tiene acciones tanto estimuladoras como
 gd; inhibitorias en la vía del fosfoinosítido.
 gd; **CALCITONINR es una huella de 6 elementos que proporciona una firma para los**
 gd; **receptores de calcitonina.** La huella se obtuvo de un alineamiento inicial de 6
 gd; secuencias: los motivos se representaron a partir de secciones conservadas dentro de
 gd; cada bucle o las regiones TM, centrándose en aquellas áreas del alineamiento que
 gd; caracterizan los receptores de calcitonina, pero que los distinguen del resto de la
 gd; familia similar a secretina - los motivos 1-3 se representaron a partir de la región N-
 gd; terminal que da lugar al primer dominio TM; el motivo 4 se encuentra en el extremo C-
 gd; terminal del segundo dominio TM después de la región de bucle; el motivo 5 es N-
 gd; terminal respecto de la séptima región TM; y el motivo 6 se representó a partir del
 gd; extremo C-terminal. Para alcanzar la convergencia, se necesitaron dos interacciones
 gd; en OWL25.2, punto en el cual se identificó un conjunto auténtico que comprendía 9
 gd; secuencias. También se encontró una sola coincidencia parcial de humano, RNCLR,
 gd; un nuevo receptor similar a calcitonina de los vasos sanguíneos pulmonares de rata
 gd; [7].
 fc; CALCITONINR2
 fl; 18
 ft; Motivo II de receptor de calcitonina - 2
 fd; KCYDRIQQLPPYEGEGPY CALR_RAT 54 1
 fd; KCYDRMEQLPPYQGEGPY CALR_RABIT 54 1
 fd; KCYDRMQQLPAYQGEGPY CALR_HUMAN 54 1
 fd; KCYDRIHQQLPSYEGEGLY CALR_MOUSE 54 1
 fd; RYDRMQQLPPYEGEGPY CALR_CAVPO 54 1
 fd; RYDRMQKLPPYQGEGLY CALR_PIG 55 1

RECEPTOR DE TIPO 1 DE LEUCOTRIENO B4

BLKPROB Versión 5/21/00.1

Base de datos=/gcg/husar/gcgdata/gcgblimps/blocksplus.dat

Steven Henikoff y Jorja G. Henikoff, Protein Family Classification Based on Searching a Database of Blocks, Genomics 19:97-107 (1994).

Cada resultado numerado consiste en uno o más bloques de un grupo PROSITE o PRINTS encontrado en la secuencia de búsqueda. Se selecciona para su análisis un conjunto de los bloques de mayor puntuación que se encuentran en el orden correcto y separados por distancias comparables a la base de datos BLOCKS. En caso de que este conjunto incluya múltiples bloques, se anota la probabilidad de que los bloques de menor puntuación soporten al bloque de mayor puntuación. Se muestran los mapas de los bloques de la base de datos y la secuencia de búsqueda:

< indica que se ha truncado la secuencia para que quepa en la página
 : indica la distancia mínima entre bloques en la base de datos
 . indica la distancia máxima entre bloques en la base de datos

Los mapas se alinean con el bloque de mayor puntuación. Se muestra el alineamiento de la secuencia de búsqueda con la secuencia más próxima a esta en la base de datos BLOCKS. Las mayúsculas en la secuencia de búsqueda indican al menos una aparición del resto en esa columna del bloque.

Búsqueda=**uro705rev.1a.74** Longitud: 74 Tipo: P C
 Tamaño=74 aminoácidos
 Bloques buscados=29068
 Alineamientos efectuados= 2896529
 Valor de corte combinado esperado para coincidencias= 0
 Valor de corte esperado de bloque para repeticiones/otros= 0

Familia	Hebra	Bloques	Valor de E combinado
<u>IPB003983</u> Señal de receptor de tipo 1 de leucotrieno B4	1	1 de 6	<u>0,0042</u>

>IPB003983 1/6 bloques Valor de E combinado= 0,0042: Firma de **receptor de tipo 1 de leucotrieno B4**
 Bloque Marco Ubicación (aa) Valor de E de bloque
 IPB003983C 0 25-41 0. 0046

Otros alineamientos comunicados:

```

|--- 141 aminoácidos---|
      IPB003983 AAA:::BB:::CC:::DDD:::EEEFF
uro705rev.1a.74_12                ::::CCC
IPB003983C <->C {202,207}:24
Q9WTK1|Q9WTK1_CAVPO207 SRRLRVRRFHRRRTGR
      || | || ||| ||
uro705rev.1a.74_12 25      lRRrRpRRpLRRRRrGR
    
```

rheu.cd.215rev.1.736

>IPB003983 1/6 bloques Valor de E combinado= 0,0094: Firma de **receptor de tipo 1 de leucotrieno B4**
 Bloque Marco Ubicación (aa) Valor de E de bloque
 IPB003983C 0 28-44 0,0096

Otros alineamientos comunicados:

```

| 141 aminoácidos |
      IPB003983 AAA:::BB:::CC:::DDD:::EEEFF
rheu.cd.215rev.1.7                ::::CCC
IPB003983C <->C {202,207}:27
LT4R1_RAT|Q9R0Q2 206 GRRLQARRFRRSRRTGR
      || ||| || |
rheu.cd.215rev.1.7 28      rRRRpARRFRaRRRvRr
    
```

zpr5.B4.12dk.209 Longitud: 209 Tipo: P

Familia	Hebra	Bloques	Valor de E combinado
<u>IPB003983</u> Señal 1 de receptor de tipo 1 de leucotrieno B4		1 de 6	<u>0,0078</u>

zpr5.B4.12dk

>IPB003983 1/6 bloques Valor de E combinado= 0,0078: Firma de **receptor de tipo 1 de leucotrieno B4**
 Bloque Marco Ubicación (aa) Valor de E de bloque
 IPB003983C 0 32-48 0,0081

Otros alineamientos comunicados:

|---141 aminoácidos---|

IPB003983 AAA:::BB:::::CCC:::DDD:::::EEEFF
zpr5.B4.12dk.209_2 ::::CCC

IPB003983C <->C (202,207):31
Q9WTK1|Q9WTK1_CAVPO207 SRRLRVRRFHRRRTGR
|| | || ||| |
zpr5.B4.12dk.209_2 32 rRRpRrRRvRRRRRwrR

AUTOANTÍGENO 1 DE SÍNDROME DE SJÖGREN/ESCLERODERMIA (AUTOANTÍGENO P27)

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine

5 Distribuido libremente bajo la licencia pública general GNU (GPL)

Archivo HMM: pfam.hmm

Archivo de secuencia: **rheu.cd.211rev.164**

Auto_anti-p27: dominio 1 de 1, de 117 a 156: puntuación -12,1, E = 4,6

```
*->eiskkmaellLkGatMLdehCpkCGtPLFrIkDgkvfCPiCe<-*
+ ++ +++ l + L++ +kC + +r + Gk fC +Ce
rheu.cd.21 117 HT-AVKGQFGLGTGRALGKALKKCAFAGLR-RKGKCFCKVCE 156
```

- #=GF ID **Auto anti-p27**
- #=GF AC PF06677.4
- #=GF DE **Autoantígeno 1 de síndrome de Sjögren/esclerodermia (Autoantígeno p27)**
- #=GF AU Moxon SJ
- #=GF SE Pfam-B 21881 (compilación 10.0)
- #=GF TP Familia
- #=GF RN [1]
- #=GF RM 9486406
- #=GF RT cDNA cloning of a novel autoantigen targeted by a minor subset of anti-centromere antibodies.
- #=GF RA Muro Y, Yamada T, Himeno M, Sugimoto K;
- #=GF RL Clin Exp Immunol 1998;111:372-376.
- #=GF DR INTERPRO; IPR009563;
- #=GF CC Esta familia consiste en varias secuencias de autoantígeno 1 de síndrome de Sjögren/esclerodermia (Autoantígeno p27). Se cree que la potencial asociación de anti-p27 con anticuerpos anti-centrómero sugiere que el autoantígeno p27 puede desempeñar un papel en la
- #=GF CC mitosis [1].

10 **VASOPRESINA**

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine Distribuido libremente bajo la licencia pública general GNU (GPL)

Archivo HMM: prints.hmm

15 Archivo de secuencia: **uro742rp.l32**

VASOPRSNV2R_6: dominio 1 de 1, de 7 a 26: puntuación 7,4, E = 9,1

```
*->RaGgrRrGrRtGsPsEGArv<-*
R rRrG t s sE A
uro742rp.l 7 RNASRRRGSSSTASTSEEASL 26
```

VASOPRSNV2R_6: dominio 1 de 1, de 7 a 26: puntuación 7,4, E = 9,1

```
*->RaGgrRrGrRtGsPsEGArv<-*
R rRrG t s sE A
zc37.B8.10 7 RNASRRRGSSSTASTSEEASL 26
```

20 **VASOPRSNV1BR_4**: dominio 1 de 1, de 130 a 149: puntuación 3,0, E = 7,1

```
*->TQAgRverrGWRTWDksSsS<-*
Q + +e R WD++
zc35s.B2.9 130 AQDWAEETYTACRYWDRPPRT 149
```

gc; **VASOPRSNV2R**
gx; PR00898
gn; COMPUESTO(8)
ga; 15-APR-1998; ACTUALIZADO 07-JUN-1999
gt; **Firma de receptor V2 de vasopresina**
gp; PRINTS; [PR00237](#) GPCRRHODOPSN; [PR00247](#) GPCRCAMP; [PR00248](#) GPCRMGR
gp; PRINTS; [PR00249](#) GPCRSECRETIN; [PR00250](#) GPCRSTE2; [PR00899](#) GPCRSTE3
gp; PRINTS; [PR00251](#) BACTRLOPSIN PRINTS; [PR00896](#) VASOPRESSINR
gp; PRINTS; [PR00752](#) VASOPRSNV1AR; [PR00897](#) VASOPRSNV1BR; [PR00665](#) OXYTOCINR
gp; INTERPRO; [IPR000161](#)
gp; 1. ATTWOOD, T.K. Y FINDLAY, J.B.C.
gr; Fingerprinting G protein-coupled receptors.
gr; PROTEIN ENG. 7(2) 195-203 (1994).
gr; 2. ATTWOOD, T.K. Y FINDLAY, J.B.C.
gr; G protein-coupled receptor fingerprints.
gr; 7TM, VOLUMEN 2, EDS. G.VRIEND Y B.BYWATER (1993).
gr; 3. BIRNBAUMER, L.
gr; G proteins in signal transduction.
gr; ANNU.REV.PHARMACOL.TOXICOL. 30 675-705 (1990).
gr; 4. CASEY, P.J. Y GILMAN, A.G.
gr; G protein involvement in receptor-effector coupling.
gr; J.BIOL.CHEM. 263(6) 2577-2580 (1988).
gr; 5. ATTWOOD, T.K. Y FINDLAY, J.B.C.
gr; Design of a discriminating fingerprint for G protein-coupled receptors.
gr; PROTEIN ENG. 6(2) 167-176 (1993).
gr; 6. WATSON, S. Y ARKINSTALL, S.
gr; Vasopressin and oxytocin.
gr; EN THE G PROTEIN-LINKED RECEPTOR FACTSBOOK, ACADEMIC PRESS, 1994, PP.284-291.

gr; Los receptores acoplados a proteína G (GPCR) constituyen una vasta familia de
gd; proteínas que abarca una gran variedad de funciones (incluyendo diversos
gd; procesos autocrinos, paracrinos y endocrinos). Muestran una considerable
gd; diversidad a nivel de secuencia, que puede usarse como base para separarlas en
gd; distintos grupos. Se usa el término clan para describir las GPCR, ya que abarcan
gd; un grupo de familias para las que hay indicaciones de relación evolutiva, pero
gd; entre las cuales no hay una similitud estadísticamente significativa en su
gd; secuencia [1,2]. Los miembros del clan conocidos en la actualidad incluyen las
gd; GPCR similares a rodopsina, las GPCR similares a secretina, los receptores de
gd; AMPc, los receptores de feromonas de emparejamiento fúngico y la familia de
gd; receptor de glutamato metabotrópico.
gd; Las GPCR similares a rodopsina representan por sí mismas una familia de proteínas
gd; ampliamente diseminada que incluye receptores de hormonas, neurotransmisores y
gd; de luz, todos los cuales transducen señales extracelulares mediante su interacción
gd; con proteínas de unión a nucleótidos guanina (G). Aunque sus ligandos activadores
gd; varían ampliamente en su estructura y carácter, las secuencias de aminoácidos de
gd; los receptores son muy similares y se cree que adoptan un armazón estructural
gd; común que comprende 7 hélices transmembrana (TM) [3-5].
gd; La vasopresina y la oxitocina son miembros de la familia de hormonas
gd; neurohipofisarias encontradas en todas las especies de mamífero [6]. Están
gd; presentes a altos niveles en la pituitaria posterior. La vasopresina tiene un papel
gd; esencial en el control del contenido de agua en el organismo, actuando en el riñón
gd; para aumentar la absorción de agua y sodio [6]. A concentraciones mayores, la
gd; vasopresina estimula la contracción del músculo liso vascular, estimula la
gd; degradación del glucógeno en el hígado, induce la activación de plaquetas y
gd; desencadena la liberación de corticotrofina de la pituitaria anterior [6]. La vasopresina
gd; y sus análogos se usan en la práctica clínica para tratar la diabetes insípida [6].
gd; El receptor V2 se encuentra a altos niveles en los epitelios osmorreguladores del
gd; tracto urinario terminal, donde estimula la reabsorción de agua [6]. También se
gd; encuentra presente a niveles menores en el endotelio y los vasos sanguíneos de
gd; algunas especies, donde induce vasodilatación [6]. En el SNC, se encuentran sitios
gd; de unión en el subículo, con niveles menores en el caudado-putamen y en las islas
gd; de Calleja [6]. El receptor está implicado en una vía efectora que forma AMPc
gd; mediante la activación de G [6].
gd; VASOPRSNV2R es una huella de 8 elementos que proporciona una firma para
gd; receptores de vasopresina V2. La huella se obtuvo de un alineamiento inicial de 4
gd; secuencias: los motivos se representaron a partir de secciones cortas conservadas
gd; que abarcan la longitud completa del alineamiento, centrándose en aquellas regiones
gd; que caracterizan a los receptores V2 de vasopresina pero que los distinguen del
gd; resto de la familia de vasopresina - los motivos 1 y 2 se encuentran en el extremo N-
gd; terminal; el motivo 3 abarca el primer bucle citoplasmático; el motivo 4 abarca el
gd; segundo bucle citoplasmático; los motivos 5 y 6 abarcan el tercer bucle citoplasmático; y
gd; los motivos 7 y 8 se encuentran en el extremo C-terminal. Para alcanzar la
gd; convergencia se necesitó una sola interacción en OWL30.1, sin identificar secuencias
gd; adicionales más allá del conjunto de inicio.
fc; VASOPRSNV2R6
fl; 20
ft; Motivo VI de receptor V2 de vasopresina - 2
fd; RAGRRRRGHRTGSPSEGAHV 088721 243 2
fd; RAGRRRRGRRTGSPSEGAHV V2R_RAT 243 2
fd; RAGGHRGGRRAGSPREGARV V2R_PIG 242 2
fd; RPPGRRRRGRRTGSPGEGAHV V2R_HUMAN 243 2
fd; RAGGCRGGHRTGSPSEGARV 077808 242 2
fd; RAGGPRRGCRPGSPAEGARV 2R_BOVIN 242 2

gc; **VASOPRSNV1BR**
gx; PR00897
gn; COMPUESTO(9)
ga; 15-APR-1998; ACTUALIZADO 07-JUN-1999

gt; **Firma de receptor V1B de vasopresina**
 gp; PRINTS; [PR00237](#) GPCRRHODOPSN; [PR00247](#) GPCRCAMP; [PR00248](#) GPCRMGR
 gp; PRINTS; [PR00249](#) GPCRSECRETIN; [PR00250](#) GPCRSTE2; [PR00899](#) GPCRSTE3
 gp; PRINTS; [PRO0251](#) BACTRLOPSIN
 gp; PRINTS; [PR00896](#) VASOPRESSINR
 gp; PRINTS; [PR00752](#) VASOPRSNV1AR; [PR00898](#) VASOPRSNV2R; [PR00665](#) OXYTOCINR
 gp; INTERPRO; [IPR000628](#)
 gr; 1. ATTWOOD, T.K. Y FINDLAY, J.B.C.
 gr; Fingerprinting G protein-coupled receptors.
 gr; PROTEIN ENG. 7(2) 195-203 (1994).
 gr; 2. ATTWOOD, T.K. Y FINDLAY, J.B.C.
 gr; G protein-coupled receptor fingerprints.
 gr; 7TM, VOLUMEN 2, EDS. G.VRIEND Y B.BYWATER (1993).
 gr; 3. BIRNBAUMER, L.
 gr; G proteins in signal transduction.
 gr; ANNU.REV.PHARMACOL.TOXICOL. 30 675-705 (1990).
 gr; 4. CASEY, P.J. Y GILMAN, A.G.
 gr; G protein involvement in receptor-effector coupling.
 gr; J.BIOL.CHEM. 263(6) 2577-2580 (1988).
 gr; 5. ATTWOOD, T.K. Y FINDLAY, J.B.C.
 gr; Design of a discriminating fingerprint for G protein-coupled receptors.
 gr; PROTEIN ENG. 6(2) 167-176 (1993).
 gr; 6. WATSON, S. Y ARKINSTALL, S.
 gr; Vasopressin and oxytocin.
 gr; EN THE G PROTEIN-LINKED RECEPTOR FACTSBOOK, ACADEMIC PRESS, 1994, PP.284-291.
 gd; **VASOPRSNV1BR** es una huella de 9 elementos que proporciona una firma para receptores
 gd; de vasopresina V1B. La huella se obtuvo de un alineamiento inicial de 3 secuencias: los
 gd; motivos se representaron a partir de secciones cortas conservadas que abarcan la longitud
 gd; completa del alineamiento, centrándose en aquellas regiones que caracterizan a los
 gd; receptores V1B de vasopresina pero que los distinguen del resto de la familia de
 gd; vasopresina - el motivo 1 se encuentra en el extremo N-terminal; el motivo 2 se encuentra
 gd; en el segundo bucle citoplásmico; el motivo 3 se encuentra en el segundo bucle externo; los
 gd; motivos 4 y 5 abarcan el tercer bucle citoplásmico; el motivo 6 se encuentra en el tercer
 gd; bucle externo; y los motivos 7-9 se encuentran en el dominio C-terminal. Para alcanzar la
 gd; convergencia se necesitó una sola interacción en OWL30.1, sin identificar secuencias
 gd; adicionales más allá del conjunto de inicio.
 fc; **VASOPRSNV1BR4**
 fl; 20
 ft; Motivo IV de receptor V1B de vasopresina - 2
 fd; TQAWRVGGGGWRTWDRPSPS V1BR_HUMAN 234 48
 fd; TQAGREERRGWRTWDKSSSS V1BR_RAT 234 48

FIRMA DE RECEPTOR 2 DE HORMONA CONCENTRADORA DE MELANINA HMMER 2.3.2 (Oct 2003)
 Copyright (C) 1992-2003 HHMI/Washington University School of Medicine Distribuido libremente bajo la licencia
 pública general GNU (GPL)

5 Archivo HMM: prints.hmm
 Archivo de secuencia: **uro742rp.133**
MCH2RECEPTOR_5: dominio 1 de 1, de 69 a 86: puntuación 5,9, E = 7,1
 ->LvqPFRLtrWRtRYKtiRin<-
 PF +t+Wrt + + n
 uro742rp.1 69 --RPFcITKWRTSFLFFKNN 86

gc; **MCH2RECEPTOR**
 gx; PR01784
 gn; COMPUESTO(9)
 ga; 25-SEP-2002
 gt; **Firma de receptor 2 de hormona concentradora de melanina**
 gp; PRINTS; [PR00237](#) GPCRRHODOPSN; [PR00247](#) GPCRCAMP; [PR00248](#) GPCRMGR
 gp; PRINTS; [PR00249](#) GPCRSECRETIN; [PR00250](#) GPCRSTE2; [PR00899](#) GPCRSTE3
 gp; PRINTS; [PRO0251](#) BACTRLOPSIN
 gp; PRINTS; [PR01507](#) MCH1RECEPTOR; [PR01783](#) MCHRECEPTOR

- gr; 1. ATTWOOD, T.K. Y FINDLAY, J.B.C.
gr; Fingerprinting G protein-coupled receptors.
gr; PROTEIN ENG. 7(2) 195-203 (1994).
- gr; 2. ATTWOOD, T.K. Y FINDLAY, J.B.C.
gr; G protein-coupled receptor fingerprints.
gr; 7TM, VOLUMEN 2, EDS. G.VRIEND Y B.BYWATER (1993).
- gr; 3. BIRNBAUMER, L.
gr; G proteins in signal transduction.
gr; ANNU.REV.PHARMACOL.TOXICOL. 30 675-705 (1990).
- gr; 4. CASEY, P.J. Y GILMAN, A.G.
gr; G protein involvement in receptor-effector coupling.
gr; J.BIOL.CHEM. 263(6) 2577-2580 (1988).
- gr; 5. ATTWOOD, T.K. Y FINDLAY, J.B.C.
gr; Design of a discriminating fingerprint for G protein-coupled receptors.
gr; PROTEIN ENG. 6(2) 167-176 (1993).
- gr; 6. CHAMBERS, J., AMES, R.S., BERGSMA, D., MUIR, A., FITZGERALD, L.R.,
gr; HERVIEU, G., DYTOKO, G.M., FOLEY, J.J., MARTIN, J., LIU, W.S., PARK, J., ELLIS,
gr; C., GANGULY, S., KONCHAR, S., CLUDERAY, J., LESLIE, R., WILSON, S. AND
gr; SARAU, H.M.
gr; Melanin-concentrating hormone is the cognate ligand for the orphan G
gr; protein-coupled receptor SLC-1.
gr; NATURE 400 261-265 (1999).
- gr; 7. SAITO, Y., NOTHACKER, H.-P., WANG, Z., LIN, S.H.S., LESLIE, F. Y
gr; CIVELLI, O.
gr; Molecular characterization of the melanin-concentrating-hormone receptor. NATURE 400 265-269
gr; (1999).
- gr; 8. SAITO, Y., NOTHACKER, H.-P. Y CIVELLI, O.

gr; Melanin-concentrating hormone receptor: an orphan receptor fits the key.
 gr; TRENDS ENDOCRINOL.METAB. 11(8) 299-303 (2000).
 gr; 9. HILL, J., DUCKWORTH, M., MURDOCK, P., RENNIE, G., SABIDO-DAVID, C., AMES,
 gr; R.S., SZEKERES, P., WILSON, S., BERGSMAN, D.J., GLOGER, I.S., LEVY, D.S.,
 gr; CHAMBERS, J.K. Y MUIR, A.I.
 gr; Molecular cloning and functional characterization of MCH2, a novel human MCH
 gr; receptor.
 gr; J.BIOL.CHEM. 276(23) 20125-20129 (2001).
 gd; Los receptores acoplados a proteína G (GPCR) constituyen una vasta familia de
 gd; proteínas que abarca una gran variedad de funciones (incluyendo diversos procesos
 gd; autocrinos, paracrinos y endocrinos). Muestran una considerable diversidad a nivel de
 gd; secuencia, que puede usarse como base para separarlas en distintos grupos. Se usa el
 gd; término clan para describir las GPCR, ya que abarcan un grupo de familias para las que
 gd; hay indicaciones de relación evolutiva, pero entre las cuales no hay una similitud
 gd; estadísticamente significativa en su secuencia [1,2]. Los miembros del clan conocidos
 gd; en la actualidad incluyen las GPCR similares a rodopsina, las GPCR similares a
 gd; secretina, los receptores de AMPc, los receptores de feromonas de emparejamiento
 gd; fúngico y la familia de receptor de glutamato metabotrópico.
 gd; Las GPCR similares a rodopsina representan por sí mismas una familia de proteínas
 gd; ampliamente diseminada que incluye receptores de hormonas, neurotransmisores y de
 gd; luz, todos los cuales transducen señales extracelulares mediante su interacción con
 gd; proteínas de unión a nucleótidos guanina (G). Aunque sus ligandos activadores varían
 gd; ampliamente en su estructura y carácter, las secuencias de aminoácidos de los
 gd; receptores son muy similares y se cree que adoptan un armazón estructural común que
 gd; comprende 7 hélices transmembrana (TM) [3-5]. La hormona concentradora de
 gd; melanina (MCH) es un péptido cíclico identificado originalmente en peces teleosteos
 gd; [6,7]. En peces, la MCH se libera de la pituitaria y provoca el aclaramiento de las células
 gd; pigmentarias de la piel mediante la agregación de pigmentos [6,8]. En mamíferos, MCH
 gd; se expresa principalmente en el hipotálamo y funciona como neurotransmisor en el
 gd; control de una serie de funciones [8]. Se cree que un papel principal de la MCH se
 gd; encuentra en la regulación de la alimentación: la inyección de MCH en cerebros de rata
 gd; estimula la alimentación; la expresión de MCH se regula positivamente en el hipotálamo
 gd; de ratones obesos y sometidos a ayuno; y los ratones que carecen de MCH son
 gd; delgados y comen menos [6]. MCH y la hormona estimulante de alfa-melanocitos (alfa-
 gd; MSH) tienen efectos antagonistas en una serie de funciones fisiológicas. La alfa-MSH
 gd; oscurece la pigmentación en peces y reduce la alimentación en mamíferos, mientras
 gd; que MCH aumenta la alimentación [6,8].
 gd; Dos receptores acoplados a proteína G, MCH1 y MCH2, se han identificado
 gd; recientemente como receptores para la hormona.
 gd; El perfil de expresión de MCH2 es similar al de MCH1, encontrándose los niveles
 gd; máximos en el cerebro. Sin embargo, la expresión de MCH2 es significativamente
 gd; menor que la de MCH1 en la pituitaria, el hipotálamo, el locus cerúleo, el bulbo raquídeo
 gd; y el cerebelo [9]. La unión de MCH al receptor provoca un aumento en el calcio
 gd; intracelular insensible a la toxina tetroférica, lo que sugiere su acoplamiento a proteínas
 gd; Gq [9].
 gd; MCH2RECEPTOR es una huella de 9 elementos que proporciona una firma para el
 gd; receptor 2 de hormona concentradora de melanina. La huella se obtuvo de un
 gd; alineamiento inicial de 5 secuencias: los motivos se representaron a partir de secciones
 gd; conservadas en los extremos N y C-terminales y las regiones de bucle, centrándose en
 gd; aquellas áreas del alineamiento que caracterizan a los receptores MCH2, pero que lo
 gd; distinguen del resto de la familia de receptores MCH - los motivos 1 y 2 abarcan el
 gd; extremo N-terminal; el motivo 3 codifica el primer bucle citoplásmico; el motivo 4 se
 gd; encuentra en el primer bucle externo; el motivo 5 abarca el segundo bucle citoplásmico,
 gd; desembocando en el dominio 4 TM; el motivo 6 se encuentra en el segundo bucle
 gd; externo; el motivo 7 abarca el tercer bucle citoplásmico; el motivo 8 está ubicado en el
 gd; extremo N-terminal del dominio 7 TM; y el motivo 9 codifica el extremo C-terminal. Para
 gd; alcanzar la convergencia, se necesitaron dos interacciones en SPTR40 22f, punto en el
 gd; cual se identificó un conjunto auténtico que comprendía 6 secuencias.

fc; MCH2RECEPTOR5

fl; 20

ft; Motivo V de receptor 2 de hormona concentradora de melanina - 2

fd;	LVQPFRLTSWRTRYKTIRIN	Q8MJ88	135	29
fd;	LVQPFRLTRWRTRYKTIRIN	Q969V1	135	29
fd;	LVQPFRLTRWRTRYKTIRIN	Q9BXA8	135	29
fd;	LVQPFRLTSWRTRYKTIRIN	Q8SQ54	135	29

fd; LVQPFRLTSWRTRYKTIRIN Q8MIN7 135 29
 fd; LVQPFRLTSWRTRYKTIRIN Q8MIP5 135 29

FIRMA DEL RECEPTOR EP1 DE PROSTANOIDE

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine

Distribuido libremente bajo la licencia pública general GNU (GPL)

5

 Archivo HMM: prints.hmm
 Archivo de secuencia: **uro742rev.107r**
PRSTNOIDEP1R_4: dominio 1 de 1, de 1 a 18: puntuación 8,4, E = 4,7

```

  *->isLGPpGGWRqAL.LAGL<-*
  ++LGP GG R+ L +AG
uro742rev. 1 MGLGPSGGNRKTLfIAGK 18

```

PRSTNOIDEP1R_4: dominio 1 de 1, de 1 a 18: puntuación 8,4, E = 4,7

```

  *->isLGPpGGWRqAL.LAGL<-*
  ++LGP GG R+ L +AG
  zc37.B8.10 1 MGLGPSGGNRKTLfIAGK 18

```

10

gc; **PRSTNOIDEP1R**
 gx; PR00580
 gn; COMPUESTO(7)
 ga; 25-SEP-1996; ACTUALIZADO 07-JUN-1999
 gt; **Firma del receptor EP1 de prostanoide**
 gp; PRINTS; [PR00237](#) GPCRRHODOPSN; [PR00247](#) GPCRCAMP; [PR00248](#) GPCRMGR
 gp; PRINTS; [PR00249](#) GPCRSECRETIN; [PR00250](#) GPCRSTE2; [PR00899](#) GPCRSTE3
 gp; PRINTS; [PR00251](#) BACTRLOPSIN
 gp; PRINTS; [PR00428](#) PROSTAGLNDNR; [PR00581](#) PRSTNOIDEP2R; [PR00582](#) PRSTNOIDEP3R
 gp; PRINTS; [PR00583](#) PRSTNOIDE31R; [PR00584](#) PRSTNOIDE32R; [PR00585](#) PRSTNOIDE33R
 gp; PRINTS; [PR00586](#) PRSTNOIDEP4R; [PR00854](#) PRSTNOIDDP; [PR00855](#) PRSTNOIDFPR
 gp; PRINTS; [PR00856](#) PRSTNOIDIPR
 gp; INTERPRO; [IPR000708](#)
 gr; 1. ATTWOOD, T.K. Y FINDLAY, J.B.C.
 gr; Fingerprinting G protein-coupled receptors.
 gr; PROTEIN ENG. 7(2) 195-203 (1994).
 gr; 2. ATTWOOD, T.K. Y FINDLAY, J.B.C.
 gr; G protein-coupled receptor fingerprints.
 gr; 7TM, VOLUMEN 2, EDS. G.VRIEND Y B.BYWATER (1993).
 gr; 3. BIRNBAUMER, L.
 gr; G proteins in signal transduction.
 gr; ANNU.REV.PHARMACOL.TOXICOL. 30 675-705 (1990).
 gr; 4. CASEY, P.J. Y GILMAN, A.G.
 gr; G protein involvement in receptor-effector coupling.
 gr; J.BIOL.CHEM. 263(6) 2577-2580 (1988).
 gr; 5. ATTWOOD, T.K. Y FINDLAY, J.B.C.
 gr; Design of a discriminating fingerprint for G protein-coupled receptors. PROTEIN ENG. 6(2) 167-176
 gr; (1993).
 gr; 6. WATSON, S. Y ARKINSTALL, S.

gr; Prostanoids.
 gr; EN THE G PROTEIN-LINKED RECEPTOR FACTSBOOK, ACADEMIC PRESS, 1994,
 gd; PP.239-251. Los receptores acoplados a proteína G (GPCR) constituyen una vasta
 gd; familia de proteínas que abarca una gran variedad de funciones (incluyendo diversos
 gd; procesos autocrinos, paracrinos y endocrinos). Muestran una considerable diversidad
 gd; a nivel de secuencia, que puede usarse como base para separarlas en distintos
 gd; grupos. Se usa el término clan para describir las GPCR, ya que abarcan un grupo de
 gd; familias para las que hay indicaciones de relación evolutiva, pero entre las cuales no
 gd; hay una similitud estadísticamente significativa en su secuencia [1,2]. Los miembros
 gd; del clan conocidos en la actualidad incluyen las GPCR similares a rodopsina, las
 gd; GPCR similares a secretina, los receptores de AMPc, los receptores de feromonas de
 gd; emparejamiento fúngico y la familia de receptor de glutamato metabotrópico.
 gd; Las GPCR similares a rodopsina representan por sí mismas una familia de proteínas
 gd; ampliamente diseminada que incluye receptores de hormonas, neurotransmisores y de
 gd; luz, todos los cuales transducen señales extracelulares mediante su interacción con
 gd; proteínas de unión a nucleótidos guanina (G). Aunque sus ligandos activadores varían
 gd; ampliamente en su estructura y carácter, las secuencias de aminoácidos de los
 gd; receptores son muy similares y se cree que adoptan un armazón estructural común
 gd; que comprende 7 hélices transmembrana (TM) [3-5].
 gd; Los prostanoides (prostaglandinas (PG) y tromboxanos (TX)) median una gran
 gd; variedad de acciones y desempeñan papeles fisiológicos importantes en los sistemas
 gd; cardiovascular e inmunitario y en la sensación de dolor en los sistemas periféricos [6].
 gd; PGI2 y TXA2 tienen acciones opuestas, que implican la interacción de las plaquetas
 gd; con el endotelio vascular, mientras que PGE2, PGI2 y PGD2 son potentes
 gd; vasodilatadores y potencian la acción de diversas autacoides para inducir la
 gd; extravasación del plasma y la sensación de dolor. Hasta la fecha, se han obtenido
 gd; pruebas de al menos 5 clases de receptores de prostanoides. Sin embargo, la
 gd; identificación de subtipos y su distribución se ve dificultada por la expresión de más de
 gd; un receptor en un tejido, junto con la escasa disponibilidad de agonistas y
 gd; antagonistas disponibles.
 gd; Los receptores EP1 median la contracción de los músculos lisos gastrointestinales en
 gd; diversas especies y la relajación de los músculos lisos de las vías respiratorias y
 gd; uterinos, especialmente en roedores [6]. Los receptores activan la vía de fosfoinosítido
 gd; mediante una proteína G insensible a la toxina tosferínica, probablemente de la clase
 gd; Gq/G11 [6].
 gd; PRSTNOIDEP1R es una huella de 7 elementos que proporciona una forma para los
 gd; receptores EP1 de prostanoides. La huella se obtuvo de un alineamiento inicial de 2
 gd; secuencias; los motivos se representaron a partir de secciones conservadas dentro de
 gd; cada bucle o las regiones N y C-terminales, centrándose en aquellas áreas del
 gd; alineamiento que caracterizan a los receptores EP1 de prostanoides, pero que los
 gd; distinguen del resto de la superfamilia similar a rodopsina - el motivo 1 se encuentra
 gd; en el extremo N-terminal; el motivo 2 abarca el primer bucle citoplasmático; el motivo 3
 gd; abarca el primer bucle externo; el motivo 4 se encuentra en el segundo bucle externo;
 gd; el motivo 5 se encuentra en el tercer bucle citoplasmático; y los motivos 6 y 7 abarcan el
 gd; extremo C-terminal. Para alcanzar la convergencia se necesitó una sola interacción en
 gd; OWL28.2, sin identificar secuencias adicionales más allá del conjunto de inicio.
 gd;
 fc; PRSTNOIDEP1R4
 fl; 17
 ft; Motivo IV de receptor EP1 de prostanoides - 2
 fd; ISLGPRGGWRQALLAGL PE2I_MOUSE 192 73
 fd; ISLGPPGGWRQALLAGL PE2I_RAT 192 73
 fd; IGLGPPGGWRQALLAGL PE2I_HUMAN 190 73

CICLINCINASA

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine

Distribuido libremente bajo la licencia pública general GNU (GPL)

Archivo HMM: prints.hmm

Archivo de secuencia: **rheu.cd.215rev.1..736**

5 **CYCLINKINASE_3:** dominio 1 de 1, de 662 a 676: puntuación 9,3, E = 3,1

->EWRs1GvqqslGwvh<-

E + Gvqq l Wvh

rheu.cd.21 662 ESSRFGVQORLPWVH 676

gc; **CICLINCINASA**
 gx; PR00296
 gn; COMPUESTO(4)
 ga; 07-OCT-1994; ACTUALIZADO 07-JUN-1999
 gt; **Firma de subunidad reguladora de cinasa dependiente de ciclina**
 gp; INTERPRO; IPR000789
 gp; PROSITE; PS00944 CKS_1; PS00945 CKS_2
 gp; PFAM; PF01111 CKS
 gr; 1. BRIZUELA, L., DRAETTA, G. Y BEACH, D.
 gr; pl3suc1 acts in the fission yeast cell division cycle as a component of the
 gr; p34cdc2 protein kinase.
 gr; EMBO J. 6 3507-3514 (1987).
 gr; 2. PARGE, H.E., ARVAI, A.S., MURTARI, D.J., REED, S.I. Y TAINER, J.A.
 gr; Human CksHs2 atomic structure: a role for its hexameric assembly in cell
 gr; cycle control.
 gr; SCIENCE 262 387-395 (1993).
 gr; 3. TANG, Y. Y REED, S.I.
 gr; The Cdk-associated protein Cks1 functions both in G1 and G2 in
 gr; Saccharomyces cerevisiae.
 gr; GENES DEV. 7 822-832 (1993).
 gd; En eucariotas, las proteína cinasas dependientes de ciclina interactúan con las ciclinas para
 gd; regular la progresión del ciclo celular y son necesarias para los estadios G1 y G2 de la división
 gd; celular [1]. Las proteínas se unen a una subunidad reguladora (subunidad reguladora de cinasa
 gd; dependiente de ciclina o CKS), que es esencial para su función [2]. Las subunidades reguladoras
 gd; existen en forma de hexámeros, formados por el ensamblaje simétrico de 3 homodímeros
 gd; entrelazados, creando una estructura infrecuente de barril beta de 12 hebras [2]. A través del
 gd; centro del barril se encuentra un túnel con un diámetro de 12A, revestido por 6 pares de hélices
 gd; expuestas [3]. Puede modelarse la unión de seis unidades de cinasa a la estructura hexamérica,
 gd; que por tanto puede actuar como interconexión para la multimerización de proteínas cinasas
 gd; dependientes de ciclina [2,3].
 gd; CICLINCINASA es una huella de 4 elementos que proporciona una firma para
 gd; subunidades reguladoras de cinasa dependiente de ciclina. La huella se obtuvo de
 gd; un alineamiento inicial de 4 secuencias: los motivos se representaron a partir de
 gd; regiones conservadas que abarcan prácticamente la longitud completa del
 gd; alineamiento, los motivos 1, 2 y 4, que abarcan las regiones codificadas por los
 gd; patrones CKS 1 (PS00944) y CKS_2 (PS00945) de PROSITE. Para alcanzar la
 gd; convergencia, se necesitaron dos interacciones en OWL24.0, punto en el cual se
 gd; **identificó** un conjunto auténtico que comprendía 5 secuencias.
 fc; CYCLINKINASE3
 fl; 15
 ft; Motivo III de subunidad reguladora de cinasa dependiente de ciclina - 2
 fd; EWRRLGVQQLGWVH CKS2_XENLA 42 7
 fd; EWRNLGVQSSQGWVH CKS1_HUMAN 42 7
 fd; EWRRLGVQQLGWVH CKS2_HUMAN 42 7
 fd; EWRRLGVQQLGWVH CKS2_MOUSE 42 7
 fd; EWRSIGVQSSHGWIH CKS1_PATVU 42 7
 fd; EWRSIGVQSSRGWIH CKS1_DROME 41 7
 fd; EWRGLGVQSSQGWVH CKS1_PHYPO 42 7
 fd; EWRQLGVQSSQGWVH CKS1_LEIME 67 7
 fd; EWRAIGVQSSRGWVH O23249 40 7
 fd; EWRGLGITQSLGWQH O60191 73 16
 fd; EWRGLGITQSLGWEM CKS1_SCHPO 69 16
 fd; EWRGLGITQSLGWEH CKS1_YEAST 73 16
 fd; EWRSLGIQQSPGWMH CKS1_CAEEL 44 7

FIRMA DE RECEPTOR ACTIVADO POR PROLIFERADOR DE PEROXISMAS (RECEPTOR NUCLEAR 1C)

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine Distribuido libremente bajo la licencia pública general GNU (GPL)

5

Archivo HMM: prints.hmm
 Archivo de secuencia: rheu.cd.215rev.1.736
PROXISOMEPAR_7: dominio 1 de 1, de 721 a 733: puntuación 8,0, E = 5,7
 ->KtEtdasLHPLLq<-
 K + sLHPLL
 rheu.cd.21 721 KVQAGHSLHPLLs 733

gc; **PROXISOMEPAR**
gx; PR01288
gn; COMPUESTO(7)
ga; 19-FEB-2000
gt; **Firma de receptor activado por proliferador de peroxisomas (receptor nuclear 1C)**
gp; PRINTS; [PR00398](#) STRDHORMONER; [PR00047](#) STROIDFINGER
gp; PRINTS; [PR01289](#) PROXISOMPAAR; [PR01290](#) PROXISOMPABR; [PR01291](#) PROXISOMPAGR
gr; 1. NUCLEAR RECEPTORS NOMENCLATURE COMMITTEE
gr; A unified nomenclature system for the nuclear receptor superfamily.
gr; CELL 97 161-163 (1999).
gr; 2. NISHIKAWA, J-I., KITAURA, M., IMAGAWA, M. Y NISHIHARA, T.
gr; Vitamin D receptor contains multiple dimerisation interfaces that
gr; are functionally different.
gr; NUCLEIC ACIDS RES. 23(4) 606-611 (1995).
gr; 3. DE VOS, P., SCHMITT, J., VERHOEVEN, G. Y STUNNENBERG, G.
gr; Human androgen receptor expressed in HeLa cells activates transcription
gr; *in vitro*.
gr; NUCLEIC ACIDS RES. 22(7) 1161-1166 (1994).
gr; 4. KREY, G., KELLER, H., MAHFOUDI, A., MEDIN, J., OZATO, K., DREYER, C.
gr; Y WAHLI, W.
gr; Xenopus peroxisome proliferator activated receptors: genomic organization,
gr; response element recognition, heterodimer formation with retinoid X receptor
gr; and activation by fatty acids.
gr; J.STEROID BIOCHEM.MOL.BIOL. 47 65-73 (1993).
gr; 5. DREYER, C., KREY, G., KELLER, H., GIVEL, F., HELFTENBEIN, G.
gr; Y WAHLI, W.
gr; Control of the peroxisomal beta-oxidation pathway by a novel family of
gr; nuclear hormone receptors.
gr; CELL 68 879-887 (1992).

gd; Los receptores de esteroides o de hormonas nucleares (NR) forman una superfamilia
gd; importante de reguladores de la transcripción que están implicados en funciones fisiológicas
gd; ampliamente diversas, incluyendo el control del desarrollo embrionario, la diferenciación
gd; celular y la homeostasia [1]. Los miembros de la superfamilia incluyen los receptores de
gd; hormonas esteroideas y los receptores para hormonas tiroideas, retinoides, 1,25-dihidroxi-
gd; vitamina D3 y varios ligandos diferentes. Las proteínas funcionan como moléculas diméricas
gd; en los núcleos para regular la transcripción de genes diana de una manera sensible al
gd; ligando [2,3]. Además de los dominios de unión a ligando C-terminales, estos receptores
gd; nucleares contienen un dedo de cinc N-terminal altamente conservado que media la unión
gd; específica a secuencias de ADN diana, denominados elementos sensibles a ligando. En
gd; ausencia de ligando, se cree que los receptores de hormonas esteroideas se asocian
gd; débilmente con componentes nucleares; la unión a hormonas aumenta en gran medida la
gd; afinidad del receptor.
gd; Los NR son extremadamente importantes en investigación médica, estando implicado un
gd; gran número de estos en enfermedades tales como cáncer, diabetes, síndromes de
gd; resistencia a hormonas, etc. [1]. Aunque varios NR actúan como factores de transcripción
gd; inducibles por ligando, muchos aún no tienen un ligando definido y por consiguiente, se les
gd; denomina receptores "huérfanos". Durante la última década, se han descrito más de 300
gd; NR, muchos de los cuales se encuentran huérfanos, que no pueden nombrarse fácilmente
gd; debido a las confusiones de nomenclatura actual en la bibliografía. Sin embargo,
gd; recientemente se ha introducido un nuevo sistema en un intento de racionalizar el conjunto
gd; de nombres cada vez más complejo usado para describir a miembros de superfamilias [1].
gd; **Los receptores activados por proliferador de peroxisomas (PPAR)** son factores de
gd; transcripción activados por ligando que pertenecen a la superfamilia de receptores de
gd; hormonas nucleares. Se han aislado tres ADNc que codifican PPAR de *Xenopus laevis*:
gd; xPPAR alfa, beta y gamma [4]. Los tres xPPAR parecen estar activados tanto por
gd; proliferadores de peroxisomas sintéticos como por ácidos grasos de origen natural, lo que
gd; sugiere un modo de acción común para todos los miembros de esta subfamilia de
gd; receptores [4]. Además, la multiplicidad de los receptores sugiere la existencia de vías de
gd; señalización celular hasta ahora desconocidas para xenobióticos y supuestos ligandos
gd; endógenos [5].
gd; PROXISOME PAR es una huella de 7 elementos que proporciona una firma para receptores
gd; activados por proliferador de peroxisomas. La huella se obtuvo de un alineamiento inicial de
gd; 11 secuencias: los motivos se representaron a partir de regiones conservadas que abarcan
gd; prácticamente la longitud completa del alineamiento, centrándose en aquellas secciones
gd; que caracterizan a la familia PPAR, pero que los distinguen del resto de la superfamilia de
gd; receptores de hormonas esteroides - los motivos 1 y 2 se encuentran en C-terminal
gd; respecto del dominio de dedo de cinc; y los motivos 3-7 abarcan el supuesto dominio de
gd; unión a ligando. Para alcanzar la convergencia, se necesitaron tres interacciones en
gd; SPTR37_10f, punto en el cual se identificó un conjunto auténtico que comprendía 19
gd; secuencias. También se encontró una sola coincidencia parcial de humano, el receptor
gd; activado por proliferador de peroxisomas beta de *Xenopus*, PPAS XENLA, que no llega a
gd; coincidir con el primer motivo.

fc; PROXISOME PAR7

fl; 13

ft; Motivo VII de receptor activado por proliferador de peroxisomas - 3

fd; KTETDMSLHPLLQ	018924	486	16
fd; KTETDMSLHPLLQ	Q15832	486	16
fd; KTETDMSLHPLLQ	PPAT_HUMAN	456	16
fd; KTETDMSLHPLLQ	062807	485	16
fd; KTETDMSLHPLLQ	018971	486	16
fd; KTETDMSLHPLLQ	PPAT_RABIT	456	16
fd; KTETDMSLHPLLQ	077815	485	16
fd; KTETDMSLHPLLQ	088275	456	16
fd; KTETDMSLHPLLQ	PPAT_MOUSE	456	16
fd; KTETDMSLHPLLQ	Q15180	487	16
fd; KTEADMCLHPLLQ	PPAT_XENLA	458	16
fd; KTETDAALHPLLQ	PPAR_XENLA	455	16
fd; KTESDAALHPLLQ	PPAR_HUMAN	449	16
fd; KTESDAALHPLLQ	PPAR_RAT	449	16
fd; KTESDAALHPLLQ	PPAR_MOUSE	449	16
fd; KTETETSLHPLLQ	PPAS_HUMAN	422	16
fd; KTESDAALHPLLQ	PPAR_CAVPO	448	15
fd; KTESETLLHPLLQ	PPAS_MOUSE	421	16
fd; KTESETLLHPLLQ	Q62879	421	16

FIRMA DE RECEPTOR ML MUSCARÍNICO

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine

Distribuido libremente bajo la licencia pública general GNU (GPL)

5

Archivo HMM: prints, hmm

Archivo de secuencia: **rheu.cd.215rev.1.736**

MUSCRINICM1R_4: dominio 1 de 2, de 161 a 177: puntuación 0,9, E = 98

->KmPmvdpEAqAPtKqPPk<-

K P vD q t qPP

rheu.cd.21 161 KHPTVDFMVQINT-QPPF 177

gc; **MUSCRINICM1R**

gx; PR00538

gn; COMPUESTO(6)

ga; 01-JUN-1996; ACTUALIZADO 07-JUN-1999

gt; **Firma de receptor Ml muscarínico**

gp; PRINTS; [PR00237](#) GPCRRHODOPSN; [PR00247](#) GPCRCAMP; [PR00248](#) GPCRMGR

gp; PRINTS; [PR00249](#) GPCRSECRETIN; [PR00250](#) GPCRSTE2; [PR00899](#) GPCRSTE3

gp; PRINTS; [PR00251](#) BACTRLOPSIN

gp; PRINTS; [PR00243](#) MUSCARINICR; [PR00539](#) MUSCRINICM2R; [PR00540](#) MUSCRINICM3R

gp; PRINTS; [PR00541](#) MUSCRINICM4R; [PR00542](#) MUSCRINICM5R

gp; INTERPRO; [IPR002228](#)

gr; 1. ATTWOOD, T.K. Y FINDLAY, J.B.C.

gr; Fingerprinting G protein-coupled receptors.

gr; PROTEIN ENG. 7(2) 195-203 (1994).

gr; 2. ATTWOOD, T.K. Y FINDLAY, J.B.C.

gr; G protein-coupled receptor fingerprints.

gr; 7TM, VOLUMEN 2, EDS. G.VRIEND Y B.BYWATER (1993).

gr; 3. BIRNBAUMER, L.

gr; G proteins in signal transduction.

gr; ANNU.REV.PHARMACOL.TOXICOL. 30 675-705 (1990).

gr; 4. CASEY, P.J. Y GILMAN, A.G.

gr; G protein involvement in receptor-effector coupling.

gr; J.BIOL.CHEM. 263(6) 2577-2580 (1988).

gr; 5. ATTWOOD, T.K. Y FINDLAY, J.B.C.

gr; Design of a discriminating fingerprint for G protein-coupled receptors.

gr; PROTEIN ENG. 6(2) 167-176 (1993).

gr; 6. KERLAVAGE, A.R., FRASER, C.M., CHUNG, F-Z. Y VENTER, J.C.

gr; Molecular structure and evolution of adrenergic and cholinergic receptors.

gr; PROTEINS 1 287-301 (1986).

gr; 7. WATSON, S. Y ARKINSTALL, S.

gr; Acetylcholine

gr; EN THE G PROTEIN-LINKED RECEPTOR FACTSBOOK, ACADEMIC PRESS, 1994, PP.7-18.

gd; Los receptores acoplados a proteína G (GPCR) constituyen una vasta familia de proteínas
 gd; que abarca una gran variedad de funciones (incluyendo diversos procesos autocrinos,
 gd; paracrinos y endocrinos). Muestran una considerable diversidad a nivel de secuencia, que
 gd; puede usarse como base para separarlas en distintos grupos. Se usa el término clan para
 gd; describir las GPCR, ya que abarcan un grupo de familias para las que hay indicaciones de
 gd; relación evolutiva, pero entre las cuales no hay una similitud estadísticamente significativa
 gd; en su secuencia [1,2]. Los miembros del clan conocidos en la actualidad incluyen las
 gd; GPCR similares a rodopsina, las GPCR similares a secretina, los receptores de AMPc, los
 gd; receptores de feromonas de emparejamiento fúngico y la familia de receptor de glutamato
 gd; metabotrópico.

gd; Las GPCR similares a rodopsina representan por sí mismas una familia de proteínas
 gd; ampliamente diseminada que incluye receptores de hormonas, neurotransmisores y de
 gd; luz, todos los cuales transducen señales extracelulares mediante su interacción con
 gd; proteínas de unión a nucleótidos guanina (G). Aunque sus ligandos activadores varían
 gd; ampliamente en su estructura y carácter, las secuencias de aminoácidos de los
 gd; receptores son muy similares y se cree que adoptan un armazón estructural común que
 gd; comprende 7 hélices transmembrana (TM) [3-5].

gd; Los receptores muscarínicos de acetilcolina, presentes en el sistema nervioso central, las
 gd; neuronas motoras de la médula espinal y los preganglios autónomos, modulan una serie
 gd; de funciones fisiológicas, incluyendo las contracciones del músculo liso de las vías
 gd; respiratorias, el ojo e intestinal; la frecuencia cardíaca; y las secreciones glandulares. Los
 gd; receptores median la atenuación de la adenilato ciclasa, la activación de canales de calcio
 gd; y potasio y la renovación del fosfatidil inositol [6]. Esta diversidad puede ser el resultado
 gd; de la aparición de múltiples subtipos de receptor (de los cuales, en la actualidad se
 gd; conocen 5, denominados M1 a M5), que se han clasificado basándose en las diferencias
 gd; observadas en la unión de ligandos a receptores en membranas de diversos tejidos.

gd; El receptor M1 se encuentra a altos niveles en células neuronales del SNC; es
 gd; particularmente abundante en la corteza cerebral y el hipocampo [7]. Su distribución se
 gd; solapa en gran medida con la de los subtipos M3 y M4. En la periferia, los receptores M1
 gd; se encuentran en los ganglios autónomos y ciertas glándulas secretoras y también se
 gd; encuentran en líneas celulares. No se ha descrito ningún agonista auténticamente
 gd; selectivo [7].

gd; MUSCRINICM1R es una huella de 6 elementos que proporciona una forma para los
 gd; receptores M1 muscarínicos. La huella se obtuvo de un alineamiento inicial de 4
 gd; secuencias: los motivos se representaron a partir de secciones conservadas dentro de
 gd; cada bucle o las regiones N y C-terminales, centrándose en aquellas áreas del
 gd; alineamiento que caracterizan a los receptores M1, pero que los distinguen del resto de
 gd; la familia de receptores muscarínicos - el motivo 1 se encuentra en el extremo N-
 gd; terminal; los motivos 2-5 abarcan el tercer bucle citoplásmico; y el motivo 6 se encuentra
 gd; en el extremo C-terminal. Para alcanzar la convergencia se necesitó una sola
 gd; interacción en OWL28.0, sin identificar secuencias adicionales más allá del conjunto de
 gd; inicio.

fc; MUSCRINICM1R4

fl; 18

ft; Motivo IV de receptor M1 muscarínico - 2

fd;	KMPMVDPEAQAPTKQPPR	ACMI_HUMAN	303	3
fd;	KMPMVDPEAQAPTKQPPK	ACM1_MOUSE	303	3
fd;	KMPMVDSEAQAPTKQPPK	ACMI_RAT	303	3
fd;	KMPMVDPEAQAPTKQPPR	ACM1_MACMU	303	3
fd;	KMPMVDPEAQAPAKQPPR	ACM1_PIG	303	3

FIRMA DE RECEPTOR DE TIPO B2 DE ÁCIDO GAMMA-AMINO BUTÍRICO METABOTRÓPICO (GABA)
transcrito zc35s.B3.3e.172:

GABAB2RECPTR_1: dominio 1 de 1, de 111 a 129: puntuación 5,9, E = 6,4

->LAPGAWGwaRGAPRPPPs<-

+ P W + F+PPPs+

zc35s.B3.3 111 VGPEQWLFPERKPKPPPSA 129

5

gc; **GABAB2RECPTR**
 gx; PR01178
 gn; COMPUESTO(13)
 ga; 18-SEP-1999

gt; **Firma de receptor de tipo b2 de ácido gamma-aminobutírico metabotrópico**
gp; PRINTS; [PR00237](#) GPCRRHODOPSN; [PR00247](#) GPCRCAMP; [PR00249](#) GPCRSECRETIN
gp; PRINTS; [PR00250](#) GPCRSTE2; [PR00899](#) GPCRSTE3; [PR00251](#) BACTRLOPSIN
gp; RINTS; [PR00592](#) CASENSINGR; [PR00593](#) MTABOTROPICR
gp; PRINTS; [PR01176](#) GABABRECEPTR; [PR01177](#) GABAB1RECPTR
gp; INTERPRO; [IPR002457](#)
gr; 1. KAUPMANN, K., HUGGEL, K., HEID, J., FLOR, P.J., BISCHOFF, S., MICKEL, S.J.,
gr; MCMASTER, G., ANGST, C., BITTIGER, H., FROESTL, W. Y BETTLER, B.
gr; Expression cloning of GABA(B) receptors uncovers similarity to metabotropic
gr; glutamate receptors.
gr; NATURE 386 239-246 (1997).
gr; 2. KAUPMANN, K., SCHULER, V., MOSBACHER, J., BISCHOFF, S., BITTIGER, H., HEID, J.,
gr; FROESTL, W., LEONHARD, S., PFAFF, T., KARSCHIN, A. Y BETTLER, B.
gr; Human gamma-aminobutyric acid type B receptors are differentially expressed and regulate inwardly
gr; rectifying K+ channels.
gr; PROC.NATL.ACAD.SCI.U.S.A. 95(25) 14991-14996 (1998),
gr; 3. WHITE, J.H., WISE, A., MAIN, M.J., GREEN, A., FRASER, N.J., DISNEY, G.H., BARNES, A.A.,
gr; EMSON, P., FOORD, S.M. Y MARSHALL, F.H.
gr; Heterodimerization is required for the formation of a functional GABA(B)
gr; receptor.
gr; NATURE 396 679-82 (1998).
gd; GABA (ácido gamma-aminobutírico) es el principal neurotransmisor inhibitor en el
gd; cerebro y señaliza a través de sistemas receptores ionotrópicos
gd; (GABA(A)/GABA(C)) y metabotrópicos (GABA(B)) [1]. Se han clonado receptores
gd; GABA(B) y los experimentos de marcaje de fotoafinidad sugieren que
gd; corresponden a dos formas de receptor altamente conservadas en el sistema
gd; nervioso de vertebrados [1].
gd; Los receptores GABA(B) están implicados en el refinado de la transmisión sináptica
gd; inhibitora [2]. Los receptores presinápticos inhiben la liberación de neurotransmisores
gd; regulando negativamente los canales de Ca²⁺ activados por alto voltaje, mientras que
gd; los receptores prosinápticos reducen la excitabilidad neuronal activando una
gd; prominente conductancia hacia dentro rectificadora de K⁺ (Kir) que subyace a los
gd; potenciales postsinápticos inhibitorios [2]. Los receptores GABA(B) se acoplan
gd; negativamente a la adenililo ciclase y muestran similitud de secuencia con los
gd; receptores metabotrópicos para el neurotransmisor excitante, L-glutamato.
gd; Se ha identificado un nuevo subtipo del receptor GABA(B) (GABA(B)R2) mediante
gd; búsquedas exhaustivas en la base de datos EST [3]. La exploración con dos híbridos
gd; de levadura ha demostrado que el nuevo subtipo forma heterodímeros con
gd; GABA(B)R1 mediante una interacción en sus colas C-terminales intracelulares [3]. Al
gd; expresar GABA(B)R2 en células HEK293T, GABA(B)R1 se glucosila terminalmente y
gd; se expresa en la superficie celular. La expresión conjunta de los receptores produce un
gd; receptor GABA(B) completamente funcional en la superficie celular; este receptor se
gd; une a GABA con un equivalente de alta afinidad al del receptor endógeno del cerebro
gd; [3]. Dichos resultados indican que, *in vivo*, los receptores GABA(B) funcionales del
gd; cerebro pueden ser heterodímeros de GABA(B)R1 y GABA(B)R2.
gd; GABAB2RECPTTR es una huella de 13 elementos que proporciona una firma para los
gd; receptores GABA(B) de tipo 2. La huella se obtuvo de un alineamiento inicial de 2
gd; secuencias: los motivos se representaron a partir de regiones conservadas que
gd; abarcan prácticamente la longitud completa del alineamiento, centrándose en aquellas
gd; secciones que caracterizan los receptores de tipo 2, pero que los distinguen del resto
gd; de la familia de receptores GABA(B). Para alcanzar la convergencia se necesitó una
gd; sola interacción en SPTR37 10f, sin identificar secuencias adicionales más allá del
gd; conjunto de inicio.

fc; GABAB2RECPTTR1
fl; 19
ft; Motivo 1 de receptor GABAB2 - 1
fd; LAPGAWGWARGAPRPPSS 075899 35 35
fd; LAPGAWGWTRGAPRPPSS 088871 34 34

FIRMA DE ARGININA DESIMINASA

ARGDEIMINASE_6: dominio 1 de 1, de 57 a 75: puntuación 8,0, E = 6,8

->seLsrGrggprcsmplvR<-
s L+rG g pr s p++

zc35s.B3.3 57 SPLGRGAGEPRRTSTPVAA 75

gc; **ARGDEIMINASE**
 gx; PRO 14 66
 gn; COMPUESTO(6)
 ga; 08-JAN-2001
 gt; **Firma de arginina desiminasa bacteriana**

gp; PRINTS; PRO0102 OTCASE
 gp; PFAM; PF02726 Arg deiminasa
 gp; INTERPRO; IPR003876
 gr; 1. BROWN, D.M., UPCROFT, J.A., EDWARDS, M.R. Y UPCROFT, P.
 gr; Anaerobic bacterial metabolism in the ancient eukaryote *Giardia duodenalis*.
 gr; INT.J.PARASITOL. 28 149-64 (1998).
 gr; 2. HARASAWA, R., KOSHIMIZU, K., KITAGAWA, M., ASADA, K. Y KATO, I.
 gr; Nucleotide sequence of the arginine deiminase gene of *Mycoplasma hominis*.
 gr; MICROBIOL.IMMUNOL. 36 661-665 (1992).
 gr; 3. KANAOKA, M., KAWANAKA, C., NEGORO, T., FUKITA, Y., TAYA, K. Y AGUI, H.
 gr; Cloning and expression of the antitumor glycoprotein gene of *Streptococcus pyogenes* Su in
 gr; *Escherichia coli*.
 gr; AGRIC.BIOL.CHEM. 51 2641-2648 (1987).
 gr; 4. DEGNAN, B.A., PALMER, J.M., ROBSON, T., JONES, C.E., FISCHER, M., GLANVILLE, M.,
 gr; MELLOR, G.D., DIAMOND, A.G., KEHOE, M.A. Y GOODACRE, J.A.
 gr; Inhibition of human peripheral blood mononuclear cell proliferation by
 gr; *Streptococcus pyogenes* cell extract is associated with arginine
 gr; deiminase activity.
 gr; INFECT.IMMUN. 66 3050-3058 (1998).
 gd; La vía de arginina dihidrolasa (AD) se encuentra en muchos procariotas y algunos de los
 gd; eucariotas primitivos, siendo un ejemplo de estos últimos *Giardia* [1]. La vía anaeróbica de
 gd; tres enzimas degrada la L-arginina para formar 1 mol de ATP, dióxido de carbono y
 gd; amoniaco. En las bacterias más simples, la primera enzima, arginina desiminasa, puede
 gd; suponer hasta un 10 % de la proteína celular total [1].
 gd; La arginina desiminasa cataliza la conversión de la L-arginina en L-citrulina y amoniaco.
 gd; Además de producir energía mediante ATP, el amoniaco también sirve para proteger a la
 gd; bacteria contra el daño por ácidos y la citrulina generada puede usarse en otras vías
 gd; biosintéticas [2]. También se ha demostrado que una glucoproteína ácida estreptocócica
 gd; (SAGP) funciona como arginina desiminasa [3].
 gd; Recientemente, se ha descubierto otra función de esta enzima [4]. Tiene un potente efecto
 gd; antitumoral y puede inhibir la proliferación de células mononucleares de sangre periférica
 gd; estimulada por antígenos, superantígenos o mitógeno [4].
 gd; Otra función de la proteína puede ser inhibir la proliferación celular mediante la detención
 gd; del ciclo celular y la inducción de la apoptosis. Por lo tanto, se ha barajado la hipótesis de
 gd; que podría usarse arginina desiminasa recombinante como nuevo agente antitumoral [4].
 gd; ARGDEIMINASE es una huella de 6 elementos que proporciona una firma para la familia
 gd; de proteínas de arginina desiminasa bacterianas. La huella se obtuvo de un alineamiento
 gd; inicial de 4 secuencias: los motivos se representaron a partir de regiones conservadas que
 gd; abarcan la longitud completa del alineamiento (~430 aminoácidos). Para alcanzar la
 gd; convergencia, se necesitaron dos interacciones en SPTR37_10f, punto en el cual se
 gd; identificó un conjunto auténtico que comprendía 13 secuencias. También se encontraron
 gd; tres coincidencias parciales: P75475 y P75474 son arginina desiminasa de *Mycoplasma*
 gd; *pneumoniae* que coinciden con los tres primeros y los tres últimos motivos,
 gd; respectivamente; una Q48294 es una arginina desiminasa de *Halobacterium salinarium*
 gd; que coincide con los motivos 2 y 6.
 bb;

c; ARGDEIMINASE6
 fl; 19
 ft; Motivo VI de arginina desiminasa bacteriana - 2

fd;	SELSRGRGGPRCMSGMLIR	051896	388	8
fd;	SELSRGRGGPRCMSGMLIR	Q46254	392	8
fd;	SELVRGRGGPRCMSGMPFER	SAGP_STRPY	389	8
fd;	SELSRGRGGPRCMSGMSLVR	051781	389	8
fd;	GELSRGRGGPRCMSGMPLYR	086131	391	8
fd;	SELSRGRGGPRCMSGMPLVR	053088	388	8
fd;	SELGRGRGGGGHCMTCPIVR	ARCA_PSEAE	394	8
fd;	NQLSLGMGNARCMSGMPLSR	ARCA_MYCHO	385	8
fd;	SELGRGRGGGGHCMTCPIWR	031017	387	8
fd;	NQLSLGMGNARCMSGMPLSR	ARCA_MYCAR	386	8
fd;	GELGRGRGGGGHCMTCPIVR	ARCA_PSEPU	397	8
fd;	SELGTGRGGPRCMSGCPAAR	005585	381	8
fd;	SELSRGPSPLEMVCSLWR	ARCA_MYCPN	419	8

REPETICIÓN DE FACTOR DE CRECIMIENTO OPIOIDE

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine

Distribuido libremente bajo la licencia pública general GNU (GPL)

5

Archivo HMM: pfam.hmm

Archivo de secuencia: **zc37.B9.2de.p2**

OGFr III: dominio 1 de 1, de 186 a 207: puntuación 8,2, E = 3,6

->sPsEtPGPrPA..GParDEPAE<-

+ tP P PA +GP+r +P E

zc37.B9.2d 186 RAASTPVPTPAI rGPTRQDPGE 207

- #=GE ID **OGFr_III**
- #=GF AC PF04680.5
- #=GF DE **Repetición de factor de crecimiento opioide**
- #=GF PI OGFr_repeat;
- #=GF AU Waterfield DI, Finn RD
- #=GF SE Pfam-B_4529 (compilación 7.5)
- #=GF GA 33,30 0,00; 25,00 25,00;
- #=GF TC 40,70 0,30; 28,20 35,60;
- #=GF NC 30,90 18,10; 17,10 16,10;
- #=GF TP Repetición
- #=GF BM hmmbuild -FHMM_Is.ann SEED.ann
- #=GF BM hmmbuild --seed 0 HMM_Is
- #=GF BM hmmbuild -f -FHMM fs.ann SEED.ann
- #=GF BM hmmbuild --seed 0 HMM_fs
- #=GF AM globalfirst
- #=GF RN [1]
- #=GF RM 11890982
- #=GF RT The biology of the opioid growth factor receptor (OGFr).
- #=GF RA Zagon IS, Verderame MF, McLaughlin PJ;
- #=GF RL Brain Res Brain Res Rev 2002;38:351-376.
- #=GF DR INTERPRO; IPR006770;
- #=GF CC Repetición rica en prolina encontrada únicamente en el receptor de factor de
- #=GF CC crecimiento opioide [1].

FIRMA DE MOLÉCULA DE ADHESIÓN CD36

HMMER 2.3.2 (Oct 2003)

Copyright (C) 1992-2003 HHMI/Washington University School of Medicine Distribuido libremente bajo la licencia pública general GNU (GPL)

Archivo HMM: prints.hmm

Archivo de secuencia: **zc3rll.B4.10d.p1**

15

CD36ANTIGEN_3: dominio 1 de 1, de 11 a 29: puntuación 6,3, E = 7,7

->WiFDvqnPdevaknsskikvkqR<-

vq P+e sst +v+qR

zc3rll.B4. 11 ---NVQDPPEE-QNESSRFVQQR 29

- gc; **CD36ANTIGEN**
- gx; PR01610
- gn; COMPUESTO(13)
- ga; 23-DEC-2001
- gt; **Firma de molécula de adhesión CD36**

gp; PRINTS; PRO1609 CD36FAMILY; PRO1611 LIMPII
gp; MIM; 173510
gr; 1. OKUMURA, T. Y JAMIESON, G.A.
gr; Platelet glycoprotein. Orientation of glycoproteins on the human platelet surface.
gr; J.BIOL.CHEM. 251 5944-5949 (1976).
gr; 2. NICHOLSON, A.C., FEBBRAIO, M., HAN, J., SILVERSTEIN, R.L. Y HAJJAR,
gr; D.P.
gr; CD36 in atherosclerosis. The role of a class B macrophage scavenger receptor.
gr; ANN.N.Y.ACAD.SCI. 902 128-131 (2000).
gr; 3. SILVERSTEIN, R.L. Y FEBBRAIO, M.
gr; CD36 and atherosclerosis.
gr; CURR.OPIN.LIPIDOL. 11 483-491 (2000).
gr; 4. SAVILL, J., HOGG, N., REN, Y. Y HASLETT, C.
gr; Thrombospondin cooperates with CD36 and the vitronectin receptor in macrophage
gr; recognition of neutrophils undergoing apoptosis.
gr; J.CLIN.INVEST. 90 1513-1522 (1989).
gr; 5. TANDON, NN., KRALISZ, U. Y JAMIESON, GA.
gr; Identification of glycoprotein IV (CD36) as a primary receptor for platelet-collagen
gr; adhesion.
gr; J.BIOL.CHEM. 264 7576-7583 (1989).
gr; 6. MCGREGOR, J.L., CATIMEL, B., PARMENTIER, S., CLEZARDIN, P.,
gr; DECHAVANNE, M. Y LEUNG, L.L.
gr; Rapid purification and partial characterization of human platelet glycoprotein IIIb.
gr; Interaction with thrombospondin and its role in platelet aggregation.
gr; J.BIOL.CHEM. 264 501-506 (1989).
gr; 7. BARNWELL, J.W., ASCH, A.S., NACHMAN, R.L., YAMAYA, M., AIKAWA, M. E
gr; INGRAVALLO, P.
gr; A human 88-KD membrane glycoprotein (CD36) functions *in vitro* as a receptor for
gr; a cytoadherence ligand on Plasmodium falciparum-infected erythrocytes.
gr; J.CLIN.INVEST. 84 765-772 (1989).
gr; 8. BULL, H.A., BRICKELL, P.M. Y DOWD, P.M.
gr; Src-related protein tyrosine kinases are physically associated with the surface
gr; antigen CD36 in human dermal microvascular endothelial cells.
gr; FEBS LETT. 351 41-44 (1994).
gr; 9. MIYAOKA, K., KUWASAKO, T., HIRANO, K., NOZAKI, S., YAMASHITA, S.
gr; Y MATSUZAWA, Y.
gr; CD36 deficiency associated with insulin resistance.
gr; LANCET 357 686-687 (2001).
gr; *CD36 es una glucoproteína transmembrana altamente glucosilada de 88 kda [1]*
gr; *expresada por monocitos, macrófagos, plaquetas, células endoteliales microvasculares*
gd; *y el tejido adiposo [2], Es un receptor multifuncional que se une a LDL oxidadas*
gd; *(OxLDL), ácidos grasos de cadena larga, fosfolípidos aniónicos, células apoptóticas,*
gd; *trombospondina (TSP), colágeno y eritrocitos infectados por Plasmodium falciparum [2].*
gd; *CD36 tiene numerosas funciones celulares. Es un receptor depurador de tipo B,*
gd; *que desempeña un papel importante en la captación de OxLDL por los macrófagos*
gd; *[3]. Después, los macrófagos ricos en lípidos se diferencian en células espumosas*
gd; *y contribuyen a la formación de lesiones ateroscleróticas [3]. Además, CD36 de*
gd; *macrófagos, junto con TSP y la integrina alfa beta3, puede fagocitar a los*
gd; *neutrófilos apoptóticos [4]. Además, la proteína es uno de los receptores de*
gd; *colágeno en la adhesión y agregación plaquetaria [5,6]. CD36 también puede*
gd; *mediar la citoadherencia de eritrocitos infectados por Plasodium falciparum al*
gd; *endotelio de vénulas poscapilares de diferentes órganos [7]. Además, el CD36*
gd; *citoplásmico desempeña un papel importante en la transducción de señales,*
gd; *interactuando con las tirosina cinasas de la familia Src [8]. La deficiencia de CD36*
gd; *en las poblaciones asiática y africana se ha asociado con resistencia a la insulina*
gd; *[9].*
gd; *CD36 es una huella de 13 elementos que proporciona una firma para las*
gd; *moléculas de adhesión CD36. La huella se obtuvo de un alineamiento*
gd; *inicial de 4 secuencias, centrándose en aquellas secciones que*
gd; *caracterizan a las moléculas de adhesión a CD36 pero que las distinguen*
gd; *del resto de la familia de CD36: el motivo 1 abarca el primer supuesto*
gd; *dominio N-terminal TM; los motivos 2-12 se encuentran en el dominio*
gd; *extracelular; y el motivo 13 abarca el segundo supuesto dominio C-*
gd; *terminal TM. Para alcanzar la convergencia, se necesitaron dos*
gd; *interacciones en SPTR40_18f, punto en el cual se identificó un conjunto*
gd; *auténtico que comprendía 6 secuencias.*

bb;
 fc; CD36ANTIGEN3
 fl; 23
 ft; Motivo III de molécula de adhesión a CD36 - 2
 fd; WVFQVQNPPEEVAKNSSKIKVIQR CD36_RAT 65 18
 fd; WIFDVQNPDDVAKNSSKIKVKQR CD36_MOUSE 65 18
 fd; WIFDVQNPDEVTVNSSKIKVKQR D36_BOVIN 65 18
 fd; WIFDVQNPQEVMMNSSNIQVKQR D36_HUMAN 65 18
 fd; WIFDVQNPDEVAVNSSKIKVKQR D36_MESAU 65 18
 fd; WIFDVQNPPEEVAKNSSKIKVKQR 035754 66 18

FIRMA DE PROTEÍNA PROTELIPÍDICA DE MIELINA (PLP)

HMMER 2.3.2 (Oct 2003)
 Copyright (C) 1992-2003 HHMI/Washington University School of Medicine
 Distribuido libremente bajo la licencia pública general GNU (GPL)

Archivo HHM:	prints.hm
Archivo de secuencia:	zc312.B11.20d.trrev4_8009.sreformat

5 MYELINPO_5: dominio 1 de 1, de 70 a 91: puntuación -0,1, E = 9,3

-> GVVIGAIIGGvLGvLLIMIIYLv <.

IG iIGGv G VLL + +L +

zc312.B11.70

--MLGRIIGGV-GCVLLELXGLGVR

91

gc; MYELINPLP
 gx; PR00214
 gn; COMPUESTO (7)
 ga; 11-JUL-1994; ACTUALIZADO 07-JUN-1999
 gt; firma de proteína proteolipídica de mielina (PLP)
 gp; INTERPRO; IPR001614
 gp; PROSITE: PS00575 MYELIN_PLP_1: PS01004 MYELIN_PLP_2
 gp; BLOCKS; BL00575
 gp; PFAM: PF01275 Myelin_PLP
 gr; 1. SAKAMOTO, Y., KITAMURA, K., YOSHIMURA, K., NISHIJIMA, T. Y UYEMURA, K.
 gr; Complete amino acid sequence of PO protein in bovine peripheral nerve
 gr; myelin.
 gr; J.BIOL.CHEM. 262 4208-4214 (1987).
 gr; 2. SHAW, S.Y., LAURSEN, R.A. Y LEES, M.B.
 gr; Identification of thiol groups and a disulfide crosslink site in bovine myelin
 gr; proteolipid protein.
 gr; FEBS LETT. 250 306-310 (1989).
 gr; 3. DIEHL, H.H., SCHAICH, M., BUDZINSKI, R.M. Y STOFFEL, W.
 gr; Individual exons encode the integral membrane domains of human myelin
 gr; proteolipid protein.
 gr; PROC.NATL.ACAD.SCI.U.S.A. 83 9807-9811 (1986).
 gd; La vaina de mielina es una membrana multicapa, única en el sistema nervioso, que
 gd; funciona como aislante para aumentar en gran medida la velocidad de la conducción
 gd; de impulsos axonales [1]. La proteína proteolipídica de mielina (PLP) es la principal
 gd; proteína encontrada en la vaina de los nervios del sistema nervioso central [2].
 gd; Traspasa 4 veces la membrana [3] y se cree que está implicada en la formación o el
 gd; mantenimiento de la estructura multilamelar. La proteína contiene varios restos de
 gd; cisteína, algunos implicados en la formación de enlaces disulfuro, estando otros
 gd; palmitoilados [2]. Las mutaciones en PLP dan como resultado trastornos neurológicos,
 gd; tales como la enfermedad de Pelizaeus-Merzbacher en seres humanos, 'jimpy' en
 gd; ratones y 'perro temblador' en perros.
 gd; MYELINPLP es una huella de 7 elementos que proporciona una firma para las
 gd; proteínas proteolipídicas de mielina. La huella se obtuvo de un alineamiento inicial de
 gd; 4 secuencias: los motivos 1, 2, 5 y 7 codifican los 4 dominios transmembrana (TM) - el
 gd; motivo 4 incluye la región codificada por el patrón MYELIN_PLP_1 (PS00575) de
 gd; PROSITE, que está ubicada entre los segmentos segundo y tercero TM y contiene 2
 gd; restos de Cys que están palmitoilados; el motivo 7 incluye parte de la región
 gd; codificada por el patrón MYELIN_PLP_2 (PS01004) de PROSITE. Para alcanzar la

gd; convergencia, se necesitaron dos interacciones en OWL23.2, punto en el cual se
 gd; identificó un conjunto auténtico que comprendía 9 secuencias. También se
 gd; encontraron varias coincidencias parciales, todas las cuales son mutantes de
 gd; eliminación o fragmentos de PLP de mielina.
 gd; Una actualización en SPTR37_9f identificó un auténtico grupo de 8 secuencias y 9
 gd; coincidencias parciales.

CHLAMIDIAOM

HMMER 2.3.2 (Oct 2003)

Copyright © 1992-2003 HHMI/Washington University School of Medicine

Distribuido libremente bajo la licencia pública general GNU (GPL)

5

HMMfile: prints.hmm
 Archivo de secuencia: rheu.cd.212rp.365_22305.sreformat

CHLAMIDIAOM3_3: dominio 1 de 1, de 88 a 100: puntuación 4,6, E= 9,7

```

          *->CgsYvPsCskpcG<-*
          C+Y+C k . G
rheu. Cd.21      88      CTGYTEFCAKYTG      100

```

gr; 3. BACHMAIER, K., NEU, N., DE LA MAZA, L.M., PAL, S., HESSEL, A. Y
 gr; PENNINGER, J.M.
 gr; Chlamydia infections and heart disease linked through antigenic mimicry
 gr; SCIENCE 283 1335-1339 (1999).

bb;
 bb;

gd; La proteína rica en tres cisteínas (que también se cree que es una
 gd; lipoproteína) forma la matriz extracelular de la membrana externa de
 gd; Chlamydia [1]. Están implicadas en la integridad estructural esencial de
 gd; la fase tanto de cuerpo elemental (EB) como de cuerpo reticulado (RB).
 gd; Ya que estas bacterias carecen de la capa de peptidoglucano común
 gd; para la mayoría de microbios gramnegativos, dichas proteínas son muy
 gd; importantes en la patogenicidad del organismo.
 gd; La mayor de estas es la proteína principal de la membrana externa (MOMP) y
 gd; supone aproximadamente un 60 % de la proteína total de la membrana [2],
 gd; OMP2 es la segunda más grande, con una masa molecular de 88kDa,
 gd; mientras que la proteína OMP3 tiene ~15kDa [1]. Se cree que MOMP
 gd; desencadena la respuesta inmunitaria más fuerte y recientemente se ha
 gd; relacionado con la enfermedad cardíaca por su similitud de secuencia con
 gd; una alfa miosina específica de músculo cardíaco murino [3].
 gd; La familia OMP3 desempeña un papel estructural en la membrana
 gd; externa durante la etapa EB de la célula de Chlamydia y diferentes
 gd; variantes biológicas muestran un pequeño, aunque muy significativo,
 gd; cambio a nivel de carga del péptido [1]. Los miembros de esta familia
 gd; incluyen *C. trachomatis*, *C. pneumoniae* y *C. psittaci*.

gd; CHLAMIDIAOM3 es una huella de 3 elementos que proporciona una firma para
 gd; la familia de proteína 3 de la membrana externa rica en cisteína (OMP3) de
 gd; Chlamydia. La huella se obtuvo de un alineamiento inicial de 3 secuencias: los
 gd; motivos se representaron a partir de regiones conservadas que abarcan la
 gd; longitud completa del alineamiento (~90 aminoácidos). Para alcanzar la
 gd; convergencia, se necesitaron dos interacciones en SPTR37_10f, punto en el
 gd; cual se identificó un conjunto auténtico que comprendía 8 secuencias.

10 La presente invención también se refiere a un cebador oligonucleotídico que comprende o consiste en parte de un poli(ácido nucleico) como se ha definido anteriormente, siendo dicho cebador capaz de actuar como cebador para secuenciar de manera específica o amplificar de manera específica el poli(ácido nucleico) de HCR de virus TT de la invención y las secuencias de ADN celular (del hospedador).

15 El término "cebador" se refiere a una secuencia de oligonucleótido de ADN capaz de actuar como punto de inicio para la síntesis de un producto de extensión de cebador que es complementario a la hebra de ácido nucleico que se va a copiar. La longitud y la secuencia del cebador ha de ser tal que permita el cebado de la síntesis de los productos de extensión. Preferentemente, el cebador tiene aproximadamente 5-50 nucleótidos. La longitud y

secuencia específicas dependerán de la complejidad de las dianas de ADN o ARN necesarias, así como las condiciones de uso del cebador, tales como la temperatura y la fuerza iónica.

El hecho de que los cebadores de amplificación no tengan que coincidir exactamente con la secuencia de molde para garantizar una amplificación adecuada se ha documentado ampliamente en la bibliografía. El procedimiento de amplificación usado puede ser reacción en cadena de la polimerasa (PCR), reacción en cadena de la ligasa (LCR), amplificación basada en la secuencia de ácido nucleico (NASBA), sistemas de amplificación basados en transcripción (TAS), amplificación por desplazamiento de hebra (SDA) o amplificación mediante Q β replicasa o cualquier otro procedimiento adecuado para amplificar moléculas de ácido nucleico usando extensión de cebadores. Durante la amplificación, pueden marcarse de manera conveniente los productos amplificados usando cebadores marcados o incorporando nucleótidos marcados.

Los marcadores pueden ser isotópicos (³²P, ³⁵S, etc.) o no isotópicos (biotina, digoxigenina, etc.). La reacción de amplificación se repite entre 20 y 70 veces, ventajosamente entre 25 y 45 veces.

Puede usarse cualquiera de una serie de reacciones de secuenciación conocidas en la técnica para secuenciar directamente la información genética vírica y determinar el ORF traduciendo la secuencia de la muestra en la secuencia de aminoácidos correspondiente. Las reacciones de amplificación ilustrativas incluyen las basadas en técnicas desarrolladas por Sanger o Maxam y Gilbert. También se contempla que puedan utilizarse una serie de procedimientos de secuenciación automatizados cuando se llevan a cabo los presentes ensayos, incluyendo secuenciación por espectrometría de masas (véase, por ejemplo: la publicación PCT WO 94/16101). Será evidente para un experto en la materia que, por ejemplo, la aparición de solo dos o tres bases de nucleótidos ha de determinarse en la reacción de secuenciación.

Preferentemente, estos cebadores tienen de 5 a 50 nucleótidos de longitud, más preferentemente, de aproximadamente 10 a 25 nucleótidos. Los más preferidos son cebadores que tienen una longitud de al menos 13 bases. Preferentemente, un cebador desvelado en el presente documento tiene una secuencia de nucleótidos como la mostrada en la tabla 2.

Tabla 2

Cebadores usados para generar genomas de VTT-HD completos y genomas subvíricos de μVTT-HD mediante amplificación por PCR de larga distancia				
VTT	Cebador	Número de nucleótidos	Secuencia	
VTT-jt34f (n.º de ref. AB064607)	jt34f-1s	223-247	5'-GGCCGGGCCA TGGGCAAGGC TCTTA-3'	
	jt34f-2as	195-222	5'-AGTCAAGGGG CAATTCGGGC TCGGGACT-3'	
	jt34f-5s	205-222	5'-CAATTCGGGC TCGGGACT-3'	
	jt34f-6as	186-204	5'-ACACACCGCA GTCAAGGGG-3'	
	jt34f-7s	205-223	5'-CAATTCGGGC TCGGGACTG-3'	
	jt34f-8as	181-204	5'-AGTTTACACA CCGCAGTCAA GGGG-3'	
	VTT-HD1 (n.º de ref. AJ620222)	th25-1s	126-156	5'-CCGCAGCGAG AACGCCACGG AGGGAGATCC T-3'
		tth25-2as	95-125	5'-ACTTCCGAAT GGCTGAGTTT TCCACGCCCG T-3'
VTT-HD3 (n.º de ref. AJ620231)	tth8-1s	133-164	5'-AGAGGAGCCA CGGCAGGGGA TCCGAACGTC CT-3'	

(continuación)

Cebadores usados para generar genomas de VTT-HD completos y genomas subvéricos de μVTT-HD mediante amplificación por PCR de larga distancia			
VTT	Cebador	Número de nucleótidos	Secuencia
	tth8-2as	102-132	5'-CTTACCGACT CAAAAACGAC GGGCAGGCGC C
VTT-HD4 (n.º de ref. AJ620226)	tth4-1s	129-156	5'-CAGCGAGAAC GCCACGGAGG GAGATCCT- 3'
	tth4-2as	101-128	5'-GAATGGCTGA GTTTTCCACG CCCGTCCG-3'
VTT-t3pb (n.º de ref. AF247138)	t3pb-1s	209-226	5'-CAATTCGGGC ACGGGACT-3' *
	t3pb-2as	185-208	5'-AGTTTACACA CCGAAGTCAA GGGG-3'
* A - La secuencia de VTT-t3pb tiene una T en esta posición			

La presente invención también se refiere a una sonda oligonucleotídica que comprende o consiste en parte de un poli(ácido nucleico) de virus TT reorganizado como se ha definido anteriormente, siendo dicha sonda capaz de actuar como sonda de hibridación para la detección específica de un ácido nucleico de VTT de acuerdo con la invención.

El término "sonda" se refiere a oligonucleótidos específicos de secuencia monocatenarios que tienen una secuencia que es complementaria a la secuencia diana del poli(ácido nucleico) de VTT que se va a detectar.

Preferentemente, estas sondas tienen de aproximadamente 5 a 50 nucleótidos de longitud, más preferentemente, de aproximadamente 10 a 25 nucleótidos. Las más preferidas son sondas que tienen una longitud de al menos 13 bases.

La sonda puede estar marcada o unida a un soporte sólido.

La expresión "soporte sólido" puede referirse a cualquier sustrato al que pueda acoplarse a una sonda oligonucleotídica, a condición de que conserve sus características de hibridación y a condición de que el nivel de fondo de hibridación permanezca bajo. Normalmente, el sustrato sólido será una placa de microtitulación, una membrana (por ejemplo, nilón o nitrocelulosa) o una microesfera (perla). Antes de la aplicación a la membrana o de la fijación, puede ser conveniente modificar la sonda de ácido nucleico para facilitar su fijación o mejorar la eficacia de la hibridación. Dichas modificaciones pueden abarcar la adición de colas de homopolímero, acoplamiento con diferentes grupos reactivos, tales como grupos alifáticos, grupos NH₂, grupos SH, grupos carboxílicos o acoplamiento con biotina o haptenos.

Los oligonucleótidos de acuerdo con la presente invención, usados como cebadores o sondas, también pueden contener o consistir en análogos de nucleótidos, tales como fosforotioatos, alquilfosforotioatos o ácidos nucleicos peptídicos o pueden contener agentes intercalantes. Estas modificaciones necesitarán adaptaciones respecto de las condiciones en las que debería usarse el oligonucleótido para obtener la especificidad y sensibilidad necesaria. Sin embargo, los eventuales resultados serán esencialmente los mismos a los obtenidos con los oligonucleótidos no modificados.

Puede ser ventajosa la introducción de estas modificaciones para influenciar positivamente características tales como la cinética de hibridación, la reversibilidad de la formación de híbridos, estabilidad biológica de las moléculas de oligonucleótido, etc. Los poli(ácidos nucleicos) de la invención pueden estar comprendidos en una composición de cualquier tipo. Dicha composición puede ser para uso diagnóstico, terapéutico o profiláctico.

En el presente documento también se desvelan variantes de secuencia de los poli(ácidos nucleicos) seleccionados de cualquiera de las secuencias de nucleótidos con dichas variantes que contienen eliminaciones y/o inserciones de uno o más nucleótidos, especialmente inserciones o eliminaciones de 1 o más codones, principalmente en las extremidades de los oligonucleótidos (3' o 5') o sustituciones de algunos nucleótidos no esenciales por otros (incluyendo nucleótidos modificados y/o inosina).

Las secuencias de poli(ácido nucleico) de VTT reorganizadas de acuerdo con la presente invención que son similares a las secuencias mostradas en la figura 1 pueden caracterizarse y aislarse de acuerdo con cualquiera de las técnicas conocidas en la técnica, tales como amplificación mediante cebadores específicos de secuencia, hibridación con sondas específicas de secuencia en condiciones más o menos rigurosas, determinación de secuencia de la información genética de VTT, etc.

La presente invención también se refiere a un vector de expresión recombinante que comprende un poli(ácido nucleico) de VTT reorganizado de la invención como se ha definido anteriormente, ligado funcionalmente a elementos de control de la transcripción y traducción procarióticos, eucarióticos o víricos.

El término "vector" puede comprender un plásmido, un cósmido, un cromosoma artificial, un fago o un virus o un animal transgénico no humano. Particularmente útiles para el desarrollo de vacunas pueden ser moléculas recombinantes de virus TT, BCG o vectores adenovíricos, así como virus recombinantes de la viruela aviar.

La expresión "expresado de forma recombinante" empleada en el contexto de la presente invención, se refiere al hecho de que los polipéptidos de la presente invención se producen mediante procedimientos de expresión recombinantes ya sea en procariotas o eucariotas inferiores o superiores, como se describe más detalladamente a continuación.

La expresión "eucariota inferior" se refiere a células hospedadoras tales como levaduras, hongos y similares. Los eucariotas inferiores son por lo general (pero no necesariamente) unicelulares. Eucariotas inferiores preferidos son levaduras, particularmente especies dentro de *Saccharomyces*, *Schizosaccharomyces*, *Kluiveromyces*, *Pichia* (por ejemplo, *Pichia pastoris*), *Hansenula* (por ejemplo, *Hansenula polymorph*), *Schwanomyces*, *Schizosaccharomyces*, *Yarrowia*, *Zygosaccharomyces* y similares. *Saccharomyces cerevisiae*, *S. carlsbergensis* and *K. lactis* son los hospedadores de levadura empleados más comúnmente, y son hospedadores fúngicos convenientes.

La expresión "eucariota superior" se refiere a células hospedadoras obtenidas a partir de animales superiores, tales como mamíferos, reptiles, insectos y similares. Células hospedadoras eucariotas superiores preferidas actualmente se obtienen a partir de células de hámster chino (por ejemplo, CHO), mono (por ejemplo, COS y células Vero), riñón de hámster bebé (BHK), riñón de cerdo (PK15), células de riñón de conejo 13 (RK13), la línea celular de osteosarcoma humano 143 B, la línea celular humana HeLa y líneas celulares de hepatoma humano tales como Hep G2, y líneas celulares de insecto (por ejemplo, *Spodoptera frugiperda*). Las células hospedadoras se pueden proporcionar en cultivos en suspensión o en matraz, cultivos de tejidos, cultivos de órganos y similares. Como alternativa, las células hospedadoras también pueden ser animales transgénicos no humanos.

El término "procariotas" se refiere a hospedadores tales como *E. coli*, *Lactobacillus*, *Lactococcus*, *Salmonella*, *Streptococcus*, *Bacillus subtilis* o *Streptomyces*. También estos hospedadores se contemplan dentro de la presente invención.

La expresión "célula hospedadora" se refiere a células que se pueden haber utilizado o se han usado como receptores para un vector recombinante u otro polinucleótido de transferencia, e incluyen la progenie de la célula original que se ha transfectado.

Se entiende que la progenie de una única célula parental puede que no sea necesariamente completamente idéntica en la morfología o en el complemento de ADN genómico o total, a la parental original, debido a una mutación o recombinación natural, accidental o deliberada.

El término "replicón" es cualquier elemento genético, por ejemplo, un plásmido, un cromosoma, un virus, un cósmido, etc., que se comporta como a una unidad autónoma de replicación de polinucleótidos dentro de una célula, es decir, es capaz de una replicación bajo su propio control.

El término "vector" es un replicón que comprende además secuencias que proporcionan una replicación y/o una expresión de un marco de lectura abierto deseado.

La expresión "elemento de control" se refiere a secuencias de polinucleótidos que son necesarias para efectuar la expresión de secuencias codificantes a las que están ligadas. La naturaleza de tales secuencias de control difiere dependiendo del organismo hospedador; en procariotas, tales secuencias de control incluyen generalmente un promotor, un sitio de unión al ribosoma, sitios de corte y empalme y terminadores; en eucariotas, generalmente, tales secuencias de control incluyen promotores, sitios de corte y empalme, terminadores y, en algunos casos, potenciadores. La expresión "elementos de control" se entiende que incluye, como mínimo, todos los componentes cuya presencia es necesaria para la expresión, y también pueden incluir componentes adicionales cuya presencia es ventajosa, por ejemplo, secuencias líder que gobiernan la secreción.

El término "promotor" es una secuencia de nucleótidos que se compone de secuencias de consenso que permiten la unión de la ARN polimerasa con el molde de ADN, de manera que esa producción de ARNm se inicia en el sitio normal de iniciación de la transcripción para los genes estructurales adyacentes.

La expresión "ligado funcionalmente" se refiere a una yuxtaposición en la que los componentes descritos de este modo están en una relación que les permite actuar de la manera pretendida. Una secuencia de control "ligada funcionalmente" a una secuencia codificante, está ligada de tal manera que la expresión de la secuencia codificante

se consigue en condiciones compatibles con las secuencias de control.

El segmento del ADN de VTT reorganizado que codifica la secuencia deseada insertada en la secuencia del vector, se puede fijar a una secuencia señal. Dicha secuencia señal puede ser la que proviene de una fuente no VTT, pero las estructuras artificiales particularmente preferidas de acuerdo con la presente invención, contienen secuencias señal que aparecen en el genoma de VTT antes de los respectivos puntos de inicio de las proteínas.

Los eucariotas superiores se pueden transformar con vectores, o pueden estar infectados con un virus recombinante, por ejemplo, un virus vaccinia recombinante. Las técnicas y los vectores para la inserción de ADN extraño en virus vaccinia son bien conocidos en la técnica, y utilizan, por ejemplo, la recombinación homóloga. Una amplia variedad de secuencias promotoras víricas, posiblemente secuencias terminadoras y secuencias de adición de poli(A), posiblemente secuencias potenciadoras y posiblemente secuencias de amplificación, todas necesarias para la expresión en mamíferos, están disponibles en la técnica. El virus vaccinia es particularmente preferido ya que el virus vaccinia detiene la expresión de las proteínas de células hospedadoras. Para una vacunación de los seres humanos, la viruela aviar y el virus Ankara modificado (AMV) son vectores particularmente útiles.

También se conocen vectores de transferencia de expresión de insectos obtenidos a partir del baculovirus, virus de la poliedrosis nuclear *Autographa californica* (AcNPV), que es un vector de expresión vírica independiente auxiliar. Los vectores de expresión obtenidos a partir de este sistema normalmente utilizan el potente promotor del gen de polihedrina vírica para dirigir la expresión de genes heterólogos. Diferentes vectores así como procedimientos para la introducción de ADN heterólogo en el sitio deseado de baculovirus están disponibles para el experto en la técnica, para la expresión de baculovirus. También diferentes señales para una modificación posterior a la traducción, reconocidas por células de insectos, son conocidas en la técnica.

La presente invención también se refiere a una célula hospedadora tal y como se ha definido anteriormente, transformada con un vector recombinante como se ha definido anteriormente.

La presente invención también se refiere a un kit diagnóstico para su uso en determinar la presencia de un poli(ácido nucleico) de virus TT de la invención, comprendiendo dicho kit un cebador y/o una sonda de la invención.

Como alternativa, la presente invención también se refiere a un procedimiento para la detección de un poli(ácido nucleico) de VTT reorganizado de acuerdo con la invención presente en una muestra biológica, que comprende: (a) opcionalmente extraer una muestra de poli(ácido nucleico), (b) amplificar el poli(ácido nucleico) como se ha descrito anteriormente con al menos un cebador como se ha definido anteriormente, opcionalmente un cebador marcado y (c) detectar los poli(ácidos nucleicos) amplificados.

La expresión "poli(ácido nucleico)" también puede citarse como hebra de analito y corresponde a una molécula de poli(ácido nucleico) mono o bicatenaria.

El término "marcado" se refiere al uso de ácidos nucleicos marcados. Esto puede incluir el uso de nucleótidos marcados incorporados durante la etapa de polimerasa de la amplificación o cebadores marcados o mediante cualquier otro procedimiento conocido por el experto en la materia.

La presente invención también se refiere a un procedimiento para la detección de un poli(ácido nucleico) de VTT reorganizado de acuerdo con la invención presente en una muestra biológica, que comprende: (a) opcionalmente extraer una muestra de poli(ácido nucleico), (b) hibridar el poli(ácido nucleico) como se ha descrito anteriormente con al menos una sonda como se ha definido anteriormente y (c) detectar los poli(ácidos nucleicos) hibridados.

Las condiciones de hibridación y lavado deben entenderse como rigurosas y son generalmente conocidas en la técnica (por ejemplo, Maniatis y col., *Molecular Cloning: A Laboratory Manual*, Nueva York, Cold Spring Harbor Laboratory, 1982). Sin embargo, según la solución de hibridación (SSC, SSPE, etc.), estas sondas deben hibridarse a su temperatura adecuada para lograr una especificidad suficiente.

Según la solución de hibridación (SSC, SSPE, etc.), estas sondas deben hibridarse de manera rigurosa a su temperatura adecuada para lograr una especificidad suficiente. Sin embargo, mediante una ligera modificación de las sondas de ADN, ya sea añadiendo o eliminando uno o unos pocos nucleótidos en sus extremos (3' o 5') o sustituyendo algunos nucleótidos no esenciales (es decir, nucleótidos no esenciales para discriminar entre tipos) por otros (incluyendo nucleótidos modificados o inosina) puede hacerse que estas sondas o variantes de las mismas hibriden específicamente en las mismas condiciones de hibridación (es decir, la misma temperatura y la misma solución de hibridación). Asimismo, el cambio de la cantidad (concentración) de sonda usada puede ser beneficioso para obtener resultados de hibridación más específicos. Cabe destacar en este concepto, que las sondas de la misma longitud, independientemente de su contenido de GC, hibridarán específicamente aproximadamente a la misma temperatura en las soluciones TMACI.

Los procedimientos de ensayo adecuados a efectos de la presente invención para detectar híbridos entre las sondas oligonucleotídicas y las secuencias de poli(ácido nucleico) en una muestra pueden comprender cualquiera de los formatos de ensayo conocidos en la técnica, tales como el formato de transferencia por puntos convencional, hibridación de tipo sándwich o hibridación inversa. Por ejemplo, la detección puede lograrse usando un formato de transferencia por puntos, uniéndose la muestra amplificada sin marcar a una membrana, incorporándose la membrana con al menos una sonda marcada en condiciones de hibridación y lavado adecuadas y controlándose la presencia de sonda unida.

Un procedimiento alternativo y preferido es un formato de transferencia por puntos "inversa", en la que la secuencia amplificada contiene un marcador. En este formato, las sondas oligonucleotídicas no marcadas se unen a un soporte

sólido y se exponen a la muestra marcada en condiciones de hibridación y lavado posterior rigurosas adecuadas. Cabe destacar que también puede usarse otro procedimiento de ensayo que se base en la formación de un híbrido entre los poli(ácidos nucleicos) de la muestra y las sondas oligonucleotídicas de acuerdo con la presente invención.

5 Finalmente, en el presente documento se desvela un proceso para la replicación y propagación *in vitro* torque teno virus (VTT), preferentemente un VTT reorganizado de acuerdo con la presente invención, que comprende las siguientes etapas:

- (a) transfectar ADN de VTT linealizado en células 293TT que expresan altos niveles del antígeno T grande SV40, preferentemente al menos niveles como los comunicados en Buck y col. (2004);
- 10 (b) recoger las células y aislar las células que muestran la presencia de ADN de VTT;
- (c) cultivar las células obtenidas en la etapa (b) durante al menos tres días, preferentemente al menos una semana o más, dependiendo de las condiciones experimentales y el tipo de VTT concreto; y
- (d) recoger las células de la etapa (c).

Los siguientes ejemplos ilustran la invención.

Ejemplo 1

15 Materiales y procedimientos

(A) Aislamiento y caracterización de virus TT

El aislamiento de los aislados de virus TT VTT-HD3a (tth8, n.º de referencia AJ620231) y VTT-HD1a (tth25, n.º de ref. AJ620222) se ha descrito con anterioridad (Jelcic y col., 2004). Se clonaron las secuencias genómicas de longitud completa de VTT-HD3a y VTT-HD1a en el vector pUC18 usando las enzimas de restricción Sail (Leppik y col., 2007) y EcoRI, respectivamente. Se identificaron secuencias de VTT adicionales en muestras humanas mediante amplificación de ADN anidada usando los cebadores NG472/NG352 y NG473/NG351 como se ha descrito con anterioridad (Peng y col., 2002; Leppik y col., 2007). La disponibilidad limitada de ADN para una serie de muestras de biopsia y suero requirió antes de la amplificación el uso de amplificación de círculo rodante con un kit TempliPhi (GE Healthcare). Se clonaron y secuenciaron todos los productos amplificados (Leppik y col., 2007).
 20 Posteriormente, las muestras que portaban ADN de virus TT se sometieron a amplificación PCR de larga distancia usando la enzima *Taq* de TaKaRa LA (TAKARA BIO INC., Japón) y los cebadores respectivos que se habían diseñado basándose en las secuencias de ADN de VTT identificadas inicialmente. Estos cebadores consecutivos incluyeron las siguientes combinaciones: tth25-1s y tth25-2as, jt34f-1s y jt34f-2as, jt34f-7s y jt34f-8as, jt34f-5s y jt34f-6as, tth4-1s y tth4-2as, t3pb-1s y t3pb-2as, así como tth8-1s y tth-2as (tabla 2). La amplificación por PCR larga se
 25 llevó a cabo usando una reacción de aterrizaje por etapas como se ha descrito anteriormente (Leppik y col., 2007) con la excepción de combinaciones de cebadores t3pb-1/2, jt34f-5/6 y tth4. Las condiciones de la PCR para la amplificación por PCR con los cebadores t3pb-1/2 y jt34f-5/6 fueron una desnaturalización inicial a 94°C durante 1 min, seguido de 30 ciclos de 94°C durante 30 s, hibridación a 65°C durante 1 min y elongación a 72°C durante 4 min con una elongación final a 72°C durante 10 min. Las condiciones de la PCR para la amplificación con los cebadores
 30 tth4 fueron similares, salvo por que la hibridación se llevó a cabo a 68°C. Todos los amplicones obtenidos en el intervalo de 3,8 kb se eluyeron y purificaron tras la electroforesis en gel, se clonaron en el vector pCR2.1 (TA-Cloning-Kit, Invitrogen) y se propagaron en células competentes NovaBlue Singles (Merck Chemicals, R. U.). Todos los genomas de longitud completa se secuenciaron a lo largo de ambas hebras. Se obtuvo un total de 53 genomas de longitud completa.

(B) Análisis de secuencia y filogenia

Las secuencias de ADN se compararon con secuencias de VTT en todos los bancos de datos usando el paquete informático HUSAR (Jelcic y col., 2004). El ICTV clasificó recientemente los virus TT en la familia *Anelloviridae* basándose en la secuencia de ADN del marco abierto de lectura 1 grande (ORF1) (Biagini y de Micco, 2010). La
 45 caracterización de los genomas de los aislados obtenidos reveló la reorganización de secuencias en la región ORF1. Los genomas de longitud completa del género Alfatorque virus y los aislados se sometieron, por tanto, a análisis filogenéticos como se ha descrito anteriormente (Jelcic y col., 2004). El árbol filogenético (figura 4) se presentó usando el programa Treeview de la University of Glasgow. Los ORF traducidos se analizaron respecto de proteínas homólogas y dominios funcionales usando ProtSweep (del Val y col., 2004).

(C) Cultivo celular y transfección

50 Se mantuvo la estirpe celular de riñón embrionario humano 293TT (Buck y col. 2004) en DMEM complementado con suero de ternero fetal al 10 %, Glutamax al 1 %, aminoácidos no esenciales al 1 % (ambos de Invitrogen, Karlsruhe, Alemania) y 400 µg/ml de higromicina B (Roche Diagnostics, Mannheim). El ADN de virus linealizado (2 µg por pocillo en placas de 6 pocillos) se transfectó el células cultivadas sin higromicina B usando el reactivo Lipofectamine (Invitrogen) según las instrucciones del fabricante (Fei y col., 2005). El medio de cultivo (2 ml) se complementó con
 55 800 µl de Opti-MEM antes de la incubación durante 4 horas a 37°C. Los cultivos transfectados se incubaron posteriormente con medio fresco que contenía higromicina B y se propagaron cuando se alcanzó la confluencia. Se transfectaron los genomas de longitud completa de 12 aislados de VTT, se mantuvieron y se recogieron en paralelo

en todos los instantes. Los genomas de virus TT incluyeron VTT-HD14a, VTT-HD14b, VTT-HD14c, VTT-HD14e, VTT-HD15a, VTT-HD16a, VTT-HD20a, VTT-HD3a, VTT-HD1a, VTT-HD23a, VTT-HD23b y VTT-HD23d (tabla 3).

Tabla 3

<i>Genomas de TT de longitud completa (3,8 kb)</i>		<i>Genomas subv3ricos</i>
tth25	HD1a	μ TTV-HD1 – zpr9.B1.6 (621nt)
tth3	HD1b	
tth9	HD1c	
tth16	HD1d	
tth17	HD1e	
tth26	HD1f	
tth27	HD1g	
tth31	HD1h	
tth5	HD2a	
tth14	HD2b	
tth29	HD2c	
tth8	HD3a	
tth7	HD3b	
tth13	HD3c	
tth19	HD3d	
tth22g4	HD3e	
tth23	HD3f	
tth4	HD4	
tth10	HD5a	
tth11g2	HD5b	
tth18	HD5c	
tth21	HD5d	
tth6	HD6a	
tth20	HD6b	
t32c2	HD7	
t32b8	HD8	
sle1957	HD9	
sle1931	HD10a	
sle1932	HD10b	
sle2045	HD10c	
sle2037	HD11	
sle2065	HD12a	
sle2057	HD12b	
sle2058	HD12c	
sle2061	HD12d	
sle2072	HD12e	
gB20.33	HD13a	
gB20.58	HD13b	
gB21.51	HD13c	
gbDhDi33.32	HD14a	μ TTV-HD14.1 – zpr4.B5.20 (719nt)
gbCuCv33.2	HD14b	μ TTV-HD14.2 – zpr4.B6.125 (1224nt)
gbDhDi33.31	HD14c	
gbDhDi33.33	HD14d	
gbDhDi33.35	HD14e	
gbDhDi32.36	HD14f	

(continuación)

<i>Genomas de TT de longitud completa (3,8 kb)</i>		<i>Genomas subvéricos</i>
<i>gbDfDg33.45</i>	<i>HD14g</i>	
<i>gbDfDg33.48</i>	<i>HD14h</i>	
<i>gbDfDg33.49</i>	<i>HD14i</i>	
<i>gbCsCt38.1</i>	<i>HD15b</i>	
<i>gbCsCt38.2</i>	<i>HD15a</i>	μ TTV-HD15 – zpr5.B4.12 (913nt)
<i>gbCsCt38.4</i>	<i>HD15c</i>	
<i>gbCsCt38.6</i>	<i>HD15d</i>	
<i>gbCsCt43.2</i>	<i>HD16a</i>	
<i>gbCsCt43.1</i>	<i>HD16b</i>	
<i>gbCsCt43.3</i>	<i>HD16c</i>	
<i>gbCsCt43.5</i>	<i>HD16d</i>	
<i>gbCsCt43.6</i>	<i>HD16e</i>	
<i>gbCuCv43.1</i>	<i>HD16f</i>	
<i>gbCuCv43.4</i>	<i>HD16g</i>	
<i>gbDhDi43.1</i>	<i>HD16h</i>	
<i>gbDhDi43.4</i>	<i>HD16i</i>	
<i>gbDhDi43.6</i>	<i>HD16j</i>	
<i>gbDhDi43.7</i>	<i>HD16k</i>	
<i>gbDhDi43.22</i>	<i>HD16l</i>	
<i>uro702</i>	<i>HD17</i>	
<i>uro703</i>	<i>HD18a</i>	
<i>uro705</i>	<i>HD18b</i>	
<i>rheu242</i>	<i>HD19</i>	
<i>uro960</i>	<i>HD20a</i>	
<i>uro742</i>	<i>HD20b</i>	
<i>uro745</i>	<i>HD20c</i>	
<i>uro746</i>	<i>HD20d</i>	
<i>uro953</i>	<i>HD20e</i>	
<i>uro958</i>	<i>HD20f</i>	
<i>rheu111</i>	<i>HD21</i>	
<i>rheu112</i>	<i>HD22</i>	
<i>rheu215</i>	<i>HD23a</i>	
<i>rheu210</i>	<i>HD23b</i>	μ TTV-HD23.1 – zpr12.B2.22 (401nt)
<i>rheu211</i>	<i>HD23c</i>	μ TTV-HD23.2 – zpr12.B5.24 (642nt)
<i>rheu212</i>	<i>HD23d</i>	
<i>rheu213</i>	<i>HD23e</i>	
<i>rheu214</i>	<i>HD23f</i>	
<i>rheu231</i>	<i>HD24b</i>	
<i>rheu232</i>	<i>HD24a</i>	
<i>rheu234</i>	<i>HD24c</i>	
<i>rheu236</i>	<i>HD24d</i>	
<i>rheu238</i>	<i>HD24e</i>	
<i>rheu241</i>	<i>HD24f</i>	

El ADN de virus se liberó del vector antes de la transfección. Los controles incluyeron transfección solo con vector y células transfectadas con 1 x TE. Las células transfectadas y el medio de cultivo se congelaron a -80°C y se tomaron muestras para la extracción de ADN y ARN en cada punto de tiempo durante la propagación. El ADN se extrajo con fenol-cloroformo-alcohol isoamílico y el ARN usando el kit RNeasy Mini (Qiagen, Hilden, Alemania). Se controló la replicación del ADN de virus y se demostró mediante amplificación por PCR larga como se ha descrito anteriormente. Todos los experimentos de transfección se llevaron a cabo 3 veces a intervalos de 6 semanas entre

las transfecciones primarias. Las células congeladas o las preparaciones de virus purificadas se pasaron entre 4 a 6 veces.

(D) Propagación de virus, purificación y microscopía electrónica

5 Se recogieron las células transfectadas de los matraces mediante agitación, seguido de centrifugación durante 10 min a 200 g. Los sedimentos celulares se resuspendieron en DPBS-Mg (Invitrogen) y se separaron en gradientes de Optiprep a etapas del 27-33-39 % (Sigma, St. Louis, MO) durante 3,5 h a 234.000 g (Buck y col., 2005). Los gradientes se fraccionaron y exploraron respecto de la presencia de ADN de virus mediante electroforesis en gel de alícuotas lisadas. Las alícuotas se lisaron con proteinasa K, EDTA 0,25 mM y SDS al 0,5 % durante 10 min a 56°C inmediatamente antes de cargarlas en el gel. El sobrenadante de las células resuspendidas se filtró, como alternativa, a través de un filtro de 0,22 µm. Las alícuotas de fracciones de gradiente, así como los sobrenadantes filtrados, se congelaron a -80°C para su uso como inóculo. Las alícuotas filtradas se sedimentaron. Los sedimentos se sometieron a tinción negativa y se visualizaron mediante microscopía electrónica. Los genomas de µVTT subvíricos se transfectaron en 293TT del mismo modo que los genomas de longitud completa. Los cultivos se propagaron a lo largo de varias semanas. Las células se retiraron parcialmente raspando una parte de las monocapas de células, mientras que se permitía el recrecimiento de las demás células. Las células retiradas se sedimentaron y el sobrenadante se filtró a través de un filtro de 0,22 µm antes de su visualización por microscopía electrónica. Los sedimentos celulares se trataron como se ha descrito anteriormente antes de la centrifugación y separación mediante gradientes de Optiprep. Se lisaron alícuotas y se visualizó el ADN después de la electroforesis en gel.

(E) Análisis de transcripción

20 Se analizaron transcritos de genomas de longitud completa de VTT-HD usando dos estrategias diferentes. Los productos de RACE 5' y 3' se generaron a partir de ADNc tanto mono como bicatenario. Los ADNc monocatenarios preparados para 5'-RACE y 3'- RACE se sintetizaron, respectivamente, a partir de 1 µg de ARN total purificado en una mezcla de reacción de 10 µl usando el kit de amplificación de ADNc SMART-er™RACE (Clontech n.º de cat. 634923) en el que el ARN se retrotranscribe mediante la retrotranscriptasa SMARTScribe™ a 42°C durante 90 min. Se usó el cebador A de 3'RACE-CDS para la síntesis del ADNc preparado para 3'RACE, mientras que se usaron el cebador A de 5'RACE-CDS y el oligonucleótido SMARTer IIA para la síntesis del ADNc preparado para 5'-RACE. El ADNc bicatenario se sintetizó de manera concomitante. En este caso, el ADNc monocatenario de longitud completa se sintetizó inicialmente usando el kit de síntesis de ADNc por PCR SMARTer™ (Clontech, n.º de cat. 634925) según el protocolo del fabricante. El ARN total purificado (1 µg) se transcribió usando la retrotranscriptasa SMARTScribe™ y los cebadores 3'SMART CDS PrimerIIA y el oligonucleótido SMARTer IIA. Estos cebadores contienen en ambos casos una serie de nucleótidos no de molde, creando de este modo un molde extendido. La amplificación de la segunda hebra de ADNc se obtuvo mediante amplificación por PCR de larga distancia (LD PCR) con el cebador IIA de 5'PCR y la mezcla de polimerasa Advantage 2 (Clontech n.º de cat. 639201). La amplificación por PCR se llevó a cabo del siguiente modo: 15 s a 95°C, 30 s a 65°C y 3 min a 68°C por ciclo y variando el número de ciclos para determinar las condiciones óptimas.

La amplificación por PCR 5' y 3'-RACE se llevó a cabo usando ADNc preparado para 5'-RACE o preparado para 3'-RACE, respectivamente o molde de ADNc bicatenario en ambos casos. La RACE-PCR se llevó a cabo usando mezcla de polimerasa Advantage 2, una mezcla universal de cebador A (UPM) del kit de amplificación de ADNc SMARTer™RACE y cebadores directos e inversos que se ajustan a los respectivos tipos de VTT (tabla 4).

40

Tabla 4

Posiciones de nucleótidos de los cebadores usados para la amplificación por PCR en RACE			
VTT	cebador	Número de nucleótidos	transcrito
VTT-HD14b	1-f1	716-743	+
	1-f3	2886-2912	+
	1-r1	757-730	+
	1-r2	3521-3492	+
VTT-HD14c	2-f1	716-743	+
	2-f2	3054-3082	+
	2-r3	2912-2885	+
VTT-HD14a	3-f1	717-744	+
	3-f2	2890-2917	+
	3-f3	3496-3521	+

(continuación)

Posiciones de nucleótidos de los cebadores usados para la amplificación por PCR en RACE			
VTT	cebador	Número de nucleótidos	transcrito
VTT-HD14e	3-r1	745-720	-
	3-r2	2914-2887	+
	4-f1	2887-2914	+
	4-f2	3494-3519	+
	4-f3	3053-3080	+
	4-r1	757-730	+
	4-r2	2911-2884	+
VTT-HD15a	5-f1	125-149	+
	5-f2	2807-2834	+
	5-f3	3388-3415	+
	5-r1	224-197	+
	5-r2	3014-2987	+
	5-r3	3425-3398	-
VTT-HD16a	6-f1	100-127	+
	6-f2	3145-3172	-
	6-f3	3564-3591	-
	6-r1	3204-3182	+
	6-r2	3443-3418	-
VTT-HD20a	7-f1	314-341	+
	7-f2	3025-3052	+
	7-r1	227-200	+
	7-r2	743-716	+
	7-r3	3332-3305	-
VTT-HD23b	10-f1	113-139	-
VTT-HD23d	10-f3	3121-3148	+
	11-f1	126-148	+
	11-f2	354-381	+
	11-f3	3397-3422	-
	11-r1	226-199	+
	11-r2	3653-3626	+
VTT-HD23a	11-r3	3327-3302	+
	12-f1	126-148	+
	12-f2	354-381	+
	12-r2	3177-3150	+
	12-r3	3326-3301	+

Las condiciones de amplificación fueron: 29 ciclos de 30 s a 94°C, hibridación durante 30 s a 68°C y elongación durante 3 min a 72°C, con una extensión final durante 15 min a 72°C. Todos los productos se analizaron mediante electroforesis en gel, se purificaron después de la elución del gel, se clonaron en el vector pCR2.1 (Invitrogen, n.º de cat. K2020-40) y se secuenciaron. Se llevaron a cabo dos controles adicionales para controlar la amplificación no específica. En un control, la amplificación se llevó a cabo usando únicamente un cebador específico para VTT y en el segundo usando el cebador UPM solo. No se detectaron productos en cualquiera de estos.

Ejemplo 2

Demostración de la persistencia del ADN de VTT en células procedentes de líneas de cultivo de tejidos obtenidas a partir de tumores malignos

Las líneas celulares obtenidas a partir de tumores malignos poseen una ventaja sobre el material de una biopsia de

5 tumor primario. Comúnmente representan preparaciones puras de células de cáncer, mientras que los materiales primarios generalmente están contaminados por células mesenquimatosas normales, por células del sistema hematopoyético y células epiteliales normales. Por otra parte, un inconveniente de las líneas de cultivo de tejidos puede surgir de la selección de clones específicos que crecen en condiciones de cultivo de tejidos y la adquisición de modificaciones genéticas secundarias durante el curso de un cultivo a largo plazo. Además, los sueros de ternera fetal pueden suponer un riesgo debido a la introducción de virus del ganado que sobreviven después de procedimientos de inactivación del suero (por ejemplo, poliomavirus bovino); véase la tabla 5 que resume estas ventajas/inconvenientes.

Tabla 5

Análisis de biopsias de tumores primarios frente a líneas celulares establecidas para secuencias relacionadas con el VTT			
Biopsias		Líneas celulares	
Ventaja	Inconveniente	Ventaja	Inconveniente
Materiales auténticos	<p>Contaminadas por mezcla por adición de células normales</p> <p>Búsqueda complicada de secuencias de VTT por la presencia uniforme de VTT en la sangre periférica</p> <p>Disponibilidad limitada</p>	<p>Preparaciones puras de células de cáncer</p> <p>Disponibles en cantidades ilimitadas</p>	<p>Selección de clones específicos adaptados a las condiciones de cultivo de tejidos</p> <p>Cambios genéticos secundarios durante el cultivo a largo plazo</p> <p>El uso de suero de ternera fetal plantea el riesgo de contaminaciones con virus de ganado</p>

10 Los intentos de encontrar ADN de VTT en materiales tumorales primarios humanos adolece de un inconveniente: la pluralidad de genotipos de VTT en el material humano. Esto hace que sea prácticamente imposible identificar un genotipo específico como un agente etiológico para un tipo de cáncer humano. Por estos motivos, se iniciaron estudios de la persistencia de secuencias de ADN de VTT en células obtenidas a partir de líneas de cultivo de tejidos de cáncer. Hasta ahora, los resultados han sido extremadamente sorprendentes: los cebadores para la PCR usados para descubrir regiones del marco de lectura abierto de VTT han sido del todo infructuosos. Sin embargo, otras combinaciones de cebadores, que descubrirían exclusivamente una región reguladora corta, rica en GC del genoma VTT de aproximadamente 71 bases, detectaron esa secuencia en un mayor número de líneas celulares (Figura 1). Esta región reguladora está muy conservada entre diferentes genotipos de VTT y no está presente en la base de datos del genoma humano.

20 En una primera serie de experimentos, la misma secuencia fue descubierta en una serie de líneas celulares adicionales. Estas incluyen las siguientes líneas:

- *MCF7* (línea de cáncer de mama);
- *HAK-1*, *KMH-2*, *L1236* (todas líneas de linfoma de Hodgkin negativas para el virus Epstein-Barr);
- *Y69* (linfoma B negativa para el virus Epstein-Barr)
- 25 • *HSB-2* (leucemia linfocítica aguda);
- *P3HR-1* (linfoma de Burkitt positiva para el virus Epstein-Barr);
- *BJAB* (linfoma de Burkitt negativa para el virus Epstein-Barr);
- *Ng* (linfoblastos B inmortalizados con EBV de un paciente con esclerosis múltiple).

30 Además de estas 9 líneas positivas, dos líneas celulares de melanoma (IGL y KR, Fig. 1) y ADN de la placenta humana eran negativos en los experimentos iniciales. Curiosamente, después de la eliminación del ADN enrollado de las células L1236 y el tratamiento con ARNasa de la solución restante, además del ADN mitocondrial, se convirtieron en bandas visibles dos bandas débiles de tamaño similar entre las posiciones 4,3-6,6 kb (marcador del tamaño del ADN de cadena doble) en los geles de agarosa (Fig. 2). Un análisis de estas secuencias reveló de nuevo la presencia de la región reguladora de VTT. La nucleasa de semilla de Mungo, que digiere selectivamente el ADN de cadena sencilla, suprimió completamente las bandas que contenían ADN celular procedentes de cuatro biopsias de esclerosis múltiple, en contraste con el ADN de cadena doble de control, lo que subraya la naturaleza de una sola hebra del primero. Estudios similares se están realizando actualmente para aislados de ADN tumoral.

Ejemplo 3

Análisis de secuencias de ADN de VTT quimérico/células hospedadoras truncadas

Inicialmente, todos los intentos fallaron para el uso de cebadores con orientación hacia fuera que comenzaban dentro de la región reguladora, con el fin de encontrar ADN del virus TT flanqueante, que rodeaba esta región. Invariablemente, sin embargo, el ADN celular humano se pudo demostrar en los respectivos clones (Fig. 3).

- 5 Los genes humanos en esos clones y sus disposiciones dentro del ADN episómico de una sola hebra, controlados obviamente por la región de 71 bases de VTT, están siendo analizados actualmente. Los datos disponibles indican una variación sustancial en la captación de genes de células hospedadoras comúnmente truncadas. Su posible conversión en oncogenes estimulantes del crecimiento o en funciones que interfieren con genes supresores tumorales, requiere pruebas funcionales que se están investigando actualmente. Lo mismo explica las secuencias de virus VTT reorganizadas. Algunos de los datos disponibles se presentan en las Figuras 7, 8, 9 y 11 a 13.

Ejemplo 4

Identificación y caracterización de genomas de VTT

15 La amplificación de la región rica en GC conservada corta de los virus TT en suero y muestras de biopsia dio lugar a la identificación de ADN de VTT en la mayoría de casos. La posterior amplificación del genoma completo es necesaria para identificar tipos específicos de VTT ya que muchos comparten homología de ADN exacta en los 71 pb amplificados que se encuentran en la región de control, pero difieren en hasta un 60-80 % en la identidad de secuencia en el resto de sus genomas. Se diseñó una serie de combinaciones de cebadores consecutivos en secuencias obtenidas durante el transcurso de las investigaciones (tabla 2). La amplificación por PCR de larga distancia se llevó a cabo en muestras positivas a ADN de VTT. Se clonaron y secuenciaron amplicones en el intervalo de 3 a 4 kb. Las muestras positivas para ADN de VTT se originaron de sujetos sanos, así como de pacientes con leucemia, esclerosis múltiple, artritis reumatoide y enfermedad renal. Se ha descrito previamente parte de estos datos (Leppik y col., 2007; Sospedra y col., 2005; de Villiers y col., 2009).

25 Se caracterizó un total de 53 genomas de ADN de longitud completa. Se identificaron hasta 12 aislados de longitud completa tras secuenciar 19 genomas de una sola biopsia. La organización del genoma de diferentes aislados de un tipo de VTT varió a pesar de la baja diversidad de nucleótidos (en el intervalo del 1 - 4 %). Aunque estaba implicado principalmente el marco de lectura abierto OFR1, también se observaron diferencias dentro de la región no codificante y otros genes. Estos datos confirmaron observaciones más tempranas (Jelcic y col., 2004; Leppik y col., 2007; de Villiers y col., 2009). Las modificaciones en el ORF1 incluyeron codones de parada prematura que daban lugar a ORF más pequeños individuales en esta región, una diversidad de secuencia considerable en la región hipervariable (Nishizawa y col., 1999; Jelcic y col., 2004) o la ausencia de un codón de parada, que daba como resultado un ORF1 más grande que el presente en el prototipo (figura 16). La clasificación oficial de la familia *Anelloviridae* se basa en comparaciones de las secuencias de ADN del ORF1 (Biagini y de Micco, 2010). Debido a las modificaciones en el ORF1 en los aislados obtenidos, se incluyeron las secuencias genómicas de longitud completa en los análisis filogenéticos aquí presentados. El objetivo de este análisis fue tener una visión panorámica de los aislados de VTT-HD en relación con las especies de VTT establecidas (figura 18). Todos los aislados previos también se incluyen en este árbol (Jelcic y col., 2004; Leppik y col., 2007; de Villiers y col., 2009).

Ejemplo 5

Replicación *in vitro* de VTT-HD

40 Se ha informado repetidamente en el pasado de intentos para asociar la infección por torque teno virus con la patogénesis de una enfermedad específica. Se han analizado muestras de una gran variedad de enfermedades. Las investigaciones *in vitro* se vieron dificultadas por los intentos negativos para identificar un sistema de cultivo celular en el que estos virus pueden propagarse fácilmente a lo largo de periodos de tiempo más largos. Las partículas víricas se caracterizaron inicialmente con la ayuda de gradientes de densidad y agregados de inmunoglobulina (revisado en Okamoto, 2009) y posteriormente se visualizaron a partir de sueros y heces (Itoh y col., 2000). Los torque teno virus se producen principalmente en células del sistema hematopoyético (Okamoto, 2009). Los primeros aislados se obtuvieron del bazo de un paciente con linfoma de Hodgkin (Jelcic y col., 2004). Por lo tanto, se usó la línea celular L428 en los intentos iniciales para demostrar la replicación y la transcripción *in vitro* de VTT-HD3a. Se logró la replicación del genoma de longitud completa durante hasta 7 días después de la transfección del ADN vírico linealizado (Leppik y col., 2007). Para prolongar este periodo de replicación, se transfectaron genomas de VTT de longitud completa en la estirpe celular de riñón embrionario humano 293TT que se había modificado por ingeniería genética para expresar altos niveles del antígeno T grande de SV40 (Buck y col., 2004). En segundo lugar, se decidió incluir 12 aislados de longitud completa en este estudio para determinar si 1) las variaciones en el ORF1 podrían influenciar la replicación y la formación de partículas víricas, 2) los tipos divergentes de VTT varían en su modo de replicación. Se tomó un cuidado extremo para propagar los 12 aislados en paralelo para excluir las variaciones en la medida de lo posible, lo que puede producirse durante la manipulación.

Se seleccionaron los siguientes aislados para la transfección y la propagación: VTT-HD3a (Leppik y col., 2007) y VTT-HD1a (Jelcic y col., 2004). VTT-HD1a es el más estrechamente relacionado con la especie VTT3 (hel32) y VTT-HD3a a la especie VTT12 (ct44f) (figura 4). VTT-HD16a (relacionado con la especie VTT22), VTT-HD15a

(relacionado con la especie VTT12), VTT-HD14a, VTT-HD14b, VTT-HD14c y VTT-HD14e (relacionados con la especie VTT29) se aislaron de biopsias cerebrales de pacientes con esclerosis múltiple. VTT-HD20a (relacionado con la especie VTT13-) se originó de tejido renal y VTT-HD23a, VTT-HD23b y VTT-HD23d (relacionados con la especie VTT3) se amplificaron de suero extraído de pacientes con artritis reumatoide. Las secuencias de VTT-HD14a, VTT-HD14b, VTT-HD14c y VTT-HD14e varían entre un 1-2 % en sus genomas de longitud completa. El prototipo es VTT-HD14a con un ORF1 intacto de 648 aminoácidos (aa) de tamaño. El ORF1 de VTT-HD14b tiene un tamaño de 660aa compartiendo identidad solo 554aa con ORF1 de VTT-HD14a, mientras que el resto de los ORF indican fusión al ORF4 (según de Schmidt y Noteborn, 2009). Del mismo modo, VTT-HD14c ORF1 tiene 712aa y constituye el ORF1 (primeros 645aa) fusionado al ORF5. El ORF1 de VTT-HD14e se interrumpe, dando como resultado 2 ORF de 467aa y 179aa de tamaño. Los genomas de VTT-HD23b, VTT-HD23d y VTT-HD23a varían tan solo en un 1-3 % de identidad de secuencia, pero sus genes de ORF1 difieren del siguiente modo: VTT-HD23a ORF1 como prototipo tiene un tamaño de 736aa, la secuencia de ADN del ORF1 de VTT-HD23b varía respecto de la de VTT-HD23a en la región hipervariable en un 18,4 % (34,2 % de los aminoácidos). Las secuencias de ADN de VTT-HD23b y VTT-HD23d difieren tan solo en un 1 % en su identidad general, pero el ORF1 de VTT-HD23d está interrumpido, dando como resultado 2 ORF de 307aa y 365aa de tamaño (figura 16).

Las transfecciones se efectuaron en células 293TT semiconfluentes. La naturaleza de esta estirpe celular, con sus múltiples células redondeadas unidas a la monocapa, no permite una identificación de corte clara de los efectos citopáticos. Las células se pasaron cuando alcanzaron la confluencia o cuando las células comenzaron a desprenderse de la superficie. Se agitaron los matraces para desprender todas las células. Se centrifugaron las células y se congelaron las alícuotas, así como se usaron para la extracción de ADN y ARN y los análisis por microscopía electrónica. Las células infectadas congeladas se usaron inicialmente para volver a infectar nuevos cultivos de 293TT, ya que la reinfección fracasaba si las células se habían tripsinizado en el momento de la recogida. La replicación vírica se llevó a cabo llevando a cabo una PCR de larga distancia en ADN extraído de células infectadas. Los periodos entre la reinfección y la recogida de las células varió entre 3 a 7 días, dependiendo de la densidad del cultivo. No se observaron diferencias morfológicas evidentes entre los cultivos de diferentes aislados de VTT. La reinfección durante el transcurso de un experimento se llevó a cabo varias veces usando alícuotas de células congeladas. No se ha descrito con anterioridad la propagación de virus TT *in vitro*. La digestión con enzimas de restricción se llevó a cabo en ADN celular obtenido de las muestras transfectadas inicialmente para eliminar cualquier ADN vírico generado por bacterias residuales. Los resultados de la amplificación por PCR larga indicaron la replicación de novo de ADN vírico. Los ejemplos de estos amplicones de ADN de VTT usando ADN celular infectado como molde se representan en la figura 19.

La amplificación por PCR de larga distancia de las moléculas de ADN de longitud completa indicó diferencias considerables entre los cultivos. Se necesitaron segundas rondas de amplificación (usando los mismos cebadores que en la primera ronda) en todos los cultivos infectados con aislados de biopsias cerebrales, es decir, VTT-HD16a, HD15a y los 4 aislados individuales de VTT-HD14 (figura 19A), a pesar de su divergencia (45-50 % de homología de nucleótidos) según los análisis filogenéticos (figura 18). Las modificaciones en el ORF1 no parecían tener influencia en la amplificación o la propagación, según se visualizó en la amplificación del ADN de longitud completa (Fig. 21A a-c). Se observaron amplicones de ADN adicionales de diversos tamaños en los cultivos infectados por HD15a. La aparición de estas moléculas aumentó durante la posterior propagación con una reducción concomitante en el genoma de longitud completa (figura 21A a-c carril 5). Los presentes inventores habían comunicado moléculas subvíricas de una naturaleza similar en muestras de suero humano (Leppik y col., 2007). En ocasiones también se observaron amplicones similares de diversos tamaños en cultivos infectados por VTT-HD16a (carril 6) y raramente en cultivos de VTT-HD14 (carriles 1-4).

Se observaron grandes diferencias en el comportamiento de los otros 6 aislados. Esta variación también fue evidente entre experimentos y pases (figura 19B b1, b2, b3) reflejando una alta sensibilidad aparente para modificaciones muy leves en las condiciones de cultivo. El genoma de longitud completa de replicación inicial (3,8 kb) se perdió durante la propagación (figura 19B a-c) en la aparición con amplicones subgenómicos prominentes con diversos tamaños en células infectadas por VTT-HD20a, VTT-HD3a y VTT-HD1a (carriles 7-9, figura 19B). Las cantidades de ADN de entrada usadas para la amplificación por PCR de larga distancia, así como de amplicones cargados en los geles fueron iguales para todos los cultivos. El alto nivel de amplicones de ADN de los aislados VTT-HD23b, VTT-HD23d y VTT-HD23a después de una sola ronda de PCR de larga distancia puede indicar, por lo tanto, un potencial de replicación más fuerte durante los pases tempranos.

Debido a las diferencias observadas entre los dos grupos de aislados, se investigó si podían observarse variaciones durante el muestreo en serie. Se propagaron pases equivalentes de VTT-HD14e y VTT-HD23b en paralelo y se tomaron muestras a diario. La amplificación de larga distancia indicó una replicación constante de VTT-HD14e (visible tras dos rondas de amplificación de ADN) lo que contrasta con la replicación decreciente de VTT-HD23b (visible después de una sola ronda de amplificación de ADN) que se perdió tras 10 días en cultivo (figura 19C). Estos cultivos no se pasaron y no se percibieron diferencias morfológicas entre cultivos.

Ejemplo 6

Formación, replicación y caracterización *in vitro* de moléculas subvíricas de μ VTT

La apariencia de amplicones de ADN más pequeños de un tamaño constante en los cultivos de los aislados de VTT-HD14b, VTT-HD14c, VTT-HD14d y VTT-HD14e, así como VTT-HD1a y los 3 aislados de VTT-HD23, ya se percibió poco después de la transfección y se mantuvo durante los pases (figura 19A y B). Se clonaron y se caracterizaron. Estas moléculas de ADN subvívico (μ VTT-HD14, 719 bases de tamaño) de VTT-HD14b y los 3 aislados de VTT-HD14 fueron idénticas en su secuencia de ADN y representaron moléculas subgenómicas reorganizadas que se originan del genoma de VTT-HD14 parental (figura 20A). Del mismo modo, se originó una molécula de ADN subvívico reorganizada (μ VTT-HD1, 621 bases) a partir del genoma de VTT-HD1a (figura 20B). De forma interesante, la replicación de μ VTT-HD1 se mantuvo durante los pases, a pesar de la desaparición del genoma de VTT-HD1a de longitud completa. Esta presencia o ausencia de las moléculas subvívicas en los cultivos de VTT-HD23 indica una posible influencia de las condiciones de cultivo. En este caso, estas moléculas tenían un tamaño que varió entre 400 a 900 bases con un nivel aumentado de 642 y 401 moléculas de base. La caracterización de las moléculas clonadas indicó un proceso de maduración evolutivamente evidente como un segmento de la molécula subvívica de 401 bases (μ VTT-HD23.1) que se duplicó en el ADN subvívico de 642 bases (μ VTT-HD23.2; figura 20C). Estaban presentes múltiples versiones de este segmento en las moléculas más grandes. Los genomas subvívicos que se originan a partir de los cultivos de VTT-HD23b, VTT-HD23d así como VTT-HD23a, tenían una secuencia de ADN idéntica. La transfección de estas moléculas subvívicas reorganizadas en células 293TT dio como resultado la replicación de sus genomas (figura 21) y se visualizó tras la amplificación por la PCR. De forma interesante, la μ VTT respectiva reaccionó exactamente del mismo modo que los genomas parentales, es decir, la replicación del ADN genómico inicial de μ VTT-HD15 fue fuerte, pero posteriormente se visualizó únicamente después de la amplificación por PCR anidada (figura 21). Fueron visibles estructuras similares a proteínas con un tamaño de 10 nm mediante microscopía electrónica después de la filtración (0,22 μ m) del medio de cultivo de estos cultivos celulares (figura 22).

Ejemplo 7

Purificación de partículas viroides (genomas completos y μ VTT)

Se iniciaron intentos de purificar partículas de virus tras la segunda ronda de reinfecciones. Los extractos celulares en bruto se centrifugaron en gradientes de Optiprep de etapas al 27-33-39 % (Buck y col., 2005). Se lisaron alícuotas de las fracciones de gradiente antes de la separación por electroforesis en gel. Las fracciones de gradiente que indicaron ADN de virus se congelaron a -80°C y se usaron para reinfecciones adicionales. Fueron claramente visibles dos bandas de ADN al nivel de 2 kb y 1,0 kb del marcador de tamaño de ADN bicatenario (figura 8A). Los tamaños exactos de estas moléculas de ADN no pudieron determinarse ya que no hay disponibles marcadores de ADN bicatenario adecuados. Además, las suspensiones celulares se filtraron a través de un filtro de 0,22 μ m antes de la centrifugación de gradiente. La tinción negativa de estas muestras indicó partículas viroides con un tamaño de aproximadamente 30 nm (figura 8). De manera similar, se observaron estructuras de proteína (aproximadamente 10 nm de tamaño) después de la filtración del medio de cultivo tras la propagación de los genomas de μ VTT-HD (figura 22). Estos filtrados se lisaron y el ADN se separó en geles de agarosa (figura 22).

Ejemplo 8

*Transcripción *in vitro**

Se han comunicado patrones de transcripción detallados de VTT para los aislados VTT-P1C1 (Müller y col., 2008), VTT-HEL32 (Qiu y col., 2005; Kakkola y col., 2009) y VTT-HD3a (Leppik y col., 2007). Se han comunicado con anterioridad tres especies principales de ARNm (1,0, 1,2 y 3,0 kb) en células de médula ósea (Okamoto y col., 2000a) y en células COS1 (Kamahora y col., 2000). Las predicciones de uso de codones de iniciación de acuerdo con las reglas de Kozak (Jelcic y col., 2004) en combinación con el uso de sitios aceptores y donadores de corte y empalme alternativo (Leppik y col., 2007) indicó el ambiente de mecanismos no conservados durante la transcripción de torque teno virus. La transcripción de los aislados se investigó usando ADN tanto mono como bicatenario como moldes para el mapeo 3' y 5'RACE. El ADNc bicatenario reduce la posibilidad de formación de híbridos no específicos. Además, se seleccionaron cebadores (directos e inversos) que estaban ubicados en las regiones intergénicas, en lugar de los cebadores específicos de genes comúnmente usados. Esto se llevó a cabo con la intención de abarcar la expresión de cualquier gen no predicho en el genoma de VTT. Se extrajo el ARN de todos los cultivos en el día 7 después de la transfección. El ARN de las transfecciones de control con solo vector se incluyó para controlar la amplificación de falsos positivos. Los análisis de transcripción se repitieron para controlar un punto de tiempo adecuado para recoger el ARNm mediante la extracción de ARN 48 después de la transfección en el caso del aislado VTT-HD14e. Los patrones de transcripción observados no diferían entre el día 2 y el día 7. Todos los resultados obtenidos en los análisis de transcripción se presentan en la figura 17.

Se aislaron transcritos abundantes a partir de cultivos infectados con VTT-HD23. Sus patrones de transcripción, así como los de VTT-HD20a, VTT-HD15a, VTT-HD16a fueron en general similares a los patrones de transcripción descritos previamente (revisado en Kakkola y col., 2009). Una excepción es la ausencia de un transcrito del ORF1 de longitud completa de todos los aislados. Esto es sorprendente, en vista del hecho de que las partículas viroides se producen de manera concomitante. Los transcritos que abarcan secciones del gen del ORF1 (ya sean los extremos 5' o 3') y que podrían codificar proteínas más pequeñas, estaban presentes (ejemplos en la figura 17). Los análisis *in silico* de supuestas proteínas revelaron información adicional de la que se había comunicado hasta la

fecha. Los ejemplos son (fusiones) de corte y empalme entre el ORF2 u ORF2a con el ORF1 o con el ORF5 en VTT-HD16a (6.3s.2, 6.3s.3, 6.3s.9), El corte y empalme entre ORF1 y ORF5 es otra posibilidad (6.3.7). También pueden expresarse transcritos cortos que abarcan la región de ORF2 en VTT-HD20a en forma de una proteína ORF1 más pequeña (7.3.5, 7.3.4, 7.5.13) (figura 17). Además, se obtuvieron transcritos usando cebadores (directos o inversos) ubicados en la región de control. Se efectuaron dos observaciones. Los cebadores inversos dieron como resultado transcritos de corte y empalme o sin corte y empalme que abarcaban regiones extendidas del genoma (12.5.19, 12.5.20, 12.5.21, 5.5s.16, 5.5s.17, 5.5s.18, 5.5s.19) o transcritos de diversas longitudes que no tenían capacidad codificante (5.5s.12, 5.5s.13, 5.5s.14, 5.5s.15, 11.5.7, 11.5.8, 11.5.9). La amplificación con cebadores directos en esta región dio como resultado otros transcritos cortos no codificantes o transcritos de corte y empalme con capacidad codificante incluso tan distantes como la ORF5 (4.3.4, 3.3.1,3.3.2) (figura 17).

Lista de referencias

1. Belotserkovskii, B.P., Liu, R., Tornaletti, S., Krasilnikova, M.M., Mirkin, S.M. y Hanawalt, P.C. 2010. Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *Proc. Natl. Acad. Sci USA*. 107:12816-12821.
2. Biagini, P., y P. de Micco. 2010. La famille des Anelloviridae: virus TTV et genres apparentés. *Virologie* 14:3-16.
3. Biagini, P., Charrel, R.N., de Micco, P., y X. de Lamballerie. 2003. Association of TT virus primary infection with rhinitis in a newborn. *Clin. Infect. Dis.* 36:128-129.
4. Buck, C.B., Pastrana, D.V., Lowy, D.R., y J.T. Schiller. 2004. Efficient intracellular assembly of papillomaviral vectors. *J. Virol.*78:751-757.
5. Buck, C.B., Pastrana, D.V., Lowy, D.R., y J.T. Schiller. 2005. Generation of HPV pseudovirions using transfection and their use in neutralization assays. *Methods Mol. Med.* 119:445-462.
6. Del Val, C., Mehrle, A., Falkenhahn, M., Seiler, M., Glattig, K-H., Poustka, A., Suhai, S., y S. Wiemann. 2004. High-throughput protein analysis integrating bioinformatics and experimental assays. *Nucleic Acid Res.* 32:742-748.
7. de Schmidt, M.H., y M.H.M. Noteborn. 2009. Apoptosis-inducing proteins in chicken anemia virus and TT virus. *Curr. Topics Microbiol. Immunol.* 331:131-149.
8. de Villiers, E-M., Kimmel, R., Leppik, L., y K. Gunst. 2009. Intragenomic rearrangement in TT viruses: a possible role in the pathogenesis of disease. *Curr. Topics Microbiol. Immunol.*331:91-107.
9. de Villiers, E-M., Schmidt, R., Delius, H., y H. zur Hausen. 2002. Heterogeneity of TT virus related sequences isolated from human tumor biopsy specimens. *J. Mol. Med.* 80:44-50.
10. Fei, J-W., Wei, Q-X., Angel, P., y E-M. de Villiers. 2005. Differential enhancement of a cutaneous HPV promoter by p63, Jun and mutant p53. *Cell Cycle* 4:689-696.
11. Garbuglia, A.R., Iezzi, T., Capobianchi, M.R., Pignoloni, P., Pulsoni, A., Sourdís, J., Pescarmona, E., Vitolo, D., y F. Mandelli. 2003. Detection of TT virus in lymph node biopsies of B-cell lymphoma and Hodgkin's disease, and its association with EBV infection. *Int. J. Immunopathol. Pharmacol.* 16:109-118.
12. Itoh, Y., Takahashi, M., Fukuda, M., Shibayama, T., Ishikawa, T., Tsuda, F., Tanaka, T., Nishizawa, T., y H. Okamoto. 2000. Visualization of TT virus particles recovered from the sera and feces of infected humans. *Biochem. Biophys. Res. Commun* 279:718-724.
13. Jelcic, I., Hotz-Wagenblatt, A., Hunziker, A., zur Hausen, H., y E-M. de Villiers. 2004. Isolation of multiple TT virus genotypes from spleen biopsy tissue from a Hodgkin's disease patient: Genome reorganization and diversity in the hypervariable region. *J. Virol.* 78:7498-7507.
14. Jeske, H. 2009. Geminiviruses. *Curr Top Microbiol Immunol.* 331:185-226
15. Kakkola, L., Bondén, H., Hedman, L., Kivi, N., Moisala, S. Julin, J., Ylä-Liedenpohja, Miettinen, S., Kantola, K., Hedman, K., y M. Söderlund-Venermo. 2008. Expression of all six human Torque teno virus (TTV) proteins in bacteria and in insect cells, and analysis of their IgG responses. *Virology* 382:182-189.
16. Kakkola, L., Hedman, K., Qiu, J., Pintel, D., y M. Söderlund-Venermo. 2009. Replication of and protein synthesis by TT viruses. *Curr. Topics Microbiol. Immunol.*331: 53-64.
17. Kakkola, L., Tommiska, J., Boele, L.C.L., Miettinen, S., Blom, T., Kekarainen, T., Qiu, J., Pintel, D., Hoeben, RC., Hedman, K., y M. Söderlund-Venermo. 2007. Construction and biological activity of a full-length molecular clone of human Torque teno virus (TTV) genotype 6. *FEBS. J.* 274:4719-4730.
18. Kamada, K., Kamahora, T., Kabat, P., y S. Hino. 2004. Transcriptional regulation of TT virus: promoter and enhancer regions in the 1.2-kb noncoding region. *Virology* 321:341-348.
19. Kamahora, T., Hino, S., y H. Miyata. 2000. Three spliced mRNAs of TT virus transcribed from a plasmid containing the entire genome in COS1 cells. *J. Virol* 74:9980 - 9986.
20. Kanda, Y., Tanaka, Y., Kami, M., Saito, T., Asai, T., Izutsu, K., Yuji, S., Ogawa, S., Honda, H., Mitani, K., Ciba, S., Yasaki, Y., y H. Hirai. 1999. TT virus in bone marrow transplant recipients. *Blood* 93: 2485-2490.
21. Kazi, A., Miyata, H., Kurokawa, K., Khan, M.A., Kamahora, T., Katamine, S., y S. Hino. 2000. High frequency of postnatal transmission of TT virus in infancy. *Arch. Virol.* 145:535-540.
22. Kovacs, E., Tompa, P., Liliom, K., y L. Kalmar. 2010. Dual coding in alternative reading frames correlates with intrinsic protein disorder. *Proc. Natl. Acad. Sci. U. S. A.* 107:5429-5434
23. Leppik, L., Gunst, K., Lehtinen, M., Dillner, J., Streker, K., y E-M. de Villiers. 2007. *In vivo* and *in vitro* intragenomic rearrangement of TT viruses. *J Virol* 81:9346-9356.
24. Maggi, F., Andreoli, E., Riente, L., Meschi, S., Rocchi, J., Delle Sedie, A., Vatteroni, ML., Ceccherini-Nelli, L.,

- Specter, S., y M. Bendinelli. 2007. Torquetenovirus in patients with arthritis. *Rheumatology* 46:885-886.
25. Maggi, F., Focosi, D., Albani, M., Lanini, L., Vatteroni, ML, Petrini, M., Ceccherini-Nelli, L., Pistello, M., y M Bendinelli. 2010. Role of hematopoietic cells in the maintenance of chronic human torquetenovirus plasma viremia. *J. Virol.* 84:6891-6893.
- 5 26. Maggi, F., Fornai, C., Vatteroni, ML., Siciliano, G., Menichetti, F., Tascini, C., Specter, S., Pistello, M., y M. Bendinelli. 2001a. Low prevalence of TT virus in the cerebrospinal fluid of viremic patients with central nervous system disorders. *J. Med. Virol.* 65:418-422
27. Maggi, F., Fornai, C., Zaccaro, L., Morrica, A., Vatteroni, M.L., Isola, P., Marchi, S., Ricchiuti, A., Pistello, M., y M. Bendinelli. 2001b. TT virus (TTV) loads associated with different peripheral blood cell types and evidence for
- 10 TT replication in activated mononuclear cells. *J. Med. Virol.* 64:190-194.
28. Maggi, F., Pifferi, M., Fornai, C., Andreoli, A., Tempestini, E., Vatteroni, M., Presciuttini, S., Marchi, S., Pietrobelli, A., Boner, A., Pistello, M., y M. Bendinelli. 2003a. TT virus in the nasal secretions of children with acute respiratory disease: relations to viremia and disease severity. *J. Virol.* 77:2418-2425.
- 15 29. Maggi, F., Pifferi, M., Tempestini, E., Fornai, C., Lanini, L., Andreoli, E., Vatteroni, M., Presciuttini, S., Pietrobelli, A., Boner, A., Pistello, M., y M. Bendinelli. 2003b. TT virus loads and lymphocyte subpopulations in children with acute respiratory diseases. *J. Virol* 77:9081-9083.
30. Mariscal, L.F., Lopez-Alcorocho, J.M., Rodriguez-Inigo, E., Ortiz-Movilla, N., de Lucas, S., Bartolome, J., y V. Carreno. 2002. TT virus replicates in stimulated but not in nonstimulated peripheral blood mononuclear cells. *Virology* 301:121-129.
- 20 31. Müller, B., März, A., Doberstein, K., Finsterbusch, T., y A. Mankertz. 2008. Gene expression of the human Torque Teno Virus isolate P/1C1. *Virology* 381:36-45.
32. Nawaz-ul-Rehman, M.S., y C.M. Fauquet. 2009. Evolution of geminiviruses and their satellites. *FEBS Letter* 583:1825-1832.
- 25 33. Nishizawa, T., Okamoto, K., Konishi, H., Yoshikawa, H., Miyakawa, Y., y M. Mayumi. 1997. A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. *Biochem. Biophys. Res. Commun.* 241:92-97.
34. Ninomiya, M., Nishizawa, T., Takahashi, M., Lorenzo, F.R., Shimosegawa, T., y H. Okamoto. 2007. Identification and genomic characterization of a novel human torque teno virus of 3.2kb. *J. Gen. Virology* 88:1939-1944.
- 30 35. Ninomiya, M., Takahashi, M., Nishizawa, T., Shimosegawa, T., y H. Okamoto. 2008. Development of PCR assays with nested primers specific for differential detection of three human anelloviruses and early acquisition of dual or triple infection during infancy. *J. Clin. Microbiol.* 46:507-514.
36. Okamoto, H. 2009. History of discoveries and pathogenicity of TT viruses. *Curr. Top. Microbiol. Immunol.* 331:1-20.
- 35 37. Okamoto, H., Nishizawa, T., Tawara, A., Takahashi, M., Kishimoto, J., Sai, T., y Y. Sugai. 2000a. TT virus mRNAs detected in the bone marrow cells from an infected individual. *Biochem. Biophys. Res. Commun.* 279:700-707.
38. Okamoto, H., Takahashi, M., Kato, N., Fukuda, M., Tawara, A., Fukuda, S., Tanaka, T., Miyakawa, Y., y M. Mayumi. 2000b. Sequestration of TT virus of restricted genotypes in peripheral blood mononuclear cells. *J. Virol.* 74:10236-10239.
- 40 39. Okamoto, H., Takahashi, M., Nishizawa, T., Tawara, A., Sugai, Y., Sai, T., Tanaka, T., y F. Tsuda. 2000c. Replicative forms of TT virus DNA in bone marrow cells. *Biochem. Biophys. Res. Commun.* 270:657-662.
40. Okamoto, H., Ukita, M., Nishizawa, T., Kishimoto, J., Hoshi, Y., Mizuo, H., Tanka, T., Miyakawa, Y., y M. Mayumi. 2000d. Circular double-stranded forms of TT virus DNA in the liver. *J. Virol.* 74:5161-5167.
- 45 41. Paprotka, T., Metzler, V., y H. Jeske. 2010. The first DNA 1-like a satellite in association with New World begomovirus in natural infections. *Virology* 404:148-157.
42. Patil, B.L, y C.M. Fauquet. 2010. Differential interaction between cassava mosaic geminivirus and geminivirus satellites. *J. Gen. Virol.* 91:1871-1882.
- 50 43. Peng, Y.H., Nishizawa, T., Takahashi, T., Ishikawa, T., Yoshikawa, A., y H. Okamoto. 2002. Analysis of the entire genomes of thirteen TT virus variants classifiable into the fourth and fifth genetic groups, isolated from viremic infants. *Arch. Virol.* 147:21-41.
44. Pifferi, M., Maggi, F., Andreoli, E., Lanini, L., Marco, ED., Fornai, C., Vatteroni, ML., Pistello, M., Ragazzo, V., Macchia, P., Boner, A., y M. Bendinelli. 2005. Associations between nasal torquetenovirus load and spitometric indices in children with asthma. *J. Infect. Dis.* 192:1141-1148.
- 55 45. Qiu, J., Kakkola, L., Cheng, F., Ye, C., Söderlund-Venermo, M., Hedman, K., y D.J. Pintel. 2005. Circovirus TT virus genotype 6 expresses six proteins following transfection of a full-length clone. *J. Virol.* 79:6506-6510.
46. Ryabova, L.A., Pooggin, M., y T. Hohn. 2006. Translation reinitiation and leaky scanning in plant viruses. *Virus Res.* 119:52-62.
- 60 47. Saunders, K., Bedford, I.D., Briddon, R.W., Markham, P.G., Wong, S.M., y J. Stanley. 2000. A unique virus complex causes *Ageratum* yellow vein disease. *Proc. Natl. Acad. Sci. USA* 97:6890-6895.
48. Shiramizu, B., Yu, Q., Hu, N., Yanagihara, R., y V.R. Nerurkar. 2002. Investigation of TT virus in the etiology of pediatric acute lymphoblastic leukaemia. *Pediatr. Hematol. Oncol.* 19:543-551.
49. Sospedra, M., Zhao, Y., zur Hausen, H., Muraro, P.A., Hamashin, C., de Villiers, E.M., Pinilla, C., y R. Martin. 2005. Recognition of conserved amino acid motifs of common viruses and ist role in autoimmunity. *PLoS Pathog.* 1:e41.
- 65 50. Stanley, J. 2004. Subviral DNAs associated with geminivirus disease complexes. *Vet. Microbiol* 98:121-129.

51. Takahashi, M., Asabe, S., Gotanda, Y., Kishimoto, J., Tsuda, F., y H. Okamoto. 2002. TT virus is distributed in various leukocyte subpopulations at distinct levels, with the highest viral load in granulocytes. *Biochem. Biophys. Res. Commun.* 290:242-248.
- 5 52. Takahashi, K., Iwasa, Y., Hijikata, M., y S. Mishiro. 2000. Identification of a new human DNA virus (TTV-like mini virus, TLMV) intermediately related to TT virus and chicken anemia virus. *Arch. Virol.* 145:979-993.
53. Zhong, S., Yeo, W., Tang, M., Liu, C., Lin, X.R., Ho, W.M., Hui, P., y P.J. Johnson. 2002. Frequent detection of the replicative form of TT virus DNA in peripheral blood mononuclear cells and in bone marrow cells in cancer patients. *J. Med. Virol.* 66:428-434.
- 10 54. zur Hausen H., y E-M. de Villiers. 2005. Virus target cell conditioning model to explain some epidemiologic characteristics of childhood leukemias and lymphomas. *Int. J. Cancer* 115:1-5.

- (15) aerolisina que tiene la secuencia de aminoácidos de WDKRYIPGEVKWWDWNWTIQ;
 (16) orexina que tiene la secuencia de aminoácidos de
 MNLPsAKVSWAAYTL LLLLLLPPALLSLGVDAQPLPDCCRQKTCSC
 RLYELLHGAGNHAAGILTLGKRRPGPPGLQGRLQRLQASGNHAAGILT
 5 MGRRAGAELEPRLCPGRRCLAAAASALAPRGRSRV;
 (17) receptor GIP que tiene la secuencia de aminoácidos de PRLGPYIGDQTLTLWNQALAA;
 (18) prion que tiene la secuencia de aminoácidos de SNGGSRYPGQGSPGNGRYPPQ;
 (19) neurotensina que tiene la secuencia de aminoácidos de METSSPWPPRSP;
 (20) firma de la familia de receptor nuclear huérfano (receptor nuclear A4) que tiene la secuencia de
 10 aminoácidos de PVNLLNALVRAHVDSTP;
 (21) firma de factor neurotrófico derivado de cerebro (BDN) que tiene la secuencia de aminoácidos de
 PLLFLLEEKYNYLDAAN;
 (22) calcitonina que tiene la secuencia de aminoácidos de KCYDRMQQLPPYELEGY;
 (23) receptor de tipo I de leucotrieno B4 seleccionada entre las secuencias de aminoácidos
 15 SRRLRVRRFHRRTGR y GRRLQARRFRSRRTGR;
 (24) autoantígeno 1 de síndrome de Sjögren/esclerodermia (Autoantígeno p27) que tiene la secuencia de
 aminoácidos de EISKKMAELLLKGATMLDEHCPKCGTPLFRLKDGKVFCEP;
 (25) vasopresina que tiene la secuencia de aminoácidos de RAGRRRGRRTGSPSEGARV;
 (26) firma de receptor 2 de hormona concentradora de melanina que tiene la secuencia de aminoácidos de
 20 LVQPFRLTRWRTRYKTIRIN;
 (27) firma del receptor EP1 de prostanoides que tiene la secuencia de aminoácidos de
 ISLGPPGGWRQALLAGL;
 (28) ciclinina que tiene la secuencia de aminoácidos de EWRS LGVQQSLGWVH;
 (29) firma de receptor activado por proliferador de peroxisomas (receptor nuclear 1C) que tiene la secuencia
 25 de aminoácidos de KTETDASLHPLLQ;
 (30) firma de receptor M1 muscarínico que tiene la secuencia de aminoácidos de
 KMPMVDPEAQAPTKQPPK;
 (31) firma de receptor de tipo B2 de ácido gamma-aminobutírico (GABA) metabotrópico que tiene la
 secuencia de aminoácidos de LAPGAWGWARGAPRPPSS;
 30 (32) firma de arginina desaminasa que tiene la secuencia de aminoácidos de SELSRGRGGPRCMSMPLVR;
 (33) repetición de factor de crecimiento opioide que tiene la secuencia de aminoácidos de
 SPSETGPRPAGPARDEPAE;
 (34) firma de molécula de adhesión CD36 que tiene la secuencia de aminoácidos de
 WIFDVQNPDEVAKNSSKIKVKQR;
 35 (35) firma de proteína protelipídica de mielina (PLP) que tiene la secuencia de aminoácidos de
 GVVLGAIIGGVLGVLLLVLLLYLV; y
 (36) chlamidiaom que tiene la secuencia de aminoácidos de CGSYVPSCKPCG.
2. El poli(ácido nucleico) de virus TT reorganizado de la reivindicación 1, que está presente como un episoma mono
 o bicatenario extracromosómico.
- 40 3. El poli(ácido nucleico) de virus TT reorganizado de la reivindicación 1 o 2 que es un ADN monocatenario.
4. Un cebador oligonucleotídico mostrado en la figura 10 y que tiene una secuencia de nucleótidos seleccionada
 entre el grupo que consiste en CAGCGAGAACGCCACGGAGG GAGATCCT;
 CGGACGGGCGTGGAAAACAGCCATTC;
 CGGACTGGCCGGGCTAT; y
 45 AGCCCGAATTGCCCTTGA.
5. Un vector de expresión que comprende un poli(ácido nucleico) de virus TT reorganizado de una cualquiera de las
 reivindicaciones 1 a 3 unido operativamente a elementos de control de la transcripción y la traducción procarióticos,
 eucarióticos o víricos.
6. El vector de expresión de la reivindicación 5 que es un cromosoma artificial.
- 50 7. Una célula hospedadora que comprende un vector de expresión de acuerdo con la reivindicación 5 o 6.
8. Un procedimiento de detección de un poli(ácido nucleico) de VTT reorganizado de la reivindicación 1 en una
 muestra biológica, que comprende: (a) opcionalmente extraer una muestra de poli(ácido nucleico), (b) hibridar el
 poli(ácido nucleico) comprendido en dicha muestra con al menos una sonda que hibrida específicamente con el
 poli(ácido nucleico) de la reivindicación 1, en el que dicha sonda comprende una parte del poli(ácido nucleico) de
 55 acuerdo con la reivindicación 1, estando dicha sonda opcionalmente marcada y (c) detectar el poli(ácido nucleico)
 hibridado.

Figura 1

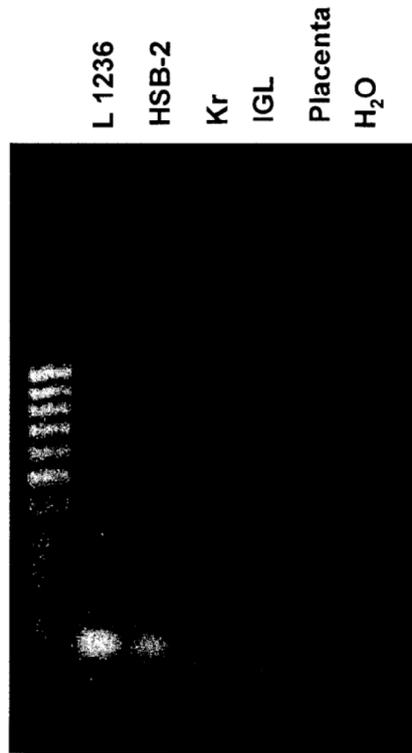


Figura 2

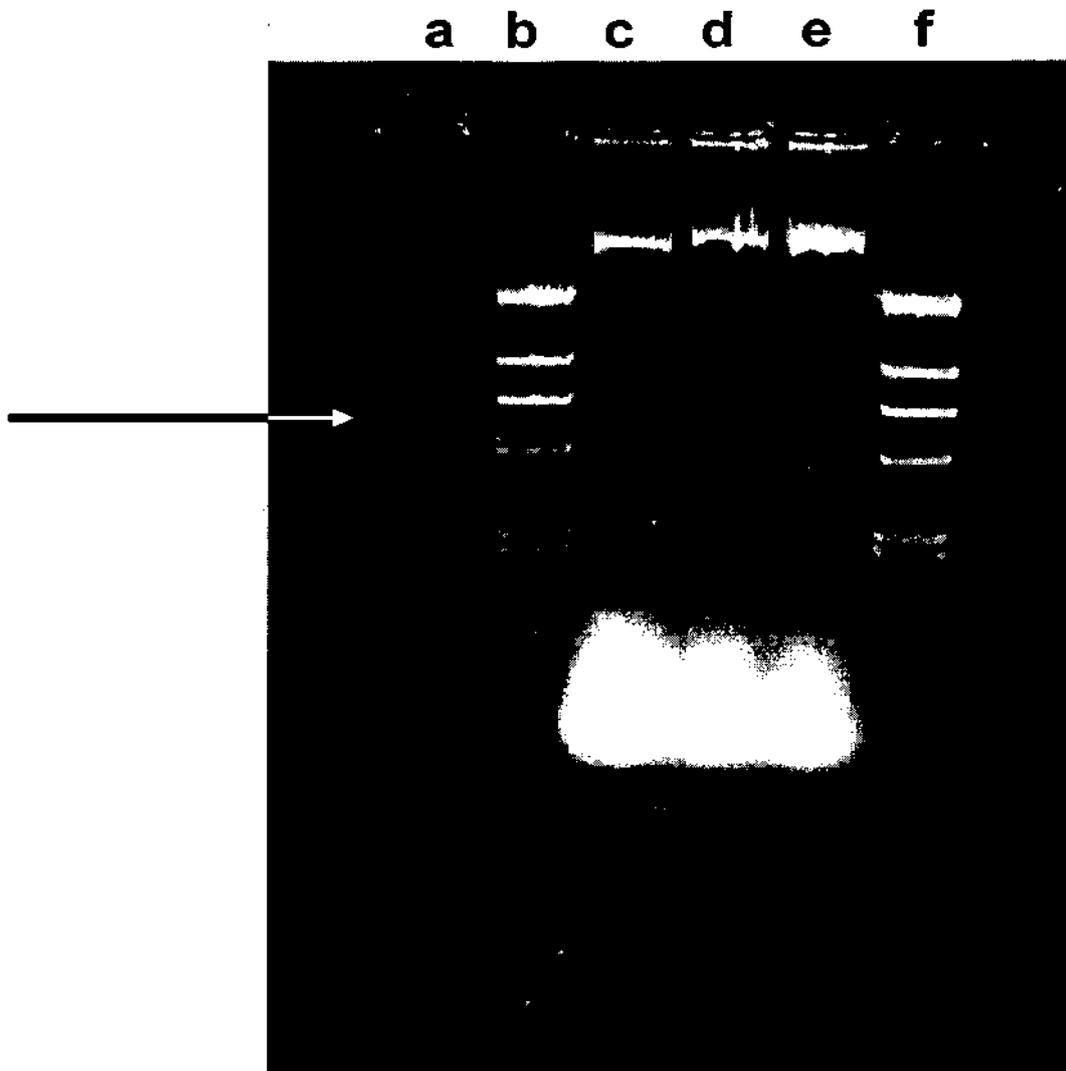


Figura 3

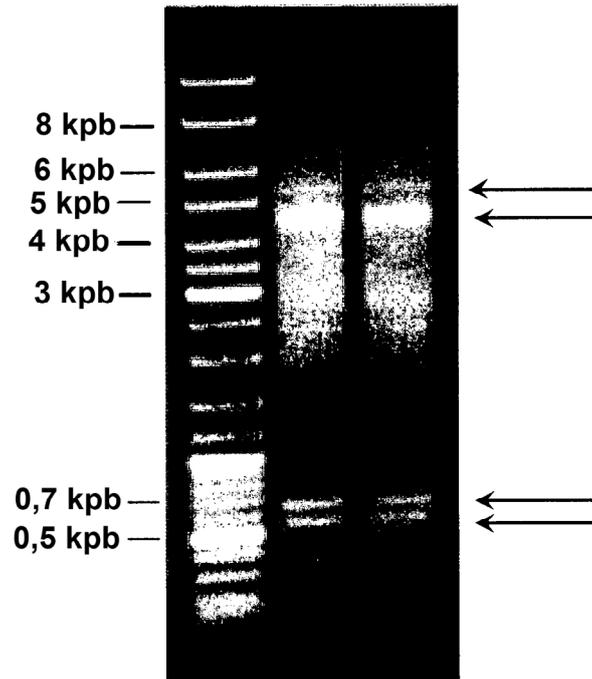


Figura 4

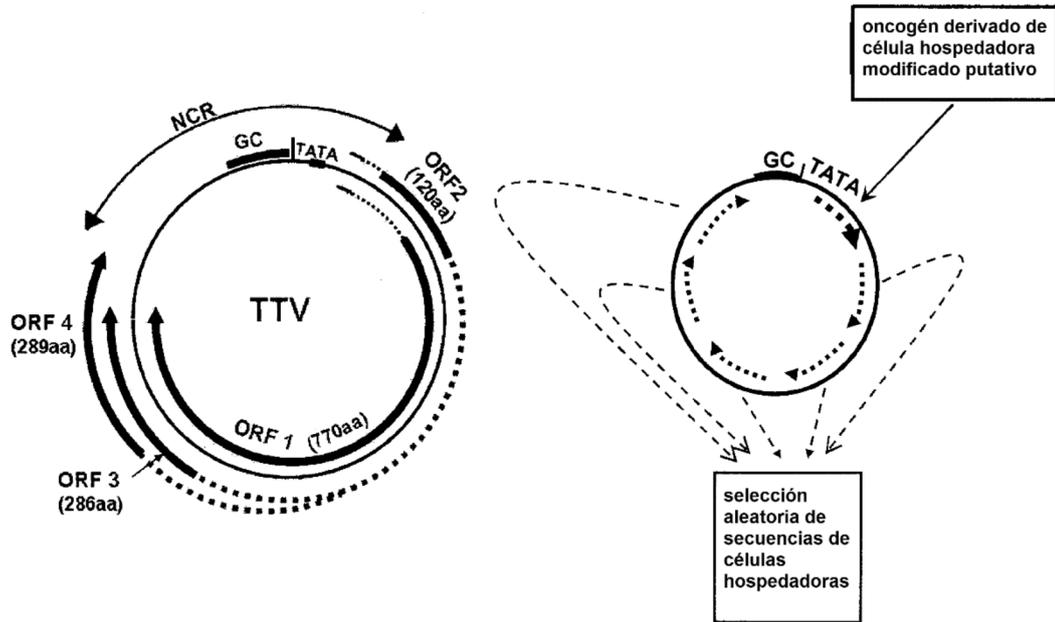


Figura 5

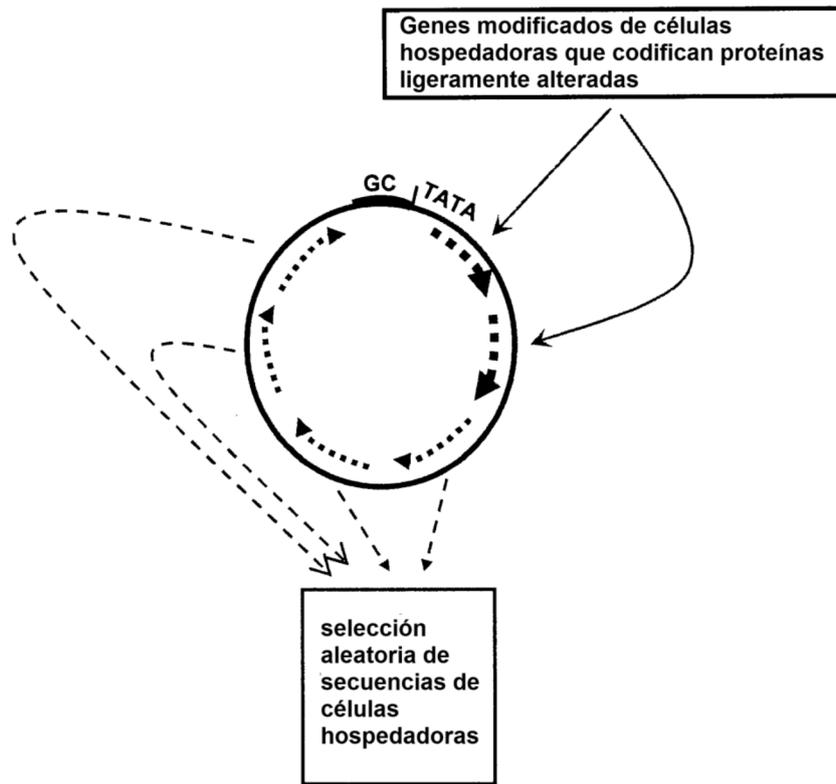


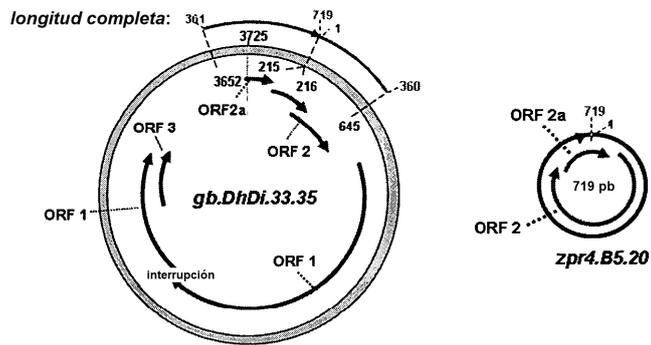
Figura 6

zyb2 CGGGTGCCG~~A~~ AGGTGAGTTT ACACACCGCA GTCAAGGGGC AATTCGGGCT CGGGACTGGC CGGGC~~C~~ATGG G
zyb9 CGGGTGCCG~~A~~ AGGTGAGTTT ACACACCGCA GTCAAGGGGC AATTCGGGCT CGGGACTGGC CGGGC~~I~~ATGG G
zkb5 CGGGTGCCG~~I~~ AGGTGAGTTT ACACACCGCA GTCAAGGGGC AATTCGGGCT CGGGACTGGC CGGGC~~I~~ATGG G
zkb69 CGGGTGCCG~~G~~ AGGTGAGTTT ACACACCGCA GTCAAGGGGC AATTCGGGCT CGGGACTGGC CGGGC~~I~~ATGG G

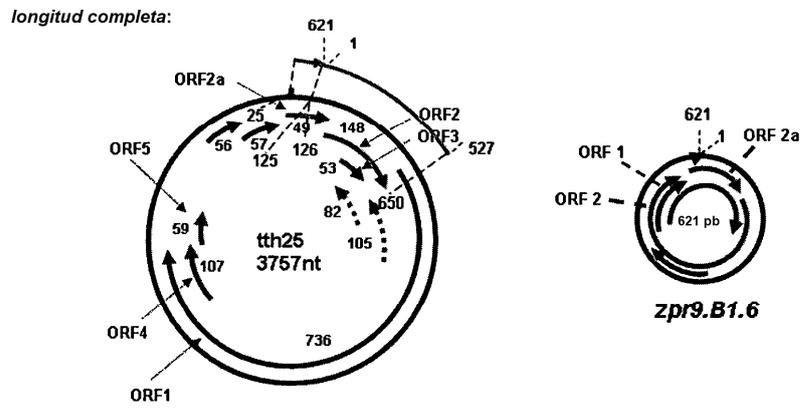
↑ ↑

Figura 7:

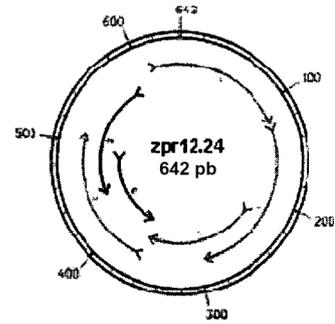
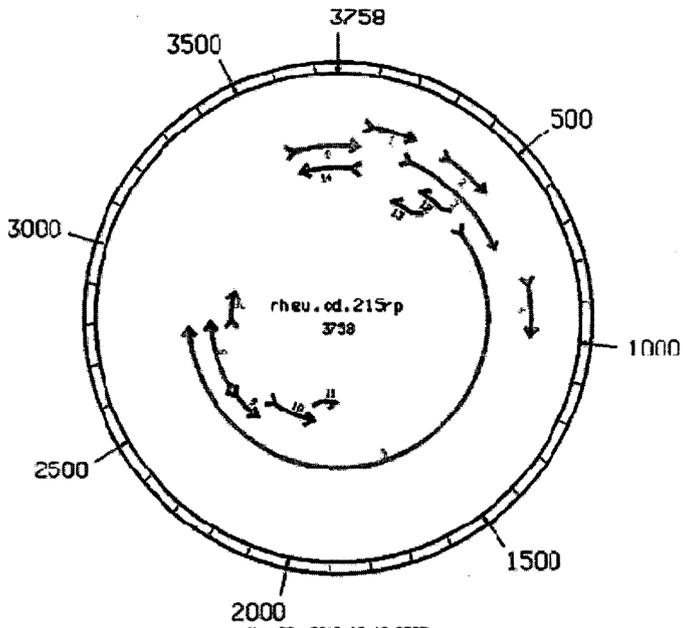
A



B



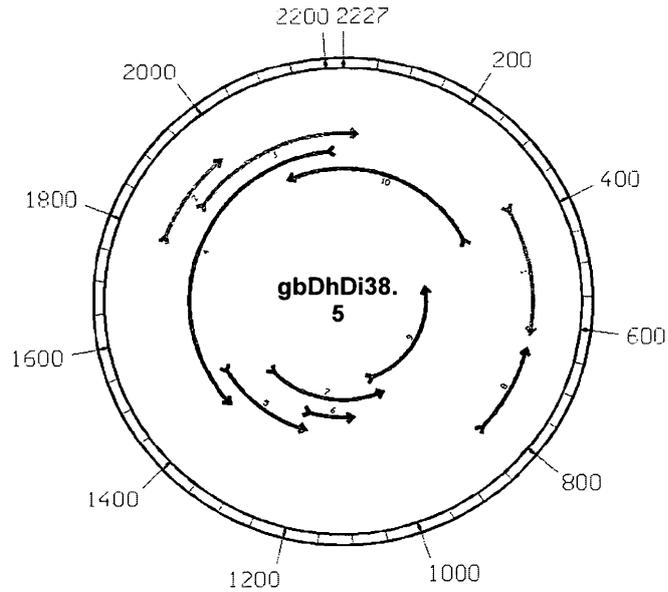
C



May 20, 2010 18:13 CEST
Name From to Strand Color FromSymbol ToSymbol style

Figura 8

A



gb38.5
gbDhDi38.5

WV13038 Klon6
pcr larga según Templphi
cebador : t3pb1+2 aso. : 65 °C
Longitud : 2227

```

1  CAATTCGTGC ACGGGACTAC AAGGAAAGGG GTTGACCCCC ACCCTCCCCC
51  GCCATGCCCA GGAGGGTGCA GACACAAC TG GAAAGGTGCT AGAGACCCCC
101 GGGGGAGGCT GGGCCAGCAC CAGGCATTGG GGGGCAGGTT CCCGTCTCTA
151 CACCCAGCC CCAGGCGGAC AGCGCGTGCC CCTCCCGCTG CCCACCTGT
201 CACCCACCTG CTGGCCCCGG GCTGTCTCTG CTCCTGGCTC CCCTCCCAGC
251 TCGTCCCCA GCTGCCTCTC CAGGGAGGAG TGACAGCTGG CCTGTGCCAC
301 ACCCTCGAGC CCCCCGGAC TACCCCTCC CTGGGCAGG ACCCCTGCCT
351 GTGGCACAAC CAAGGGGCCT GCTGATGGGG GCTCATGTGA GCAGTGCCCC
    
```

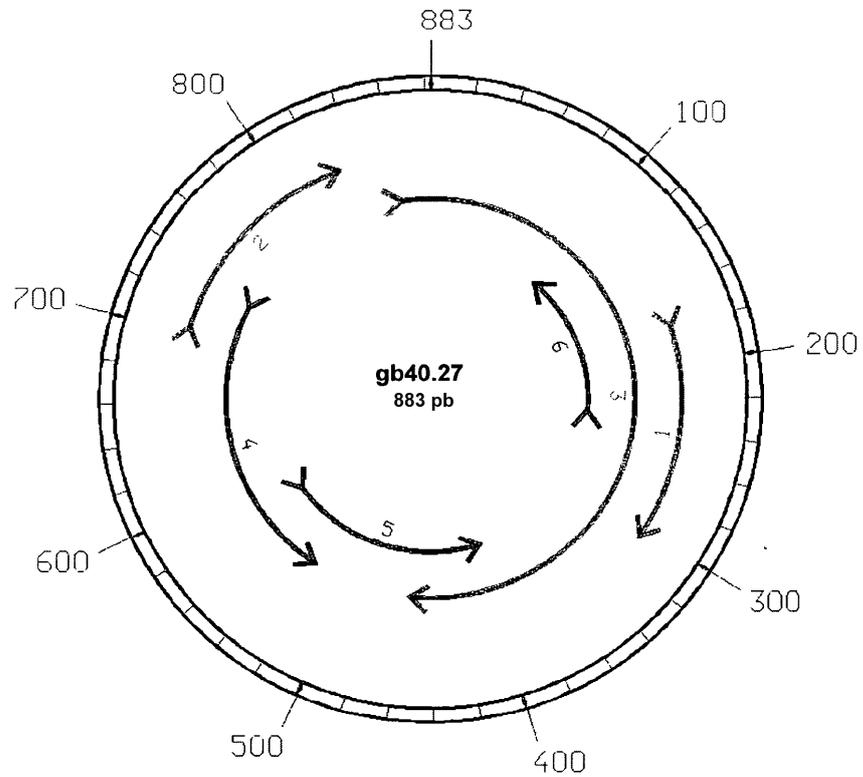
ES 2 685 605 T3

401 AGCTGTGGGT GTGGGTGCTG CCAGCTGCCA CCGCCTTTGC CCTGGTTTCC
 451 CAGATAGACC CCGACCCACA CTCGGAAGCT GTATCATGAA CGCTGTGGTG
 501 GGCGGCTGGT GGGGAGCGGG GFTGCCGTCC CACTACCCTC TGGGAGCCTC
 551 AGCCATGAAG GGCCCCTGTG GGCACCTTTT CCCGGCACAC GGTGCTGTGT
 601 TTCTCCACTC TTGGGCTCTG CAGTGACTTG AGGGGTCAAG TCTATGATCC
 651 CACGGGAGGC TGGGCTAATG AGGGGACCAG AGACCTCAGT GCTGTGCAGG
 701 GAGTCCTGAA CCACCCTGGT GGAAGGCCCA GCCCAACTCC CCAGTCCTCC
 751 CGCCAGCTCC CTGTGGTGTC CAGGAGACCT GTGGTCAGGC CTGGAGGAGA
 801 AGCTCCTCCT CCCCTCGACA TCCTCCCTGC AGCCCTTGCT CTTACCAGA
 851 GCTCCTGAC TCCCCAGGAC CCCAGAGAGG ACTGACCCTC TCCAGCCGAC
 901 CTCTGGGCTC AGGACAGCTG GCGGGGGCAG CCACAGGAGC TGCCTGTAGG
 951 GAGCAGAGTC AGGACGGGGA CCGAGCCGGA CACCCATTCT GGAAGTGTCT
 1001 GCACTTCCAG GCAGGGGAAG GACGGCAGTG GGTAGCTGGG AGTGCTGGGC
 1051 CGAAGATGGG CATTGTCAGG CCCTCAGTGG GGACTGGGAG GTAGAGGTGG
 1101 GGAGGTCTGT GGAGGAAGGA GAAGAAGGGC CAGTGTCCCG AGTTGGGGGT
 1151 GGTGGCAGT GGACGAGGCC GACAGGAACA GACCTGAGCT TGGGGAGCTC
 1201 CACTCAGAAC GAGGCATCCT TCAGGGTTCT GTGCATACTG GTGTCCCTGG
 1251 CTGGGGGCCG GGCCCCGAAG TGGAGCCTGG GACTGTGAGG GTGGGGGGGG
 1301 TGTGCTGGGG TGGGAGGTGG ATGGAGCCCC CCCTCCACCG CCTGGCCGCT
 1351 TGGGCTGAAC CTTGGACTTC GGAGCCGGAA CAGACATAGG AAATGGCCTA
 1401 ACTGCATTTG CGCAGGAACA CCAAATCCCT CGCAGCTGCA CGGGGCTGAG
 1451 CCAGGGCCAC GGGCGGGGTC GGCATCCCA GAGTCCGAC AGCTCCGTGG
 1501 TGTATGCCAA GGGGCCTGGG CCGCTGACCG AGGGGCGCCT TCCCAGGCC
 1551 AGAGGCCCCC ACCCCACCCC AGGAGAGCTG CCCCCCTTTC AGTTCCCAGA
 1601 ACGGAGCCCG GCTGTGGAAT AGTGATGCGG TGAGGTCATG GGGAGGGGGC
 1651 CCGCATGACT CATATCCTGG GGTAGGGGAA AGGGAGGAGA CGGAGAAGGG
 1701 GCCCAGAGGC CTCCACGTCC TCAGCTCTGC TGGGTCAGAG GCCAGGGGCT
 1751 GGCGGGGCTT CTCCTCAGCA CTGGGTTTTA GGGGAGACAC CAGGAGATGC
 1801 TFACTCTGCA TCCCCACTCT GTCCCCCAGG CCCCTAGCCA GGGAGAGCTC
 1851 AGTCAGAGTG ATCCTCCAGG GGCCAGCTC TGCATGGATG ATGTTCCAG
 1901 AGTACACACC TGGGCCTCGT GCCAGGGCCG GCACCGCCGT TGTCAGGGCT

ES 2 685 605 T3

1951 ATGGCAAGGC AAACAGTCAA TGTTTGCCTC ACTAAAGTGA GGCTGCAGCA
2001 CCCTGAAGGG ATCCCTGGAG GGGGACGTGG TCCCCTTGTT CCCAAGCTTG
2051 TCTGCACATG CACGTGGATG TCAAGGGTTC CCGTGTGTGA GCACATGCAT
2101 ATTTGTATGT GCATGGGGTG CGGGCATGTG TGCCTGTGTG GCCGGAGCGT
2151 GGGCTCGTGG AGAATGTGTG TGAGTTGGGT GTGCACCTGC ATGTGCCCCA
2201 GGCCTAGGGA GTCCCGTGCC CGAATTG

B



gb40.27

gbDfDg40.27.seq (secuencia celular de gb40 más cerebro de ms adicional)

Longitud : 883

pcr larga según Templphi

cebador: hel 1 + hel 2 aso 68°C (anidada)

vector: pcr2.1 invitrogen

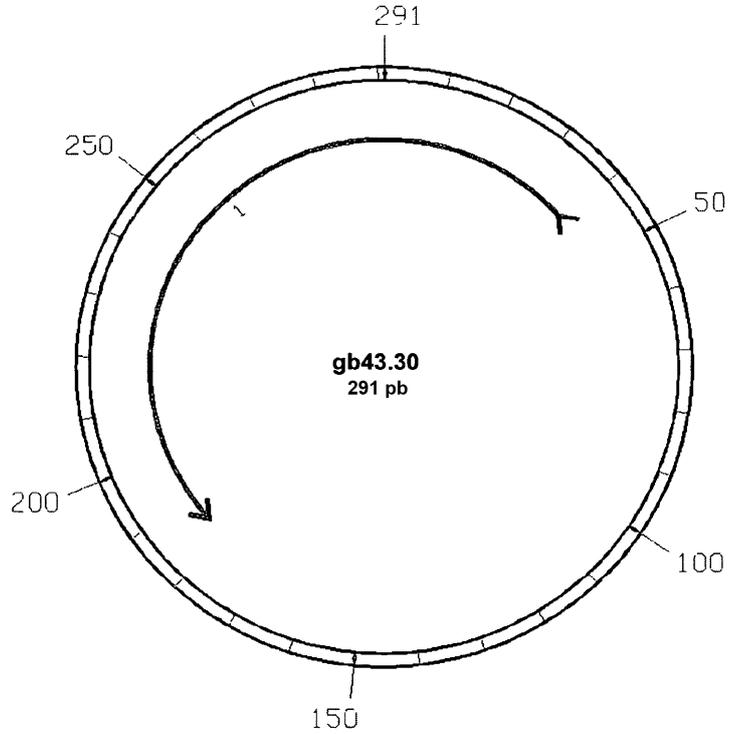
```

1  CGGGACTGGC CGGGCTATGC CCCAGACACA CTCACGTAGG GGTGTCCGGC
51  CTGGCAGCCC AGGACCATGG TCTGCAGGET TTCCTCTCGG CCATTCAGGA
101 CAACCCTAGT CTCCAGGGAA TAGCGCTGGT GTCGCCTATC AGCCGTGAAG
151 GTCTCCTGCA GGAGGAGGCT CTGCGGGATG GGCAGGTGCA ATGGGTGCCT
    
```

ES 2 685 605 T3

201 GGTGTGCAGA GGGAAAAACA GGCCAAAGCC ATTAAAGCAG CTGGCAGTGC
251 CAGGGGACAA TTGTGCCCCA CGGTCTCAGC CTGGGCCTGT CACGAGCTTG
301 CAGAGTTAAG ACTCTGCCAC AGAGAAGAGA ACATCAGGAC ACCTGGCAGC
351 CCTATGCTTT ACAATGTGGC ATCCAGAACC CTTACCACC TCACTGTGCC
401 AGAGAAGTGG GCATGGCTGG GGTCCCCGTC GCCATTGAC AGCAAAGACC
451 CAAGAGGATA GATGACACAC AGCATCTGGT GTCACACAGA CTGGGATTAG
501 AATCCAGGCA CGGTCTTTCA CTAGCTGTGT GACCTTGGGA AAAGGACTTG
551 ACTGTTCTGT GCCTCAGTTT CCCCATCTGT AAAACGGAGG CTAATAAAT
601 ACTGATCGGA CACAGTGGTC AGGGTTAGAG ATAACATACA TGAAACGACC
651 ACAAGCTCCC CAAGGGCAA GGTTCCTGAC ATFCGGTTC TCTGCCATTT
701 TCCATGTGCC CAGAAGAGCA CTGGTCCAT AGTATGTGCT CAATGAATGT
751 AAATGGGATA AAAACACGAA CGAACACTCT GCCAACGATG CTGCTGTTC
801 TTTGTCATCA CTGCTTCTGT TTAGGCTGTA GCTGACTTAT CTAAGGCCAT
851 ACAGCTGCTC AATGCATAGC CCGGCCAGTC CCG

C



gb43.30

gbDhDi43.30.seq (secuencia celular de gb43 más cerebro de ms adicional)

Longitud: 291

pcr larga según Temp.

cebador: t3pb 1+2 aso. 65%

vector : pCR2.1, Invitrogen

```

1 CCCCTTGACT TCGGTGTGTA AACTTGTGGT ATAGAACATG ATGTTTTAAG
51 ATACATGTAC ATTGTGGAAT GGCTTGATCA TGCTAATTAA CATATGAATT
101 ACCTCACTTA GCTATCTTTT TTATGGTGAA AGCACTTAAA ATCTACCCTC
151 AGCAGTTTC AAGTACAAA TACATTTCTA TTAACATAG TCACCATGTT
201 GTACAATAAA TCTCTTGAAT TTATTCCTCC TGCCTAACTG ACATTTTGTA
251 TCCTTTGACT GATCTCTCTC CCCAGTCCCG TGCCCGAATT G
    
```

ES 2 685 605 T3

Negrita - secuencia del cebador
Cursiva - continuación de la secuencia celular

BLASTN2 de: /home/vir088/ttbrain/gbDhDi43.30.seq desde: 1 a: 291 9 de Junio, 2010
12:21

comparada con la base de datos : **nrnuc**

>>>nrnuc:GI_225543527 Gi|225543527|ref|NG_011635.1| **Miosina IIIA de Homo sapiens**
(MYO3A), RefSeqGene en el cromosoma 10. 0/0
Longitud = 285464

Puntuación = 525 bits (265), Esperada = e-146
Identities = 268/269 (99%)
Hebra = Plus / Plus

```
CCCTTGACTTCG
Query: 14  gtgtgtaaacttgggtatagaacatgatgttttaagatacatgtacattgtggaatgcc 73
          |||  |||
Sbjct: 276074 gtgtataaacttgggtatagaacatgatgttttaagatacatgtacattgtggaatgcc 276133
          TAATTGACAAAAC

Query: 74      ttgatcatgctaattaacatatgaattacctcacttagctatctttttatggtgaaagc 133
          |||
Sbjct: 276134 ttgatcatgctaattaacatatgaattacctcacttagctatctttttatggtgaaagc 276193

Query: 134     acttaaaatctaccctcagcagttttcaagtacacaatacatttctattaactatagtc 193
          |||
Sbjct: 276194 acttaaaatctaccctcagcagttttcaagtacacaatacatttctattaactatagtc 276253

Query: 194     ccatgttgtaacaataatctcttgaatatttcctcctgcctaactgacattttgtatcc 253
          |||
Sbjct: 276254 ccatgttgtaacaataatctcttgaatatttcctcctgcctaactgacattttgtatcc 276313

Query: 254     tttgactgatctctctcccagtcccgtg 282      CCCGAATTG
          |||
Sbjct: 276314 tttgactgatctctctcccagtcccgtg 276342 ACCAGTGCCT
```

>>>nrnuc:GI_14018255 Gi|14018255|emb|AL162503.12| Secuencia de ADN humano
del clon RP11-420F12 sobre el cromosoma 10 Contiene el gen GAD2 para la
glutamato descarboxilasa 2 (islotos pancreáticos y cerebro 65kDa), el extremo 3'
del gen MYO3A para la miosina IIIA y dos islas CpG, . . . 0/0

Longitud = 175594

Puntuación = 525 bits (265), Esperada = e-146
Identities = 268/269 (99%)
Hebra = Plus / Plus

```
Query: 14      gtgtgtaaacttgggtatagaacatgatgttttaagatacatgtacattgtggaatgcc 73
          |||  |||
Sbjct: 35112 gtgtataaacttgggtatagaacatgatgttttaagatacatgtacattgtggaatgcc 35171

Query: 74      ttgatcatgctaattaacatatgaattacctcacttagctatctttttatggtgaaagc 133
          |||
Sbjct: 35172 ttgatcatgctaattaacatatgaattacctcacttagctatctttttatggtgaaagc 35231
```

ES 2 685 605 T3

```

Query: 134  acttaaaatctaccctcagcagttttcaagtacacaatacatttctattaactatagtc 193
            |
Sbjct: 35232 acttaaaatctaccctcagcagttttcaagtacacaatacatttctattaactatagtc 35291

Query: 194  ccatggtgtacaataaatctcttgaattttattcctcctgcctaactgacattttgtatcc 253
            |
Sbjct: 35292 ccatggtgtacaataaatctcttgaattttattcctcctgcctaactgacattttgtatcc 35351

Query: 254  tttgactgatctctctccccagtcctcg 282
            |
Sbjct: 35352 tttgactgatctctctccccagtcctcg 35380
    
```

Marco de lectura abierta : 49aa

(**Péptido**) **FASTA** de: gbDhDi43.30rev.49.pep desde: 1 a : 49 9 de Junio, 2010

REFORMAT de: gbDhDi43.30rev.49.pep check: 3689 desde: 1 a : 49 9 de Junio,

PARA: **SwissProtPlus**:* Secuencias : 11,223,768 Símbolos : 3,635,054,084 Tamaño de palabra: 2

```

SPTREMBL:Q9WB12_9VIRU  Begin: 19  End: 60
! Q9wb12 SubName: Full=ORF2; Flags: F... 84 109 108 181.5 0.3
SPTREMBL:Q9WB09_9VIRU  Begin: 19  End: 54
! Q9wb09 SubName: Full=ORF2; Flags: F... 91 104 176.0 0.61
SPTREMBL:Q9WB02_9VIRU  Begin: 19  End: 59
! Q9wb02 SubName: Full=ORF2; Flags: F... 82 82 100 168.8 1.5
SPTREMBL:Q9WSW4_9VIRU  Begin: 32  End: 73
! Q9ws4 SubName: Full=ORF2; Flags: F... 82 82 97 162.5 3.4
SPTREMBL:Q9WAY4_9VIRU  Begin: 19  End: 54
! Q9way4 SubName: Full=ORF2; Flags: F... 84 84 94 160.8 4.2
SPTREMBL:Q9WB10_9VIRU  Begin: 19  End: 54
! Q9wb10 SubName: Full=ORF2; Flags: F... 84 84 94 160.1 4.6
SPTREMBL:Q9WAZ2_9VIRU  Begin: 19  End: 54
! Q9waz2 SubName: Full=ORF2; Flags: F... 84 84 94 160.1 4.6
SPTREMBL:O70807_9VIRU  Begin: 23  End: 58
! O70807 SubName: Full=Putative uncha... 84 84 94 158.2 5.9
\\End of List
    
```

gbDhDi43.30rev.49.pep
SPTREMBL:Q9WB12_9VIRU

```

ID  Q9WB12_9VIRU          Sin revisar;          150 AA.
AC  Q9WB12;
DT  01-NOV-1999,         integrado en      UniProtKB/TrEMBL.
DT  01-NOV-1999,         versión de secuencia 1.
DT  09-FEB-2010,         versión de entrada 20.
DE  SubName: Full=ORF2; . . .
    
```

```

PUNTUACIONES Init1: 84  Initn: 109  Opt: 108  puntuac.z : 181.5 E(): 0.3
>>SPTREMBL:Q9WB12_9VIRU          (150 aa)
    initn: 109 init1: 84 opt: 108 puntuac.Z: 181.5 esperada(): 0.3
Puntuación de Smith-Waterman : 108; 53.5% de identidad en un solapamiento de 43 aa
(6-46:19-60)
    
```

ES 2 685 605 T3

```

                                10      20      30      40
gbDhDi43.30r                MFYTTSLHTEVKGQFGHGTGERDQSKDKCQ--LGRRNKFKRFIVQH
                                |||| | |||| | || | : : | | : || : | | : |
Q9WB12_9VIRU AQTQRRVIPASRGRVPEVSLHTXVKGQFGLGTG-RAMGKALKKDMFLGKLYKKRALS LH
                                10      20      30      40      50

```

```

gbDhDi43.30r GDYS
                                |
Q9WB12_9VIRU GLRTPPEAKPPAMSWRPPVHNPRIERNLWEAFFRIHASSCGGHLVGH LTVLARRYGAPP
                                60      70      80      90      100     110

```

```

ID   Q9WB12_9VIRU                Sin revisar. ;          150 AA.
AC   Q9WB12;
DT   01-NOV-1999,               integrado en UniProtKB/TrEMBL.
DT   01-NOV-1999,               versión de secuencia 1.
DT   09-FEB-2010,              versión de entrada 20.
DE   SubName: Full=ORF2;
DE   Flags: Fragmento;
OS   Torque teno virus.
OC   Virus;                     virus de ADNss ; Anelloviridae; Anelloviridae sin clasificar.
OX   NCBI_TaxID=68887;
RN   [1]
RP   SECUENCIA DE NUCLEÓTIDOS.
RC   TEJIDO=Suero;
RX   MEDLINE=99335592; PubMed=10405352; DOI=10.1006/viro.1999.9797;
RA   Hijikata M., Takahashi K., Mishiro S.;
RT   "Complete circular DNA genome of a TT virus variant (isolate name
RT   SANBAN) and 44 partial ORF2 sequences implicating a great degree of
RT   diversity beyond genotypes.";
RL   Virology 260:17-22 (1999) .
CC   -----
CC   Derechos de autor de UniProt Consortium, véase http://www.uniprot.org/terms
CC   Distribuido bajo la licencia de Creative Commons Attribution-NoDerivs
CC   -----
DR   EMBL; AB024379; BAA77446.1; -; Genomic DNA.
DR   InterPro; IPR004118; Gyrovir_VP2/TT_ORF2.
DR   Pfam; PF02957; TT_ORF2; 1.
PE   4:                          Previsto;
FT   NON_TER                      1          1
FT   NON_TER                      150        150
SQ   SECUENCIA 150 AA; 16415 MW; A4A7DF15855FC40D CRC64;
      AQTQRRVIPASRGRVPEVSLHTXVKGQFGLGTGRAMGKALKKDMFLGKLYKKRALS LH
      LRTPEAKPPAMSWRPPVHNPRIERNLWEAFFRIHASSCGGHLVGH LTVLARRYGAPP
      PPAPGAPRPA LKRQLALPAP PADPQQANPT

```

Figura 9

A

hod11

hodL.VvWw.1.seq Longitud: 639
Hodgkin L1236+TPA

Cebador de PCR: hlf1+Lr1 GCI 59°C
Vector pGEM T-Easy

```

1  CCCCTTGACT GCGGTGTGTA AAGCGCCCA GCCTGTGCCT GCACAGTGCC
51  TGTGTGGTGT GAACCCATGA CCAGGCCTCT GGAGGGAAG AAGGTTAGGC
101 TTAGTGGACA CCAGCTTTC TAAGGTGGGT CTTAGACCAA CTCATTA AAA
151 TGGCAGGATG GGCTTTTGTG CTGTATTCT TGGGATTTTC AAGATGCCCC
201 ACACAGCAGA AGGGATGTGC ATTTTTTCT CTGCCCTGAG TTGTTGATA
251 AAAATCAGTG ACCTCGTCT CCACTTAGAA CTCCTGAA CTGCACTCGG
301 TGTCTAGGAC TGTGGGGAA GGAAGTGAAG AGCCAGCATG TAGTCTCCTC
351 TGGACTCTTA CAGGATCTGT CCACCTCTGG GCTCTTTATG TAGGGGAAGG
401 TGTGAGCTCC TGGGAGTACT CCTGATAGAG GACTGTTTCC CTGAAAACCT
451 CAGCAGTGTG TGAGGCCCTA GCAGGGGAA CCCAGACCCC GCCTGCCAAA
501 GCCCCTAATC CTCAGGGCT ATTATCAGCA GCCTAAGCGC CTTAGGGTGG
551 CCAGAGTCCA GCCCAGCAAG CAGCAAAGTC AGCAGCCTCC TCGCCCTATC
601 CTCTCCATGC CCCGGGGCAC TCCAGTCCCG ACCGAATTG
    
```

Cebadores en negrita

Continuación de la secuencia celular en cursiva

BLASTN2 de: /home/vir088/ttmixture/hodL.VvWw.1.seq desde: 1 a: 639
comparada con la base de datos: nrnuc ..

Base de datos : nrnuc

>>> <u>nrnuc:GI_18121492</u>	Gi 18121492 emb AL513485.10 Human DNA sequ...	690	0.0
>>> <u>nrnuc:GI_149944898</u>	Gi 149944898 gb AC198797.3 MACACA MULATTA...	509	e-141
>>> <u>nrnuc:GI_55416061</u>	Gi 55416061 gb AC121551.11 Mus musculus ch...	84	1e-12
>>> <u>nrnuc:GI_241752248</u>	Gi 241752248 ref XM_002400984.1 Ixodes sc...	46	0.26
>>> <u>nrnuc:GI_291191454</u>	Gi 291191454 gb GU722348.1 Torque teno vi...	44	1.0
>>> <u>nrnuc:GI_291191453</u>	Gi 291191453 gb GU722347.1 Torque teno vi...	44	1.0
>>> <u>nrnuc:GI_291191452</u>	Gi 291191452 gb GU722346.1 Torque teno vi...	44	1.0
>>> <u>nrnuc:GI_217416834</u>	Gi 217416834 gb FJ426280.1 Torque teno vi...	44	1.0

```

>>>nrnuc:GI_18121492 Gi|18121492|emb|AL513485.10| Secuencia de ADN humano
del clon RP11-48O20 sobre el cromosoma 1 Contiene el gen TAGLN2 para la
transgelina 2, el gen IGSF9 para el miembro 9 de la superfamilia de inmu-
globulinas, el gen SLAMF9 para el miembro 9 de la familia SLAM, un gen
novedoso y . . . 0/0
Longitud = 80398
    
```


B

hoht33

hoHT.vAf.33a.seq Longitud: 3387
 Linfoma HSB+TPA, Templyphi-RCA

Cebador de PCR HLF1+Hr4 60°C

Vector pCR2.1

```

1  AATTCGGTCG GGACTGGCAG AGTGACGCTC AGTTCAGCCT GACAGCAGGG
51  TGATTGAAGG GGCCAGATAC CCCAGCAGGG CCTGAGGCCA GAACACAGCA
101 TAGGCTGGCT CTGATGGGTG GAGGAGGTGG CCAGGCATCA TCTGGAGCTT
151 GGAGTTGAGA ACATCTGTGA CTCCTCCTTC AGGAGGGTGC TCTAGGAGTT
201 GAGAGCATCC TAGGTAGGAC CATACATCTA CCCCCATCCT AGTTCCTCCTC
251 AGCCTCTCTT TTCAGCTCCA GGTCTACCTT AAGGGACCTA GGACACCTGG
301 GCTGGGGCAT AACAGGACTT GGTTTTATGT AAAGGAGCTG GGAAGAGACT
351 GAGATAACAG AGGGCTGCAA GGAGAGAGAC AGAGAGAGAA GAACCTGCCA
401 GAAGAAGCTC CTCAGCAATC CACTAAGCCC TGATCTTTGC CTCACTGCCT
451 GTCCTTCCC ATCCGCTCTT CTGCTCTCTC AATCTCTGCC TTCAGAAAT
501 TTGGTGCATA TTGGAATAGG GAGGAATAGA AGCACCTGG GTGGAGCTCT
551 GGGCTTGGCT GTGCACGAGC TTTCAGTGGG TGGTTTGTG GTCTCCAAAG
601 ATGACCTCC ATTAGTCATG CTTCTCGGTG TTTGTCTCA GGTAGTCTCA
651 TCCCATCTTG AGTCTGGGCT TGCCCTGTGA CTCACTTTAA CCACAAGAAT
701 GTGGCAGAAA GGATGTTGTG CCAGTTCTAG AACTAAGCCT TCAGAAAGCC
751 TAGCACCTTC TGCTTTTAGG AGCACTGAGC CCCCATGTTA GAAGTCCACT
801 TTTATACTCT GCTCTGGAGA CTAGCAGAAT TAGAAATGCA CTGCTGAATG
851 CTGCTCGAGA GACTAATGGA GAGGCCATGT GAATAAGGAG GCCTGAAACT
901 ACATGGAGAT AGAGGGCCAG CCACCCAGC ACCACGGCTC AGCTGTGCCT
951 CCCAGCCATC TCTGCCAGTC CTCAGGGCT ATGAGTGAAC CATCTTGGAT
1001 GTTCTAGCTC GGTGGAGCCC CCAGGTGATT GCAGCCTCAG CCACCATCTG
1051 ACTGTAGCTG CATGAGAGGC CCCCAGTGGG ACCAGCAGGA CTGCCAAGCT
1101 GAGCCCTGCC CACCCACAGA ACTGTGAGAA ATAAAAAAT GGTGTTTCC
1151 TTAAGCCATT AAGTTTGGG ATGATTGTGTT ACTACAATT GATAACTGAT
1201 ACAGTCTGTC TTTAGGGAAA ACAAGGATA ACTCTGGGCT CCAGGTGTCT
1251 TCTATAGGAT GAATGGGACT TGGTTGCTGA CAAGCTGACA AGTTTGAGCA
1301 TGAACCTCT TTTTTTTTT GGAGAAGGAA TTTTGCTCTT GTTATCCAGG
1351 CTGGAATACA GTGGTGGAT CTCGGCCAA GGCAACCTCT GCCTCCTGGG
1401 TTCAAGCAAT TCTCCTGCCT CAGCCTCCTG AGTAGCTGGG ATTACAGGCA
1451 CCCACCACTA CACCTGGCTC FTTTTTTTTT TTGFATTTT AGTAGAGACA
1501 GGTTFTCATT ATGTTGGCCT GGTGAGGTTT TGAACCTCTG ACCTCAGGTG
1551 ATCCACCTGC CTTGGCCTCC TAAATGCTG GGATTACAGG TGTGAGCCAC
    
```

ES 2 685 605 T3

1601 CGTGCCTGGC CTGAGCATGA AACTTTTATG CTCAAACATT AAAGTGTA
1651 CACTCACCAG CTCAGCTGAA TAAGAAGTTC TGGGGGCAAG GCCCAGGAAT
1701 CTACAGTTTA GTAAGTGCCC CCACCACTGG ACCCTGGGAA AGTGGACTGC
1751 ATTTTGAAAA ACTCTAGATC AGTTGATACC CAGGAGTCTC CATAACACTA
1801 AGTTGTAATA CCTCAGTGTG AATTAGTCTG ATGCAGCTCT TCTTAGAGGT
1851 CATTGACAGA GGGCAAGACA TTCCAAAAG GAAGGAATAG CCAATATGGA
1901 ATGACAGGTG GATTGGATGA CCCICTAITA TTAGTTTCA ACCTGCCCTT
1951 CCTGCCTTCC CTCCCACAAA TTCCTTTCA GATCCTCCGT CCTAATCCTC
2001 TTCGATAGTT CAITGTTCTT CTGCAGACAG AGCAGCGAAG TGTATCTGT
2051 TGTACCCACT ATGACTAGTT GATGGTGCAT GGCTTCCATG GAGCAGTCT
2101 GTGATCCATT AGTCATGGAG CAGTGTGTG ATCCATTGTC ATGTCTGCCA
2151 TGAACACTGG AAGGGCAGT GGTAAATGACA GCCTCTTACA TTTGCCAECT
2201 CTGCCCAACA TTCTTCCAG TGTGGGAAA GCCTTTGCTT ATTCATATCC
2251 TTCTGGAAA GCTTTGTTC TCCATTTTAC ATTTTAAAT TTTCTCATTT
2301 TTATGGTGCA CCATGGATAC CACCTGTCCA TATAGCTGGC TTCTGATTTT
2351 TCCAGATGAA AGTAATCCTT CCTCTCCTAA CCTCCCATGA CACCTAACCT
2401 GGCACCTATT TACGGTGTTC AGCTCCTTCT CTTGACGTT CTCATTGTTT
2451 TCCTCTCATC TTCTCCCAG GAATGGATTC CCCGCCAAGG GAGGTACCAG
2501 GTCAGTTTCT TCTTTGTGCA ACAGGGTGTG CCTGATGAGC ACAACCTGG
2551 AACAAAGTGT TGTAGGGCTG GTGGCATCT GGTTCCTCTG GGTGTGTGT
2601 AGCCTGAGCC GGGGGCAA TGGGTGTTG TTTTCTGAA GAAGCCAGCC
2651 GTTCTGTGGC AGATGTGGT GAGGGGGTT GGGAGTAGT ATCATGGAGA
2701 GGCTGGGATC CTATCTATCT CTTCCCCTG CTTGAAGGGC AACTTGGGAG
2751 AAGCTCAAGA GGGAGGAGT GACTGCAGAA GCTGGGATC CTGCATAACT
2801 CTCAGTTCA AGCATCACTG CTTTAGGGCC CTGGGGCCCT ATGTGTAGT
2851 CAAGAAAGGG AGATAGAGAG AGAAGAGAGA GAGAGGAGAG AGAGAGAGAG
2901 AGAGAGAAGA CAGAGGAGAG AGAGAGAGAA GARAGAGGAG AGAGAGAGAA
2951 GAGAGAGCAG AGAGAGAGAG CATGCTGTCA GTGAGGTGGC CCTAAGCCCT
3001 CTTGAAATA ACTTGGAGGC ACTGTGGGT GGCTCTGAGG TGCTGAGGTA
3051 TACCTGTAGT GGGCTAGGA CCTTCCAAC CTGGTCTGA AGGTTGAGGC
3101 AACCTTGGGT GTACCTGCTG GTGAGCTGAG AGCCCTGGGG ACCTTGGCA
3151 GACATTCCCA CCCCTGAGC CTGGAGGGTT TGCAATGAGT GAGGCTGTCC
3201 TGCTCATCAC TACGTCTCTT GGGACAGCAC ATTGCCGTGT CTGAACAGGC
3251 ATTCAGTTG GATTGTGGA ATCACTGTTG GTGAGGAGGG CAAGTGGCAA
3301 CAGAAATGGG GGTGTGCTCC CCCAGTTCC TCAGTACAA TCTCCATGAC
3351 CTCTACACT GCCCTGGGCC CAGTCCCAGC CGAATTG

C

ES 2 685 605 T3

hoht22

hoHT.vAf.22a.seq Longitud: 1790

Lymphom HSB+TPA, Templiphi-RCA

Cebador PCR HLF1+Hr4 60°C

Vector pCR2.1

```
1 CAATTCGGTC GGGACTGGGG AGCTGTGAGA AAGAGAAGAG AAGGTCAGAT
51 CAGGAACATT ACACAGAAGT CGGCAAACT GGAACGAGGA GGGAAAGAAA
101 TGAGCGAGTC TGACACTCAG TCCATCCTAG TTCCTATCAC ACAGGGAGGG
151 ACATTGCCAT GCACATCCCC ACAGAGATGC ACCGTGTAAG GGGTCGAGGC
201 AGATCCTGTC CACTATTGCC AGCTCTGAGG TGATCAAATT GTGCTGCCC
251 AGGGTAACCC GGTTGACCTA AACCAACCCA CTCCCTTGCA CATCTTAGGT
301 GTTCTGAGT CAGCAAGGCT GAGGAAGCCA CTCCAGCCAA AATCCCTTGT
351 GCGATCTTCA AGCCCAATC ACAGGCAATG ACAAGGCCAT GTCTGGCTGG
401 CCTCATGGGG ACTGCCCTCC CTCACCAGA CCTAGAACAC AGGCAATGCT
451 CAGCAGCGTT CTGAGAAGAG CTGAGGTCAA GAACTCCAAC CCCACGCAAC
501 CCAGACCTGA TACAAACAGA CACCCATTTG CACTCCTAAC CCTTGAGCCT
551 CTATTTCCAG ACCTCCTCAC TGGGTCTCAG CTGAGAACCC ACTTTTAGCC
601 AAGCATCTTT AGTTCAGAGT TCCTCGCAGT GAGGGGATCC CTCCTCTGCC
651 TTGCTGTCTG TGCTGCATCC ATTATACCCT CACACCGTGC TACTCAGCAG
701 GGGAGAAATG GAGCCCTGGG GAGCCGGCAC TTTTCTCTFC TGCCCTTTC
751 TTGCCTTGCC TCAGGAAGGG GAAAAACTCT GGGTTGTTTT AGTTTGATCC
801 CCTGTCTTAA GTGACCACAG GAACACTAGG CAGTGAGTAC ATATGGATTC
851 TTAGCAGAGA GCTGACAAGT CTTCAGAAAC ATAGAAAACA TAGAAGCTTT
901 GAGTGAGGAG ATCAGAATGT AATTAGGAGT TTCTTTTGA GCAAACCCCA
951 CCCAAGAGA GTGAGCCCAA GTTCTTGAAG GCCCACCTGA GCAGATGACA
1001 CCAGCGTCTT CACTATGGCC ACAGTTGTGG GTGAGCCAGC CATTGTGGGG
1051 GCAGCTCCAC AGGTAGGACT CGTGTCTTGA GCAGCGCACA TCATCCAGGA
1101 CAATGGGTCC TGAGCCCTGG CCAAACCTGG CATTTCCTGG GGCTGACATG
1151 GCCCAGCCAC AGCCCGGCTG CCTGCAGACC ACATTGGCAT CATTGGTGTG
1201 CCAGTAGTCA TCACACACGG TGCCCCAGGA GCCTCGGTAT AGGACCTCCA
1251 CTCGGCCTCG ACACCTGTG CCTCCATTCA CCAGCCTCAG GGCCAAACTG
1301 GATTCAGATC CTACAGGGGA ACACAAGAAC CTTTCATCCA TCCCTATCAT
1351 GAGGTCAAGA ATCTAAGGTA AGTCCACAC TCAGGGTACT TCCTAATGAA
1401 CTAAGTCACC TAGGCAGGCA GTCACCTTG CATATGACTA CAGACTAGGC
1451 TTCATACCCG TGAAAGTAGC ACTGATAACC TACTCTGCCC AGGTCTATGG
1501 GTGCTCAACT TTTGGGAAG CACCTGTGAC CCCAGTGGAT GTGATGGGAA
1551 TGGATGCCCC ACTCCCAGT TGGGTACACA GAGGATGGAG CTGCTCAGCT
1601 CCAGATGGCA GGCCAGACC CCTCCCTTAT TCAGGAGCAT GGTCCATCT
1651 GGGATCTGAC TGGCAGAGTA CCAGAGATGG CAGGGATGAG GTCCCCATAG
1701 GATTAGGGAG ACCCCAGGG CTGTCTTGA GCCCATAGAT AAGGATCTTT
```

1751 TCTGACCACT TGGAACAGGA TCCCAGTCCC GACCGAATTG

Figura 10

DhDi:

t3pb-1.prime Longitud: 18 cebador directo de pcr larga
1 caattcgggc acgggact

t3pb-2.prime Longitud: 24 cebador inverso de pcr larga
1 ccccttgact tcggtgtgta aact

cd:

tth4prime1.seq Longitud: 28 cebador directo de pcr larga consecutivo con prime2
1 CAGCGAGAAC GCCACGGAGG GAGATCCT

tth4prime2.seq Longitud: 28 cebador inverso de pcr larga consecutivo con prime1
1 CGGACGGGCG TGGAAACTC AGCCATTC

DfDg:

hel32-1.prime Longitud: 18 cebador directo de pcr larga
1 cgggactggc cgggctat

hel32-2.prime Longitud: 19 cebador inverso de pcr larga
1 agcccgaatt gcccttga

Figura 11

A**ttgb33.35**

gbDhDi33.35.seq (genoma completo de una biopsia de ms)
 pcr larga , Templphi
 cebador: T3PB 1 + T3PB 2 aso.:65 C
 vector:TA-cloning pCR2.1 invitrogen

Longitud: 3725

```

1  ATTTTGTGCA GCCCGCCAAT TTCTGTTCAA ACAGACCAAT CAGGACCTTC
51  TACGTGCACT TCCTGGGGCG TGCTACGAG GTCTATATAA GCAACAGCGG
101 TGACGAATGG TAGAGTTTTT CTTCGCCGT CCGCGCGGAG AGCGCGAGCG
151 AAGCGAGCGA TCGAGCGTCC CGTGGGGGGG TGCCGTAGGT GAGTTACAC
201 ACCGAAGTCA AGGGGCAATT CGGGCACGGG ACTGGCCGGG CTATGGGCAA
251 GGCTCTTAAA AAATTCCTCC GCTCTGCTCT CCGGCAGGAC ACAAGTCAT
301 GCCGTGGAGA CCGCCGGTCC ATAACGTGCC AGGTAGAGAG AATCAATGGT
351 TTGCAGCGTT CTTTCACGGT CATGCTGCTT TCTGCGGGTG TGGTGACCCT
401 GTTGGGCATC TTAACGGCAT TGCTCCTCGC TTTCTAACG CCGTCCACC
451 GAGACCACCT CCAGGGCTAG ACCAGCTTAA TCCCGAGGGC CCGCAGGTC
501 CCGGAGGGCC CCCCGCCATC TTGCCAGCTC TGCCGGCCCC GGCAGACCCT
551 GAACCGGCAC CACGGCGTGG TGGTGGGGCA GATGGAGCGG CCGCCGCTGG
601 GGCCGCCGCC GACGCAGACC ATACCGGGTA CGAAGAAGGA GACCTAGAAG
651 ATCTTTTCGC CGCCCGGGCC GAGGACGATA TGTGAGTAGG CCGAGGCGCC
701 GCCGCTACTA CAGGCGCAGA CTGAGACGGG GCAGACCCAG AGGGCGACGA
751 AAGAGACACA GACGACTCT AGTAGTGAGG CAGTGGCAAC CTGACGTTGT
801 TAAAAAGTGT AAAATAACAG GATGGATGCC TCTTATAATC TGTGGCTCTG
851 GAAGCACACA GATGAACTTT ATAACTCACA TGGACGATAC TCCCCCTATG
901 GGATACACCT ACGGGGGCAA CTTTGTAAT GTAACTTTCA GTCTAGAGGC
951 CATCTATGAA CAATTCCTGT ACCACAGAAA CAGGTGGTCC AGGTCTAACC
1001 ATGACTTAGA CCTGGCCAGA TACCAAGGAA CCACTCTAAA ACTTTACAGA
1051 CACCAAACCG TGGACTATAT AGTTAGCTAC AACAGAACAG GCCCCTTAC
1101 TATAAGTGAA ATGACTTACA TGAGCACACA CCCGGCTCTC ATGCTACTAC
1151 AAAAACATAG AATAGTTGTA CCCAGCTTCA GAACCAAGCC AAAAGGCAAA
1201 AGAGCCATAA AAATTAGAAAT AAGGGCCCCA AAATAATGC TCACCAAGTG
1251 GTACTTTACA AAAGACATTT GCTCCATGGG CCTCTTTCAA CTAATGGCAA
1301 CAGCTGCAGA ACTTACAAC CCATGGCTCA GAGACACCAC AAAAAGCCCA
1351 GTAATTGGCT TCAGAGTCTT AAAAAACAGC TTATACACAT GCCTTTCCAA
1401 CTTAAAAGAC CAAGCAATAC AAGGTGAAAG AAAGACTGTA CAAAATAGAT

```

ES 2 685 605 T3

1451 TACACCCAGA AACCTACAT GGCACAGGAC CTAATGCTAA AGGCTGGGAA
 1501 TACACATACA CAAAACATAAT GGCATCTACA TACTACTCAG CCAACAGAAA
 1551 CAGCACCTAC AACTGGCAA ACTATCARAC TAACTATGCA AACACATATA
 1601 CAAAATTTAA AGAAAAAAGA ACAGCAAAC TAACTTAAT TAAAGCAGAA
 1651 TACCTATATC ATTACCCTAA CAATGTCACA CAATCTGACT TTATATTAGA
 1701 CTACACACTA ACACCCGACT GGGGCATATA CAGCCCTAC TACCTAACAC
 1751 CCACCAGAAT TAGCCTAGAC TGGGACACAC CATGGACATA TGTAAAGATAC
 1801 AACCCACTAT CAGACAAAGG CATAGGTAAC AGAATATATG CACAGTGGTG
 1851 CTCAGAAAAA TCTAGTAAAT TAGACACCAC AAAGAGCAAG TGCATACTAA
 1901 GAGACTTCCC ACTGTGGGCC ATGGCCTATG GCTACTGTGA CTGGGTGGTG
 1951 AAGTGCACAG GAGTGTCCAG TGCTTGGACA GACATGAGAA TAGCCATTAT
 2001 ATGTCCCTAC ACAGAACCAG CACTTATAGG GTCAACAGAA GACGTAGGCT
 2051 TCATTCCAGT AAGTGACACC TTTTGCAACG GAGACATGCC GTTCTTGCA
 2101 CCATACATAC CTATTACATG GTGGATTAAAG TGGTACCCCA TGTATTACACA
 2151 CCAAAGGAA GTTCTTGAGG CAATAGTTAA CTGTGGACCG TTTGTACCCC
 2201 GAGACCAAAC TTCCCAGCT TGGGAATAAC CATGGGTAC AAAATGGATT
 2251 GGAANTGGG CGGCTCTCCC CTGCCTCAC AGGCAATCGA CGACCCCTGC
 2301 CAGAAGTCCA CCCACGAACT TCCGACCCC GATAGACACC CTCGCATGTT
 2351 ACAAGTCTCT GACCCGACAA AGCTCGGACC GAAGACAGTT TTTCACAAT
 2401 GGGACTGGAG ACGTGGGATG CTTAGCAAAA GAAGTATTA AAGACTCCAA
 2451 GAAGACTCAA CAGACGATGA ATATGTTGCA GGACCCTTAC CAAGAAAAAG
 2501 AAACAAGTTC GATACTCGAG TCCAAGGCC TCCAAACCCCA GAAAAAGAAA
 2551 GTTACACTTT ACTCCAAGCC CTCCAAGAGT CGGGGCAAGA GAGCAGCTCA
 2601 GAGGACCAAG AACAAGCACC CCAAGAAAAA GAGGACCAGA AGGAAGCGCT
 2651 CATGGAGCAG CTCAGCTCC AGAAACACCA CCAGCGAGTC CTCAGCGAG
 2701 GCCTCAAACCT CCTCCTCGGA GACGTGCTCC GACTCCGGAG AGGAGTCCAC
 2751 TGGGACCCCC TCCTGTCTTA ATTCAAGGTC CCAGTATCCC AGACCTGCTT
 2801 TTCCCTAACA CACAAAAAAA AAAACGATTT TCCAACACTAG ACTGGGTGTG
 2851 CGAGTAGCAG CTGGCCAAAT GGATGGATCG GCCCTTGGCG CACTACCCAT
 2901 CAGACCCCCC TCACTACCCC TGGCTACCAA AAAAGCCTCC TACCCCTCCT
 2951 ACATGTAGAG TAAGTTTCAA ATTAAAGCTC AATGACTAAA ATTCAAGGCC
 3001 GTGGGTGTTT CACTTCATCG GTGICTACCT CTAAAAGTCA CTAAGCACTC
 3051 CGAGCGTAAG CGAGGAGTGC GACCCCTCG CCGGTAGCA ACTTCCTCGG
 3101 GGTCCGGCGC TACGCCTTCG GCTGCGCGG GCGCCTCGGA CCCCCCTCG
 3151 ACCCGAATCG CTCGCGCGAT TCGGACCTGC GGCTCGGGG GGGTCGGGG
 3201 CTTTACTAAA CAGACTCTGA GGTCCGTTG GACTGAGG GGGTGAACAG
 3251 CAACGAAAGT GAGTGGGGCC AAACCTCGCC ATAAGGCCTT TAACTTTGGG
 3301 TCGCTTGTC A GCAGCTTCCG GGTCCGCTG GAGGCCGCCA TTTTACATTC
 3351 GSCCGCCATT FTAGGCCCTC GCGGGCCTCC ATAGTCGCAC ATCAGTGACG

ES 2 685 605 T3

```
3401 TCACGGCAGC CATCTTGGCT GTGACGTCAA CGTCACGTGG GGAGGACGGC
3451 GTGTAACCCG GAAGTCATCC TCATCACGCG ACCTGACGTC ACGGCCGCCA
3501 TTTTGTGCTG TCCGCCATCT TGTGACTTCC TTCCGCTTTT TGTAATAAAAA
3551 AGAGGAAGTG TGACGTAGCG GCGGGGGGGn nnnnnnnnnn nnnnnnnCGC
3601 CACCAGGGGG CGCTACGCGC CCCCCCCC GCATGTGCGG GTCCCCCCCC
3651 TCGGGGGGGG CTCCGCCCCC CCGGCCCCCC CCCGGGCTAA ATACACCGCG
3701 CATGCGGGC CACGCCCCCG CCGCC
```

B

zpr4.20

zpr4.20.seq (molécula subvívica , ttgb33.35)
293TT +4 (gbDhDi33.35) PCR
Tr9.7 A/9nested B5

en pCR2.1, Nova Blue

zpr4.B5.20.seq Longitud: 719

```
1 CAATTCGGC ACGGGACTGG CCGGGCTATG GGCAAGGCTC TTAATAAATT
51 CCCCCGCTCT GCTCTCCGGC AGGACACAAA GTCATGCCGT GGAGACCGCC
101 GGTCCATAAC GTGCCAGGTA GAGAGAATCA ATGGTTTGCA GCGTTCCTTC
151 ACGGTCATGC TGCTTTCTGC GGGTGTGGTG ACCCTGTGG GCATCTTAAC
201 GGCATTGCTC CTCGCTTTC TAACGCCGGT CCACCGAGAC CACCTCCAGG
251 GCTAGACCAG CTTAATCCCG AGGGCCCGC AGGTCCCGGA GGGCCCCCGC
301 CCATCTTGCC AGCTCTGCCG GCCCCGGCAG ACCCTGAACC GGCACCACGG
351 CGTGGTGGTG GGCAGATGG AGGCGCCGCC GCTGGGGCCG CCGCCGACGC
401 AGACCATAAC GGTACGAAG AAGGAGACCT CGGGGGGGGC TCCGCCCCCC
451 CGGCCCCCCC CCGGGCTAAA TACACCGCGC ATGCGCGGCC ACGCCCCCGC
501 CGCCATTTTG TGCAGCCCGC CAATTTCTGT TCAAACAGAC CAATCAGGAC
551 CTTCTACGTG CACTTCCTGG GCGTGTCTA CGAGGTCTAT ATAAGCAACA
601 GCGGTGACGA ATGGTAGAGT TTTTCTTCGC CCGTCCCGCG CGAGAGCGCG
651 AGCGAAGCGA GCGATCGAGC GTCCCGTGGG CGGGTGCCGT AGGTGAGTTT
701 ACACACCGAA GTCAGGGG
```

Figura 12

A

tth25 (genoma completo)
Longitud 3758

```

1  AAGTACGTCA CTAACCACGT GACTCCCGCA GGCCAACCAG AGTCTACGTC
51  GTGCACTTCC TGGGCATGGT CTACATCATA ATATAAGAAC GTGCACTTCC
101 GAATGGCTGA GTTTTCCACG CCCGTCCGCA GCGAGAACGC CACGGAGGGA
151 GATCCTCGCG TCCCGAGGGC GGGTGCCGGA GGTGAGTTA CACACCCGAG
201 TCAAGGGGCA ATTCCGGCTC GGGACTGGCC GGGCCCCGGG CAAGGCTCTT
251 AAAAAATGCG TTTTCGCAGG GTTGCCGAGA AAAGGAAAGT GCTTTTGCAA
301 ACTGTGCCAG CTGCAAAGAA GGCTAGGCGG CTTCTAGGTA TGTGGCAGCC
351 CCCCACGCAC AATGTCCCGG GCATCGAGAG AAAGTGTAC GAGAGCTGTT
401 TTAGATCCCA CGCTGCTGTT TGTGGCTGTG GCGATTTTGT TGGCCATCTT
451 AATCATCTGG CAACTACTCT GGGTCGTCTT CCGCGTCTG GGGCCCCAGG
501 CGGACCCCGC ACGCCGCAA TAAGAAACCT GCCAGCGCTC CGGGCGCCCC
551 AGGGCGAGCC CGGTGACAGA GCGCCATGGC ATGGGGCTTC TGGGCGCGAC
601 GCCGCGGTG GAGACGATGG AGAGCGCGGC GCAGACGGTG GAGACCCGCG
651 AGACGTAGGA GACGACGCC TACTCGCCGC TTTGAGCTC GTCGAAGAGT
701 AAGGAGGCGC GGGGGGAGGT GGCGCAGACG CTACAGAAA TGGCGACGGG
751 GCAGACGCAG ACGGACTCAT AGAAAAAGA TAGTCATAAA ACAGTGGCAA
801 CCAAACTTTA TAAGACGCTG CTACGTCATA GGTACTTAC CACTATATAT
851 CTGCGCGCAA AATACAACCG CCCAGAACTT TGCCACTCAC TCGGACGACA
901 TGATAAGCAA AGGACCGTAC GGGGGGGGCA TGACTACCAC CAAATTCACT
951 CTGAGAATAC TGTACGACGA GTTACCAGG TTTATGAACT TTTGGACTGT
1001 CAGTAACGAA GACCTAGACC TGTGTAGATA CGTGGGCTGC AAATAATAT
1051 TTTTAAACA CCCACGGTG GACTTTATAG TACAGATAAA CACTCAGCCT
1101 CCTTCTTAG ACACGCACCT CACCGGGCC AGCATAACCC CGGGCATCAT
1151 GATGCTCAGC AAGAGACACA TACTAATACC CTCTCTAAG ACCCGGCCCA
1201 GCAGAAAACA CAGGGTGGTC GTCAGGGTGG GCGCCCCAAG ACTTTTTCAG
1251 GACAAGTGGT ACCCCAGTC AGACCTGTGT GACACAGTTC TGCTTCCAT
1301 ATTTGCAACC GCCTGCGACT TGCAATATCC GTTCGGCTCA CCACTAACTG
1351 ACAACCCTTG CGTCAACTC CAGATCCTGG GGGCCAGTA CAAAAACAC
1401 CTTAGTATTA GCTCCACTAT GGATCAAAC AACGAAAACC ATTATAAGA
1451 AAACCTATTT AACAAAACG AACTATACAA CACCTTTCAA ACCATAGCTC
1501 AGCTTAAAGA GACAGGACAC ATTTACAGCA TTAGTCCTAC TTGGAATGAA
1551 GTCCAGAATT CAACAACACT TACTAAAGGA GGTGACAATG CCACTCAGAG
1601 TAGAGACACT TGGTATAAAG GAAATACATA CAACGAGAAG ATATGCGAGT

```

ES 2 685 605 T3

```

1651 TAGCACAAAT AACAGAAAC AGATTTAAAA ATGCAACCAA AGGAGCACTA
1701 CCAAACCTACC CCACAATAAT GTCCACAGAC CTATATGAAT ACCACTCAGG
1751 CATACTACTC AGCATATATC TATCAGCTGG CAGGAGCTAC TTTGAAACCA
1801 CCGGGGCCTA CTCTGACATT ATATACAACC CTTTCACAGA CAAAGGCACA
1851 GGCAACATAA TCTGGATAGA CTACCTCACA AAAGAAGACA CCATTTTGTG
1901 GAAAAACAAA AGCAAATGCG AGATAATGGA CATGCCCTTG TGGGCGGCCT
1951 GCACAGGATA CACAGAGTTT TGTGCAAAGT ATACAGGCGA CTCTGCCATT
2001 ATCTACAATG CAAGAATACT CATAAGATGC CCATAGACTG AGCCCATGTT
2051 AATAGACCAC TCAGACCCAA ACAAAGCTT CGTCCCTAC TCATTTAACT
2101 TTGGCAACGG AAAGATGCCC GGAGGACGCT CCAACGTGCC CATAAGAATG
2151 AGAGCCAAGT GGTACGTGAA CATATTCCAC CAAAAGAAG TATTAGAGAG
2201 CATAGTACAG TCCGACCGT TTGGGTACAA GGGCGACATA AGATCAGCTG
2251 TACTAGCCAT GAAATACAGA TTTCACTGGA AGTGGGCGG AAACCTATA
2301 TCCAACAGG TCGTCAGGAA TCCCTGCTCC AACTCCAGCT CCTCCGCGC
2351 CCATAGAGGA CCTCGCAGCG TACAAGCGT TGACCCGAAA TACAATACCC
2401 CAGAGGTAC GTGGCACTCG TGGGACATTA GACGAGGACT CTTTGGCAAA
2451 GCAGGTATTA AAAGAATGCA ACAGGAATCA GATGCTCTT ACATTCCTCC
2501 AGGACCAATC AAGAGACCTC GCAGGGACAC CAACGCCAA GACCCAGAAG
2551 AGCAAAACGA AAGCTCAGGT TTCAGAGTCC AGCAGGACT CCCGTGGGTG
2601 CACTCCAGCC AAGAGACCCA AAGCTCCCA GAAGAGACGG AGCGCCAGG
2651 GTCGGTACAA GACCAACTAC TCCTCCAGCT CCGAGAGCAG CGAGTTCTCC
2701 GACTCCAGCT CCAGCAACTC GCAACCCAAG TCCTCAAAGT CCAAGCAGGG
2751 CACAGCCTAC ACCCCCTATT ATCTTCCCAA GCATAAACAA AGCCTTTATG
2801 TTTGAGCCCC AGGGTCTTAA ACCCATAACG GGTACAACG ACTGGTAGA
2851 AGAGTACACT CCTTGCAAT TCTGGGACAG ACCCCGAGA AAGTACACA
2901 CAGACATACC CTTTACCCC TGGGCACCA AACCCCAACA GCAAGTCAGG
2951 GTGCCTTTA AACTCAACTT TCAATAAAA TTCTAGGCCG TGGGAGTTTC
3001 ACTTGTGGT GTCTGCTTCT TAAGTCCGC AAGCACTCC AGCGCCAGG
3051 AGGAGTGCGA CCCCCCTCC GGTAGCAACG CCTTCGGAGC CGCGCGCTAC
3101 GCCTFCGGCT GCGCGCGCA CCTCAGACCC CCCCTCCACC CGAAACGCTT
3151 GCGCGTTTCG GACCTTCGGC GTCGGGGGG TCGGGAGCTT TATTAACAG
3201 ACTCCGAGTT GCCATTGGAC ACTGGAGCTG TGAATCAGTA ACGAAAGTGA
3251 GTGGGGCCAG ACTTCGCCAT AGGGCCTTA TCTTCTCGCC ATTGGATAGT
3301 GTCGGGGTC GCCGTAGGCT TCGGCTCGT TTTTAGGCCT TCCGACTAC
3351 AAAATGGCG GTTTTAGTGA CGTCACGGC GCCATTTAA GTAAGCGGA
3401 AGCAGTCCA CTTTCTACA AATGGCGGC GGAGCACTTC CGGCTTGCCC
3451 AAAATGGCG GCAAGCTCTT CCGGTAAAG GGTACAGCAG TACGTCACAA
3501 GTCACCTGAC TGGGGAGGG TCACAACCG GAAGCCCTCC TCAGTCACGT
3551 GGCTGTTTAC GTGGTTGCTA CGTCAICGGC GCCATCTGT GTCGCAAAAT

```

ES 2 685 605 T3

```
3601 GCGCGACAAC TTCCGCTTTT TTAATAAAG GCGCGAAAAA ACGGCGGGCG
3651 CGGCGCGCGC GCTGTGCGCG CGCGCCGGGG GGGCGCCAGC GCCCCCCCCC
3701 CCGCGCATGC GCGGGTCCCC CCCCCGCGG GGGGCTCCGC CCCCCGGCCC
3751 CCCCCCG
```

B

zpr9.6

zpr9.B1.6.seq (molécula subvirica, tth25)
293TT+9(tth25smfr3)PCR
Tr11.8 C/6
En pCR2.1,ONESHOT Topo10F'

```
zpr9.B1.6.seq Longitud: 621
 1 CCGCAGCGAG AACGCCACGG AGGAGATCC TCGCGTCCCG AGGCGGGTG
 51 CCGGAGGTGA GTTTACACAC CGCAGTCAAG GGGCAATTCG GGCTCGGGAC
101 TGGCCGGGCC CCGGGCAAGG CTCTTAAAAA ATGCGTTTTT GCAGGGTTGC
151 CCAGAAAAGG AAAGTGCTTT TGCAAACTGT GCCAGCTGCA AAGAAGGCTA
201 GCGGCTTCT AGGTATGTGG CAGCCCCCA CGCACATGT CCCGGGCATC
251 GAGAGAACT GGTACGAGAG CTGTTTTAGA TCCCACGCTG CTGTTTGTGG
301 CTGTGGCGAT TTTGTTGGCC ATCTTAATCA TCTGGCAACT ACTCTGGGTC
351 GTCCTCCGCG TCCTGGGCC CCAGCGGAC CCCGCACGCC GCAAATAAGA
401 AACCTGCCAG CGCTCCCGC GCCCCAGGC GAGCCCGTG ACAGAGCGCC
451 ATGGCATGGG GCTTCTGGG CCGACGCCG CGGTGGAGAC GATGGAGAGC
501 GCGGCGAGA CGGTGGAGAC CCCGAGGCC AACCAGAGTC TACGTCGTGC
551 ACTTCCTGGG CATGGTCTAC ATCATAATAT AAGAACGTGC ACTTCCGAAT
601 GGCTGAGTTT TCCACGCCG T
```

Figura 13

A

ttrh215
rheu.cd.215.seq
Genomiphi tth4-cebador
aso. 78°+ . taq polimerasa adicional (progr.tth4*78)
vector:TA-cloning pcr2.1 invitrogen

rheu.cd.215rp.seq Longitud : 3758

```

1 AAAGTACGTC ACTAACCACG TGACTCCCAC AGGCCARCCA CAGTCTACGT
51 CGTGCATTTC CTGGGCATGG TCTACATCAT AATATAAGAA GGCGCAC TTC
101 CGAATGGCTG AGTTTTCCAC GCCCGTCCGC AGCGAGAACG CCACGGAGGG
151 AGATCCTCGC GTCCCGAGGG CGGGTGCCGG AGGTGAGTTT ACACACCGCA
201 GTC AAGGGGC AATTCGGGCT CGGGACTGGC CGGGCCCTGG GCAAGGCTCT
251 TAAAAAATGC GCTTTCGCAG GGTTCGGGAG AAAAGGAAAG TGCTTCTGCA
301 AACTCTGCGA GCTGCAAAGC AGGCTAGGCG GCTTCTAGGT ATGTGGCAGC
351 CCCCCGCGCA CAATGTCCCC GGCATCGAGA GAAACTGGTA CGAGAGCTGC
401 TTCAGGTCTC ACGCTGCTGT TTGTGGCTGT GGCACITTTG TTGGCCATAT
451 TAATCATTTC GCAACTACTC TGGGTCGTCC TCCGCGTCCT GGGCCCCCAG
501 GCGGACCCCG CACGCCGCAA ATAAGAAACC TGCCAGCGCT CCCGGCGCCC
551 CAGGGCCGAG CCGGTGACAG AGGCCCATGG CGTGGGGTTT CTGGGGCCGA
601 CCGCCCGGTT GGAGACGGTG GAGAGCCCGG CGCAGACGGT GGAGACCCCG
651 GAGACGTAGG AGACGACGCC CTGCTCGCCG CTTTCGAGCT CGTCGAAGAG
701 TAAGGAGACG CGGGGGGAGG TGGCGCAGAC GCTACAGAAA ATGGCGACGG
751 GGCAGACGCA GACGGACTCA CAGAAAAAAG ATAATTATTA AACAGTGGCA
801 ACCAAACTTT ATTAGACGCT GCTACATAAT AGGATGCCTA CCTCTCGTTT
851 TCTGTGGCGA AAATACAACC GCCCAGAACT ATGCCACTCA CTCAGACGAT
901 ATGATAAGCA AAGGACCGTA CGGGGGGGGC ATGACTACCA CGAARTTCAC
951 TCTGAGAATA CTGTACGACG AGTTTACCAG GTTTATGAAC TTTTGGACTG
1001 TCAGTAACGA AGACCTAGAC CTGTGTAGAT ACGTGGGCTG CAAACTGATA
1051 TTTTTTAAAC ACCCCACGGT GGACTTTATG GTACAGATRA AACTCAGCC
1101 TCCTTTCTTA GACACAAGCC TCACCGCGGC CAGCATAAC CCGGGCATCA
1151 TGATGCTCAG CAAGAGACGC ATATTAATAC CCTCTCTAAA GACCCGGCCG
1201 AGCAGAAAAC ACAGGGTGGT CGTCAGGGTG GGCGCCCAA GACTTTTTCA
1251 GGACAAGTGG TACCCAGT CAGACCTATG TGACACAGTT CTGCTTTCCA
1301 TATTTGCAAC CGCCCGCAC TTGCAATATC CGTTCGGCTC ACCACTAACT
1351 GACAACCCCT CCGTCAACTT CCAGATCCTG GGGCCCCAGT AAAAAAACA
1401 CCTTAGTATT AGCTCCACTA TGGATGATAC TAACAAACAG CACTATAACA
1451 GCAACTTATT TAATAAAACT GCACTATACA ACACCTTTCA AACCATAGCC

```

ES 2 685 605 T3

1501 CGGCTTAAAG AGACAGGACA AACTGCAAAC ATTAGTCCAA GTTGGAGTGA
 1551 AGTACAAAAC ACAAACTAC TAGATCACAC AGGTGCTAAT GCAACTGCCA
 1601 GCAGAGACAC TTGGTACAAG GGAACACAT ACAATGACTA CATACAACAG
 1651 TTAGCAGAGA AAACAAGAGA AAGGTTAAA AAAGCAACAA TGTCCAGCACT
 1701 ACCAAACTAC CCCACAATAA TGTCCACAGA CTTATACGAA TACCACTCAG
 1751 GCATNTACTC CAGCATATTT CTATCAGCTG GCAGGAGCTA CTTTGA AAC
 1801 ACTGGGGCCT ACTCTGACAT TATATACAAC CCTTTGACAG ACAAAGGCAC
 1851 AGGCAACATA ATCTGGATAG ACTACCTTAC AAAAGACGAC ACAATCTTTG
 1901 TAAAAAACA AAGCAAATGT GAGATAATGG ACATGCCCTT GTGGGGGGCC
 1951 GGCACAGGAT ACACAGAGTT TTGTGCAAAG TACACAGGAG ACTCTGCCAT
 2001 TATTTACAAT GCCAGAATAC TCATAAGATG CCCATACACT GAACCCATGC
 2051 TAATAGACCA CTCAGACCCA AACAAAGGCT TTGTACCGTA CTCATTTAAC
 2101 TTTGGCAACG GAAAGATGCC GGGAGGACG TCCAACGTGC CCATAAGAAT
 2151 GAGAGCCAAG TGGTACGTAA ACATATTTCCA CCAAAAAGAA GTATTGGAGA
 2201 GCATAGTACA GTCCGGACCG TTCGGGTACA GGGGCGACAT AAANTCAGCT
 2251 GTACTGTCCA TGAANTACAG ATTTCACTGG AAATGGGGCG GAAACCCAT
 2301 ATCCAACACG GTCGTCAGGA ATCCCTGCTC CAACTCCAGC ACCTCCGGCG
 2351 CCCATAGAGG ACCTCGCAGC GTACAAGCGG TTGACCCGAA ATACAATACC
 2401 CCAGAAGTCA CTTGGCACTC GTGGGACATC AGACGAGGAC TCTTTGGCAA
 2451 AGCAGGTATT AAAAGAATGC AACAGAATC AGATCCTCTT TACGTTCCTG
 2501 CAGGACCACT CAAGAGGCCT CGCAGAGACA CCAACGCCCA AGACCCGGAA
 2551 AAGCAAAACG AAAGCTCAGC TTTCCGAGTC CAGCAGCGAC TCCCGTGGGT
 2601 CCACTCCAGC CAAGAGACGC AAAGCTCCGA AGAAGAGACG CAGGCGCAGG
 2651 GGTCCGTACA AGACCAACTA CTCCTCCAGC TCCGAGAGCA GCGAGTACTC
 2701 CGACTCCAGC TCCAACAACG CGCACCCCAA GTCCTCAAAG TTCAAGCAGG
 2751 ACACAGCCTA CACCCCTAT TATCTCCCA AGCATAAACA AAGCCTATAT
 2801 GTTTGAACCC CAGGGTCCTA AACCCATACA GGGGTACAAC GATTCGCTAG
 2851 AGGAGTACAC TAGTTGCAAG TTCCGGGACA GACCCCGGAG AATGCTACAC
 2901 ACAGACTTAC CCTTTTACCC CTGGGCACCA AAACCCCAAG ACCAAGTCAG
 2951 GGTAACCTTT AAACCAACT TTCATAAAA ATTCTAGGCC GTGGGACTTT
 3001 CACTTGTCCG TGFTGCTTC TTAAGGTCGC CAAGCACTCC GAGCGTCAGC
 3051 GAGGAGTGCG ACCCCCCCC TCGGTAGCAA CGCCTTCGGA GCCGCGGCT
 3101 ACGCCTTCGG CTGCGCGCGG CACCTCAGAC CCCCCCTCA CCCGRAACGC
 3151 TTGCGCGTTT CGGACCTTCG GCSTCGGGG GGTCCGGAGC TTTATTAAC
 3201 AGACTCCGAG TTGCCATTGG ACACTGGAGC TGTGAATCAG TAACGAAAGT
 3251 GAGTGGGGCC AGACTTCGCC ATAGGGCCTT TATCTTCTCG CCATTGGATA
 3301 GTGTCCGGGG TTGCCGTAGG CTTCCGGCCT GTTTTAGGC CTTCGGACT
 3351 ACAAAAATGG CGGATTTGT GACGTCACGG CCGCCATTT AAGTAAGGCG
 3401 GAAGCAGCTC CACCCTCTCA CATAATGGCG GCGGAGCACT CCCGGCTTG

ES 2 685 605 T3

```
3451 CCAAAATGGC GGGCAAGCTC TTCCGGGTCA AAGTTGGCA GCTACGTCAC
3501 AAGTCACCTG ACTGGGGAGG AGTTACATCC CGGAAGTTCT CCTCGGTAC
3551 GTGACTGTAC ACGTACTGC TACGTCATTG ACGCCATCTT GTGTCACAAA
3601 ATGGCGGTGC ACTTCCGCTT TTTGAAAAA AGGCGCGAAA AAACGGCGGC
3651 GGGCGCGCGC GCGCTGCGCG CGCGCGCCGG GGGGCGCCA GCGCCCCCC
3701 CCCCgcgcAT GCACGGGTCC CCCCCCCAC GGGGGGTCC GCCCCCGGC
3751 CCCCCCC
```

B

zpr12.24
zpr12.24.seq (molécula subvívica, ttrh215)
Longitud: 642

```
1 CAGCGAGAAC GCCACGGAGG GAGATCCTCG CGTCCCAGG GCGGGTGCCG
51 GAGGTGAGTT TACACACCGC AGTCAAGGGG CAATTCGGGC TCGGGACTGG
101 CCGGGCCCCG GGCAAGGCTC TTAAAAATG CGCTTTCGCA GGGTTGCTGA
151 GAAAAGGAAA GTGCTTCTGC AAAGTGTGCG AGCTACACAG AAGACTAGGC
201 GGCTTCTAAG CCGCCACAG GGGCATGTCT ACATGCTTCC GCAGCGAGAA
251 CGCCACGGAG GGAGATCCTC GCGTCCCAG GCGGGTGCC GGAGTGAGT
301 TTACACACCG CAGTCAAAGG GCAATTCGGG CTCGGGACTG GCCGGGCCCC
351 GGGCAAGGCT CTAAAAAAT GCGCTTTCGC GGGTTGCTG AAAAAAGGAA
401 AGTGCTTCTG CAAACTGTGC GAGTACACA GAAGACTAGG CGGCTTCTAG
451 GTATGTGGCA GCCCCCCGTG CACAATGTCC CCGGCATCTT ATTAGTACT
501 TGGCGTTGTA GATAATGGCA GAGTCTCCAG TGTACTTTGC ACAGAACTCT
551 GTGTATCCTG TGCAGGCCG CACAGGGGC ATGTCTACAT CATAATATAA
601 TAAGGCGCAC TTCCGAATGG CTGAGTTTC CACGCCCGTC CG
```

Figura 14

Marcos de lectura abierta de 71 nt (HCR) :

```

zyb2.1.pep  RVPKVSLHTA VKGQFGLGTG RAM
zyb9.1.pep  RVPKVSLHTA VKGQFGLGTG RAM
zkb69.1.pep RVPEVSLHTA VKGQFGLGTG RAM

zyb2.3.pep  GAEGEFTHRS QGAIRARDWP GHG
zyb9.3.pep  GAEGEFTHRS QGAIRARDWP GYG
zkb5.3.pep  GAVGEFTHRS QGAIRARDWP GYG
zkb69.3.pep GAGGEFTHRS QGAIRARDWP GYG
    
```

zyb2.1.pep

nucléotidos 1-71
Longitud: 23aa

RVPK~~V~~SLHTA VKGQFGLGTG RAM

BlastP2 de: zyb2.1.pep

comparado con la base de datos: **uniprot** ..

>>>sptrembl:Q9WSW0_9VIRU	Q9wsw0	SubName: Full=ORF2; Flags: Fragm...	<u>49</u>	1e-04
>>>sptrembl:Q9WB09_9VIRU	Q9wb09	SubName: Full=ORF2; Flags: Fragm...	<u>48</u>	4e-04
>>>sptrembl:Q9WSW2_9VIRU	Q9wsw2	SubName: Full=ORF2; Flags: Fragm...	<u>48</u>	4e-04
>>>sptrembl:Q9WSX0_9VIRU	Q9wsx0	SubName: Full=ORF2; Flags: Fragm...	<u>47</u>	4e-04
>>>sptrembl:Q9WB10_9VIRU	Q9wb10	SubName: Full=ORF2; Flags: Fragm...	<u>47</u>	5e-04
>>>sptrembl:Q9WAZ2_9VIRU	Q9waz2	SubName: Full=ORF2; Flags: Fragm...	<u>47</u>	5e-04
>>>sptrembl:O70807_9VIRU	O70807	SubName: Full=Putative uncharact...	<u>47</u>	6e-04
>>>sptrembl:Q9WAY4_9VIRU	Q9way4	SubName: Full=ORF2; Flags: Fragm...	<u>47</u>	7e-04
>>>sptrembl:Q9WB02_9VIRU	Q9wb02	SubName: Full=ORF2; Flags: Fragm...	<u>47</u>	8e-04
>>>sptrembl:Q9WSW4_9VIRU	Q9wsw4	SubName: Full=ORF2; Flags: Fragm...	<u>46</u>	0.001
>>>sptrembl:Q9WB12_9VIRU	Q9wb12	SubName: Full=ORF2; Flags: Fragm...	<u>46</u>	0.001
>>>sptrembl:Q9WSW6_9VIRU	Q9wsw6	SubName: Full=ORF2; Flags: Fragm...	<u>45</u>	0.002
>>>sptrembl:B3FWR6_9VIRU	B3fwr6	SubName: Full=ORF2; Flags: Fragm...	<u>36</u>	0.94

>>>>sptrembl:Q9WSW0_9VIRU Q9wsw0 SubName: Full=ORF2; Flags:
Fragmento; : 2/2010
Longitud = 204

Puntuación = 49.3 bits (116), Esperada = 1e-04, Método: ajuste de matriz composicional
Identities = 23/23 (100%), Positivos = 23/23 (100%)

Query: 1 RVPK~~V~~SLHTAVKQFGLGTGRAM 23
RVPK~~V~~SLHTAVKQFGLGTGRAM
Sbjct: 27 RVPK~~V~~SLHTAVKQFGLGTGRAM 49

>>>>sptrembl:Q9WB09_9VIRU Q9wb09 SubName: Full=ORF2; Flags:
Fragmento; : 2/2010
Longitud = 138

Puntuación = 47.8 bits (112), Esperada = 4e-04, Método: ajuste de matriz composicional
Identities = 22/23 (95%), Positivos = 22/23 (95%)

Query: 1 RVPK~~V~~SLHTAVKQFGLGTGRAM 23
RVPK~~V~~SLHTAVKQFGLGTGRAM
Sbjct: 14 RVPK~~V~~SLHTEVKQFGLGTGRAM 36

ID **Q9WSW0_9VIRU** Sin revisar; 204 AA.

ES 2 685 605 T3

AC Q9WSW0;
 DT 01-NOV-1999, integrado en UniProtKB/TrEMBL.
 DT 01-NOV-1999, versión de secuencia 1.
 DT 09-FEB-2010, versión de entrada 23.
 DE SubName: Full=ORF2;
 DE Flags: Fragmento;
 OS Torque teno virus.
 OC Virus; virus de ADNss; Anelloviridae; Anelloviridae sin clasificar.
 OX NCBI_TaxID=68887;
 RN [1]
 RP SECUENCIA DE NUCLEÓTIDOS
 RC CEPA =KC205/1-12G; TEJIDO=Sero;
 RX MEDLINE=99335592; PubMed=10405352; DOI=10.1006/viro.1999.9797;
 RA Hijikata M., Takahashi K., Mishiro S.;
 RT "Complete circular DNA genome of a TT virus variant (isolate name
 RT SANBAN) and 44 partial ORF2 sequences implicating a great degree of
 RT diversity beyond genotypes."
 RL Virology 260:17-22(1999).
 CC -----
 CC Derechos de autor de UniProt Consortium, véase <http://www.uniprot.org/terms>
 CC Distribuido bajo la licencia de Creative Commons Attribution-NoDerivs
 CC -----
 DR EMBL; [AB024383](#); [BAA77450.2](#); -; Genomic DNA.
 DR InterPro; [IPR004118](#); Gyrovir_VP2/TT_ORF2.
 DR Pfam; [PF02957](#); TT_ORF2; 1.
 PE 4: Previsto;
 FT NON TER 1 1
 SQ SECUENCIA 204 AA; 21953 MW; 6352C96D2AC0DF21 CRC64;
 CTSEWLSFPR PSAAAXPRRV IPASRWRVVK VSLHTAVKGO FGLGTGRAMG KALKVFILKM
 HFSRISRKR KVLLPALPAP PPRQLLMWQ PPIQNGTQLD RHWFESVWRS HAAAYCGCGDC
 VGHLQHLAAN LGRPPHPQPP REQHPPIRG LPALPAPPSN RNSWFGTGGD AAGEQAGGSR
 GAGDGGDGEL ADDDLXDAAL LVEE

ID Q9WB09_9VIRU Sin revisar ; 138 AA.
 AC Q9WB09;
 DT 01-NOV-1999, integrado en UniProtKB/TrEMBL.
 DT 01-NOV-1999, versión de secuencia 1.
 DT 09-FEB-2010, versión de entrada 22.
 DE SubName: Full=ORF2;
 DE Flags: Fragmento;
 OS Torque teno virus.
 OC Virus; virus de ADNss; Anelloviridae; Anelloviridae sin clasificar.
 OX NCBI_TaxID=68887;
 RN [1]
 RP SECUENCIA DE NUCLEÓTIDOS
 RC TEJIDO=Sero;
 RX MEDLINE=99335592; PubMed=10405352; DOI=10.1006/viro.1999.9797;
 RA Hijikata M., Takahashi K., Mishiro S.;
 RT "Complete circular DNA genome of a TT virus variant (isolate name
 RT SANBAN) and 44 partial ORF2 sequences implicating a great degree of
 RT diversity beyond genotypes."
 RL Virology 260:17-22(1999).
 CC -----
 CC Derechos de autor de UniProt Consortium, véase <http://www.uniprot.org/terms>
 CC Distribuido bajo la licencia de Creative Commons Attribution-NoDerivs
 CC -----
 DR EMBL; [AB024376](#); [BAA77443.1](#); -; Genomic DNA.
 DR InterPro; [IPR004118](#); Gyrovir_VP2/TT_ORF2.
 DR Pfam; [PF02957](#); TT_ORF2; 1.
 PE 4: Previsto;
 FT NON TER 1 1
 FT NON TER 138 138
 SQ SECUENCIA 138 AA; 15494 MW; 2DF27B3A4F0CA641 CRC64;

ES 2 685 605 T3

AVKPRREISA SRGRVPKVL HTEVKGQFGL GTGRAMGKAL KKSMEFGRHY RKKRALSLCA
 VRTTKKACKL LIVMWTPPRN DQQLYNQWY SSVLSSHAAM CGCPDAIAHL SHLAFVFRAP
 QNPPPPGPQR NLPLRRLP

zyb2.3.pep

nucleótidos 3-71
 Longitud: 23aa

GAEGEFTHRS QCAIRARDWP GHG

BlastP2 de: zyb2.3.pep desde: 1 a: 23
 comparado con la base de datos: **uniprot** ..

>>>sptrembl:Q98Y39 9VIRU	Q98y39 SubName: Full=ORF2;. 2/2010	50	7e-05
>>>sptrembl:Q9WAY7 9VIRU	Q9way7 SubName: Full=ORF2; Flags: Fragm...	50	7e-05
>>>sptrembl:Q786D4 9VIRU	Q786d4 SubName: Full=ORF2;. 2/2010	50	9e-05
>>>sptrembl:Q70738 9VIRU	Q70738 SubName: Full=ORF2, ORF1 genes;...	50	9e-05
>>>sptrembl:Q9JG33 9VIRU	Q9jg33 SubName: Full=Putative uncharact...	50	9e-05
>>>sptrembl:Q90363 9VIRU	Q90363 SubName: Full=ORF2 protein;. 2/2010	50	9e-05
>>>sptrembl:Q9WFY6 9VIRU	Q9wfy6 SubName: Full=ORF2;. 2/2010	50	1e-04
>>>sptrembl:Q9JGT0 9VIRU	Q9jgt0 SubName: Full=PORF2a;. 12/2009	50	1e-04
>>>sptrembl:Q9YKL2 9VIRU	Q9ykl2 SubName: Full=ORF2 protein;. 2/2010	50	1e-04
>>>sptrembl:Q9JGS7 9VIRU	Q9jgs7 SubName: Full=PORF2a;. 12/2009	49	1e-04
>>>sptrembl:Q9DYC0 9VIRU	Q9dyc0 SubName: Full=Putative uncharact...	49	1e-04
>>>sptrembl:Q9W7S4 9VIRU	Q9w7s4 SubName: Full=Putative uncharact...	49	2e-04
>>>sptrembl:Q77S01 9VIRU	Q77s01 SubName: Full=Putative uncharact...	49	2e-04
>>>sptrembl:Q9JGT3 9VIRU	Q9jgt3 SubName: Full=PORF2a;. 12/2009	46	0.001
>>>sptrembl:Q9JGS4 9VIRU	Q9jgs4 SubName: Full=PORF2a;. 12/2009	46	0.001
>>>sptrembl:Q9YR02 9VIRU	Q9yr02 SubName: Full=Putative uncharact...	45	0.003
>>>sptrembl:B2YFW4 9VIRU	B2yfw4 SubName: Full=ORF2; Flags: Fragm...	34	3.7

>>>>sptrembl:Q98Y39 9VIRU Q98y39 SubName: Full=ORF2;. 2/2010
 Longitud = 202

Puntuación = 50.1 bits (118), Esperado = 7e-05, Método: ajuste de matriz composicional
 Identidades = 22/23 (95%), Positivos = 23/23 (100%)

Query: 1 GAEGEFTHRSQCAIRARDWPGHG 23
 GAEGEFTHRSQCAIRARDWPG+G
 Sbjct: 24 GAEGEFTHRSQCAIRARDWPGYG 46

Q98Y39 9VIRU

Sin revisar ; 202 AA.

AC Q98Y39;
 DT 01-JUN-2001, integrado en UniProtKB/TrEMBL.
 DT 01-JUN-2001, versión de secuencia 1.
 DT 09-FEB-2010, versión de entrada 17.
 DE SubName: Full=ORF2;
 OS Torque teno virus.
 OC Virus; virus de ADNss; Anelloviridae; Anelloviridae sin clasificar.
 OX NCBI_TaxID=68887;
 RN [1]
 RP SECUENCIA DE NUCLEÓTIDOS.
 RC CEPA =TWH;
 RA He H.-T., Luo K.-X., Xiao H., Liu D.-X.;
 RT "Complete circular genome of TT virus isolated from feces of a
 RT hepatitis patient."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
 CC -----
 CC Derechos de autor de UniProt Consortium, véase <http://www.uniprot.org/terms>
 CC Distribuido bajo la licencia de Creative Commons Attribution-NoDerivs
 CC -----
 DR EMBL; [AF351132](#); [AAK29446.1](#); -; Genomic_DNA.
 DR InterPro; [IPR004118](#); Gyrovir_VP2/TT_ORF2.
 DR InterPro; [IPR013267](#); TTV_ORF2a.

ES 2 685 605 T3

DR Pfam; [PF02957](#); TT_ORF2; 1.
 DR Pfam; [PF08197](#); TT_ORF2a; 1.
 PE 4: Previsto;
 SQ SECUENCIA 202 AA; 21437 MW; 105B9ED104956EDE CRC64;
 MAEFSTPVRS GEATEGDHRV PRAGAEGEFT HRSQGAIRAR DWPGYGGGSE KSMFIGRHYR
 KKRALSLCAV RTTKKACKLL IVMWTPPRND QQYLNWQWYS SVLSSHASMC GCPDAVAHLI
 NLASVLRAPQ NPPPPGPORN LPLRRLPALP AAPEAPGDRA PWPMAGGAEG ENGGAGGDAD
 HGGAAGGPED ANLLDAVAAA ET
 //

zyb9.1.pep

nucleótidos 1-71
 Longitud: 23aa

RVPKVS LHTA VKGQFGLGTG RAM

BlastP2 de: zyb9.1.pep desde: 1 a: 23
 comparado con la base de datos: **uniprot** ..

>>>sptrembl:Q9WSW0_9VIRU	Q9wsw0	SubName: Full=ORF2; Flags: Fragm...	<u>49</u>	1e-04
>>>sptrembl:Q9WB09_9VIRU	Q9wb09	SubName: Full=ORF2; Flags: Fragm...	<u>48</u>	4e-04
>>>sptrembl:Q9WSW2_9VIRU	Q9wsw2	SubName: Full=ORF2; Flags: Fragm...	<u>48</u>	4e-04
>>>sptrembl:Q9WSX0_9VIRU	Q9wsx0	SubName: Full=ORF2; Flags: Fragm...	<u>47</u>	4e-04
>>>sptrembl:Q9WB10_9VIRU	Q9wb10	SubName: Full=ORF2; Flags: Fragm...	<u>47</u>	5e-04
>>>sptrembl:Q9WAZ2_9VIRU	Q9waz2	SubName: Full=ORF2; Flags: Fragm...	<u>47</u>	5e-04
>>>sptrembl:O70807_9VIRU	O70807	SubName: Full=Putative uncharact...	<u>47</u>	6e-04
>>>sptrembl:Q9WAY4_9VIRU	Q9way4	SubName: Full=ORF2; Flags: Fragm...	<u>47</u>	7e-04
>>>sptrembl:Q9WB02_9VIRU	Q9wb02	SubName: Full=ORF2; Flags: Fragm...	<u>47</u>	8e-04
>>>sptrembl:Q9WSW4_9VIRU	Q9wsw4	SubName: Full=ORF2; Flags: Fragm...	<u>46</u>	0.001
>>>sptrembl:Q9WB12_9VIRU	Q9wb12	SubName: Full=ORF2; Flags: Fragm...	<u>46</u>	0.001
>>>sptrembl:Q9WSW6_9VIRU	Q9wsw6	SubName: Full=ORF2; Flags: Fragm...	<u>45</u>	0.002
>>>sptrembl:B3FWR6_9VIRU	B3fwr6	SubName: Full=ORF2; Flags: Fragm...	<u>36</u>	0.94

>>>>sptrembl:Q9WSW0_9VIRU Q9wsw0 SubName: Full=ORF2; Flags:
 Fragmento ; . 2/2010
 Longitud = 204
 Puntuación = 49.3 bits (116), Esperado = 1e-04, Método: ajuste de matriz composicional.
 Identidades = 23/23 (100%), Positivos = 23/23 (100%)
 Query: 1 RVPKVS LHTAVKQFGLGTGRAM 23
 RVPKVS LHTAVKQFGLGTGRAM
 Sbjct: 27 RVPKVS LHTAVKQFGLGTGRAM 49

>>>>sptrembl:Q9WB09_9VIRU Q9wb09 SubName: Full=ORF2; Flags:
 Fragmento ; . 2/2010
 Longitud = 138
 Puntuación = 47.8 bits (112), Esperado = 4e-04, Método: ajuste de matriz composicional.
 Identidades = 22/23 (95%), Positivos = 22/23 (95%)
 Query: 1 RVPKVS LHTAVKQFGLGTGRAM 23
 RVPKVS LHTAVKQFGLGTGRAM
 Sbjct: 14 RVPKVS LHTAVKQFGLGTGRAM 36

Q9WSW0_9VIRU Sin revisar ; 204 AA.
 AC Q9WSW0;
 DT 01-NOV-1999, integrado en UniProtKB/TrEMBL.
 DT 01-NOV-1999, versión de secuencia 1.
 DT 09-FEB-2010, versión de entrada 23.
 DE SubName: Full=ORF2;
 DE Flags: Fragmento;

ES 2 685 605 T3

OS Torque teno virus.
 OC Virus; virus de ADNss ; Anelloviridae; Anelloviridae sin clasificar.
 OX NCBI_TaxID=68887;
 RN [1]
 RP SECUENCIA DE NUCLEÓTIDOS.
 RC CEPA =KC205/1-12G; TEJIDO=Suero;
 RX MEDLINE=99335592; PubMed=10405352; DOI=10.1006/viro.1999.9797;
 RA Hijikata M., Takahashi K., Mishiro S.;
 RT "Complete circular DNA genome of a TT virus variant (isolate name
 RT SANBAN) and 44 partial ORF2 sequences implicating a great degree of
 RT diversity beyond genotypes.";
 RL Virology 260:17-22(1999).
 CC -----
 CC Derechos de autor de UniProt Consortium, véase <http://www.uniprot.org/terms>
 CC Distribuido bajo la licencia de Creative Commons Attribution-NoDerivs
 CC -----
 DR EMBL; [AB024383](#); [BAA77450.2](#); -; Genomic DNA.
 DR InterPro; [IPR004118](#); Gyrovir_VP2/TT_ORF2.
 DR Pfam; [PF02957](#); TT_ORF2; 1.
 PE 4: Previsto;
 FT NON TER 1 1
 SQ SECUENCIA 204 AA; 21953 MW; 6352C96D2AC0DF21 CRC64;
 CTSEWLSFPR PSAAAXPRRV IPASRWRVVK VSLHTAVKGG FGLGTGRAMG KALKVFILKM
 HFSRISRSKR KVLPLPALPAP PPRQLLMWQ PPIQNGTQLD RHWFESVWRS HAAYCGCGDC
 VGHLOHLAAN LGRPPHPQPP REQHPPIRG LPALPAPPSN RNSWPGTGGD AAGEQAGGSR
 GAGDGGDGGEL ADDDLXDAAA LVEE

Q9WB09_9VIRU Sin revisar; 138 AA.
 AC Q9WB09;
 DT 01-NOV-1999, integrado en UniProtKB/TrEMBL.
 DT 01-NOV-1999, versión de secuencia 1.
 DT 09-FEB-2010, versión de entrada 22.
 DE SubName: Full=ORF2;
 DE Flags: Fragmento;
 OS Torque teno virus.
 OC Virus; virus de ADNss ; Anelloviridae; Anelloviridae sin clasificar.
 OX NCBI_TaxID=68887;
 RN [1]
 RP SECUENCIA DE NUCLEÓTIDOS.
 RC TEJIDO=Suero;
 RX MEDLINE=99335592; PubMed=10405352; DOI=10.1006/viro.1999.9797;
 RA Hijikata M., Takahashi K., Mishiro S.;
 RT "Complete circular DNA genome of a TT virus variant (isolate name
 RT SANBAN) and 44 partial ORF2 sequences implicating a great degree of
 RT diversity beyond genotypes.";
 RL Virology 260:17-22(1999).
 CC -----
 CC Derechos de autor de UniProt Consortium, véase <http://www.uniprot.org/terms>
 CC Distribuido bajo la licencia de Creative Commons Attribution-NoDerivs
 CC -----
 DR EMBL; [AB024376](#); [BAA77443.1](#); -; Genomic DNA.
 DR InterPro; [IPR004118](#); Gyrovir_VP2/TT_ORF2.
 DR Pfam; [PF02957](#); TT_ORF2; 1.
 PE 4: Previsto;
 FT NON TER 1 1
 FT NON TER 138 138
 SQ SECUENCIA 138 AA; 15494 MW; 2DF27B3A4FOCA641 CRC64;
 AVKPREISA SRGRVVKVSL HTEVKGQFGL GTGRAMGKAL KKSMEFGRHY RKKRALSICA
 VRTTKKACKL LIVMWTTPRN DQOYLNQWY SSVLSSHAAM CGCPDAIAHL SHLAFVFRAP
 QNPPPPGPQR NLPLRRLP

zyb9.3.pep

nucleótidos 3-71
 Longitud: 23aa

GAEGEFTHRS QGAIRARDWP GYG

BlastP2 de: zyb9.3.pep desde: 1 a: 23
 comparado con la base de datos : uniprot ..

>>>sptrembl:Q9WAY7_9VIRU	Q9way7	SubName: Full=ORF2; Flags: Fragm...	52	2e-05
>>>sptrembl:Q98Y39_9VIRU	Q98y39	SubName: Full=ORF2;. 2/2010	52	2e-05
>>>sptrembl:Q9Wfy6_9VIRU	Q9wfy6	SubName: Full=ORF2;. 2/2010	52	3e-05
>>>sptrembl:Q9JGT0_9VIRU	Q9jgt0	SubName: Full=PORF2a;. 12/2009	52	3e-05
>>>sptrembl:Q786D4_9VIRU	Q786d4	SubName: Full=ORF2;. 2/2010	52	3e-05
>>>sptrembl:Q70738_9VIRU	Q70738	SubName: Full=ORF2, ORF1 genes;...	52	3e-05
>>>sptrembl:Q9JG33_9VIRU	Q9jg33	SubName: Full=Putative uncharact...	52	3e-05
>>>sptrembl:Q90363_9VIRU	Q90363	SubName: Full=ORF2 protein;. 2/2010	52	3e-05
>>>sptrembl:Q9YKL2_9VIRU	Q9ykl2	SubName: Full=ORF2 protein;. 2/2010	51	3e-05
>>>sptrembl:Q9JGS7_9VIRU	Q9jgs7	SubName: Full=PORF2a;. 12/2009	51	3e-05
>>>sptrembl:Q9DYC0_9VIRU	Q9dyc0	SubName: Full=Putative uncharact...	51	4e-05
>>>sptrembl:Q9W7s4_9VIRU	Q9w7s4	SubName: Full=Putative uncharact...	50	5e-05
>>>sptrembl:Q77s01_9VIRU	Q77s01	SubName: Full=Putative uncharact...	50	5e-05
>>>sptrembl:Q9JGS4_9VIRU	Q9jgs4	SubName: Full=PORF2a;. 12/2009	48	3e-04
>>>sptrembl:Q9JGT3_9VIRU	Q9jgt3	SubName: Full=PORF2a;. 12/2009	48	3e-04
>>>sptrembl:Q9YR02_9VIRU	Q9yr02	SubName: Full=Putative uncharact...	47	7e-04
>>>sptrembl:B2YFW4_9VIRU	B2yfw4	SubName: Full=ORF2; Flags: Fragm...	35	3.2

>>>>sptrembl:Q9WAY7_9VIRU Q9way7 SubName: Full=ORF2; Flags:
 Fragmento 2010
 Longitud = 138
 Puntuación = 52.0 bits (123), Esperado = 2e-05, Método: ajuste de matriz composicional.
 Identidades = 23/23 (100%), Positivos = 23/23 (100%)
 Query: 1 GAEGEFTHRSQGAIRARDWPGYG 23
 GAEGEFTHRSQGAIRARDWPGYG
 Sbjct: 15 GAEGEFTHRSQGAIRARDWPGYG 37

Q9WAY7_9VIRU Sin revisar; 138 AA.
 AC Q9WAY7;
 DT 01-NOV-1999, integrado en UniProtKB/TrEMBL.
 DT 01-NOV-1999, versión de secuencia 1.
 DT 09-FEB-2010, versión de entrada 23.
 DE SubName: Full=ORF2;
 DE Flags: Fragmento ;
 OS Torque teno virus.
 OC Virus; virus de ADNss ; Anelloviridae; Anelloviridae sin clasificar.
 OX NCBI_TaxID=68887;
 RN [1]
 RP SECUENCIA DE NUCLEÓTIDOS.
 RC TEJIDO=Suro;
 RX MEDLINE=99335592; PubMed=10405352; DOI=10.1006/viro.1999.9797;
 RA Hijikata M., Takahashi K., Mishiro S.;
 RT "Complete circular DNA genome of a TT virus variant (isolate name
 RT SANBAN) and 44 partial ORF2 sequences implicating a great degree of
 RT diversity beyond genotypes.";
 RL Virology 260:17-22(1999).
 CC -----
 CC Derechos de autor de UniProt Consortium, véase <http://www.uniprot.org/terms>
 CC Distribuido bajo la licencia de Creative Commons Attribution-NoDerivs
 CC -----

ES 2 685 605 T3

DR EMBL; [AB024348](#); [BAA77415.1](#); -; Genomic_DNA.
 DR InterPro; [IPR004118](#); Gyrovir_VP2/TT_ORF2.
 DR InterPro; [IPR013267](#); TTV_ORF2a.
 DR Pfam; [PF02957](#); TT_ORF2; 1.
 DR Pfam; [PF08197](#); TT_ORF2a; 1.
 PE 4: Previsto;
 FT NON_TER 1 1
 FT NON_TER 138 138
 SQ SECUENCIA 138 AA; 15416 MW; 93F7D9685085141D CRC64;
 SGEATEGDLR VPRAGAEGEF THRSQGAIRA RDWPGYGQGS EKSMFIGRHY RKKRALSLCA
 VRTTKKACKL LIVMWTPPRN DQQYLNQWY SSVLSSHAAM CGCPDAVAHF NHLAAVLRAP
 QNPPPPGPQR NPLRRLP

zkb5.3.pep

nucleótidos 3-71
 Longitud: 23aa

GAVGEFTHRS QCAIRARDWP GYG

BlastP2 de: zkb5.3.pep

comparado con la base de datos: **uniprot** ..

>>>sptrembl:Q98Y39 9VIRU	Q98y39	SubName: Full=ORF2;. 2/2010	<u>49</u>	1e-04
>>>sptrembl:Q9WAY7 9VIRU	Q9way7	SubName: Full=ORF2; Flags: Fragm...	<u>49</u>	1e-04
>>>sptrembl:Q9JGT0 9VIRU	Q9jgt0	SubName: Full=PORF2a;. 12/2009	<u>49</u>	1e-04
>>>sptrembl:Q90363 9VIRU	Q90363	SubName: Full=ORF2 protein;. 2/2010	<u>49</u>	1e-04
>>>sptrembl:Q9JG33 9VIRU	Q9jg33	SubName: Full=Putative uncharact...	<u>49</u>	1e-04
>>>sptrembl:Q786D4 9VIRU	Q786d4	SubName: Full=ORF2;. 2/2010	<u>49</u>	1e-04
>>>sptrembl:Q70738 9VIRU	Q70738	SubName: Full=ORF2, ORF1 genes;...	<u>49</u>	1e-04
>>>sptrembl:Q9YKL2 9VIRU	Q9ykl2	SubName: Full=ORF2 protein;. 2/2010	<u>49</u>	1e-04
>>>sptrembl:Q9Wfy6 9VIRU	Q9wfy6	SubName: Full=ORF2;. 2/2010	<u>49</u>	1e-04
>>>sptrembl:Q9JGS7 9VIRU	Q9jgs7	SubName: Full=PORF2a;. 12/2009	<u>49</u>	2e-04
>>>sptrembl:Q9DYC0 9VIRU	Q9dyc0	SubName: Full=Putative uncharact...	<u>49</u>	2e-04
>>>sptrembl:Q9W7S4 9VIRU	Q9w7s4	SubName: Full=Putative uncharact...	<u>49</u>	2e-04
>>>sptrembl:Q77S01 9VIRU	Q77s01	SubName: Full=Putative uncharact...	<u>49</u>	2e-04
>>>sptrembl:Q9JGS4 9VIRU	Q9jgs4	SubName: Full=PORF2a;. 12/2009	<u>48</u>	3e-04
>>>sptrembl:Q9JGT3 9VIRU	Q9jgt3	SubName: Full=PORF2a;. 12/2009	<u>48</u>	3e-04
>>>sptrembl:Q9YR02 9VIRU	Q9yr02	SubName: Full=Putative uncharact...	<u>47</u>	7e-04

>>>>sptrembl:Q98Y39 9VIRU Q98y39 SubName: Full=ORF2;. 2/2010
 Longitud = 202
 Puntuación = 49.3 bits (116), Esperado = 1e-04, Método: ajuste de matriz composicional.
 Identidades = 22/23 (95%), Positivos = 22/23 (95%)
 Query: 1 GAVGEFTHRSQCAIRARDWPGYG 23
 GA GEFTHRSQCAIRARDWPGYG
 Sbjct: 24 GAEGEFTHRSQCAIRARDWPGYG 46

Q98Y39 9VIRU Sin revisar ; 202 AA.
 AC Q98Y39;
 DT 01-JUN-2001, integrado en UniProtKB/TrEMBL.
 DT 01-JUN-2001, versión de secuencia 1.
 DT 09-FEB-2010, versión de entrada 17.
 DE SubName: Full=ORF2;
 OS Torque teno virus.
 OC Virus; virus de ADNss ; Anelloviridae; Anelloviridae sin clasificar.
 OX NCBI_TaxID=68887;
 RN [1]
 RP SECUENCIA DE NUCLEÓTIDOS.
 RC CEPA =TWH;

ES 2 685 605 T3

```

RA He H.-T., Luo K.-X., Xiao H., Liu D.-X.;
RT "Complete circular genome of TT virus isolated from feces of a
RT hepatitis patient.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
-----
CC Derechos de autor de UniProt Consortium, véase http://www.uniprot.org/terms
CC Distribuido bajo la licencia de Creative Commons Attribution-NoDerivs
-----
DR EMBL; AF351132; AAK29446.1; -; Genomic DNA.
DR InterPro; IPR004118; Gyrovir_VP2/TT_ORF2.
DR InterPro; IPR013267; TTV_ORF2a.
DR Pfam; PF02957; TT_ORF2; 1.
DR Pfam; PF08197; TT_ORF2a; 1.
PE 4: Previsto ;
SQ SECUENCIA 202 AA; 21437 MW; 105B9ED104956EDE CRC64;
MAEFSTPVRS GEATEGDHRV PRAGAEGEFT HRSQGAIARR DWPGYGQGS KSMFIGRHYR
KKRALSILCAV RTTKKACKLL IVMWTPPRND QQYLNWQWYS SVLSSHASMC GCPDAVAHLI
NLASVLRAPQ NPPPPGPPQRN LPLRRLPALP AAPEAPGDRA PWPMAAGGAE GENGAGGDAD
HGGAAGGPED ANLLDAVAAA ET
//

```

zkb69.1.pep

nucleótidos 1-71
Longitud: 23aa

RVPEVSLHTA VKGQFGLGTG RAM

```

BlastP2 de: zkb69.1.pep desde: 1 a: 23
comparado con la base de datos: uniprot
>>>sptrembl:Q9WAZ2_9VIRU Q9waz2 SubName: Full=ORF2; Flags: Fragm... 49 1e-04
>>>sptrembl:Q9WB10_9VIRU Q9wb10 SubName: Full=ORF2; Flags: Fragm... 49 1e-04
>>>sptrembl:O70807_9VIRU O70807 SubName: Full=Putative uncharacter... 49 2e-04
>>>sptrembl:Q9WAY4_9VIRU Q9way4 SubName: Full=ORF2; Flags: Fragm... 49 2e-04
>>>sptrembl:Q9WSW0_9VIRU Q9wsw0 SubName: Full=ORF2; Flags: Fragm... 48 3e-04
>>>sptrembl:Q9WB12_9VIRU Q9wb12 SubName: Full=ORF2; Flags: Fragm... 48 4e-04
>>>sptrembl:Q9WSW2_9VIRU Q9wsw2 SubName: Full=ORF2; Flags: Fragm... 48 4e-04
>>>sptrembl:Q9WSW4_9VIRU Q9wsw4 SubName: Full=ORF2; Flags: Fragm... 47 7e-04
>>>sptrembl:Q9WB09_9VIRU Q9wb09 SubName: Full=ORF2; Flags: Fragm... 47 9e-04
>>>sptrembl:Q9WSX0_9VIRU Q9wsx0 SubName: Full=ORF2; Flags: Fragm... 46 0.001
>>>sptrembl:Q9WB02_9VIRU Q9wb02 SubName: Full=ORF2; Flags: Fragm... 45 0.002
>>>sptrembl:Q9WSW6_9VIRU Q9wsw6 SubName: Full=ORF2; Flags: Fragm... 45 0.002
>>>sptrembl:B3FWR6_9VIRU B3fwr6 SubName: Full=ORF2; Flags: Fragm... 38 0.34

>>>sptrembl:Q9WAZ2_9VIRU Q9waz2 SubName: Full=ORF2; Flags:
Fragmento ; - 2/2010
Longitud = 152
Puntuación = 49.3 bits (116), Esperado = 1e-04, Método: ajuste de matriz composicional.
Identidades = 23/23 (100%), Positivos = 23/23 (100%)
Query: 1 RVPEVSLHTAVKQFGLGTGRAM 23
RVPEVSLHTAVKQFGLGTGRAM
Sbjct: 14 RVPEVSLHTAVKQFGLGTGRAM 36

Q9WAZ2_9VIRU Sin revisar ; 152 AA.
AC Q9WAZ2;
DT 01-NOV-1999, integrado en UniProtKB/TrEMBL.
DT 01-NOV-1999, versión de secuencia 1.
DT 09-FEB-2010, versión de entrada 23.
DE SubName: Full=ORF2;

```

ES 2 685 605 T3

```

DE   Flags: Fragmento;
OS   Torque teno virus.
OC   Virus;   virus de ADNss   ; Anelloviridae; Anelloviridae sin clasificar.
OQ   NCBI_TaxID=68887;
RN   [1]
RP   SECUENCIA DE NUCLEÓTIDOS.
RC   TEJIDO=Suero;
RX   MEDLINE=99335592; PubMed=10405352; DOI=10.1006/viro.1999.9797;
RA   Hijikata M., Takahashi K., Mishiro S.;
RT   "Complete circular DNA genome of a TT virus variant (isolate name
RT   SANBAN) and 44 partial ORF2 sequences implicating a great degree of
RT   diversity beyond genotypes.";
RL   Virology 260:17-22(1999).
CC   -----
CC   Derechos de autor de UniProt Consortium, véase http://www.uniprot.org/terms
CC   Distribuido bajo la licencia de Creative Commons Attribution-NoDerivs
CC   -----
DR   EMBL; AB024353; BAA77420.1; -; Genomic DNA.
DR   InterPro; IPR004118; Gyrovir_VP2/TT_ORF2.
DR   Pfam; PF02957; TT_ORF2; 1.
PE   4:   Previsto ;
FT   NON_TER   1       1
FT   NON_TER   152    152
SQ   SECUENCIA 152 AA; 16658 MW; 1DEC53175C043A17 CRC64;
    ARTPRRGVRA SRGRVPEVSL HTAVKQGQFGL GTGRAMGKAL KKAMFLGRIY RKKRRLPLSP
    LHSPPKARKL LRGMWRPPTQ NVSGQERSWY DSVFYSHAAF CGCGDCVGHSL SYLATHLGRP
    PSAQPPPQLQ PPVIRRLPAL PAPPNPSGDR AA
  
```

zkb69.3.pep

nucleótidos 3-71
Longitud: 23

GAGGEFTHRS QGAIRARDWP GYG

```

BlastP2 de: zkb69.3.pep desde: 1 a: 23
comparado con la base de datos: uniprot ..
>>>sptrembl:Q9JGT3_9VIRU Q9jgt3 SubName: Full=PORF2a;. 12/2009 50 5e-05
>>>sptrembl:Q9JGS4_9VIRU Q9jgs4 SubName: Full=PORF2a;. 12/2009 50 5e-05
>>>sptrembl:Q98Y39_9VIRU Q98y39 SubName: Full=ORF2;. 2/2010 49 1e-04
>>>sptrembl:Q9WAY7_9VIRU Q9way7 SubName: Full=ORF2; Flags: Fragm... 49 1e-04
>>>sptrembl:Q9YR02_9VIRU Q9yr02 SubName: Full=Putative uncharact... 49 1e-04
>>>sptrembl:Q786D4_9VIRU Q786d4 SubName: Full=ORF2;. 2/2010 49 1e-04
>>>sptrembl:O70738_9VIRU O70738 SubName: Full=ORF2, ORF1 genes;... 49 1e-04
>>>sptrembl:Q9JG33_9VIRU Q9jg33 SubName: Full=Putative uncharact... 49 2e-04
>>>sptrembl:Q90363_9VIRU Q90363 SubName: Full=ORF2 protein;. 2/2010 49 2e-04
>>>sptrembl:Q9JGT0_9VIRU Q9jgt0 SubName: Full=PORF2a;. 12/2009 49 2e-04
>>>sptrembl:Q9YKL2_9VIRU Q9ykl2 SubName: Full=ORF2 protein;. 2/2010 49 2e-04
>>>sptrembl:Q9WFY6_9VIRU Q9wfy6 SubName: Full=ORF2;. 2/2010 49 2e-04
>>>sptrembl:Q9JGS7_9VIRU Q9jgs7 SubName: Full=PORF2a;. 12/2009 48 3e-04
>>>sptrembl:Q9DYC0_9VIRU Q9dyc0 SubName: Full=Putative uncharact... 48 3e-04
>>>sptrembl:Q9W7S4_9VIRU Q9w7s4 SubName: Full=Putative uncharact... 48 3e-04
>>>sptrembl:Q77S01_9VIRU Q77s01 SubName: Full=Putative uncharact... 48 3e-04
>>>sptrembl:B3FWR5_9VIRU B3fwr5 SubName: Full=ORF2; Flags: Fragm... 34 5.2
>>>sptrembl:B6SFP1_9VIRU B6sfp1 SubName: Full=ORF2; Flags: Fragm... 34 5.6

>>>>sptrembl:Q9JGT3_9VIRU Q9jgt3 SubName: Full=PORF2a;. 12/2009
Longitud = 49
Puntuación = 50.4 bits (119), Esperado = 5e-05, Método: ajuste de matriz composicional.
Identidades = 23/23 (100%), Positivos = 23/23 (100%)
Query: 1 GAGGEFTHRSQGAIRARDWPGYG 23
GAGGEFTHRSQGAIRARDWPGYG
  
```

ES 2 685 605 T3

Sbjct: 24 GAGGEFTHRSQGAIARDWPGYG 46

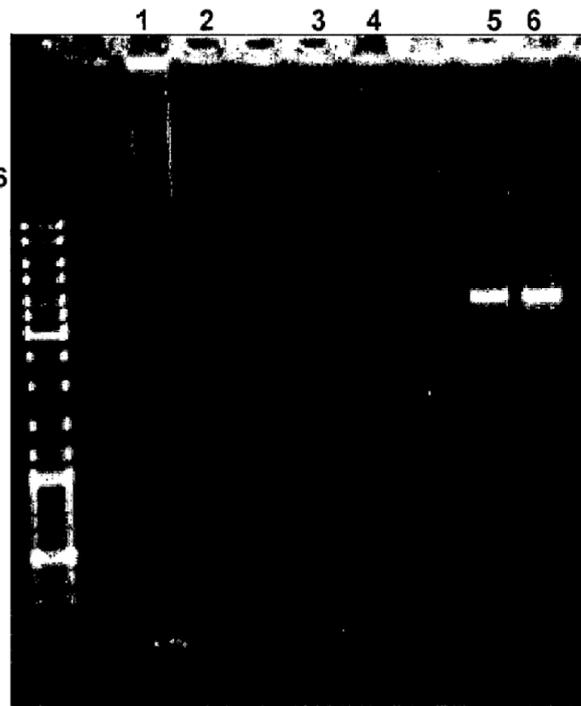
Q9JGT3_9VIRU Sin revisar ; 49 AA.
AC Q9JGT3;
DT 01-OCT-2000, integrado en UniProtKB/TrEMBL.
DT 01-OCT-2000, versión de secuencia 1.
DT 15-DEC-2009, versión de entrada 17.
DE SubName: Full=PORF2a;
GN Name=ORF2a;
OS Torque teno virus.
OC Virus; virus de ADNss ; Anelloviridae; Anelloviridae sin clasificar.
OX NCBI_TaxID=68887;
RN [1]
RP SECUENCIA DE NUCLEÓTIDOS.
RX MEDLINE=20417334; PubMed=10963344; DOI=10.1007/s007050070097;
RA Tanaka Y., Orito E., Ohno T., Nakano T., Hayashi K., Kato T.,
RA Mukaide M., Iida S., Mizokami M.;
RT "Identification of a novel 23kDa protein encoded by putative open
RT reading frame 2 of TT virus (TTV) genotype 1 different from the other
RT genotypes.";
RL Arch. Virol. 145:1385-1398(2000).

CC Derechos de autor de UniProt Consortium, véase <http://www.uniprot.org/terms>
CC Distribuido bajo la licencia de Creative Commons Attribution-NoDerivs

DR EMBL; AB030486; BAA90401.1; -; Genomic_DNA.
DR InterPro; IPR013267; TTV_ORF2a.
DR Pfam; PF08197; TT_ORF2a; 1.
PE 4: ; Previsto ;
SQ SECUENCIA 49 AA; 5118 MW; 596E44680A5D863A CRC64;
MAEFSTPVRS EGATEGIPNV PRAGAGGEFT HRSQGAIARDWPGYGQGS

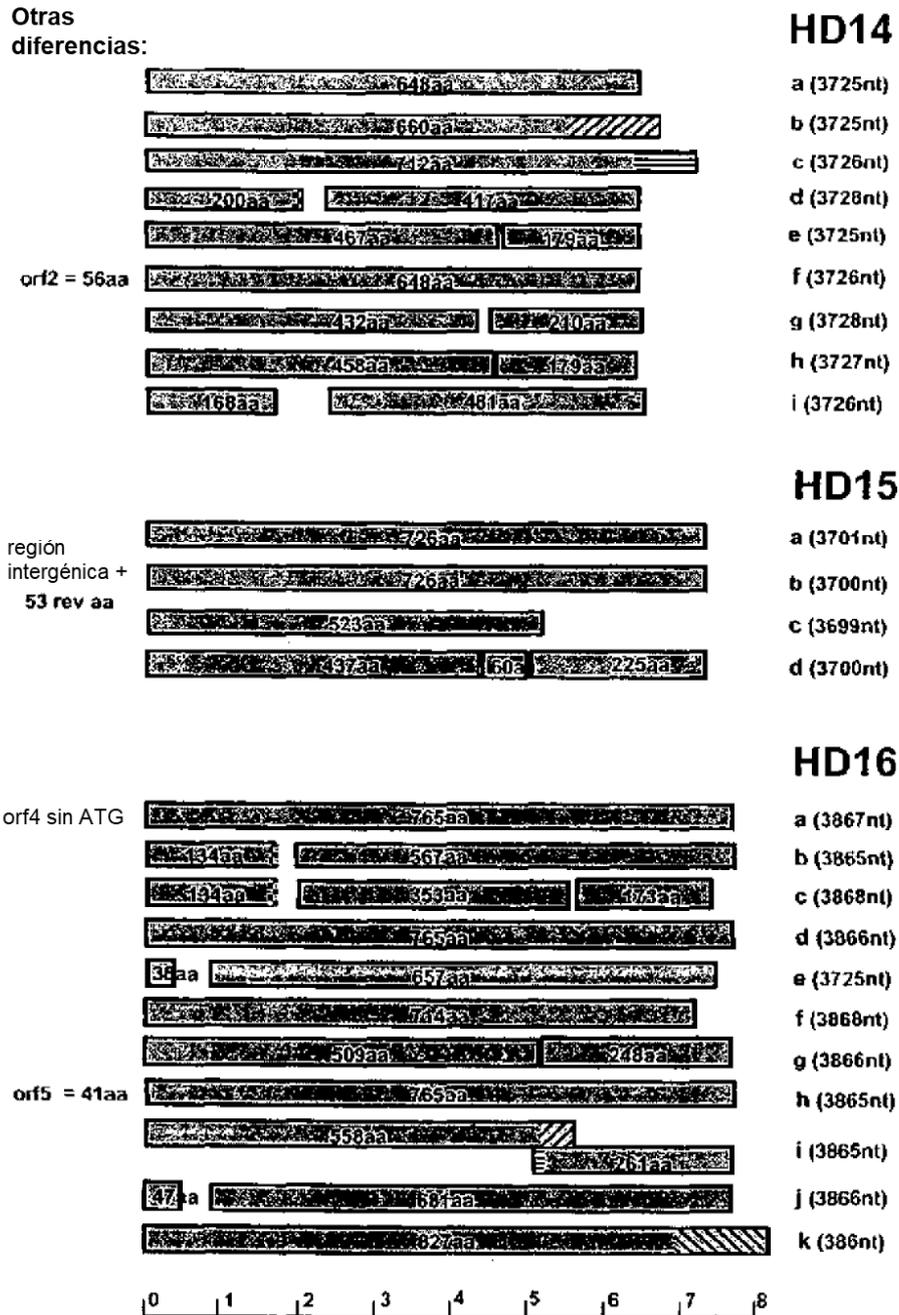
Figura 15

El ADN celular de L1236 se sometió a amplificación con templiPhi, seguida de de 2 rondas sucesivas de PCR con los cebadores jt34f 7 y 8



Sin ADN
ADN de L1236
sin digest. con MBN
sin digest. con MBN
tratado con MBN
pCR2.1 lineal
negativo para MBN
pCR2.1 lineal
digerido con MBN

Figura 16





HD20



HD23

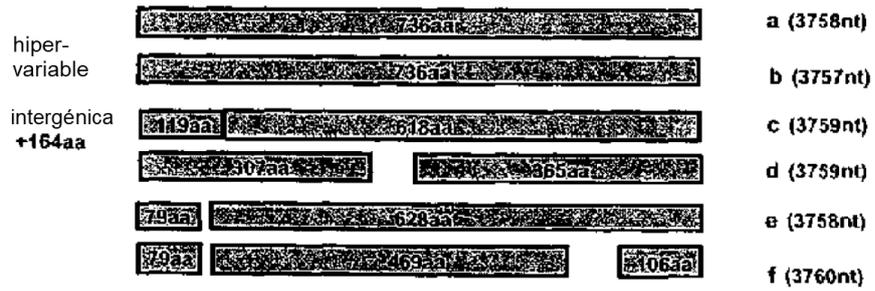
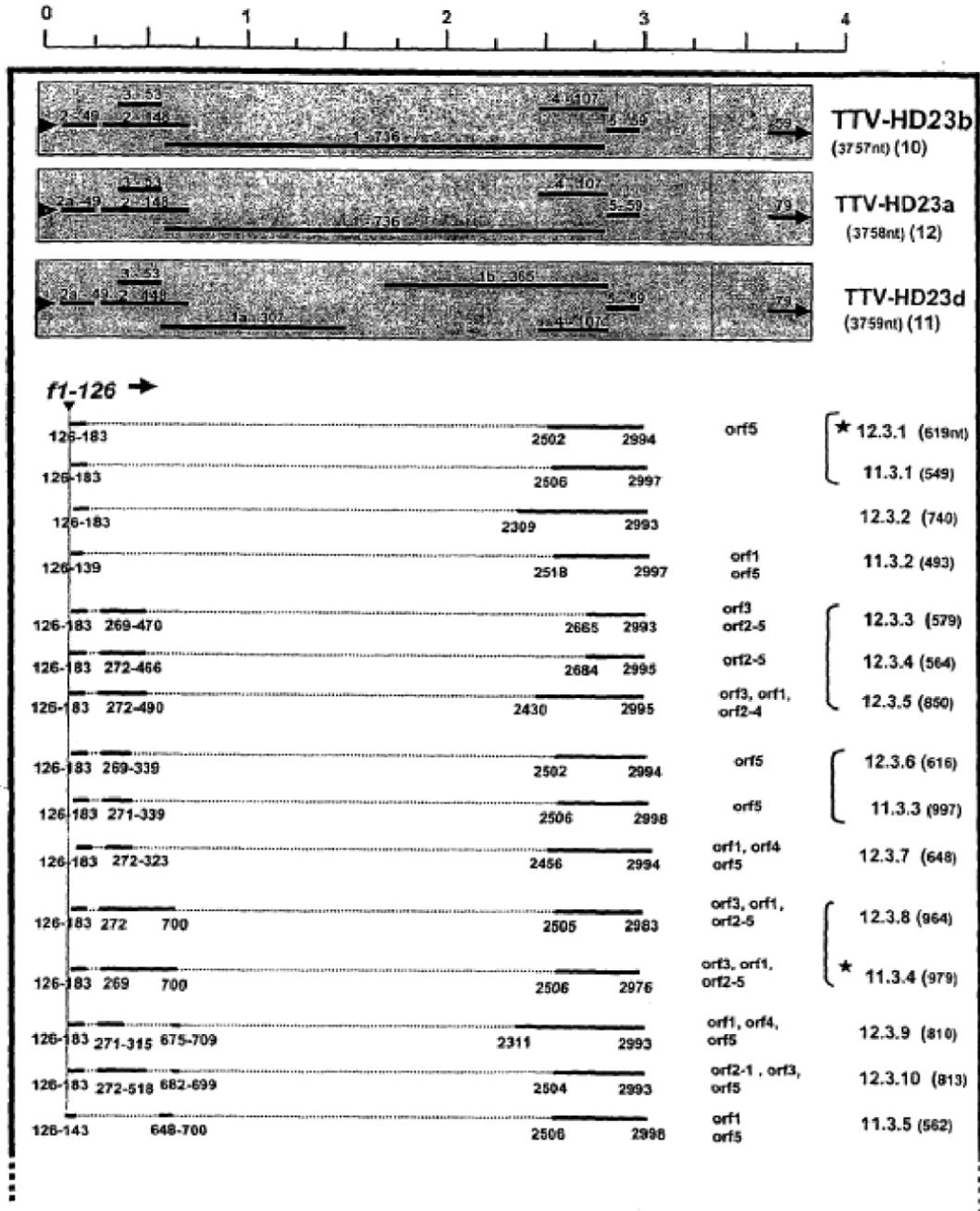
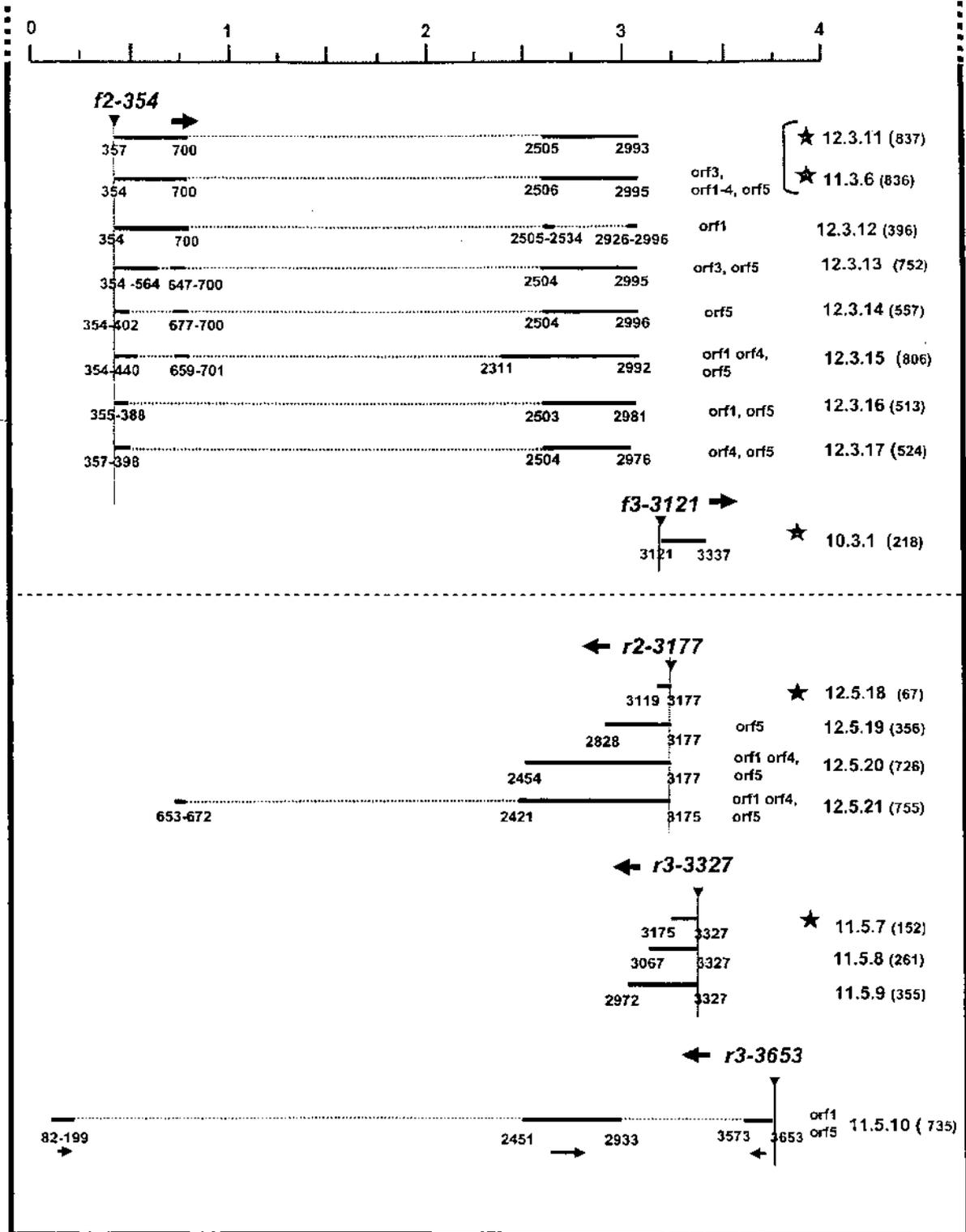
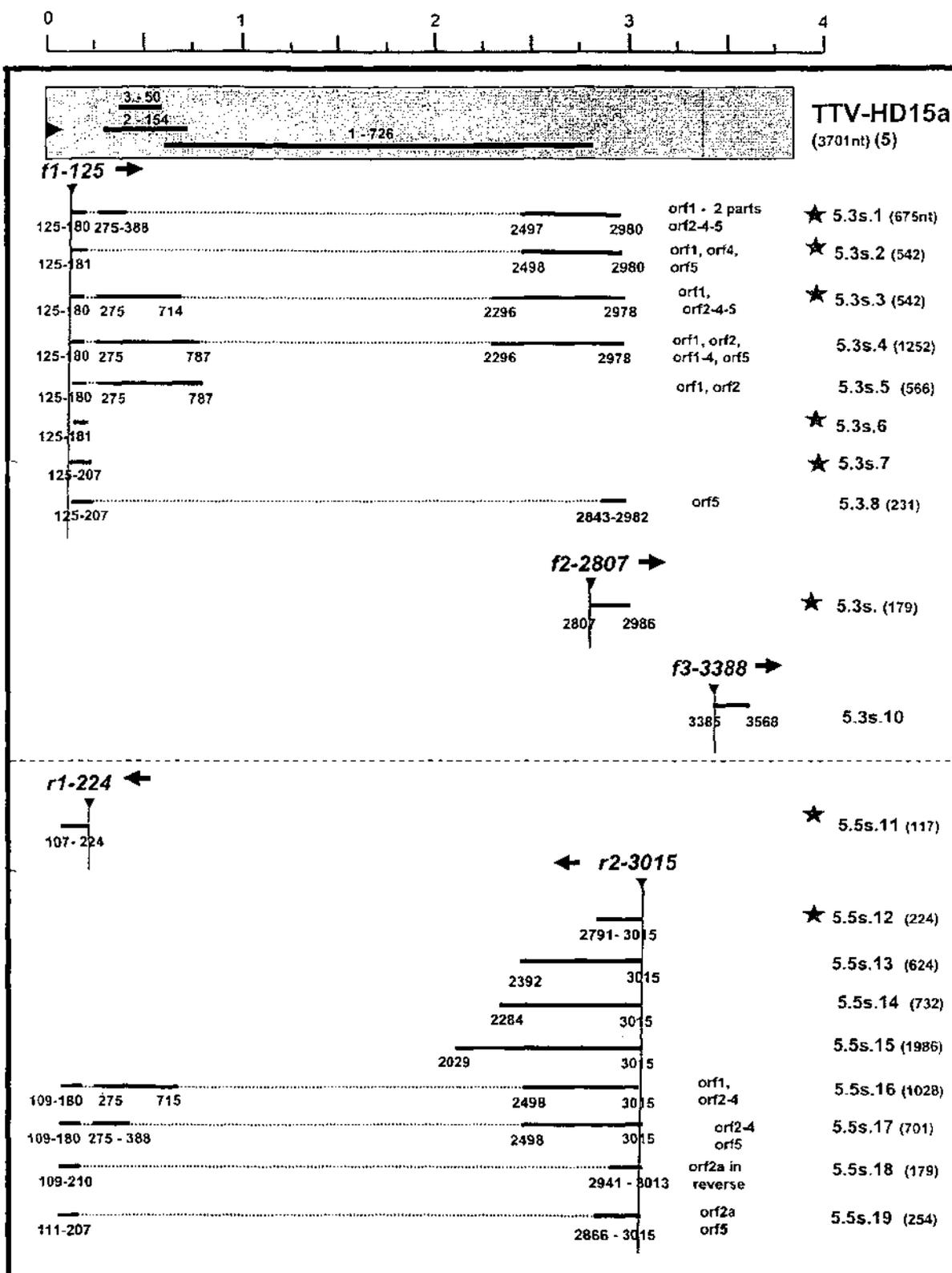
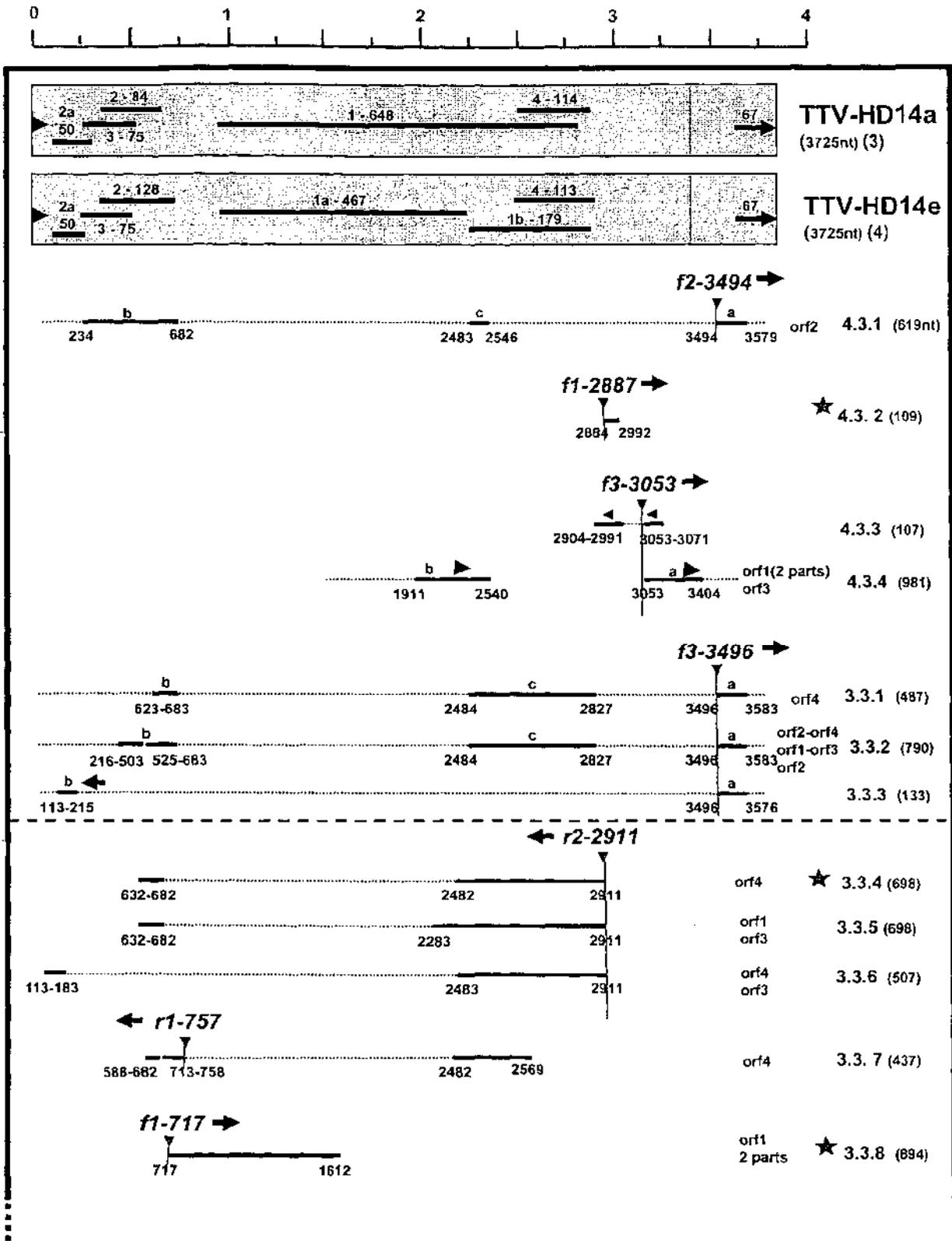


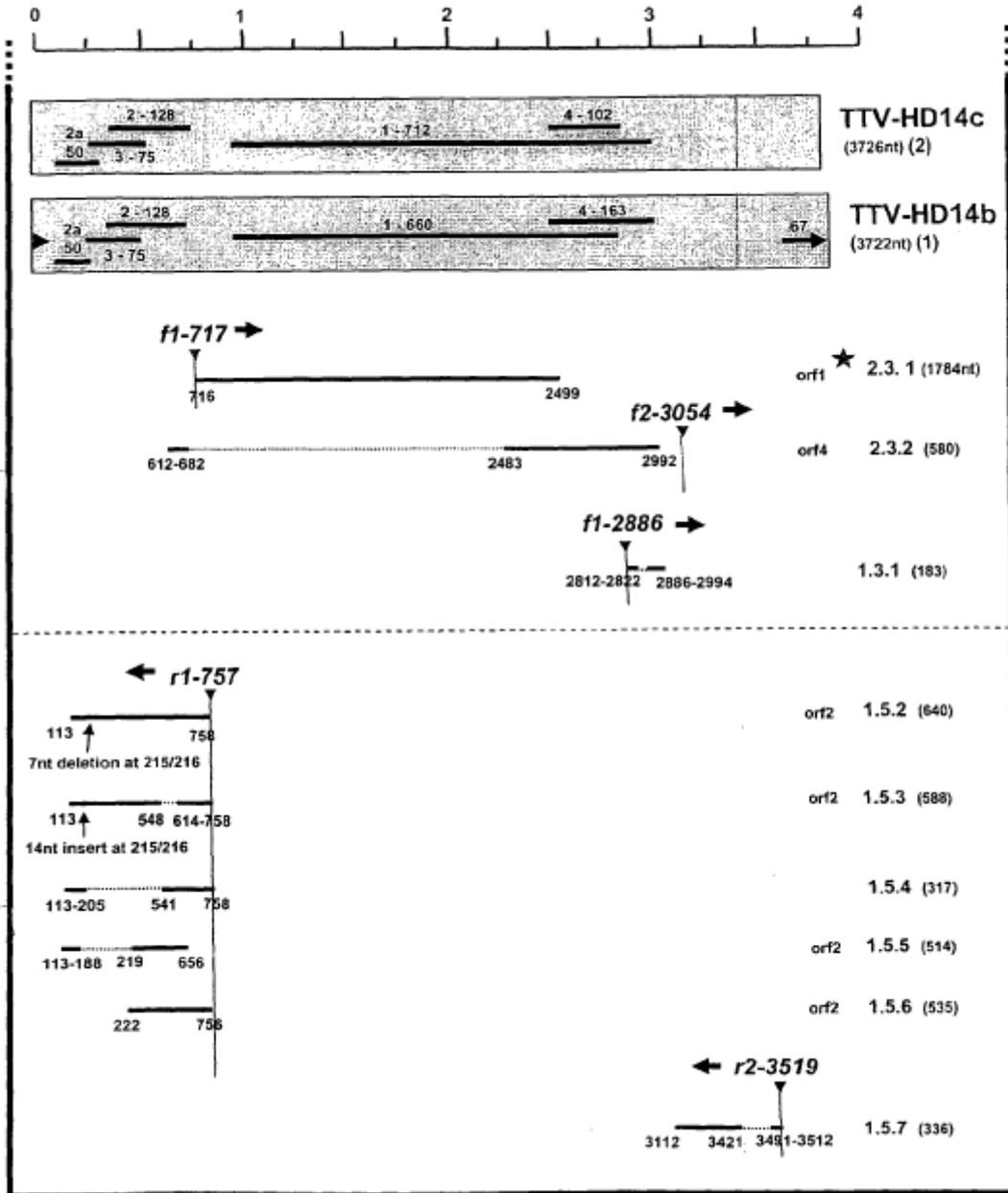
Figura 17

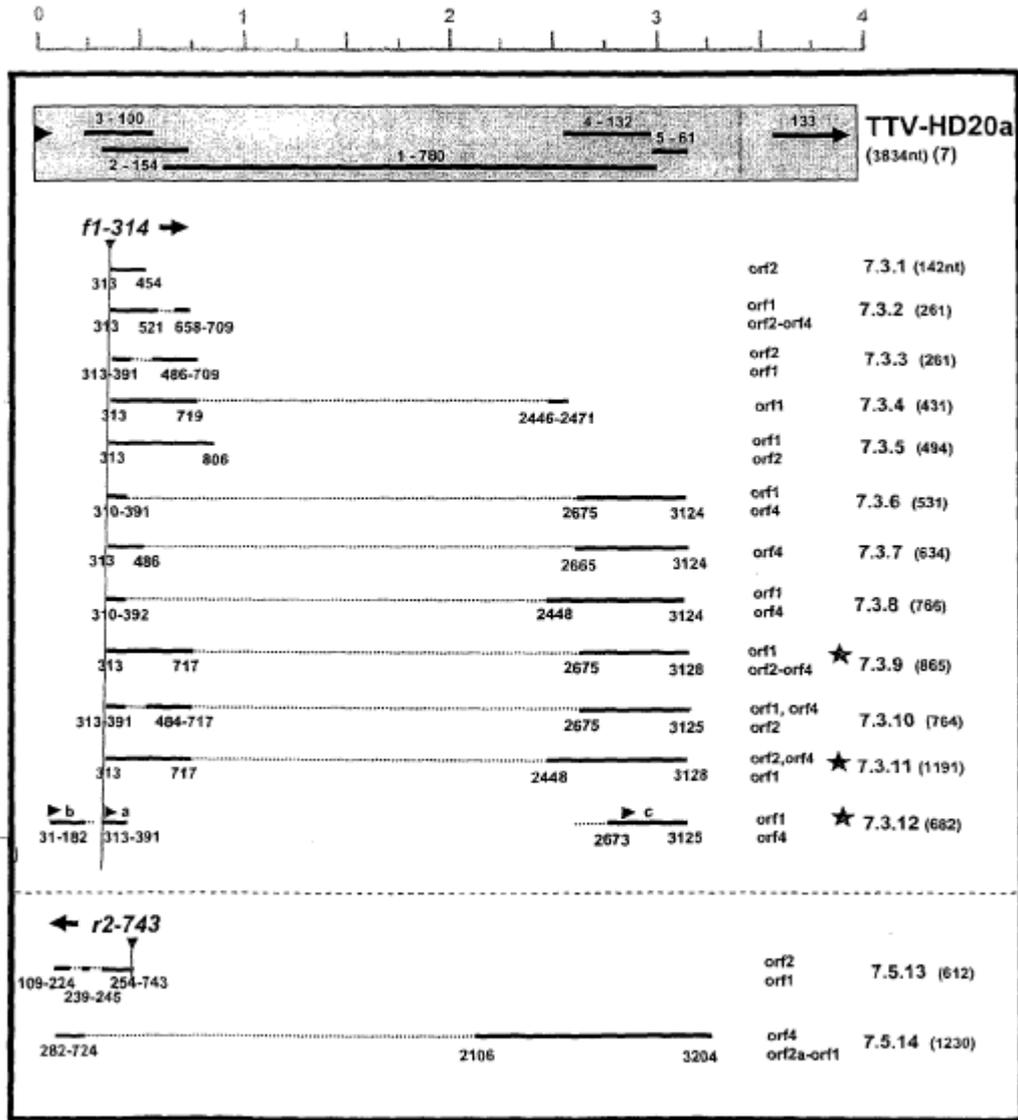












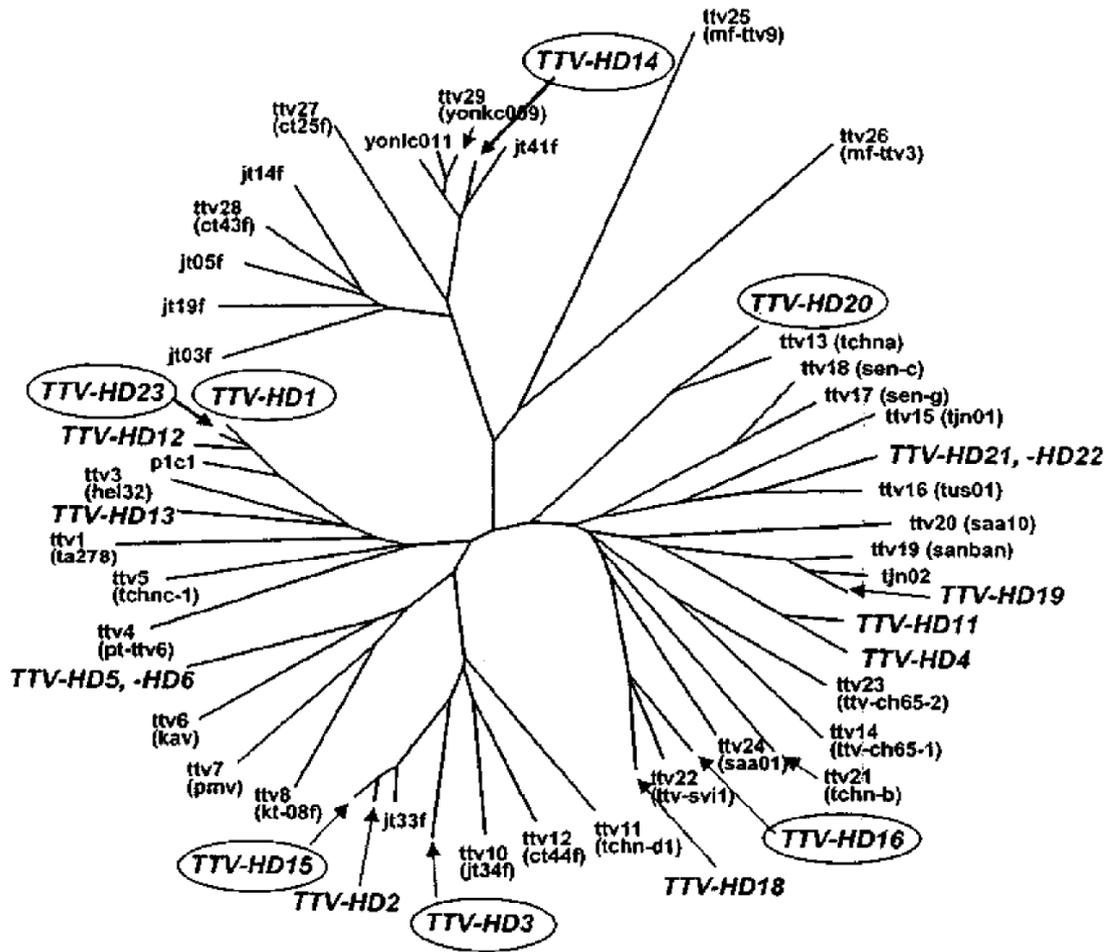
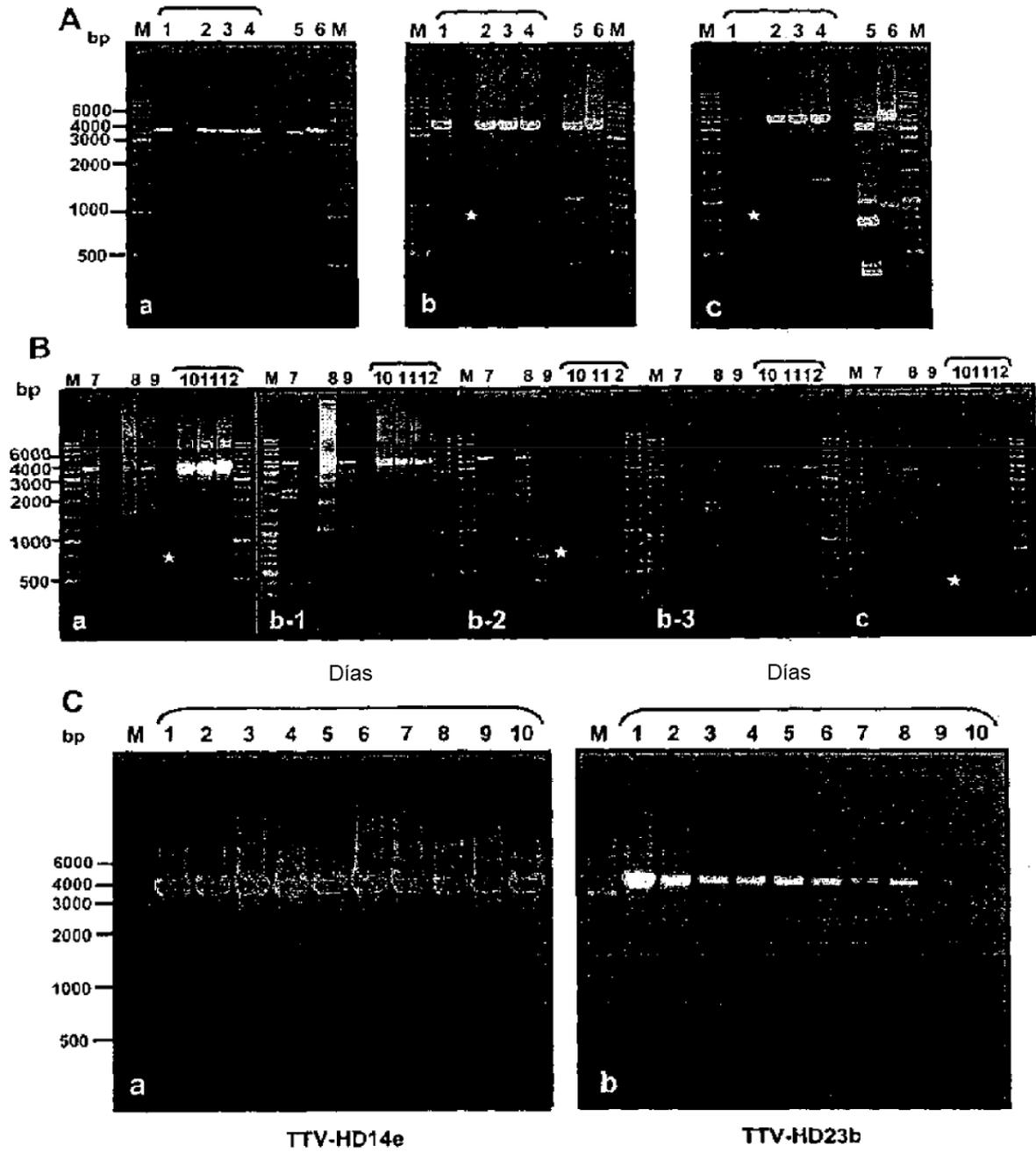


Figura 18

Figura 19



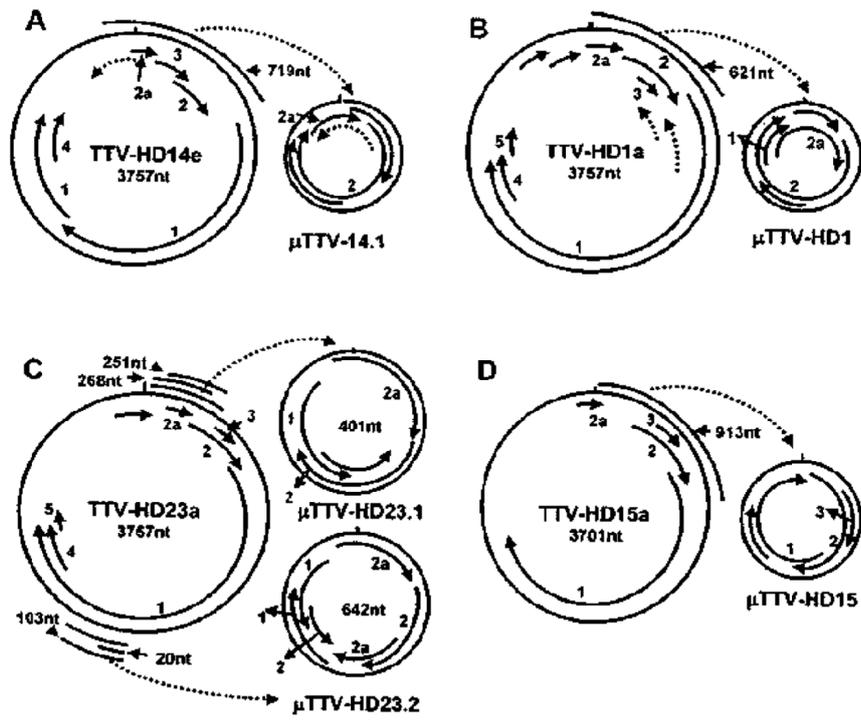


Figura 20

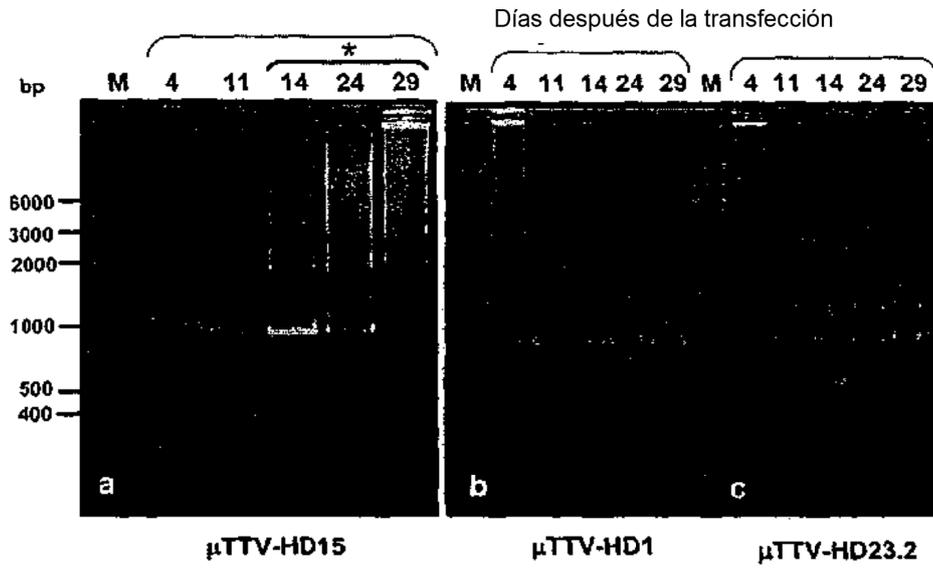


Figura 21

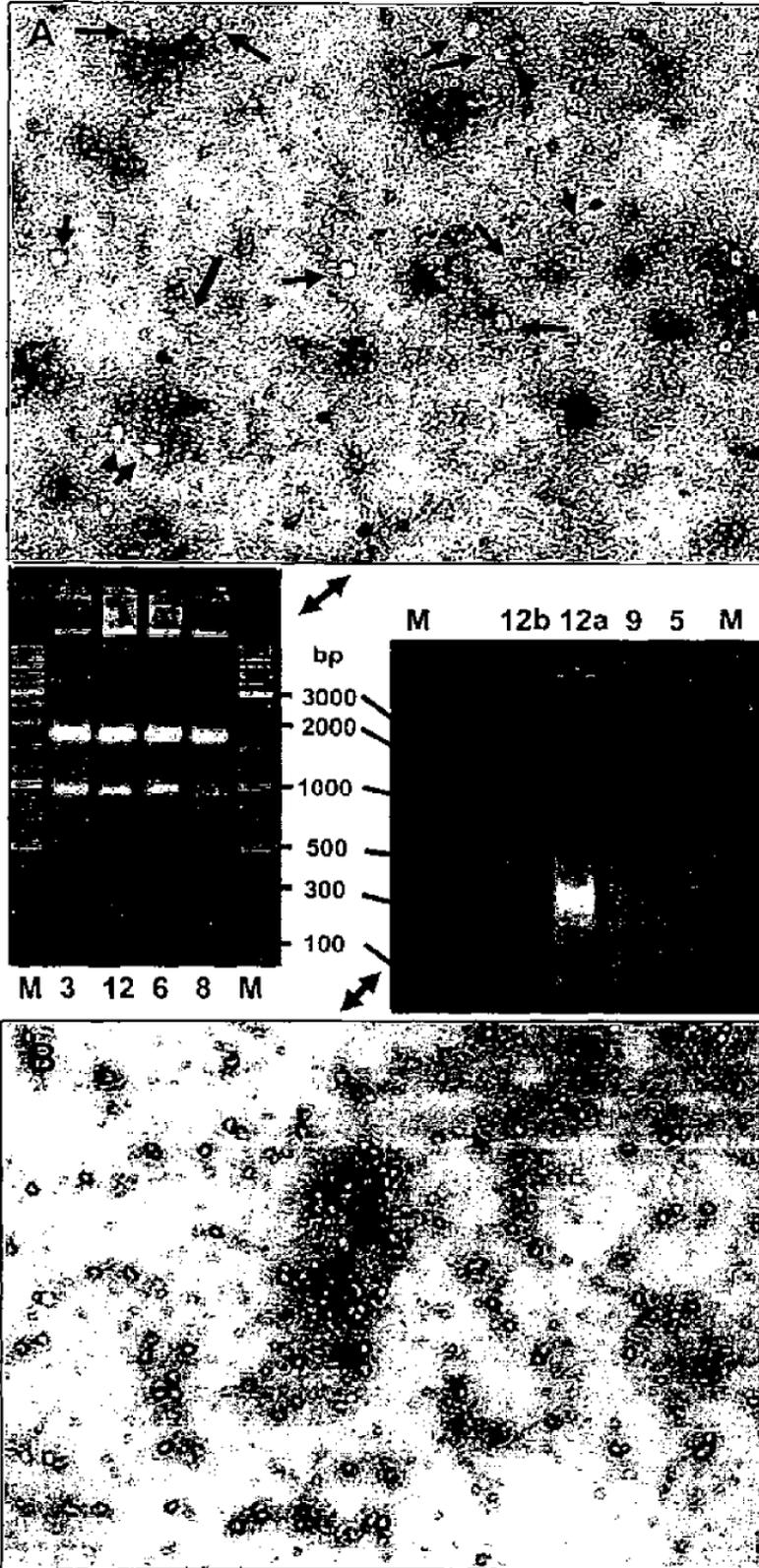


Figura 22