

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 382 827**

51 Int. Cl.:  
**C12Q 1/68** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **07756988 .7**

96 Fecha de presentación: **14.02.2007**

97 Número de publicación de la solicitud: **2121969**

97 Fecha de publicación de la solicitud: **25.11.2009**

54 Título: **Procedimiento para la detección de la miopatía equina por almacenamiento de polisacáridos**

45 Fecha de publicación de la mención BOPI:  
**13.06.2012**

45 Fecha de la publicación del folleto de la patente:  
**13.06.2012**

73 Titular/es:  
**LABOKLIN GMBH & CO. KG  
STEUBENSTRASSE 4  
97688 BAD KISSINGEN, DE**

72 Inventor/es:  
**VALBERG, Stephanie J.;  
MCCUE, Molly E. y  
MICKELSON, James R.**

74 Agente/Representante:  
**Carpintero López, Mario**

ES 2 382 827 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

Procedimiento para la detección de la miopatía equina por almacenamiento de polisacáridos

**Antecedentes de la invención**

5 La miopatía por almacenamiento de polisacáridos (MAPS) es una enfermedad muscular debilitante en muchas y diferentes razas de caballos. Los datos disponibles indican que aproximadamente el 10% de los caballos de raza Quarter Horses y el 36% de los caballos de raza Belgian Draft Horses están afectados. Los signos clínicos varían aunque pueden ir desde atrofia muscular y debilidad progresiva en las razas Draft Horse hasta espasmos musculares agudos tras el ejercicio y daño celular en los Quarter Horses y en otras razas. Todas las formas de MAPS en caballos están fundamentalmente asociadas a depósitos de un polisacárido anormal en fibras musculares esqueléticas que se aprecian por tinción histoquímica de biopsias musculares. La MAPS también se caracteriza por unos niveles de hasta cuatro veces superiores a los niveles normales de glucógeno en el músculo esquelético. Se sabe que las mutaciones en los genes del metabolismo de la glucosa y del glucógeno causan diversos tipos de enfermedades por almacenamiento de glucógeno (glucogénesis) en seres humanos y en especies animales de las cuales varias son similares histológicamente a la MAPS. Sin embargo, ninguno de estos genes parecen ser responsables de la MAPS.

El diagnóstico actual de la MAPS en caballos se basa en los signos clínicos de espasmos musculares o atrofia progresiva (dependiendo de la raza), frecuentemente con niveles séricos elevados enzimas musculares, combinado con hallazgos histopatológicos de polisacárido anormal en estos cortes de biopsias de músculo esquelético.

20 Las biopsias musculares son invasivas, precisan de personal veterinario experto para su realización, son relativamente caras para el propietario y requieren de un histopatólogo muscular especializado para su interpretación. Además, aunque el análisis de la biopsia muscular ha sido una herramienta diagnóstica extremadamente fiable, no es 100% específica ni sensible y nunca lo podrá ser.

Por consiguiente, a pesar de lo anterior, existe una necesidad en la técnica de pruebas diagnósticas adicionales para el diagnóstico de la MAPS en caballos.

**Sumario de la invención**

30 La presente invención proporciona un procedimiento para la detección de la presencia de un biomarcador asociado a la miopatía por almacenamiento de polisacáridos (MAPS). En una realización de la invención, el procedimiento implica la obtención de una muestra fisiológica de un caballo, en el que la muestra comprende ácido nucleico y la determinación de la presencia del biomarcador. Como se usa en la presente memoria, el término "muestra fisiológica" se entiende que se refiere a una muestra biológica obtenida de un mamífero que contiene ácido nucleico. Por ejemplo, una muestra fisiológica puede ser una muestra recogida de un caballo individual, tal como incluyendo, pero sin limitación, por ej., una muestra celular, tal como una célula sanguínea, por ej., un linfocito, una célula de sangre periférica, una muestra recogida de la médula espinal, una muestra de tejido, tal como tejido cardíaco o tejido muscular, por ej., músculo cardíaco o esquelético, una muestra de un órgano, por ej., hígado o piel, una muestra de pelo, por ej., una muestra de pelo con raíz y/o una muestra de líquido, tal como sangre.

40 Ejemplos de razas de caballos afectadas incluyen, pero sin limitación, Quarter Horses, Percheron Horses, Paint Horses, Draft Horses, Warmblood Horses u otras razas relacionadas o no relacionadas. El término "raza relacionada" se usa en la presente memoria para referirse a razas que están relacionadas con una raza, tal como Quarter Horse, Draft Horse o Warmblood Horse. Dichas razas incluyen, pero sin limitación, razas de cría, tales como el American Paint, el Appaloosa y el Palomino. El término "Draft Horse" incluye muchas razas incluyendo, pero sin limitación, los caballos Clydesdale, Belgian, Percheron y Shire. El término "media sangre" es también un término genérico que incluye varias razas diferentes. "Media sangre" simplemente distingue este tipo de caballo de los caballos de "sangre fría" (caballos de tiro) y de los de "sangre caliente" (los pura sangre y los árabes). El procedimiento de la presente invención también incluye caballos de razas cruzadas o mixtas.

45 El término "biomarcador" se define generalmente como un indicador biológico, tal como una característica molecular particular, que puede afectar o estar relacionada con el diagnóstico o predicción de la salud de un individuo. Por ejemplo, en determinadas realizaciones de la presente invención, el biomarcador comprende un gen mutante de la enzima glucógeno sintasa 1 (GYS1) equina, tal como el alelo polimórfico de GYS1 que tiene una sustitución de G por A en el nucleótido 926 en el exón 6. El gen GYS1 codifica para una enzima que tiene una sustitución R (arginina) por H (histidina) en el residuo aminoácido 309.

"Sonda oligonucleotídica" se puede referir a un segmento de ácido nucleico, tal como un cebador, que es útil para amplificar una secuencia del gen GYS1 que es complementaria a, e hibrida específicamente con una secuencia particular de GYS1, o con una región de ácido nucleico que flanquea a GYS1.

55 Como se usa en la presente memoria, el término "ácido nucleico" y "polinucleótido" se refiere a desoxirribonucleótidos o ribonucleótidos y polímeros de los mismos en forma monocatenaria o bicatenaria, compuesto por monómeros (nucleótidos) que contienen un azúcar, fosfato, y una base que es una purina o

5 pirimidina. Salvo que se especifique concretamente, el término abarca los ácidos nucleicos que contienen análogos conocidos de nucleótidos naturales que tienen propiedades similares a las del ácido nucleico de referencia y que se metabolizan de manera similar a los nucleótidos naturales. Salvo que se indique otra cosa, una secuencia de ácido nucleico particular también abarca implícitamente variantes modificadas conservadoramente de la misma (por ej., sustitución en un codón degenerado) y secuencias complementarias, así como la secuencia indicada explícitamente. Específicamente, las sustituciones en un codón degenerado pueden conseguirse generando secuencias en las cuales la tercera posición de uno o más codones seleccionado (o todos) está sustituida con residuos de una base mixta y/o desoxiinosina.

10 Un "fragmento de ácido nucleico" es una fracción de una molécula de ácido nucleico dada. El ácido desoxirribonucleico (ADN) en la mayoría de los organismos es el material genético, mientras que el ácido ribonucleico (ARN) interviene en la transferencia de la información contenida en el ADN en proteínas. El término "secuencia de nucleótidos" se refiere a un polímero de ADN o ARN que puede ser monocatenario o bicatenario, que contiene opcionalmente bases de nucleótidos sintéticas, no naturales o alteradas, capaces de incorporarse a polímeros de ADN o ARN.

15 Los términos "ácido nucleico", "molécula de ácido nucleico", "fragmento de ácido nucleico", "secuencia o segmento de ácido nucleico" o "polinucleótido" pueden usarse también indistintamente con gen, ADNc, ADN y ARN codificado por un gen, por ej., ADN genómico e incluso secuencias de ADN sintético. El término también incluye secuencias que incluyen cualquiera de los análogos de bases conocidos de ADN y ARN.

20 En una realización de la presente invención, el procedimiento también incluye la puesta en contacto de la muestra con al menos una sonda oligonucleotídica para formar un ácido nucleico hibridado y la amplificación del ácido nucleico hibridado. La "Amplificación" utiliza procedimientos, tales como la reacción en cadena de la polimerasa (PCR), ligación y amplificación (o reacción en cadena de la ligasa, LCR), amplificación por desplazamiento de hebra, amplificación basada en la secuencia de ácido nucleico y métodos de amplificación basados en el uso de la Q-beta replicasa. Estos métodos son bien conocidos y muy utilizados en la técnica. Los reactivos y el hardware para realizar la PCR están disponibles en el mercado. Por ejemplo, en determinadas realizaciones de la presente invención el exón 6 del gen de la enzima glucógeno sintasa 1 equina (también denominado *GSY1*), o una porción del mismo, se puede amplificar mediante PCR. En otra realización de la presente invención, se inmoviliza al menos una sonda oligonucleotídica en una superficie sólida.

30 Los procedimientos de la presente invención se pueden usar para detectar la presencia de un biomarcador asociado a la miopatía equina por almacenamiento de polisacáridos (MAPS) en un caballo, tal como un potro, por ej., un potro neonato o un potro abortado, uno de una pareja reproductora de caballos, por ej., la madre y/o el padre o cualquier caballo en cualquier etapa de la vida. El caballo puede estar vivo o muerto.

35 Además, la presente invención proporciona un procedimiento para el diagnóstico de la miopatía por almacenamiento de polisacáridos (MAPS) en un caballo, implicando el procedimiento la obtención de una muestra fisiológica del caballo, en donde la muestra comprende ácido nucleico y la detección de la presencia de un biomarcador en la muestra, en donde la presencia del biomarcador es indicativa de la enfermedad. Una realización del procedimiento implica además la puesta en contacto de la muestra con al menos una sonda oligonucleotídica para formar un ácido nucleico hibridado. Por ejemplo, en una realización, se amplifica el exón 6 del gen de la enzima glucógeno sintasa 1 equina o una porción del mismo, por ejemplo, mediante la reacción en cadena de la polimerasa, la amplificación por desplazamiento de hebra, la reacción en cadena de la ligasa, métodos de amplificación basados en el uso de la Q-beta replicasa y/o amplificación basada en la secuencia del ácido nucleico. En una realización del procedimiento, el biomarcador contiene un gen de la enzima glucógeno sintasa 1 equina que tiene una sustitución de G por A en el nucleótido 926 en el exón 6 del gen de la enzima glucógeno sintasa 1 equina o un gen que codifica para una enzima glucógeno sintasa que tiene una sustitución R por H en el residuo aminoácido 309. El procedimiento se puede usar para detectar la MAPS en un caballo.

45 Además, la presente invención proporciona un kit que comprende una prueba diagnóstica para la detección de la presencia de la MAPS equina en un caballo que comprende un envase maternal que contiene, envasado por separado, al menos una sonda oligonucleotídica capaz de formar un ácido nucleico hibridado con *GYS1* e instrucciones para el uso de la sonda de acuerdo con los procedimientos de la invención.

## 50 **Breve descripción de los dibujos**

**Figura 1.** Secuencia del ADN que codifica para la *GYS1* equina normal (SEC ID NO:1). El exón 6 está indicado en negrita. El sitio de la mutación G por A en la posición del nucleótido 926 está subrayado. Esta región de la secuencia se amplía a continuación en la Figura 2.

55 **Figura 2.** Exón de *GYS1* y secuencia de ADN flanqueante de caballos normales (SEC ID NO:2) y caballos con MAPS (SEC ID NO: 3). El exón 6 en estas secuencias de ADN de la *GYS1* equina contiene las posiciones 33 - 150. En la posición 135 una G en la secuencia de un caballo normal está sustituida por una A en la secuencia de un caballo con MAPS. Esto cambia el codón de tres bases subrayado de uno que codifica para una arginina (CGT) por uno que codifica para una histidina (CAT).

**Figura 3.** Secuencias amino de la glucógeno sintasa codificadas por el exón 6 de los genes GYS1. Las especies incluidas en el análisis son humana, caballo control, chimpancé, canina, bovina, ratón, rata, cerdo y pez cebra. Todas las especies tienen secuencias de aminoácidos idénticas en esta región de la proteína glucógeno sintasa de músculo esquelético (SEC ID NO:4), que representan los 39 aminoácidos codificados por las posiciones de nucleótidos 33 – 150 en las secuencias de ADN de la Figura 2. Sin embargo, los caballos con MAPS tienen una histidina (H) en la posición de aminoácido 34 en este exón (subrayado) (SEC ED NO:5), mientras que otras especies tienen una arginina (R). Este codón representa el número 309 en la secuencia de codificación completa.

**Figura 4.** Secuencia de ADN genómico del intrón 5, exón 6 e intrón 6 de la GYS1 equina, a partir de la cual los cebadores de PCR que amplifican la mutación de GYS1 de la MAPS serían derivados más adecuadamente (SEC ID NO:6). El exón 6 está indicado en negrita.

**Figura 5.** La secuencia nucleotídica completa que codifica para la GYS1 en la Figura 1 se tradujo para obtener esta secuencia de aminoácidos (SEC ID NO:9). El sitio de la mutación R por H en el codón 309 está subrayado.

### Descripción detallada de la invención

Los caballos afectados con miopatía equina por almacenamiento de polisacáridos (MAPS) son generalmente heterocigotos para el gen afectado.

Un "alelo" es una variante de un gen particular. Por ejemplo, la presente invención se refiere, entre otros, al descubrimiento de que algunos alelos del gen *GYS1* causan la MAPS en caballos. Un "alelo de *GYS1*" se refiere a un alelo normal del locus *GYS1*, así como a un alelo que lleva una variación(es) que predispone a un caballo a desarrollar MAPS. La coexistencia de múltiples alelos en un locus se conoce como "polimorfismo genético". Cualquier sitio en el que existen múltiples alelos como componentes estables de la población es por definición "polimórfico". Un alelo se define como polimórfico si está presente en una frecuencia de al menos el 1% en la población. Un "polimorfismo de un solo nucleótido (SNP)" es una variación de la secuencia del ADN que implica un cambio en un único nucleótido.

Los términos "proteína", "péptido" y "polipéptido" se utilizan indistintamente en la presente memoria.

La invención abarca composiciones de ácidos nucleicos aislados o sustancialmente purificados. En el contexto de la presente invención, una molécula de ADN "aislada" o "purificada" es una molécula de ADN que por intervención humana, existe fuera de su entorno natural y, por consiguiente, no es un producto natural. Una molécula de ADN aislada puede existir en forma purificada o puede existir en un entorno no natural. Por ejemplo, una molécula de ácido nucleico "aislada" o "purificada", o porción de la misma, está sustancialmente exenta de otro material celular o medio de cultivo cuando se produce por técnicas de recombinación o sustancialmente exenta de precursores químicos o de otras sustancias químicas cuando se sintetizan químicamente. En una realización, un ácido nucleico "aislado" está exento de secuencias que flanquean de forma natural al ácido nucleico (es decir, secuencias localizadas en los extremos 5' y 3' del ácido nucleico) en el ADN genómico del organismo del cual deriva el ácido nucleico. Por ejemplo, en varias realizaciones, la molécula de ácido nucleico aislada puede contener menos de aproximadamente 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0,5 kb o 0,1 kb de secuencias de nucleótidos que flanquean de forma natural a la molécula de ácido nucleico en el ADN genómico de la célula de la cual deriva el ácido nucleico. En la presente invención también se abarcan fragmentos y variantes de las secuencias de nucleótidos divulgadas y proteínas o proteínas de longitud parcial codificadas por las mismas.

Por "fragmento" o "porción" de una secuencia se entiende una secuencia de nucleótidos de longitud completa o inferior a la longitud completa de la secuencia de nucleótidos que codifica o la secuencia de aminoácidos de una polipéptido o proteína. Como esta se relaciona con una molécula de ácido nucleico, la secuencia o segmento de la invención, cuando está unida a otras secuencias para la expresión, "porción" o "fragmento" significa una secuencia que tiene, por ejemplo, al menos 80 nucleótidos, al menos 150 nucleótidos o al menos 400 nucleótidos. Si no se emplea para la expresión, una "porción" o "fragmento" significa, por ejemplo, al menos 9, 12, 15 o al menos 20, nucleótidos consecutivos, por ej., sondas y cebadores (oligonucleótidos), que corresponden a la secuencia de nucleótidos de las moléculas de ácido nucleico de la invención. En otra alternativa, los fragmentos o porciones de una secuencia de nucleótidos que son útiles como sondas de hibridación generalmente no codifican para fragmentos de proteínas que conservan la actividad biológica. Por lo tanto, los fragmentos o partes de una secuencia de nucleótidos pueden variar desde al menos aproximadamente 6 nucleótidos, aproximadamente 9, aproximadamente 12 nucleótidos, aproximadamente 20 nucleótidos, aproximadamente 50 nucleótidos, aproximadamente 100 nucleótidos o más.

Una "variante" de una molécula es una secuencia que es sustancialmente similar a la secuencia de la molécula nativa. Para las secuencias de nucleótidos, las variantes incluyen aquellas secuencias que, debido a la degeneración del código genético, codifican la secuencia de aminoácidos idéntica a la de la proteína nativa. Las variantes alélicas naturales como estas se pueden identificar mediante el uso de técnicas de biología molecular bien

conocidas como, por ejemplo, con la reacción en cadena de la polimerasa (PCR) y técnicas de hibridación. Las secuencias de nucleótidos variantes también incluyen secuencias de nucleótidos derivadas sintéticamente, tales como las generadas, por ejemplo, utilizando la mutagénesis dirigida al sitio que codifican para la proteína nativa, así como las que codifican para un polipéptido que tienen sustituciones de aminoácidos. En general, las variantes de las secuencias de nucleótidos de la invención tendrán en al menos una realización, 40%, 50%, 60% a 70%, por ej., 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% a 79%, en general al menos 80%, por ej., 81%-84%, al menos 85%, por ej., 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, a 98%, de identidad con la secuencia de la secuencia de nucleótidos (endógena) nativa.

Polinucleótidos "sintéticos" son aquellos preparados por síntesis química. Una "molécula de ADN recombinante" es una combinación de secuencias de ADN que se unen entre sí usando tecnología y procedimientos de ADN recombinante utilizados para unir entre sí secuencias de ADN como se describe, por ejemplo, en Sambrook y Russell (2001).

El término "gen" se usa en sentido amplio para referirse a cualquier segmento de ácido nucleico asociado a una función biológica. Los genes incluyen secuencias y/o secuencias reguladoras codificantes requeridas para su expresión. Por ejemplo, gen se refiere a un fragmento de ácido nucleico que expresa ARNm, ARN funcional o una proteína específica, tal como la enzima glucógeno sintasa 1, incluyendo sus secuencias reguladoras. Los genes también incluyen segmentos de ADN no expresados que, por ejemplo, forman secuencias de reconocimiento para otras proteínas. Los genes se pueden obtener de diferentes fuentes, incluyendo la clonación de una fuente de interés o sintetizando a partir de una información de una secuencia conocida o predicha y pueden incluir secuencias designadas para tener los parámetros deseados. Además, un "gen" o un "gen recombinante" se refiere a una molécula de ácido nucleico que comprende un marco de lectura abierto e incluyen al menos un exón y (opcionalmente) una secuencia intrón. El término "intrón" se refiere a una secuencia de ADN presente en un gen dado que no se traduce en una proteína y que generalmente se encuentra entre exones.

"Natural", "nativa" o "tipo silvestre" se usa para describir un objeto que se puede encontrar en la naturaleza a diferencia del producido artificialmente. Por ejemplo, una secuencia de nucleótidos presente en un organismo (incluido un virus), la cual se aísla de una fuente de la naturaleza y, la cual, no se ha modificado intencionadamente en el laboratorio, es natural. Además, "tipo silvestre" se refiere al gen normal, u organismo existente en la naturaleza sin ninguna mutación conocida.

Una enzima glucógeno sintasa 1 (GYS1) "mutante" se refiere a la proteína o fragmento de la misma que es codificada por un gen *GYS1* que tiene una mutación, por ej., la que se produce en el locus *GYS1*. Una mutación en un alelo *GYS1* puede dar lugar a una potenciación o aumento de la actividad enzimática en un caballo heterocigoto para el alelo. El incremento de la actividad enzimática se puede determinar mediante métodos conocidos en la técnica. Las mutaciones en *GYS1* pueden ser causa de la enfermedad en un caballo heterocigoto para el alelo *GYS1* mutante, por ej., un caballo heterocigoto para una mutación que da lugar a un producto génico mutante, tal como una mutación por sustitución en el exón 6 de *GYS1*, tal como el designado en la presente memoria como G926A.

"Mutaciones somáticas" son aquellas que se producen sólo en determinados tejidos, por ej., en tejido hepático y no se heredan en la línea germinal. Las mutaciones de la "línea germinal" se pueden encontrar en cualquiera de los tejidos corporales y se heredan. La presente mutación *GYSE1* es una mutación de la línea germinal.

"Homología" se refiere al porcentaje de identidad entre dos secuencias de polinucleótidos o dos secuencias de polipéptidos. Dos secuencias de ADN o polipéptidos son "homólogas" entre sí cuando las secuencias presentan al menos aproximadamente de 75% a 85% (incluyendo 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84% y 85%), al menos aproximadamente 90%, o al menos aproximadamente 95% a 99% (incluyendo 95%, 96%, 97%, 98%, 99%) de identidad con una secuencia contigua a lo largo de una longitud definida de las secuencias.

Los siguientes términos se utilizan para describir las relaciones entre secuencias entre dos o más ácidos nucleicos o polinucleótidos: (a) "secuencia de referencia", (b) "ventana de comparación", (c) "identidad de la secuencia" (d) "porcentaje de identidad de la secuencia" y (e) "identidad sustancial".

(a) Como se usa en la presente memoria, "secuencia de referencia" es una secuencia definida usada como base para la comparación de la secuencia. Una secuencia de referencia puede ser un subgrupo o la totalidad de una secuencia especificada, por ejemplo, como un segmento de un ADNc de longitud completa o secuencia génica o el ADNc completo o la secuencia génica.

(b) Como se usa en la presente memoria, "ventana de comparación" hace referencia a un segmento contiguo y especificado de una secuencia de polinucleótidos, en donde la secuencia de polinucleótidos en la ventana de comparación puede comprender adiciones o deleciones (es decir, huecos) comparado con la secuencia de referencia (que no comprende adiciones ni deleciones) para la alineación óptima de las dos secuencias. En general, la ventana de comparación es de al menos una longitud de 20 nucleótidos contiguos y opcionalmente puede ser 30, 40, 50, 100 o más larga. El experto en la materia sabe que para evitar una elevada similitud con una secuencia de referencia debido a la inclusión de huecos en la secuencia de nucleótidos, generalmente se introduce una penalización por hueco y se resta del número de apareamientos.

Los métodos para la alineación de secuencias para comparación son bien conocidos en la técnica. Por lo tanto, la determinación del porcentaje de identidad entre dos secuencias se puede realizar utilizando un algoritmo matemático.

5 Se pueden utilizar aplicaciones informáticas de estos algoritmos matemáticos para la comparación de secuencias para determinar la identidad de la secuencia. Dichas aplicaciones incluyen, sin limitación: CLUSTAL en el programa PC/Gene (Intelligenetics, Mountain View, California); el programa ALIGN (Versión 2.0) y GAP, BESTFIT, BLAST, FASTA y TFASTA del Wisconsin Genetics Software Package, Versión 8 (Genetics Computer Group (GCG), 575 Science Drive, Madison, Wisconsin, USA). Las alineaciones que utilizan estos programas pueden utilizarse con sus parámetros por defecto.

10 El software para llevar a cabo los análisis BLAST es de acceso público a través del Centro Nacional de Información Biotecnológica (véase la página web en [ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov)). Este algoritmo implica la identificación en primer lugar de pares de secuencia de altísima similitud (HSP) mediante la identificación de palabras breves de longitud  $W$  en la secuencia a analizar, (la cual coincide o satisface alguna puntuación  $T$  del umbral valorado como positivo cuando se alinea con una palabra de la misma longitud en una secuencia de una base de datos.  $T$  se refiere al umbral de puntuación de la palabra adyacente. Este acierto en la palabra adyacente inicial actúa como semilla para iniciar búsquedas de HSP más largas contenidas en las mismas. Las palabras encontradas se extienden a continuación en ambas direcciones a lo largo de cada secuencia hasta que se pueda aumentar la puntuación de la alineación acumulada. Las puntuaciones acumuladas se calculan utilizando, para secuencias de nucleótidos, los parámetros  $M$  (puntuación de recompensa para un par de residuos no coincidentes, siempre  $> 0$ ) y  $N$  (puntuación de penalización para residuos con errores de apareamiento, siempre  $< 0$ ). Para las secuencias de aminoácidos se utiliza una matriz de puntuación para calcular la puntuación acumulada. La extensión de las palabras encontradas en cada dirección se mantiene cuando la puntuación de alineación acumulada se detiene cuando la puntuación de alineación acumulada disminuye en la cantidad  $X$  desde su valor máximo alcanzado, siendo la puntuación acumulada de hasta cero o inferior debido a la acumulación de una o más alineaciones de residuos de puntuación negativa o si se alcanza el final de cualquiera de las secuencias.

Además de calcular el porcentaje de identidad de la secuencia, el algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias. Una medida de la similitud proporcionada por el algoritmo BLAST es la probabilidad de suma más pequeña ( $P(N)$ ) que proporciona una indicación de la probabilidad por la cual se produciría al azar una coincidencia entre dos secuencias de nucleótidos o aminoácidos. Por ejemplo, una secuencia de ácido nucleico a analizar se considera similar a una secuencia de referencia si la probabilidad más pequeña en una comparación de la secuencia de ácido nucleico a analizar con la secuencia de ácido nucleico de referencia es menos de aproximadamente 0,1, menos de aproximadamente 0,01 o incluso menos de aproximadamente 0,001.

Para obtener alineaciones con huecos para fines de comparación, se puede utilizar Gapped BLAST (en BLAST 2.0). Otra alternativa consiste en utilizar PSI-BLAST (en BLAST 2.0) para realizar una búsqueda iterativa que detecte relaciones distantes entre moléculas. Si se utiliza BLAST, Gapped BLAST o PSI-BLAST, se pueden usar los parámetros por defecto de los respectivos programas (por ej., BLASTN para las secuencias de nucleótidos, BLASTX para las proteínas). El programa BLASTN (para secuencias de nucleótidos) utiliza como valores por defecto una longitud de palabra ( $W$ ) de 11, una expectativa ( $E$ ) de 10, un valor de corte de 100,  $M=5$ ,  $N=-4$  y una comparación de ambas hebras. Para las secuencias de aminoácidos, el programa BLASTP utiliza como parámetros por defecto una longitud de palabra ( $W$ ) de 3, una expectativa ( $E$ ) de 10 y la matriz de puntuación BLOSUM62. Véase la página web en [ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov). La alineación también se puede llevar a cabo manualmente mediante inspección visual.

Para los fines de la presente invención, la comparación de las secuencias de nucleótidos para la determinación del porcentaje de identidad de la secuencia con las secuencias del promotor divulgadas en la presente memoria se hace preferiblemente utilizando el programa BlastN (versión 1.4.7 o posterior) con sus parámetros por defecto o con cualquier programa equivalente. Por "programa equivalente" se entiende cualquier programa de comparación de secuencias que, para cualquiera de las dos secuencias en cuestión, genera una alineación que tiene coincidencias en residuos de nucleótidos o aminoácidos idénticos y un porcentaje de identidad de secuencia idéntico cuando se compara con la correspondiente alineación generada mediante un programa BLAST.

(c) Como se usa en la presente memoria, "identidad de la secuencia" o "identidad" en el contexto de dos secuencias de ácidos nucleicos o polipéptidos hace referencia a un porcentaje especificado de residuos en las dos secuencias que son los mismos cuando se alinean para obtener la máxima correspondencia en una ventana de comparación especificada, medida mediante los algoritmos de comparación de la secuencia o por inspección visual. Cuando se utiliza el porcentaje de identidad de la secuencia en referencia a las proteínas, se reconoce que las posiciones de residuos que no son idénticos frecuentemente difieren por sustituciones conservativas de aminoácidos, en las que los residuos de aminoácidos se sustituyen por otros residuos de aminoácidos con propiedades químicas similares (por ej., carga o hidrofobicidad) y, por consiguiente, no cambian las propiedades funcionales de la molécula. Cuando las secuencias difieren en sustituciones conservativas, el porcentaje de identidad de la secuencia se puede ajustar al alza para corregir la naturaleza conservativa de la sustitución. Las secuencias que difieren mediante dichas sustituciones conservativas se dice que tienen "similitud de secuencia" o "similitud". Los medios para realizar este ajuste son bien conocidos para el experto en la materia. Habitualmente, esto implica la puntuación de una sustitución conservativa como

un apareamiento parcial en lugar de un apareamiento total, aumentando de esta manera el porcentaje de identidad de la secuencia. Así, por ejemplo, cuando a un aminoácido idéntico se le da una puntuación de 1 y a una sustitución no conservativa se le da una puntuación de cero, a una sustitución conservativa se le da una puntuación entre cero y 1. La puntuación de las sustituciones conservativas se calcula, por ejemplo,

mediante el programa PC/GENE (Intelligenetics, Mountain View, California).  
 (d) Como se usa en la presente memoria "porcentaje de identidad de la secuencia" significa el valor determinado por comparación de dos secuencias alineadas óptimamente a lo largo de una ventana de comparación, en donde la parte de la secuencia de polinucleótidos en la ventana de comparación puede comprender adiciones o deleciones (es decir, huecos) en comparación con la secuencia de referencia (la cual no comprende adiciones ni deleciones) para una alineación óptima de las dos secuencias. El porcentaje se calcula determinando el número de posiciones en las cuales existen bases de ácidos nucleicos o residuos de aminoácidos en ambas secuencias para obtener el número de posiciones coincidentes, dividiendo el número de posiciones coincidentes por el número total de posiciones en la ventana de comparación y multiplicando el resultado por 100 para obtener el porcentaje de identidad de la secuencia.

(e)(i) El término "identidad sustancial" de las secuencias de polinucleótidos significa que un polinucleótido comprende una secuencia que tiene al menos 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% o 79%; al menos 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88% o 89%; al menos 90%, 91%, 92%, 93% o 94% o incluso al menos 95%, 96%, 97%, 98% o 99% de identidad con la secuencia, comparado con una secuencia de referencia usando uno de los programas de alineación descritos utilizando los parámetros estándar. El experto en la materia reconocerá que estos valores se pueden ajustar apropiadamente para determinar la correspondiente identidad de las proteínas codificadas por dos secuencias de nucleótidos teniendo en cuenta la degeneración de los codones, la similitud entre aminoácidos, el posicionamiento del marco de lectura y similares. Una identidad sustancial de las secuencias de aminoácidos para estos fines normalmente significa una identidad de la secuencia de al menos 70%, o al menos 80%, 90% o incluso al menos 95%.

Otra indicación de que las secuencias de nucleótidos son sustancialmente idénticas es si dos moléculas hibridan entre sí en condiciones estrictas (véase más adelante). En general, se seleccionan condiciones estrictas de aproximadamente 5°C inferiores al punto de fusión térmico ( $T_m$ ) de la secuencia específica para una fuerza iónica y pH definidos. Sin embargo, las condiciones estrictas abarcan temperaturas en el intervalo de aproximadamente 1°C a aproximadamente 20°C, dependiendo del grado deseado de rigor como por otra parte se establece en la presente memoria. Los ácidos nucleicos que no hibridan entre sí en condiciones estrictas siguen siendo básicamente idénticos si los polipéptidos que codifican son sustancialmente idénticos. Esto puede ocurrir, por ej., cuando se crea una copia de un ácido nucleico utilizando la máxima degeneración de los codones permitida por el código genético. Una indicación de que dos secuencias de ácidos nucleicos son sustancialmente idénticos es cuando el polipéptido codificado por el primer ácido nucleico muestra reacción cruzada inmunológica con el polipéptido codificado por el segundo ácido nucleico.

(e) (ii) El término "identidad sustancial" en el contexto de un péptido indica que un péptido comprende una secuencia con al menos 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78% o 79%; al menos 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88% o 89%; o al menos 90%, 91%, 92%, 93%, o 94%; o incluso al menos 95%, 96%, 97%, 98% o 99% de identidad de la secuencia con la secuencia de referencia a lo largo de una ventana de comparación específica. Una indicación de que dos secuencias de péptidos son sustancialmente idénticas es que un péptido es inmunológicamente reactivo con anticuerpos obtenidos frente al segundo péptido. Por lo tanto, un péptido es sustancialmente idéntico a un segundo péptido, por ejemplo, si los dos péptidos difieren sólo por una sustitución conservativa.

Para la comparación de secuencias, habitualmente una secuencia actúa como una secuencia de referencia con la cual se comparan las secuencias a analizar. Cuando se usa un algoritmo de comparación de la secuencia, las secuencias a analizar y la de referencia se introducen en un ordenador y se designan coordenadas de la subsecuencia en caso necesario y se designan los parámetros del programa del algoritmo de la secuencia. El algoritmo de comparación de la secuencia calcula a continuación el porcentaje de identidad de la secuencias para la secuencia(s) a analizar respecto a la secuencia de referencia basándose en los parámetros del programa designados.

Como se ha señalado anteriormente, otra indicación de que dos secuencias de ácidos nucleicos son sustancialmente idénticas es que las dos moléculas hibridan entre sí en condiciones estrictas. La frase "hibrida específicamente con" se refiere a la unión, formación de dúplex o hibridación de una molécula sólo con una secuencia de nucleótidos particular en condiciones estrictas cuando la secuencia está presente en una mezcla compleja de ADN o ARN (por ej. celular total). "Se une(n) sustancialmente" se refiere a la hibridación complementaria entre un ácido nucleico sonda y un ácido nucleico diana y abarca errores de apareamiento menores que se pueden acomodar reduciendo las condiciones estrictas de los medios de hibridación para lograr la detección deseada de las secuencias de ácido nucleico diana.

"Condiciones de hibridación estrictas" y "condiciones de lavado de hibridación estrictas" en el contexto de los experimentos de hibridación de los ácidos nucleicos, tales como las hibridaciones Southern y Northern dependen de la secuencia y son diferentes para los diferentes parámetros ambientales. Secuencias más largas hibridan específicamente a temperaturas más elevadas. La  $T_m$  es la temperatura (para una fuerza iónica y pH definidos) para

la cual el 50% de la secuencia diana híbrida con una sonda perfectamente coincidente. La especificidad está generalmente en función de los lavados poshibridación, siendo los factores críticos la fuerza iónica y la temperatura de la solución de lavado final. Para los híbridos ADN-ADN, la  $T_m$  se puede calcular a partir de la ecuación de Meinkoth y Wahl:

$$T_m = 81,5^\circ\text{C} + 16,6 (\log M) + 0,41 (\%GC) - 0,61 (\% \text{ form}) - 500/L$$

5

donde M es la molaridad de los cationes monovalentes, %GC es el porcentaje de nucleótidos guanosina y citosina en el ADN, % form es el porcentaje de formamida en la solución de hibridación y L es la longitud del híbrido en pares de bases.  $T_m$  se reduce en aproximadamente  $1^\circ\text{C}$  para cada 1% de errores de apareamiento; por lo tanto, para hibridar secuencias de la identidad deseada se pueden ajustar la  $T_m$ , las condiciones de hibridación y/o lavado. Por ejemplo, si se buscan secuencias con >90% de identidad, la  $T_m$  puede reducirse en  $10^\circ\text{C}$ . En general, las condiciones estrictas se seleccionan para que sean aproximadamente  $5^\circ\text{C}$  inferiores al punto de fusión térmica ( $T_m$ ) para la secuencia específica y se complementa para una fuerza iónica y pH definidos. Sin embargo, condiciones muy estrictas pueden ser una hibridación y/o lavado a 1, 2, 3 ó  $4^\circ\text{C}$  inferior al punto de fusión térmico ( $T_m$ ), condiciones moderadamente estrictas pueden ser una hibridación y/o lavado a 6, 7, 8, 9 ó  $10^\circ\text{C}$  inferior al punto de fusión térmico ( $T_m$ ), condiciones poco estrictas pueden ser una hibridación y/o lavado a 11, 12, 13, 14, 15 o  $20^\circ\text{C}$  inferior al punto de fusión térmico ( $T_m$ ). Utilizando la ecuación, las composiciones para hibridación y lavado, y el T deseado, el experto en la materia comprenderá que las variaciones en el rigor de las soluciones de hibridación y/o lavado están intrínsecamente descritas. Si el grado deseado de errores de apareamiento da un T de menos de  $45^\circ\text{C}$  (solución acuosa) o  $32^\circ\text{C}$  (solución de formamida), se prefiere aumentar la concentración de SCC de modo que se pueda utilizar una temperatura más alta. En general, se seleccionan condiciones de hibridación y lavado muy estrictas de aproximadamente  $5^\circ\text{C}$  inferior al punto de fusión térmico ( $T_m$ ) para la secuencia específica para una fuerza iónica y pH definidos.

10

15

20

25

30

35

Un ejemplo de condiciones de lavado muy estrictas es NaCl 0,15M a  $72^\circ\text{C}$  durante aproximadamente 15 minutos. Un ejemplo de condiciones de lavado estrictas es un lavado con SSC 0,2X a  $65^\circ\text{C}$  durante 15 minutos. Frecuentemente, un lavado muy estricto va precedido de un lavado poco estricto para eliminar la señal de fondo de la sonda. Un ejemplo de lavado moderadamente estricto para un dúplex de por ej., más de 100 nucleótidos, es SSC 1X a  $45^\circ\text{C}$  durante 15 minutos. Un ejemplo de lavado poco estricto para un dúplex de por ej., más de 100 nucleótidos, es SSC 4-6X a  $40^\circ\text{C}$  durante 15 minutos. Para las sondas cortas (por ej., aproximadamente de 10 a 50 nucleótidos), condiciones estrictas son generalmente concentraciones de sal de menos de aproximadamente 1,5M, más preferiblemente aproximadamente 0,01 a 1,0M, concentración de ion Na (o de otras sales) a pH 7,0 a 8,3 y la temperatura es habitualmente al menos  $30^\circ\text{C}$  y al menos  $60^\circ\text{C}$  para las sondas largas (por ej., >50 nucleótidos). Unas condiciones estrictas también se pueden alcanzar con la adición de agentes desestabilizantes, tales como formamida. En general, una relación señal-ruido de 2X (o más) que la observada para una sonda no relacionada en el ensayo de hibridación en particular indica la detección de una hibridación específica. Los ácidos nucleicos que no hibridan entre sí en condiciones estrictas siguen siendo sustancialmente idénticos si las proteínas para las que codifican son sustancialmente idénticas. Esto ocurre, por ej., cuando se crea una copia de un ácido nucleico utilizando la máxima degeneración de codones permitida por el código genético.

40

45

Se seleccionan condiciones muy estrictas para que sean igual al  $T_m$  para una sonda en particular. Un ejemplo de condiciones estrictas para la hibridación de ácidos nucleicos complementarios que tienen más de 100 residuos complementarios en un filtro de una transferencia Southern o Northern es formamida al 50%, por ej., hibridación en formamida al 50%, NaCl 1M, SDS al 1% a  $37^\circ\text{C}$  y un lavado en SSC 0,1X a 60 a  $65^\circ\text{C}$ . Ejemplos de condiciones poco estrictas incluyen la hibridación con una solución tampón de formamida al 30 al 35%, NaCl 1M, SDS al 1% (dodecilsulfato sódico) a  $37^\circ\text{C}$  y un lavado en SSC 1X a 2X (SSC 20X = NaCl 3,0M/citrato trisódico 0,3M) a 50 a  $55^\circ\text{C}$ . Ejemplos de condiciones rigurosas moderadas incluyen la hibridación en formamida al 40 al 45%, NaCl 1,0M SDS al 1%  $37^\circ\text{C}$  y un lavado en SSC 0,5X a 1X a 55 a  $60^\circ\text{C}$ .

50

Por polipéptido "variante" se entiende un polipéptido derivado de la proteína nativa por delección (denominado truncado) o la adición de uno o más aminoácidos al extremo N terminal y/o C terminal de la proteína nativa; delección o adición de uno o más aminoácidos a uno o más sitios de la proteína nativa o sustitución de uno o más aminoácidos a uno o más sitios de la proteína nativa. Dichas variantes pueden ser el resultado, por ejemplo, de polimorfismo genético o de la manipulación humana. Métodos de este tipo de manipulaciones son generalmente conocidos en la técnica.

55

Por lo tanto, los polipéptidos de la invención se pueden alterar de varias maneras incluida las sustituciones, delecciones, truncados e inserciones de aminoácidos. Los procedimientos para este tipo de manipulaciones son generalmente conocidos en la técnica. Por ejemplo, las variantes de la secuencia de aminoácidos de los polipéptidos se pueden preparar mediante mutaciones en el ADN. Los procedimientos de mutagénesis y alteraciones de la secuencia de nucleótidos son muy conocidos en la técnica. Las sustituciones apropiadas de aminoácidos que no afectan a la actividad biológica de la proteína de interés son bien conocidas en la técnica. Se prefieren las sustituciones conservativas, tales como el intercambio de un aminoácido por otro con propiedades similares.

Por lo tanto, los genes y las secuencias nucleotídicas de la invención incluyen tanto secuencias naturales como las formas mutantes. De igual modo, los polipéptidos de la invención abarcan proteínas naturales, así como variaciones y formas modificadas de las mismas. Dichas variantes seguirán teniendo la actividad deseada. No se espera que las delecciones, inserciones y sustituciones de la secuencia de polipéptidos abarcadas en la presente memoria produzcan cambios radicales en las características del polipéptido. Sin embargo, aunque es difícil predecir anticipadamente el efecto exacto de la sustitución, delección o inserción, el experto en la materia sabrá que el efecto se evaluará mediante los ensayos rutinarios de detección.

Las sustituciones, delecciones o adiciones individuales que alteren, añadan o eliminen un único aminoácido o un pequeño porcentaje de aminoácidos (generalmente menos del 5%, más generalmente menos del 1%) en una secuencia codificada son “variaciones modificadas conservativamente”.

Las “variaciones modificadas conservativamente” de una secuencia de ácido nucleico en particular se refieren a aquellas secuencias de ácidos nucleicos que codifican para secuencias de aminoácidos idénticas o esencialmente idénticas o cuando la secuencia de ácido nucleico no codifica para una secuencia de aminoácidos, a secuencias esencialmente idénticas. Debido a la degeneración del código genético, un gran número de ácidos nucleicos funcionalmente idénticos codifican para cualquier polipéptido dado. Por ejemplo, los codones CGT, CGC, CGA, CGG, AGA y AGG codifican todos ellos para el aminoácido arginina. Por lo tanto, en cualquier posición en la que una arginina esté especificada por un codón, el codón puede alterarse por uno cualquiera de los correspondientes codones descritos sin alterar la proteína codificada. Dichas variaciones de ácidos nucleicos son “variaciones silentes”, que son una especie de las “variaciones modificadas conservativamente”. Cada secuencia de ácido nucleico descrita en la presente memoria que codifica para un polipéptido también describe cada variación posiblemente silente, salvo que se especifique de otra manera. El experto en la materia reconocerá que cada codón de un ácido nucleico (salvo ATG, que es habitualmente el único codón para metionina) se puede modificar mediante técnicas estándar para dar una molécula funcionalmente idéntica. Por consiguiente, cada “variación silente” de un ácido nucleico que codifica para un polipéptido está implícita en cada secuencia descrita.

El término “transformación” se refiere a la transferencia de un fragmento de ácido nucleico dentro del genoma de una célula hospedadora, dando lugar a una herencia genéticamente estable. Las células hospedadoras que contienen los fragmentos de ácido nucleico transformados se conocen como células “transgénicas” y los organismos que comprenden células transgénicas se conocen como “organismos transgénicos”.

Una “célula hospedadora” es una célula que se ha transformado o que es capaz de transformación mediante una molécula de ácido nucleico exógena. Por lo tanto, “transformada”, “transgénica” y “recombinante” se refiere a una célula hospedadora en la cual se ha introducido una molécula de ácido nucleico heterólogo. La molécula de ácido nucleico puede integrarse establemente en el genoma generalmente conocido en la técnica. Métodos conocidos de PCR incluyen, sin limitación, métodos que utilizan cebadores emparejados, cebadores anidados, cebadores específicos únicos, cebadores degenerados, cebadores específicos de genes, cebadores específicos del vector, cebadores parcialmente con errores de apareamiento y similares. Por ejemplo, se han obtenido células “transformadas”, “transgénicas” y “recombinantes” mediante el proceso de transformación y contienen un gen extraño integrado en su cromosoma. El término “no transformadas” se refiere a células normales que no se han obtenido por el proceso de transformación.

“Cassette de expresión” como se usa en la presente memoria, significa una secuencia de ADN capaz de dirigir la expresión de una secuencia de nucleótidos en particular en una célula hospedadora adecuada, que comprende un cebador unido operativamente a la secuencia de nucleótidos de interés y que está operativamente unido a las señales de terminación. También incluye generalmente secuencias requeridas para la traducción adecuada de la secuencia de nucleótidos. La región codificante codifica habitualmente para una proteína de interés, aunque también puede codificar para un ARN funcional de interés, por ejemplo, ARN antisentido o ARN no traducido, en la dirección sentido o antisentido. El cassette de expresión que comprende la secuencia de nucleótidos de interés puede ser quimérico, lo que significa que al menos uno de sus componentes es heterólogo con respecto a al menos uno de sus otros componentes. El cassette de expresión puede ser también uno natural, pero que se ha obtenido en una forma recombinante útil para la expresión heteróloga. La expresión de la secuencia de nucleótidos en el cassette de expresión puede estar bajo el control de un promotor constitutivo o de un promotor inducible que inicia la transcripción sólo cuando la célula hospedadora está expuesta a algunos estímulos externos en particular. En el caso de un organismo multicelular, el promotor también puede ser específico de un tejido u órgano en particular o de una etapa del desarrollo.

Dichos cassettes de expresión tendrán la región de inicio de la transcripción de la invención unida a una secuencia de nucleótidos de interés. Un cassette de expresión de este tipo contendrá una pluralidad de sitios de restricción para la inserción del gen de interés que va a estar bajo la regulación transcripcional de las regiones reguladores. El cassette de expresión puede contener adicionalmente genes marcadores seleccionables.

El cassette de transcripción incluirá en la dirección 5'-3' de transcripción, una región de inicio de la transcripción y de la traducción, una secuencia de ADN de interés y una región de terminación de la transcripción y de la traducción funcional en plantas. La región de terminación puede ser nativa con la región de inicio de la transcripción, puede ser nativa con la secuencia de ADN de interés y proceder de otra fuente.

- Los términos “secuencia de ADN heteróloga”, “segmento de ADN exógeno” o “ácido nucleico heterólogo”, se refieren cada uno de ellos a una secuencia que procede de una fuente extraña a la célula hospedadora particular o, si procede de la misma fuente, está modificada respecto a su forma original. Por lo tanto, un gen heterólogo en una célula hospedadora incluye un gen que es endógeno para la célula hospedadora en particular pero que se ha modificado mediante, por ejemplo, el uso de mutagénesis de una única hebra. Los términos también incluyen copias múltiples no naturales de una secuencia de ADN natural. Por lo tanto, los términos se refieren a un segmento de ADN que es extraño o heterólogo para la célula, u homólogo para la célula pero en una posición dentro del ácido nucleico de la célula hospedadora en la cual no se encuentra normalmente el elemento. Los segmentos de ADN exógenos se expresan para dar los polipéptidos exógenos.
- Una secuencia de ADN “homóloga” es una secuencia de ADN que está asociada naturalmente con la célula hospedadora en la cual se introduce.
- “Genoma” se refiere al material genético completo de un organismo.
- “Secuencia codificante” se refiere a una secuencia de ADN o ARN que codifica para una secuencia de aminoácidos específica y excluye las secuencias no codificantes. Por ejemplo, una “secuencia codificante” de ADN o “una secuencia codificante” de un polipéptido en particular, es una secuencia de ADN que se transcribe y se traduce en un polipéptido *in vitro* o *in vivo* cuando se coloca bajo el control de elementos reguladores apropiados. Los límites de la secuencia codificantes se determinan por un codón de inicio en el extremo 5' y un codón de parada en el extremo 3'. Una secuencia codificante puede incluir, pero sin limitación, secuencias procariontas, ADNc de ARNm eucariota, secuencias de ADN genómico de secuencias de ADN eucariota (por ej., de mamífero) e incluso secuencias de ADN sintéticas. Una secuencia de terminación de la transcripción estará habitualmente localizada en 3' respecto a la secuencia codificante. Esta puede constituir una “secuencia de codificación ininterrumpida”, es decir, que carece de un intrón, tal como en el ADNc o puede incluir uno o más intrones unidos mediante uniones de empalme adecuadas. Un “intrón” es una secuencia de ARN que está contenida en el transcrito primario, pero que se ha eliminado por corte y empalme del ARN en la célula para crear el ARNm que se puede traducir en una proteína.
- Los términos “marco de lectura abierto” y “ORF” se refieren a la secuencia de aminoácidos codificada entre los codones de inicio y finalización de la traducción de una secuencia codificante. Los términos “codón de inicio” y “codón de terminación” se refieren a una unidad de tres nucleótidos adyacentes ('codón') en una secuencia codificante que especifica el inicio y la terminación de la cadena, respectivamente, de la síntesis de proteínas (traducción del ARNm).
- El término “transcrito de ARN” se refiere al producto resultante de la transcripción catalizada por la ARN polimerasa de una secuencia de ADN. Cuando el transcrito de ARN es una copia complementaria perfecta de una secuencia de ADN, se refiere al transcrito primario o puede ser una secuencia de ARN derivada del procesamiento postranscripcional del transcrito primario y se denomina ARN maduro. “ARN mensajero” (ARNm) se refiere al ARN que está sin intrones y que se puede traducir en una proteína por la célula. “ADNc” se refiere a un ADN monocatenario o bicatenario que es complementario con y derivado del ARNm.
- El término “secuencia reguladora” es conocido en la técnica y pretende incluir promotores, intensificadores y otros elementos de control (por ej., señales de poliadenilación). Dichas secuencias reguladoras son conocidas por los expertos en la materia. Se entiende que el diseño del vector de expresión puede depender de este tipo de factores, como la elección de la célula hospedadora a transfectar y/o la cantidad de proteína de fusión a expresar.
- El término “elementos de control del ADN” se refiere colectivamente a promotores, sitios de unión a ribosomas, señales de poliadenilación, secuencias de terminación de la transcripción, dominios reguladores en dirección 5', intensificadores y similares, que en conjunto proporcionan la transcripción y traducción de una secuencia codificante en una célula hospedadora. No todas estas secuencias de control tienen que estar siempre presentes en un vector recombinante siempre que el gen deseado sea capaz de transcribirse y traducirse.
- Un elemento de control, tal como un promotor, “dirige la transcripción” de una secuencia codificante en una célula cuando la ARN polimerasa se une al promotor y transcribe la secuencia codificante en ARNm, la cual se traduce a continuación en el polipéptido codificado por la secuencia codificante.
- Una célula se ha “transformado” por un ADN exógeno cuando dicho ADN exógeno se ha introducido dentro de la membrana celular. El ADN exógeno puede integrarse o no (uniéndose covalentemente) en el ADN cromosómico pasando a formar parte del genoma de la célula. En procariontas y levaduras, por ejemplo, el ADN exógeno se puede mantener en un elemento episoma, tal como un plásmido. Con respecto a las células eucariotas, una célula transformada establemente es aquella en la cual el ADN se ha integrado en el cromosoma de modo que lo heredan las células hijas mediante la replicación cromosómica. Esta estabilidad queda demostrada por la capacidad de la célula eucariota de establecer líneas celulares o clones que tienen una población de células hija que contienen el ADN exógeno.
- “Unido operativamente” se refiere a la asociación de secuencias de ácido nucleico en un único fragmento de ácido nucleico de modo que la función de una se ve afectada por la otra, por ej., una disposición de elementos, en la que los componentes así descritos están configurados de tal forma que llevan a cabo su función habitual. Por ejemplo, se

dice que una secuencia de ADN reguladora está “operativamente unida a” o “asociada a” una secuencia de ADN que codifica para un ARN o un polipéptido si las dos secuencias están situadas de tal forma que la secuencia de ADN reguladora afecta a la expresión de la secuencia de ADN codificante (es decir, que la secuencia codificante o el ARN funcional está bajo el control transcripcional del promotor). Las secuencias codificantes pueden estar operativamente unidas a secuencias reguladoras en orientación sentido o antisentido. Los elementos de control unidos operativamente a una secuencia codificante son capaces de efectuar la expresión de la secuencia codificante. Los elementos de control deben estar contiguos a la secuencia codificante, siempre y cuando funcionen para dirigir la expresión de la misma. Por lo tanto, por ejemplo, entre un promotor y la secuencia codificante pueden existir secuencias intermedias no traducidas aunque sí transcritas y el promotor sigue considerándose “unido operativamente” a la secuencia codificante.

“Fragmento de parada de la transcripción” se refiere a las secuencias de nucleótidos que contienen una o más señales reguladoras, tales como las secuencias de la señal de poliadenilación, capaces de finalizar la transcripción. Ejemplos incluyen las regiones no reguladoras 3' de genes que codifican para la nopalina sintasa y la subunidad pequeña de la ribulosa bifsosfato carboxilasa.

“Fragmento de parada de la traducción” o “codón de parada de la traducción” se refiere a las secuencias de nucleótidos que contienen una o más señales reguladoras, tales como uno o más codones de terminación en los tres marcos, capaces de terminar la traducción. La inserción de un fragmento de parada de la traducción adyacente a o cercano al codón de inicio en el extremo 5' de la secuencia codificante tendrá como resultado la no traducción o una traducción incorrecta. El cambio de al menos un nucleótido en una secuencia de ácido nucleico puede dar lugar a la interrupción de la secuencia codificante del gen, por ej., un codón de parada prematuro. Dichos cambios en la secuencia pueden provocar una mutación en el polipéptido codificado por un gen *GYS1*. Por ejemplo, si la mutación es una mutación sin sentido, la mutación da lugar a la generación de un codón de parada prematuro, provocando la generación de un polipéptido GYS truncado.

#### **Ácidos nucleicos de la invención**

Las fuentes de secuencias de nucleótidos de las que se pueden obtener las presentes moléculas de ácido nucleico incluyen cualquier fuente procariota o eucariota. Por ejemplo, se pueden obtener de una fuente celular de un mamífero, tal como un equino. Otra alternativa es la obtención de las moléculas de ácido nucleico de la presente invención a partir de una biblioteca, tal como la biblioteca CHORI- 241 Equine BAC o la biblioteca BAC desarrollada en el INRA, Centre de Recherches de Jouy, Laboratoire de Genetique biochimique et de Cytogenetique, Departement de Genetique animale, 78350 Jouy-en-Josas Cedex, Francia.

Como se ha descrito anteriormente, los términos “aislado y/o purificado” se refieren al aislamiento *in vitro* de un ácido nucleico, por ej., una molécula de ADN o ARN de su entorno celular natural y de la asociación con otros componentes de la célula, tales como un ácido nucleico o un polipéptido, de modo que se puedan secuenciar, replicar y/o expresar. Por ejemplo, “ácido nucleico aislado” puede ser una molécula de ADN que es complementaria o que hibrida con una secuencia de un gen de interés, es decir, una secuencia de ácido nucleico que codifica para una enzima glucógeno sintasa equina y que permanece unida de forma estable en condiciones estrictas (como se ha definido en los procedimientos bien conocidos en la técnica). Por lo tanto, el ARN o el ADN se “aisla” en la medida que está libre de al menos una fuente de ácido nucleico contaminante con el cual normalmente está asociado en la fuente natural de ARN o de ADN y en una realización de la invención está sustancialmente libre de cualquier otro ARN o ADN de mamífero. La frase “libre de al menos una fuente de ácido nucleico contaminante con el cual normalmente está asociado” incluye el caso en el que el ácido nucleico se vuelve a introducir en la fuente de la célula natural pero está en una localización cromosómica diferente o si de otra manera está flanqueado por secuencias de ácido nucleico que normalmente no se encuentran en la célula, por ej., en un vector o plásmido.

Como se usa en la presente memoria, el término “ácido nucleico recombinante”, por ej., “secuencia o segmento de ADN recombinante” se refiere a un ácido nucleico, por ej., al ADN que se ha derivado o aislado de cualquier fuente celular apropiada, que puede alterarse posteriormente *in vitro*, de modo que su secuencia no es natural ni corresponde a secuencias naturales que no están situadas como estarían situadas en un genoma que no se ha transformado con ADN exógeno. Un ejemplo ADN preseleccionado “derivado” de una fuente sería una secuencia de ADN que está identificada como un fragmento útil dentro de un organismo dado y que a continuación se sintetiza en forma esencialmente pura. Un ejemplo de este tipo de ADN “aislado” de una fuente que sería una secuencia de ADN útil que se escinde o se elimina de dicha fuente por medios químicos, por ej., mediante el uso de endonucleasas de restricción, de modo que se pueda manipular, por ej., amplificar, para su uso en la invención, mediante la metodología de ingeniería genética.

Por lo tanto, la recuperación o aislamiento de un fragmento dado de ADN a partir de una digestión de restricción puede emplear la separación del digerido en poliacrilamida o agarosa por electroforesis en gel, identificación del fragmento de interés por comparación de su movilidad frente a la de los fragmentos de ADN del marcador de peso molecular conocido, retirada de la sección de gel que contiene el fragmento deseado y separación del gel del ADN. Por consiguiente, “ADN recombinante” incluye secuencias de ADN totalmente sintéticas, secuencias de ADN semisintéticas, secuencias de ADN aisladas de fuentes biológicas y secuencias de ADN derivadas de ARN, así como mezclas de las mismas.

Las moléculas de ácido nucleico que tienen sustituciones en las bases (es decir, variantes) se preparan mediante diversos procedimientos conocidos en la técnica. Estos métodos incluyen, sin limitación, aislamiento de una fuente natural (en el caso de variantes de secuencias naturales) o la preparación mediante mutagénesis mediadas por oligonucleótidos (o dirigidas al sitio), mutagénesis por PCR y mutagénesis de cassette de una versión variante o no variante preparada anteriormente de la molécula de ácido nucleico.

#### **Métodos de amplificación de ácidos nucleicos**

De acuerdo con los procedimientos de la presente invención, la amplificación del ADN presente en una muestra fisiológica puede llevarse a cabo mediante cualquier medio conocido en la técnica. Ejemplos de técnicas de amplificación adecuadas incluyen, sin limitación, la reacción en cadena de la polimerasa (incluyendo, para la amplificación del ARN, la reacción en cadena de la polimerasa con transcriptasa inversa), la reacción en cadena de la ligasa, la amplificación por desplazamiento de la hebra, la amplificación basada en la transcripción, la replicación de secuencia autosostenida (o "3SR"), el sistema de la replicasa Q $\beta$ , la amplificación basada en la secuencia del ácido nucleico (o "NASBA"), la reacción en cadena con reparación (o "RCR") y la amplificación del ADN boomerang (o "BDA").

Las bases incorporadas en la amplificación del producto pueden ser bases naturales o modificadas (modificadas antes o después de la amplificación) y las bases se pueden seleccionar para optimizar pasos de detección electroquímica posteriores.

La reacción en cadena de la polimerasa (PCR) se puede llevar a cabo de acuerdo con las técnicas conocidas. Véase, por ejemplo, las patentes US-4.683.195; US-4.683.202; US-4.800.159 y 4.965.188. En general, la PCR implica, primero, el tratamiento de una muestra de ácido nucleico (por ej., en presencia de una ADN polimerasa termoestable) con un cebador de oligonucleótidos para cada hebra de la secuencia específica a detectar en condiciones de hibridación, de modo que se sintetiza un producto de extensión de cada cebador que es complementario con cada hebra de ácido nucleico, con los cebadores suficientemente complementarios con cada hebra de la secuencia específica a hibridar, de modo que el producto de extensión sintetizado a partir de cada cebador, cuando se separa de su complementario, puede servir como molde para la síntesis del producto de extensión del otro cebador y posteriormente el tratamiento de la muestra en condiciones desnaturizantes para separar los productos de extensión del cebador de sus moldes si la secuencia o secuencias a detectar están presentes. Estos pasos se repiten cíclicamente hasta que se obtiene el grado de amplificación deseado. La detección de la secuencia amplificada se puede llevar a cabo añadiendo al producto de reacción una sonda oligonucleotídica capaz de hibridar con el producto de reacción (por ej., una sonda oligonucleotídica de la presente invención), llevando la sonda un marcador detectable y detectando posteriormente el marcador de acuerdo con las técnicas bien conocidas. Cuando el ácido nucleico a amplificar es el ARN, la amplificación se puede llevar a cabo mediante la conversión inicial en ADN por la transcriptasa inversa de acuerdo con las técnicas conocidas.

La amplificación por desplazamiento de la hebra (SDA) se puede llevar a cabo de acuerdo con las técnicas conocidas. Por ejemplo, la SDA se puede llevar a cabo con un único cebador de amplificación o con un par de cebadores de amplificación, consiguiéndose la amplificación exponencial con estos últimos. En general, los cebadores para amplificación SDA comprenden, en dirección 5' a 3', una secuencia flanqueante (cuya secuencia de ADN no es crítica), un sitio de restricción para la enzima de restricción empleada en la reacción y una secuencia de oligonucleótidos (por ej., una sonda oligonucleotídica de la presente invención) que hibrida con la secuencia diana a amplificar y/o detectar. La secuencia flanqueante, la cual sirve para facilitar la unión de la enzima de restricción al sitio de reconocimiento y que proporciona un sitio de cebado para la ADN polimerasa después de haber hecho una mella en el sitio de restricción, tiene una longitud de aproximadamente 15 a 20 nucleótidos en una realización. El sitio de restricción es funcional en la reacción de SDA. La parte de la sonda oligonucleotídica tiene una longitud de aproximadamente 13 a 15 nucleótidos en una realización de la invención.

La reacción en cadena de la ligasa (LCR) se puede llevar a cabo también de acuerdo con las técnicas conocidas. En general, la reacción se lleva a cabo con dos pares de sondas de oligonucleótidos: un par se une a una hebra de la secuencia a detectar y el otro par se une a la otra hebra de la secuencia a detectar. Cada par junto solapa completamente con la hebra a la que corresponde. La reacción se lleva a cabo primero desnaturizando (por ej., separando) las hebras de la secuencia a detectar, haciendo reaccionar a continuación las hebras con los dos pares de sondas de nucleótidos en presencia de una ligasa termoestable de modo que cada par de sondas de oligonucleótidos se ligan entre sí, separando a continuación el producto de reacción y repitiendo cíclicamente el procedimiento hasta que la secuencia se ha amplificado el grado deseado. La detección se puede llevar a cabo a continuación de la manera descrita anteriormente para la PCR.

En una realización de la invención, cada exón del gen *GYS1* se amplifica mediante PCR usando los cebadores basados en la secuencia conocida. Los exones amplificados se secuencian a continuación utilizando secuenciadores automatizados. De esta manera, los exones del gen *GYS1* de caballos con sospecha de padecer MAPS en su pedigrí se secuencian hasta encontrar una mutación. Ejemplos de dichas mutaciones incluyen las mutaciones en el exón 6 del ADN del *GYS1*. Por ejemplo, una mutación consiste en la sustitución de una G por una A en la base nucleotídica 926 del exón 6. Utilizando esta técnica, se pueden identificar mutaciones adicionales que causan la MAPS.

De acuerdo con el procedimiento diagnóstico de la presente invención, se detecta la alteración en el locus *GYS1* de tipo natural. "Alteración de un gen de tipo natural" abarca todas las formas de mutaciones incluyendo deleciones, inserciones y mutaciones puntuales en las regiones codificantes y no codificantes. Las deleciones pueden ser del gen entero o sólo de una parte del gen. Las mutaciones puntuales pueden dar lugar a codones de parada, mutaciones por desplazamiento de marco o sustituciones de aminoácidos. Los acontecimientos mutacionales puntuales se producen en regiones reguladoras, tales como en el promotor del gen, lo que conduce a pérdida o disminución de la expresión del ARN. Las mutaciones puntuales también pueden abolir el procesamiento adecuado del ARN, dando lugar a la falta de expresión del producto del gen *GYS1* o disminuir la estabilidad del ARNm o la eficacia de la traducción. La MAPS es una enfermedad causada por una mutación puntual en la posición 926 del ácido nucleico. Los caballos predispuestos o que tienen MAPS sólo necesitan tener un alelo mutado.

Las técnicas diagnósticas que son útiles en los métodos de la invención incluyen, sin limitación, la secuenciación directa del ADN, el análisis PFGE, el estudio de oligonucleótidos específicos del alelo (ASO), la transferencia de puntos (dot blot) y la electroforesis en gel con gradiente desnaturalizante, son todas bien conocidas por el experto en la materia.

Existen varios métodos que se pueden usar para detectar una variación en la secuencia de ADN. La secuenciación directa del ADN, por secuenciación manual o por secuenciación fluorescente automatizada puede detectar la variación en la secuencia. Otro método es el ensayo de polimorfismo conformacional de cadena simple (SSCA). Este método no detecta todos los cambios de una secuencia, especialmente si el tamaño del fragmento de ADN es mayor de 200 pb, pero se puede optimizar para detectar la mayoría de las variaciones de la secuencia de ADN. Esta menor sensibilidad de la detección es una desventaja, pero el mayor rendimiento que se puede conseguir con el ensayo SSCA le convierte en una alternativa atractiva y viable para dirigir la secuenciación para la detección de la mutación en un contexto de búsqueda. Los fragmentos con movilidad desplazada en los geles de SSCA se secuencian a continuación para determinar la naturaleza exacta de la variación de la secuencia del ADN. Otros métodos basados en la detección de errores de apareamiento entre las dos hebras complementarias de ADN incluyen electroforesis en gel desnaturalizante constante (CDGE), análisis de heterodúplex (HA) y escisión química de errores de apareamiento (CMC). Una vez que se conoce la mutación, se puede utilizar un método de detección específica de un alelo, tal como la hibridación de oligonucleótidos específica del alelo (ASO) para cribar rápidamente y detectar la misma mutación en grandes cantidades de otras muestras. Dicha técnica puede utilizar sondas que se marcan con nanopartículas de oro para obtener un resultado visual con color.

La detección de las mutaciones puntuales se puede llevar a cabo por clonación molecular del alelo(s) *GYS1* y secuenciación del alelo(s) utilizando técnicas bien conocidas en la técnica. Otra alternativa consiste en amplificar directamente las secuencias génicas de una preparación de ADN genómico de tejido equino utilizando técnicas conocidas. A continuación se puede determinar la secuencia de ADN de las secuencias amplificadas.

Existen seis métodos bien conocidos para un análisis más completo, aunque todavía indirecto, para confirmar la presencia de un alelo mutante: 1) ensayo de polimorfismo conformacional de cadena simple (SSCA), 2) electroforesis en gel con gradiente desnaturalizante (DGGE), 3) ensayos de protección de ARNasa, 4) hibridación de oligonucleótidos específica del alelo (ASO), 5) el uso de proteínas que reconocen errores de apareamiento de nucleótidos, tales como la proteína mutS de *E. coli*. Para la PCR específica del alelo, se utilizan cebadores que hibridan en sus extremos 3' con una mutación de *GYS1* particular. Si la mutación particular no está presente, no se observa un producto de amplificación. También se puede usar el Sistema de amplificación resistente a mutaciones (ARMS). Las inserciones y deleciones también se pueden detectar mediante clonación, secuenciación y amplificación. Además, se pueden usar sondas de polimorfismos de longitud de fragmento de restricción (RFLP) para el gen o los genes marcadores circundantes para determinar la alteración de un alelo o una inserción en un fragmento polimórfico. Se pueden usar otras técnicas para detectar inserciones y deleciones conocidas en la técnica.

En los tres primeros métodos (SSCA, DGGE y ensayo de protección de la ARNasa), aparece una nueva banda electroforética. El SSCA detecta una banda que migra diferencialmente debido a que el cambio en la secuencia produce una diferencia en el apareamiento de bases intramolecular en la hebra sencilla. La protección de ARNasa implica la escisión del polinucleótido mutante en dos o más fragmentos más pequeños. La DGGE detecta diferencias en las velocidades de migración de las secuencias mutantes comparado con las secuencias de tipo silvestre, utilizando un gel de gradiente desnaturalizante. En un ensayo de oligonucleótidos específico de alelo, se diseña un oligonucleótido que detecta una secuencia específica y el ensayo se realiza detectando la presencia o ausencia de una señal de hibridación. En el ensayo mutS, la proteína se une sólo a secuencias que contienen un error de apareamiento de nucleótidos en un heterodúplex entre las secuencias mutantes y de tipo silvestre.

Los errores de apareamiento de acuerdo con la presente invención son dúplex de ácidos nucleicos hibridados en los cuales dos hebras no son 100% complementarias. La falta de homología total puede deberse a deleciones, inserciones, inversiones o sustituciones. La detección de los errores de apareamiento puede usarse para detectar mutaciones puntuales en el gen o en su producto ARNm. Aunque estas técnicas son menos sensibles que la secuenciación, son más sencillas de realizar en un número mayor de muestras. Un ejemplo de una técnica de escisión de un error de apareamiento es el procedimiento de protección de ARNasa. En la práctica de la presente invención, el procedimiento implica el uso de una ribosonda marcada que es complementaria con la secuencia codificante del gen *GYS1* de tipo silvestre de caballo. La ribosonda y el ARNm o el ADN aislados del tejido tumoral

se unen (hibridan) entre sí y posteriormente se digieren con la enzima ARNasa A que es capaz de detectar algunos errores de apareamiento en una estructura de ARN dúplex. Si la ARNasa A detecta un error de apareamiento, lo escinde en el sitio del error de apareamiento. Por lo tanto, cuando la preparación de ARN hibridada se separa en una matriz de electroforesis, si se ha detectado un error de apareamiento y este se ha escindido por la ARNasa A, se observará un producto ARN que es más pequeño que el ARN dúplex de longitud completa para la ribosonda y el ARNm o ADN. La ribosonda no tiene por qué tener la longitud completa del ARNm ni del gen GYS1, puede ser un segmento o ambos. Si la ribosonda comprende sólo un segmento del ARNm o del gen GYS1, sería deseable utilizar varias de estas sondas para cribar toda la secuencia de ARNm a la búsqueda de errores de apareamiento.

De modo similar, se pueden utilizar sondas de ADN para detectar errores de apareamiento mediante escisión enzimática o química. Otra alternativa es detectar los errores de apareamiento por los desplazamientos en la movilidad electroforética de los dúplex mal apareados respecto a los dúplex apareados. Tanto con ribosonda como con sondas de ADN, el ARNm o el ADN celular que pudiera contener una mutación, se puede amplificar utilizando la PCR antes de la hibridación.

En la presente invención también se puede utilizar el análisis de ácidos nucleicos mediante la tecnología de microchip.

Las secuencias de ADN del gen GYS1 que se han amplificado mediante el uso de la PCR también se pueden cribar utilizando sondas específicas. Estas sondas son oligómeros de ácidos nucleicos, cada una de las cuales contiene una región de la secuencia del gen GYS1 que alberga una mutación conocida. Por ejemplo, un oligómero puede tener aproximadamente 30 nucleótidos de longitud, lo que corresponde a una parte de la secuencia del gen GYS1. Mediante el uso de una batería de este tipo de sondas específicas del alelo, se pueden cribar los productos de amplificación de la PCR para identificar la presencia de una mutación previamente identificada en el gen GYS1. La hibridación de las sondas específicas del alelo con secuencias GYS1 amplificadas se puede realizar, por ejemplo, en un filtro de nylon. La hibridación con una sonda en particular en condiciones de hibridación estrictas indica la presencia de la misma mutación en los tejidos que en la sonda específica del alelo.

La alteración de la expresión del ARNm de GYS1 se puede detectar mediante cualquier técnica conocida en la técnica. Estas incluyen análisis de transferencia Northern, amplificación mediante PCR y protección de ARNasa. La disminución de la expresión del ARNm indica una alteración del gen GYS1 de tipo silvestre.

La alteración de los genes GYS1 de tipo silvestre también se puede detectar cribando para buscar la alteración de la proteína GYS1 de tipo silvestre, o una parte de la proteína GYS1. Por ejemplo, los anticuerpos monoclonales inmunorreactivos con GYS1 (o con una parte específica de la proteína GYS1) se pueden usar para cribar un tejido. La ausencia del antígeno cognato sería indicativo de una mutación. Los anticuerpos específicos de los productos de alelos mutantes también podrían usarse para detectar el producto del gen GYS1 mutante. Dichos ensayos inmunológicos se pueden realizar en cualquier formato conveniente conocido en la técnica. Estos incluyen transferencias Western, ensayos inmunohistoquímicos y ensayos ELISA. Se puede usar cualquier medio para la detección de una proteína GYS1 alterada para detectar la alteración de los genes GYS1 de tipo silvestre. Se pueden usar ensayos funcionales, tales como la determinación de la unión a proteínas. Además, se pueden usar ensayos que detecten la función bioquímica de GYS1. La detección de un producto génico GYS1 mutante indica la alteración de un gen GYS1 de tipo silvestre.

Los genes GYS1 o los productos génicos mutantes se pueden detectar en diversas muestras fisiológicas recogidas de un caballo. Ejemplos de muestras adecuadas incluyen una muestra celular, tal como una célula sanguínea, por ej., un linfocito, una célula de sangre periférica, una muestra recogida de la médula espinal, una muestra de tejido, tal como tejido cardíaco o tejido muscular, por ej., músculo cardíaco o esquelético, una muestra de un órgano, por ej., hígado o piel, una muestra de cabello, especialmente una muestra de cabello con raíces, una muestra de líquido, tal como la sangre.

Los procedimientos de diagnóstico de la presente invención son aplicables a cualquier enfermedad equina en la cual GYS1 desempeñe algún papel. El método diagnóstico de la presente invención es útil, por ejemplo, para veterinarios, asociaciones de criadores, o criadores individuales, para que puedan decidir un ciclo apropiado de tratamiento y/o para determinar si un animal es un candidato adecuado como yegua para cría o semental.

### **Sondas oligonucleotídicas**

Como se ha señalado anteriormente, el procedimiento de la presente invención es útil para detectar la presencia de un polimorfismo en ADN equino, en particular, la presencia de una sustitución de un nucleótido G por A en la posición 926 del exón 6 de la secuencia de GYS1 equina codificante (SEC ID NO:1). Esta sustitución tiene como resultado la sustitución de un aminoácido arginina (R) en el codón 309 por una histidina (H) en la proteína glucógeno sintasa (SEC ID NO:9).

Los pares de cebadores son útiles para la determinación de una secuencia de nucleótidos de un alelo GYS1 particular utilizando PCR. Los pares de cebadores de ADN monocatenario se pueden hibridar con secuencias dentro del gen GYS1 o en sus alrededores con el fin de cebar la amplificación de la síntesis de ADN del propio gen GYS1. Un conjunto completo de estos cebadores permite la síntesis de todos los nucleótidos de las secuencias codificantes

de *GYS1*, es decir, los exones. El conjunto de cebadores permite preferiblemente la síntesis tanto de secuencias de intrón como de exón. También se pueden utilizar cebadores específicos del alelo. Dichos cebadores se hibridan sólo con alelos mutantes de *GYS1* particulares y, por lo tanto, sólo amplificarán un producto en presencia del alelo mutante como molde.

5 El primer paso del proceso implica la puesta en contacto de una muestra fisiológica obtenida de un caballo, donde esta muestra contiene ácido nucleico con una sonda oligonucleotídica para formar un ADN hibridado. Las sondas oligonucleotídicas que son útiles en los métodos de la presente invención pueden ser cualquier sonda comprendida entre aproximadamente 4 ó 6 bases hasta aproximadamente 80 ó 100 bases o más. En una realización de la presente invención, las sondas son de entre aproximadamente 10 y aproximadamente 20 bases.

10 Los propios cebadores se pueden sintetizar utilizando técnicas que son bien conocidas en la técnica. Generalmente, los cebadores se pueden obtener utilizando sintetizadores de oligonucleótidos comerciales. Dada la secuencia de la secuencia codificante expuesta en SEC ID NO:1, el diseño de cebadores particulares es bien conocido por el experto en la materia.

Las sondas oligonucleotídicas se pueden preparar teniendo una amplia variedad de secuencias de bases de acuerdo con las técnicas bien conocidas en la técnica. Bases adecuadas para preparar la sonda oligonucleotídica se pueden seleccionar de bases de nucleótidos naturales tales como la adenina, citosina, guanina, uracilo y timina y de bases de nucleótidos no naturales o "sintéticas", tales como 7-desazaguanina, 8-oxo-guanina, 6-mercaptop Guanina, 4-acetilcitidina, 5-(carboxihidroxietil)uridina, 2'-O-metilcitidina, 5-carboximetilaminometil-2-tioridina, 5-carboximetilaminometiluridina, dihidouridina, 2'-O-metilpseudouridina, p,D-galactosilqueosina, 2'-O-metilguanosina, inosina, N6-isopenteniladenosina, 1-metiladenosina, 1-metilpseudouridina, 1-metilguanosina, 1-metilinosina, 2,2-dimetilguanosina, 2-metiladenosina, 2-metilguanosina, 3-metilcitidina, 5-metilcitidina, N6-metiladenosina, 7-metilguanosina, 5-metilaminometiluridina, 5-metoxiaminometil-2-tiouridina, p,D-manosilqueosina, 5-metiloxicarbonilmetiluridina, 5-metoxiuridina, 2-metil-N6-isopenteniladenosina, N-((9-P-D-ribofuranosil-2-metilpurina-6-il)carbamoil)treonina, N-((9-P-D-ribofuranosilpurina-6-il)N-metil-carbamoil)treonina, éster metílico del ácido uridina-5-oxiacético, ácido uridina-5-oxiacético, wybutoxosina, pseudouridina, queosina, 2-tiocitidina, 5-metil-2-tiouridina, 2-tiouridina, 5-Metiluridina, N-((9-beta-D-ribofuranosilpurina-6-il)carbamoil)treonina, 2'-O-metil-5-metiluridina, 2'-O-metiluridina, wybutosina y 3-(3-amino-3-carboxipropil)uridina. Se puede emplear cualquier estructura de oligonucleótido, incluyendo ADN, ARN (aunque el ARN es menos preferido que el ADN), azúcares modificados, tales como carbociclos, y azúcares que contienen sustituciones en 2', tales como fluoro o metoxi. Los oligonucleótidos pueden ser oligonucleótidos en los que al menos uno, o todos, los internucleótidos que unen los residuos fosfato son fosfatos modificados, tales como metil-fosfonatos, metil-fosfonotiofosfatos, fosforoinorfolidatos, fosforopiperazidatos y fosforamidatos (por ej., se pueden modificar uno sí y otro no de los internucleótidos que unen residuos fosfato). El oligonucleótido puede ser un "ácido nucleico péptido" como se describe en Nielsen et al.; Science, 254, 1497-1500 (1991).

35 El único requisito es que la sonda oligonucleotídica posea una secuencia, al menos una parte de la cual sea capaz de unirse a una parte conocida de la secuencia de la muestra de ADN.

En algunas aplicaciones puede ser deseable poner en contacto la muestra de ADN con diferentes sondas oligonucleotídicas que tienen diferentes secuencias de bases (por ej., donde hay dos o más ácidos nucleicos diana en la muestra, o donde un único ácido nucleico diana hibrida con dos o más muestras en un ensayo "sándwich").

40 Las muestras de ácido nucleico proporcionadas por la presente invención son útiles para diferentes fines. Las sondas se pueden usar para detectar los productos de amplificación de PCR. También se pueden usar para detectar errores de apareamiento con el gen o el ARNm de *GYS1* utilizando otras técnicas.

### **Metodología de hibridación**

45 La muestra de ADN (o ácido nucleico) se puede poner en contacto con la sonda oligonucleotídica mediante cualquiera de los métodos conocidos por el experto en la materia. Por ejemplo, la muestra de ADN se puede solubilizar en solución y poner en contacto con la sonda oligonucleotídica solubilizando la sonda oligonucleotídica en solución con la muestra de ADN en condiciones que permitan la hibridación. Las condiciones adecuadas son bien conocidas por los expertos en la materia. Otra alternativa es solubilizar la muestra de ADN en solución con la sonda oligonucleotídica inmovilizada en un soporte sólido, donde la muestra de ADN se puede poner en contacto con la sonda oligonucleotídica sumergiendo el soporte sólido que tiene la sonda oligonucleotídica inmovilizada en el mismo en la solución que contiene la muestra de ADN.

### **Ejemplo 1: Método de detección de una mutación del ADN asociada a la miopatía equina por almacenamiento de polisacáridos**

55 La presente invención se refiere a mutaciones en el gen *GYS1* y a su uso en el diagnóstico de la MAPS y a la detección de un alelo *GYS1* mutante en un caballo.

Los presentes inventores han descubierto una mutación en el gen *GYS1* equino (que codifica para la enzima glucógeno sintasa del músculo esquelético) que está presente en muchas poblaciones de caballos afectados de

MAPS estudiados hasta la fecha. Esto fue posible derivando primero las secuencias de ADN codificantes de la proteína del gen *GYS1* equino del ARNm aislado de músculo esquelético tanto de un caballo afectado como de un caballo control. En ambos caballos, la longitud de la secuencia desde el codón de inicio (ATG) hasta el codón de parada (TAA) fue de 2.124 bases (Figura 1) y codificaría para una proteína de 737 aminoácidos. La única diferencia entre las secuencias de un caballo con MAPS y de un caballo control era una sustitución de una G por una A en el exón 6 en el nucleótido de la posición 926.

La diferencia en la secuencia del ADN en la posición 926 de la secuencia codificante de *GYS1* presente en el ARNm de músculo esquelético se confirmó posteriormente en el ADN genómico de varios caballos. En la Figura 2 se muestra una vista ampliada del exón 6 con su secuencia intrónica flanqueante del ADN genómico. La Figura 2 también muestra que el cambio de una G por una A en la secuencia de ADN provoca la sustitución de un aminoácido arginina (R) en el codón 309 por una histidina (H) en la proteína glucógeno sintasa. Por lo tanto, se hace referencia a esta mutación como la mutación ADN G926 por A926 o la mutación del aminoácido R309 por H309. Los alelos normales de este gen se pueden denominar como G926, R o R309 y los alelos mutantes como A926, H o H309.

Hasta la fecha no se ha observado ninguna otra mutación en el gen *GYS1* que cause la enfermedad por almacenamiento de glucógeno en seres humanos o en otras especies animales. El gen relacionado *GYS2* que codifica para la forma hepática de la glucógeno sintasa que se expresa en tejidos que no son el músculo, tiene varias mutaciones conocidas que conducen a la deficiencia de esta enzima y la hipoglucemia en ayunas. Sin embargo, al contrario de lo que sucede con las mutaciones de *GYS2* que reducen en gran medida la actividad de la enzima glucógeno sintasa y que tienen herencia recesiva, la mutación de *GYS1* muscular en un caballo con MAPS no reduce la actividad de la glucógeno sintasa, sino que parece dar lugar a un aumento de la actividad de la glucógeno sintasa y tiene herencia dominante (véase la Tabla 1 siguiente). Esta región de la secuencia de aminoácidos de la glucógeno sintasa muscular contenida en el exón 6 está muy conservada en todo el reino animal, lo que respalda que su mutación en los caballos con MAPS es una mutación causativa (Figura 3).

Los inventores han descubierto la mutación *GYS1* R por H en caballos Quarter Horses, Draft horses y Warmbloods afectados de MAPS (Tabla 1) y es probable que también haya que ampliar este descubrimiento a más razas de caballos. Aproximadamente el 80% de los Quarter Horses y Belgian Draft Horses fueron diagnosticados de MAPS mediante el método de biopsia de músculo y son homocigotos (tienen dos copias del alelo H; H/H) o son heterocigotos (un alelo H y un alelo R; R/H). Los caballos con MAPS con el alelo H pueden ser de cualquiera de los dos sexos, lo cual es consistente, aunque no determinante de una herencia autosómica dominante. Sólo el 4% de los Quarter Horses y el 14% de los Belgian Draft con resultados negativos en la biopsia eran heterocigotos. Los inventores creen que esto en parte refleja menos del 100% de exactitud del método diagnóstico actual, aunque podría reflejar también una penetrancia incompleta, es decir, los portadores del alelo H no siempre van a desarrollar síntomas de la enfermedad debido a otros factores genéticos y ambientales.

**Tabla 1. Frecuencias genotípicas de *GYS1* en la MAPS y caballos control de diferentes razas**

Genotipo	PSSM QH	Control QH	PSSM Belgian	Control Belgians	MAPS Warmblood	Control Warmblood
R/R	18	85	4	29	1	4
R/H	67	4	28	5	3	0
H/H	4	0	4	0	0	0

El que aproximadamente el 20% de los Quarter Horses y el 11% de los Belgians con polisacáridos anormales en las biopsias de músculo y signos clínicos de MAPS no sean portadores del alelo H de *GYS1*, sugiere que podrían existir otras causas para la MAPS. En otras palabras, la mutación de *GYS1* parece explicar la mayoría, aunque no todos los casos, de MAPS equina, y probablemente existe otro gen responsable de una forma no *GYS-1* de MAPS que requiera ser objeto de una investigación adicional.

Los investigadores han determinado la frecuencia genotípica de *GYS1* en poblaciones de caballos al azar obtenida de muestras remitidas para cumplir con los requisitos de registro de la raza. Se tomaron muestras de la raíz pilosa de una de cada 10 muestras enviadas para garantizar una distribución uniforme en todo el territorio de los EEUU. La Tabla 1 indica que la mutación de *GYS1* está muy representada en cuatro razas importantes examinadas hasta la fecha, aunque no en los pura sangre. La distribución del genotipo de *GYS1* en los Quarter Horses y en los Paint Horses es similar en el 6-7% de los heterocigotos, con pocos homocigotos para el alelo H. Sin embargo, aproximadamente el 42% de los Percheron son heterocigotos y el 14% son homocigotos para el alelo H. Puesto que el alelo *GYS1* H parece ser dominante, prevemos que aproximadamente el 7% de todos los Quarter Horses y Paint horses, 36% de todos los Belgians y el 56% de todos los Percheron son realmente genéticamente susceptibles a la MAPS.

**Tabla 2. Frecuencias genotípicas de GYS1 en poblaciones al azar de diferentes razas**

Genotipo	Quarter Horses	Paint Horses	Belgian	Percheron	Pura sangre
R/R	313 (93%)	180 (92%)	20 (61%)	22 (44%)	96 (100%)
R/H	21 (6%)	14 (7%)	13 (26%)	21 (42%)	0 (0%)
H/H	1 (<1%)	1 (<1%)	5 (10%)	7 (14%)	0 (0%)

5 La secuencia de ADN casi completa del gen *GYS1* del caballo (*GYS1* de caballo intrón 5, exón 6 e intrón 6, Figura 4) se ensambló a partir de las secuencias depositadas en el archivo de trazas del NCBI en el centro de secuenciación del Broad Institute durante su proyecto reciente para la secuenciación aleatoria del genoma equino completo (SEC ID NO:6). A continuación, se predijeron los intrones y los exones de la secuencia del gen *GYS1* de caballo a partir de las secuencias del exón del *GYS1* homólogo de otros mamíferos. El intrón 5 de esta secuencia comprende las bases 1 a 471. El exón 6 de esta secuencia está resaltado y comprende las bases 472 a 589. El intrón 6 de esta secuencia comprende las bases 590 a 886. La mutación G por A en el exón 6 que causa la mutación del aminoácido R por H en el codón 309 está subrayada y es en la base 574.

10 Utilizando la secuencia de *GYS1*, se desarrollan cebadores para PCR que pueden amplificar la mutación de *GYS1* en la MAPS. Por ejemplo, un par de cebador para PCR que se ha usado con éxito y de forma fiable para amplificar esta región de las muestras de ADN de caballo aisladas está en los intrones 5 y 6 y las localizaciones en la secuencia también se han subrayado (Figura 4). Estas secuencias son 5'-TGAAACAT- GGGACCTTCTCC-3' (SEC ID NO:7) y 5'-AGCTGTCCCCTCCCTTAGAC-3' (SEC ID NO:8). Son posibles muchos otros pares de cebadores.

15 Utilizando los cebadores para PCR anteriores para amplificar la región, se puede obtener el genotipo de cualquier caballo (G/G, G/A o A/A para la secuencia de ADN y R/R, R/H y H/H para la secuencia de aminoácidos). En este procedimiento, la enzima de restricción *HypCH4V* corta el alelo H de *GYS1* en el sitio del exón (base 574), así como en un sitio intrónico a 100 pb de distancia existente tanto en el alelo R como H que sirve para controlar la eficiencia de la enzima. Los productos se separan mediante electroforesis en gel de agarosa y se visualizan por tinción con bromuro de etidio bajo luz ultravioleta. Son posibles muchos otros métodos de detección del nucleótido G o A en esta posición de la secuencia *GYS1* equina.

20 El análisis de ADN basado en la presente invención proporciona ahora a los veterinarios y a los patólogos veterinarios un medio para determinar con mayor exactitud si un caballo con signos clínicos de MAPS tiene la forma heredable y más frecuente de la enfermedad que se puede atribuir específicamente a esta mutación en el gen *GYS1*. Todo lo que se necesita es una muestra de tejido que contenga el ADN del individuo (generalmente raíz pilosa o sangre) y una tecnología PCR y para el análisis de la secuencia adecuadas para detectar el cambio del nucleótido único G por A. Dichos cebadores para PCR están basados en el exón 6 y en sus secuencias intrónicas flanqueantes, como se muestra en la Figura 2, secuencias cercanas a esta región mostradas en la Figura 1 o en otra secuencia de ADN de los intrones de este gen.

25 Asimismo, el análisis de ADN ofrece a los propietarios y criadores un medio para determinar si se puede esperar que cualquier caballo tenga descendencia con esta forma de MAPS. Un caballo H/H produciría un 100% de potros afectados, mientras que un caballo H/R produciría un 50% de potros afectados cuando se cruzase con un caballo R/R. El cruce de caballos H/H y H/R producirían un 75% de potros afectados. Los programas de cría podrían incorporar esta información en la selección de parentales que eventualmente podría reducir e incluso eliminar esta forma de MAPS en sus rebaños.

#### LISTADO DE SECUENCIAS

<110> CONSEJO RECTOR DE LA UNIVERSIDAD DE MINNESOTA

40 <120> PROCEDIMIENTO PARA LA DETECCIÓN DE LA MIOPATÍA EQUINA POR ALMACENAMIENTO DE POLISACÁRIDOS

<130> 09531.276WO1

<140>

<141>

<160>9

45 <170> PatentIn Ver. 3.3

<210> 1

<211> 2214

<212> ADN  
<213> Equus caballus

<400> 1

```

atgcctctaa accgcacttt gtecatgtcc tcaactgccag gactggacga ttgggaggat 60
gaattcgacc tggagaatac agtgcctctc gaggtggcct gggaggtggc caacaagggtg 120
gggtggcatct acacgggtgct acagacgaag gcgaagggtga caggggatga atggggcgac 180
aactactacc tgggtgggacc atacacggag caaggcgtga ggaccaggt ggagctgctc 240
gagcccccaa ccccgccctt gaagagga cg ctggactcca tgaacagcaa gggctgcaag 300
gtgtatttcg ggcgctggct gatcgagggg ggccccctgg ttggtgctcct ggatgtgggg 360
gcctcagcct gggccctgga acgctggaag ggagagcttt gggacacctg caacatcggg 420
gtgccctggg acgaccgtga ggccaacgac gccgtccttt ttggcttctt caccacctgg 480
ttcctgggtg agttcctggc ccagagcggag gagaagccac atgtggttgc acacttccac 540
gagtggttgg cgggcatcgg gctctgcctg tgccgtgccc ggcggtgccc tgtggctaca 600
atcttcacca cccacgccac gctgctgggg cgatacctgt gtgccggtgc tgtggacttc 660
tacaacaacc tggagaattt caacgtggac aaggaagctg gtgagaggca aatttatcac 720
cgttactgca tggagcgggc ggcagcccaac tgcaactcac cgtgtcccag 780
atcacgcgca ttgaggctca gcacctactc aagaggaaac cagatatcgt gacccccaat 840
ggactgaatg tgaagaagtt ctctgccatg catgagttcc agaacctcca tgctcagagc 900
aaggccccga tccaggagtt tgtgcgtggc cttttttatg ggcacctgga ctccaacttg 960
gataagacc ttgtatttctt tatcgccggc cgctacgagt tctccaacaa gggggctgac 1020
gtcttcttgg aggccttggc ccggctcaac tatctgctca gagttaatgg cagcagcag 1080
acggtggctg ccttcttcat catgccggct cggaccaaca acttcaactg ggaaacctc 1140
aaggggcaag ccgtgcgcaa gcagctctgg gatacggcga acacagtga ggagaagttc 1200
gggaggaagc tttacgaatc cctgctggtt gggagcctcc cggacatgaa caagatgctg 1260
gacaaggagg atttcactat gatgaagaga gccatctttg ccacgcagcg gcagctcttt 1320
ccccctgtgt gcaccacaa tatgctggac gactcctcgg acctatcct gaccaccatc 1380
cgtcgaatcg gcctcttcaa tagtagtgct gacagggtea aggtgatttt ccaccagag 1440
ttcctctcct ccacgagccc cctgctcccc gtggactatg aggagtthgt ccgtggctgc 1500
caccttgggg ttttcccctc ctactatgag ccttggggct acacaccagc tgagtgcacg 1560
gttatgggca tccccagtat ctccaccaac ctctcggct tgggtgctt catggaggaa 1620
cacatcgca accectcagc ttacggcacc tacattctgg accggcgggt ccgcagcctg 1680
gatgatcct gctcgcagct tacctccttc ctctacagct tctgccagca ggcggcggg 1740
cagcgcacca tccagcggaa ccgcacggag cgcctctccg acctcttggg ctggaaatac 1800
ctaggccggg actatatgtc cgcgcgccac atggcgtgg ccaaggcctt tccagaaatc 1860
ttcaactacg agccccgcga ggtgatgcy acccagggct accgctacc acggcctgca 1920
tcgggtgcct cgtgcacctc actgtcacga cactcgagcc cgcaccagag cgaggacgag 1980
gaggagcccc gggacgtgccc gcccgatgaa gacagtgagc gctacgacga ggcagaggag 2040
gccgccaagg accggcgcaa catccgcgcc ccggagtggt cgcgtcgcgc ctctgacac 2100
tcttccacga gcgggagcaa gcgcggctcg gtggacacgg ggcctccag ctgctcagc 2160

```

```

tcttccacga gcgggagcaa gcgcggctcg gtggacacgg ggcctccag ctgctcagc 2160
acccccagcg agccccctcag ccccgccagc tccctggggc aggagcgcga ctaa 2214

```

5 <210>2  
<211> 231  
<212> ADN  
<213> Equus caballus

<400> 2

```

ttgaaacatg gggccttctc ccccatgcct agatatcgtg accccaatg gactgaatgt 60
gaagaagttc tctgccatgc atgagttcca gaacctccat gctcagagca agggccgaat 120
ccaggagttt gtgcgtggcc atttttatgg gtatgtgggc cagataccca ggtcttgaga 180
gaggtggggg ttgggtgccc agactcccgg gtctaagggg gggacagcta a 231

```

10

<210>3  
<211> 231  
<212> ADN  
<213> Equus caballus

15 <400> 3

ES 2 382 827 T3

ttgaaacatg gggccttctc ccccatgcct agatatcgtg accccaatg gactgaatgt 60  
 gaagaagttc tctgccatgc atgagttcca gaacctccat gctcagagca aggcccgaat 120  
 ccaggagttt gtgcatggcc atttttatgg gtatgtgggc cagataccca ggtcttgaga 180  
 gaggtggggg ttgggtgccc agactcccgg gtctaagggg gggacagcta a 231

<210>4

<211> 39

<212> PRT

5 <213> Equus caballus

<400> 4

Ile Val Thr Pro Asn Gly Leu Asn Val Lys Lys Phe Ser Ala Met His  
 1 5 10 15  
 Glu Phe Gln Asn Leu His Ala Gln Ser Lys Arg Arg Ile Gln Glu Phe  
 20 25 30  
 Val His Gly His Phe Tyr Gly  
 35

<210>5

<211> 39

10 <212> PRT

<213> Organismo desconocido

<220>

<223> Descripción del organismo desconocido: secuencia de aminoácidos de GYS desconocida de diferentes especies

15 <400> 5

Ile Val Thr Pro Asn Gly Leu Asn Val Lys Lys Phe Ser Ala Met His  
 1 5 10 15  
 Val Arg Gly His Phe Tyr Gly  
 35

<210>6

<211>886

<212> ADN.

20 <213> Equus caballus

<400> 6

gtggggcata ggagggggtc ccgagacatg aatgaggtag ttgacggctc acgccaaggc 60  
 cegtaccaat tttcttattg taaagagcag agaactacaa ctcccagaag accctaagtt 120  
 gatggcctgc atatgaggct gccgggctgt accctagagg ctaatgggag gtgtagtctc 180  
 tttgacattc atttcagaaa agggaaaatt attttattgg ctaaggtgga tattggtctt 240  
 ctgaaagagg atggaggttc actggatttt gtttttgaga aaagcagtgt tagaaccagt 300  
 tattcttccc aggctatagc atcaattctc aatggatgtg agaactacaa ctcccagaag 360  
 gcctctggct ggccagctcc ttaatattcc tgggaattacc ttcatttatt cattcaacca 420  
 cattgatctg aatgaagaag tgaacatgg gaccttctcc occatgccta gatatacgtga 480  
 ccccgaatgg actgaatgtg aagaagttct ctgccatgca tgagttccag aacctccatg 540  
 ctcagagcaa ggcccgaatc caggagtctg tgcgtggcca tttttatggg tatgtgggcc 600  
 agatacccag gtcttgagag aggtgggggt tgggtgcccc gactcccggg tctaagggag 660  
 gggacagctg ggggctcaga ctcttgagtt cctgagcgtc agaacaatgg ctgagtattg 720  
 caggagatcc ccgtttaagg agaaaagagg cacaagtccc atgagtccca tgatgcctct 780  
 gtactgtcta agatgttctt tgcccctgtg gcttctgggg tttgtagttt tgagccaggg 840  
 ctagaggggtg aggatcttgg cttcaccctg tcttgtggct ttttag 886

<210>7  
<211> 20  
<212> ADN  
<213> Secuencia artificial

5 <220>  
<223> Descripción de la secuencia artificial: Cebador sintético

<400> 7  
tgaaacatgg gaccttctcc 20

10 <210>8  
<211> 20  
<212> ADN  
<213> Secuencia artificial

<220>  
<223> Descripción de la secuencia artificial: Cebador sintético

15 <400> 8  
agctgtcccc tcccttagac 20

<210>9  
<211>737  
<212> PRT  
20 <213> Equus caballus

<400> 9

ES 2 382 827 T3

Met Pro Leu Asn Arg Thr Leu Ser Met Ser Ser Leu Pro Gly Leu Asp  
1 5 10 15  
Asp Trp Glu Asp Glu Phe Asp Leu Glu Asn Thr Val Leu Phe Glu Val  
20 25 30  
Ala Trp Glu Val Ala Asn Lys Val Gly Gly Ile Tyr Thr Val Leu Gln  
35 40 45  
Thr Lys Ala Lys Val Thr Gly Asp Glu Trp Gly Asp Asn Tyr Tyr Leu  
50 55 60  
Val Gly Pro Tyr Thr Glu Gln Gly Val Arg Thr Gln Val Glu Leu Leu  
65 70 75 80  
Glu Pro Pro Thr Pro Ala Leu Lys Arg Thr Leu Asp Ser Met Asn Ser  
85 90 95  
Lys Gly Cys Lys Val Tyr Phe Gly Arg Trp Leu Ile Glu Gly Gly Pro  
100 105 110  
Leu Val Val Leu Leu Asp Val Gly Ala Ser Ala Trp Ala Leu Glu Arg  
115 120 125  
Trp Lys Gly Glu Leu Trp Asp Thr Cys Asn Ile Gly Val Pro Trp Tyr  
130 135 140  
Asp Arg Glu Ala Asn Asp Ala Val Leu Phe Gly Phe Leu Thr Thr Trp  
145 150 155 160  
Phe Leu Gly Glu Phe Leu Ala Gln Ser Glu Glu Lys Pro His Val Val  
165 170 175  
Ala His Phe His Glu Trp Leu Ala Gly Ile Gly Leu Cys Leu Cys Arg  
180 185 190  
Ala Arg Arg Leu Pro Val Ala Thr Ile Phe Thr Thr His Ala Thr Leu  
195 200 205  
Leu Gly Arg Tyr Leu Cys Ala Gly Ala Val Asp Phe Tyr Asn Asn Leu  
210 215 220  
Glu Asn Phe Asn Val Asp Lys Glu Ala Gly Glu Arg Gln Ile Tyr His  
225 230 235 240  
Arg Tyr Cys Met Glu Arg Ala Ala Ala His Cys Thr His Val Phe Thr  
245 250 255  
Thr Val Ser Gln Ile Thr Ala Ile Glu Ala Gln His Leu Leu Lys Arg  
260 265 270  
Lys Pro Asp Ile Val Thr Pro Asn Gly Leu Asn Val Lys Lys Phe Ser  
275 280 285  
Ala Met His Glu Phe Gln Asn Leu His Ala Gln Ser Lys Ala Arg Ile  
290 295 300

ES 2 382 827 T3

Gln Glu Phe Val Arg Gly His Phe Tyr Gly His Leu Asp Phe Asn Leu  
 305 310 315 320  
 Asp Lys Thr Leu Tyr Phe Phe Ile Ala Gly Arg Tyr Glu Phe Ser Asn  
 325 330 335  
 Lys Gly Ala Asp Val Phe Leu Glu Ala Leu Ala Arg Leu Asn Tyr Leu  
 340 345 350  
 Leu Arg Val Asn Gly Ser Glu Gln Thr Val Val Ala Phe Phe Ile Met  
 355 360 365  
 Pro Ala Arg Thr Asn Asn Phe Asn Val Glu Thr Leu Lys Gly Gln Ala  
 370 375 380  
 Val Arg Lys Gln Leu Trp Asp Thr Ala Asn Thr Val Lys Glu Lys Phe  
 385 390 395 400  
 Gly Arg Lys Leu Tyr Glu Ser Leu Leu Val Gly Ser Leu Pro Asp Met  
 405 410 415  
 Asn Lys Met Leu Asp Lys Glu Asp Phe Thr Met Met Lys Arg Ala Ile  
 420 425 430  
 Phe Ala Thr Gln Arg Gln Ser Phe Pro Pro Val Cys Thr His Asn Met  
 435 440 445  
 Leu Asp Asp Ser Ser Asp Pro Ile Leu Thr Thr Ile Arg Arg Ile Gly  
 450 455 460  
 Leu Phe Asn Ser Ser Ala Asp Arg Val Lys Val Ile Phe His Pro Glu  
 465 470 475 480  
 Phe Leu Ser Ser Thr Ser Pro Leu Leu Pro Val Asp Tyr Glu Glu Phe  
 485 490 495  
 Val Arg Gly Cys His Leu Gly Val Phe Pro Ser Tyr Tyr Glu Pro Trp  
 500 505 510  
 Gly Tyr Thr Pro Ala Glu Cys Thr Val Met Gly Ile Pro Ser Ile Ser  
 515 520 525  
 Thr Asn Leu Ser Gly Phe Gly Cys Phe Met Glu Glu His Ile Ala Asp  
 530 535 540  
 Pro Ser Ala Tyr Gly Ile Tyr Ile Leu Asp Arg Arg Phe Arg Ser Leu  
 545 550 555 560  
 Asp Asp Ser Cys Ser Gln Leu Thr Ser Phe Leu Tyr Ser Phe Cys Gln  
 565 570 575  
 Gln Ser Arg Arg Gln Arg Ile Ile Gln Arg Asn Arg Thr Glu Arg Leu  
 580 585 590  
 Ser Asp Leu Leu Asp Trp Lys Tyr Leu Gly Arg Tyr Tyr Met Ser Ala  
 595 600 605

ES 2 382 827 T3

Arg His Met Ala Leu Ala Lys Ala Phe Pro Glu His Phe Thr Tyr Glu  
 610 615 620  
 Pro Arg Glu Ala Asp Ala Thr Gln Gly Tyr Arg Tyr Pro Arg Pro Ala  
 625 630 635 640  
 Ser Val Pro Pro Ser Pro Ser Leu Ser Arg His Ser Ser Pro His Gln  
 645 650 655  
 Ser Glu Asp Glu Glu Glu Pro Arg Asp Val Pro Pro Asp Glu Asp Ser  
 660 665 670  
 Glu Arg Tyr Asp Glu Asp Glu Glu Ala Ala Lys Asp Arg Arg Asn Ile  
 675 680 685  
 Arg Ala Pro Glu Trp Pro Arg Arg Ala Ser Cys Thr Ser Ser Thr Ser  
 690 695 700  
 Gly Ser Lys Arg Gly Ser Val Asp Thr Gly Pro Ser Ser Ser Leu Ser  
 705 710 715  
 Thr Pro Ser Glu Pro Leu Ser Pro Ala Ser Ser Leu Gly Glu Glu Arg  
 725 730 735  
 Asn

**REIVINDICACIONES**

- 5 1. Un procedimiento para la detección de la presencia de la miopatía por almacenamiento de polisacáridos (MAPS) en un caballo, que comprende la identificación en una muestra de ácido nucleico de un nucleótido 926 de caballo de la SEC ID NO: 1, en el que la presencia de un nucleótido adenina (A) en el nucleótido 926 en uno o en ambos alelos es indicativa de que el caballo tiene predisposición a padecer o padece MAPS.
2. El procedimiento de la reivindicación 1, que comprende además la puesta en contacto de la muestra con al menos una sonda oligonucleotídica para formar un ácido nucleico hibridado y la amplificación del ácido nucleico hibridado.
3. El procedimiento de la reivindicación 2, en el que se amplifica el exón 6 de la enzima glucógeno sintasa 1 o una parte de la misma.
- 10 4. El procedimiento de la reivindicación 2, en el que la amplificación del ácido nucleico hibridado se lleva a cabo mediante la reacción en cadena de la polimerasa, la amplificación por desplazamiento de mella, la reacción en cadena de la ligasa o la amplificación basada en la secuencia del ácido nucleico.
5. El procedimiento de acuerdo con la reivindicación 2, en el que al menos una sonda oligonucleotídica se inmoviliza sobre una superficie sólida.
- 15 6. Un procedimiento para la detección de un alelo de la miopatía por almacenamiento de polisacáridos (MAPS) en un caballo, que comprende la identificación en una muestra de ácido nucleico del nucleótido 926 de caballo de la SEC ID NO: 1, en el que la presencia de un nucleótido adenina (A) en el nucleótido 926 en uno o en ambos alelos es indicativa de que el caballo tiene predisposición a padecer o padece MAPS.
- 20 7. El procedimiento de la reivindicación 6, que comprende además la puesta en contacto de la muestra con al menos una sonda oligonucleotídica para formar un ácido nucleico hibridado y la amplificación del ácido nucleico hibridado.
8. El procedimiento de la reivindicación 7, en el que se amplifica el exón 6 de la enzima glucógeno sintasa 1 o una parte de la misma.
9. El procedimiento de la reivindicación 7, en el que la amplificación del ácido nucleico hibridado se lleva a cabo mediante la reacción en cadena de la polimerasa, la amplificación por desplazamiento de mella, la reacción en cadena de la ligasa o la amplificación basada en la secuencia del ácido nucleico.
- 25 10. El procedimiento de acuerdo con la reivindicación 7, en el que al menos una sonda oligonucleotídica se inmoviliza sobre una superficie sólida.
11. Un procedimiento para la detección de la presencia de un biomarcador asociado a la miopatía equina por almacenamiento de polisacáridos (MAPS) , que comprende la determinación de la presencia del biomarcador en una muestra fisiológica de un caballo, en el que la muestra comprende ácido nucleico y en el que el biomarcador representa un polimorfismo/mutación en el gen *GYS1* equino codificante identificado por genotipificación.
- 30 12. El procedimiento de la reivindicación 11, que comprende además la puesta en contacto de la muestra con al menos una sonda oligonucleotídica para formar un ácido nucleico hibridado y la amplificación del ácido nucleico hibridado.
- 35 13. El procedimiento de la reivindicación 12, en el que se amplifica el exón 6 de la enzima glucógeno sintasa 1 (*GYS1*) o una parte de la misma.
14. El procedimiento de la reivindicación 12, en el que la amplificación del ácido nucleico hibridado se lleva a cabo mediante la reacción en cadena de la polimerasa, la amplificación por desplazamiento de mella, la reacción en cadena de la ligasa o la amplificación basada en la secuencia del ácido nucleico.
- 40 15. El procedimiento de acuerdo con la reivindicación 12, en el que al menos una sonda oligonucleotídica se inmoviliza sobre una superficie sólida.
16. El procedimiento de la reivindicación 11, en el que el biomarcador comprende un gen *GYS1* equino que tiene una A en el nucleótido 926 del exón 6.
- 45 17. El procedimiento de la reivindicación 11, en el que el gen *GYS1* codifica para una enzima que tiene un H en residuo aminoácido 309.
18. Un procedimiento para el diagnóstico de la miopatía por almacenamiento de polisacáridos (MAPS) en un caballo, que comprende la detección de la presencia de un biomarcador en una muestra fisiológica del caballo, en el que la presencia del biomarcador es indicativa de la enfermedad y el biomarcador representa un polimorfismo/mutación en el gen *GYS1* equino codificante identificado por genotipificación.
- 50 19. El procedimiento de la reivindicación 18, en el que la muestra comprende ácido nucleico.

20. El procedimiento de la reivindicación 18, que comprende además la puesta en contacto de la muestra con al menos una sonda oligonucleotídica para formar un ácido nucleico hibridado y la amplificación del ácido nucleico hibridado.
- 5 21. El procedimiento de la reivindicación 20, en el que se amplifica el exón 6 de la enzima glucógeno sintasa 1 (*GYS1*) o una parte de la misma.
22. El procedimiento de la reivindicación 21, en el que la amplificación del ADN hibridado se lleva a cabo mediante la reacción en cadena de la polimerasa, la amplificación por desplazamiento de mella, la reacción en cadena de la ligasa o la amplificación basada en la secuencia del ácido nucleico.
- 10 23. El procedimiento de la reivindicación 18, en el que el biomarcador comprende un gen *GYS1* equino que tiene una A en el nucleótido 926 del exón 6.
24. El procedimiento de la reivindicación 23, en el que el gen *GYS1* codifica para una enzima que tiene un H en el residuo aminoácido 309.
25. El procedimiento de la reivindicación 18, en el que el biomarcador es una proteína GYS 1.
- 15 26. El procedimiento de la reivindicación 25, en el que la proteína *GYS1* codifica para una enzima que tiene un H en el residuo aminoácido 309.

Figura 1.

```

1   ATGCCTCTAA ACCGCACTTT GTCCATGTCC TCACTGCCAG GACTGGACGA
51  TTGGGAGGAT GAATTCGACC TGGAGAATAC AGTGCTCTTC GAGGTGGCCT
101 GGGAGGTGGC CAACAAGGTG GGTGGCATCT ACACGGTGCT ACAGACGAAG
151 GCGAAGGTGA CAGGGGATGA ATGGGGCGAC AACTACTACC TGGTGGGACC
201 ATACACGGAG CAAGGCGTGA GGACCCAGGT GGAGCTGCTC GAGCCCCCAA
251 CCCC GGCCCT GAAGAGGACG CTGGACTCCA TGAACAGCAA GGGCTGCAAG
301 GTGTATTTTCG GCGCTGGCT GATCGAGGGG GGCCCCCTGG TGGTGCTCCT
351 GGATGTGGGG GCCTCAGCCT GGGCCCTGGA ACGCTGGAAG GGAGAGCTTT
401 GGGACACCTG CAACATCGGG GTGCCCTGGT ACGACCGTGA GGCCAACGAC
451 GCCGTCTTT TTGGCTTCCT CACCACCTGG TTCCTGGGTG AGTTCCTGGC
501 CCAGAGCGAG GAGAAGCCAC ATGTGGTTGC ACACTTCAC GAGTGGTTGG
551 CGGGCATCGG GCTCTGCCTG TGCCGTGCCG GCGGGCTGCC TGTGGCTACA
601 ATCTTCACCA CCCACGCCAC GCTGCTGGGG CGATACCTGT GTGCCGGTGC
651 TGTGGACTTC TACAACAACC TGGAGAATT CAACGTGGAC AAGGAAGCTG
701 GTGAGAGGCA AATTATCAC CGTACTGCA TGGAGCGGGC GGCAGCCCAC
751 TGCACTCACG TCTTCACTAC CGTGTCCCAG ATCACCGCCA TTGAGGCTCA
801 GCACCTACTC AAGAGGAAAC CAGATATCGT GACCCCAAT GGACTGAATG
851 TGAAGAAFTT CTCTGCCATG CATGAGTTCC AGAACCTCCA TGCTCAGAGC
901 AAGGCCCGAA TCCAGGAGTT TGTGCGTGGC CATTTTTATG GGCACCTGGA
951 CTTCAACTTG GATAAGACCC TGTATTTCTT TATCGCCGGC CGCTACGAGT
1001 TCTCCAACAA GGGGGCTGAC GTCTTCTGG AGGCCTTGGC CCGGCTCAAC
1051 TATCTGCTCA GAGTAAATGG CAGCGAGCAG ACCGTGGTCG CCTTCTTCAT
1101 CATGCCGGCT CGGACCAACA ACTTCAACGT GGAAACCCTC AAGGGGCAAG
1151 CCGTGCGCAA GCAGCTCTGG GATACGGCGA ACACAGTGA GGAGAAGTTC
1201 GGGAGGAAGC TTTACGAATC CCTGCTGGTT GGGAGCCTCC CGGACATGAA
1251 CAAGATGCTG GACAAGGAGG ATTTCACTAT GATGAAGAGA GCCATCTTTG
1301 CCACGCAGCG GCAGTCTTTT CCCCCTGTGT GCACCCACAA TATGCTGGAC
1351 GACTCCTCGG ACCCTATCCT GACCACCATC CGTCGAATCG GCCTCTTCAA
1401 TAGTAGTGCT GACAGGGTCA AGGTGATTTT CCACCCAGAG TTCCTCTCCT
1451 CCACGAGCCC CCTGCTCCCC GTGGACTATG AGGAGTTTGT CCGTGGCTGC
1501 CACCTTGGGG TTTTCCCCTC CTACTATGAG CCTTGGGGCT ACACACCAGC
1551 TGAGTGCACG GTTATGGGCA TCCCCAGTAT CTCCACCAAC CTCTCCGGCT
1601 TCGGCTGCTT CATGGAGGAA CACATCGCAG ACCCCTCAGC TTACGGCATC
1651 TACATTCTGG ACCGGCGGTT CCGCAGCCTG GATGATTCCT GCTCGCAGCT
1701 TACCTCCTTC CTCTACAGCT TCTGCCAGCA GAGCCGGCGG CAGCGCATCA
1751 TCCAGCGGAA CCGCACGGAG CGCCTCTCCG ACCTTCTGGA CTGGAAATAC
1801 CTAGGCCGGT ACTATATGTC CGCGCGCCAC ATGGCGCTGG CCAAGGCCTT
1851 TCCAGAACAT TTCACCTAG AGCCCCGCGA GGCTGATGCG ACCCAGGGCT
1901 ACCGCTACCC ACGGCCTGCA TCGGTGCCTC CGTCGCCCTC ACTGTCACGA
1951 CACTCGAGCC CGCACCAGAG CGAGGACGAG GAGGAGCCCC GGGACGTGCC
2001 GCCCGATGAA GACAGTGAGC GCTACGACGA GGACGAGGAG GCCGCCAAGG
2051 ACCGGCGCAA CATCCGCGCC CCGGAGTGGC CGCGTCGCGC CTCCTGCACC
2101 TCTTCCACGA GCGGGAGCAA GCGCGGCTCG GTGGACACGG GGCCCTCCAG
2151 CTCGCTCAGC ACCCCAGCG AGCCCCTCAG CCCCGCCAGC TCCCTGGGCG
2201 AGGAGCGCAA CTAA (SEC ID NO:1)

```

Figura 2.

	1	11	21	31	41
Normal	TTGAAACATG	GGGCCTTCTC	CCCCATGCCT	AGATATCGTG	ACCCCCAATG
MAPS	TTGAAACATG	GGGCCTTCTC	CCCCATGCCT	AGATATCGTG	ACCCCCAATG
	51	61	71	81	91
Normal	GACTGAATGT	GAAGAAGTTC	TCTGCCATGC	ATGAGTTCCA	GAACCTCCAT
MAPS	GACTGAATGT	GAAGAAGTTC	TCTGCCATGC	ATGAGTTCCA	GAACCTCCAT
	101	111	121	131	141
Normal	GCTCAGAGCA	AGGCCCGAAT	CCAGGAGTTT	GTGCGTGGCC	ATTTTTATGG
MAPS	GCTCAGAGCA	AGGCCCGAAT	CCAGGAGTTT	GTGCATGGCC	ATTTTTATGG
	151	161	171	181	191
Normal	GTATGTGGGC	CAGATACCCA	GGTCTTGAGA	GAGGTGGGGG	TTGGGTGCCC
MAPS	GTATGTGGGC	CAGATACCCA	GGTCTTGAGA	GAGGTGGGGG	TTGGGTGCCC
	201	211	221	231	
Normal	AGACTCCCGG	GTCTAAGGGG	GGGACAGCTA	A ( SEC ID NO:2)	
MAPS	AGACTCCCGG	GTCTAAGGGG	GGGACAGCTA	A ( SEC ID NO:3)	

Figura 3.

	1	11	21	31
Caballos con MAPS	IVTPNGLNVK	KFSAMHEFQN	LHAQSKRRIQ	EFVHG <sup>U</sup> HFYC
(SEC ID NO:4)				
Resto de las especies	IVTPNGLNVK	KFSAMHEFQN	LHAQSKRRIQ	EFVRG <sup>U</sup> HFYC
(SEC ID NO:5)				

Figura 4.

1 GTGGGGCATA GGAGGGGCTC CCGAGACATG AATGAGGTAG TTGACGGTCT  
 51 ACGCCAAGGC CCGTACCAAT TTTCTTATTG TAAAGAGCAG AGAACTACAA  
 101 CTCCCAGAAG ACCCTAAGTT GATGGCCTGC ATATGAGGCT GCCGGGCTGT  
 151 ACCCTAGAGG CTAATGGGAG GTGTAGTTTC TTTGACATTC ATTTTCAGAAA  
 201 AGGGAAAATT ATTTTATTGG CTAAGGTGGA TATTGGTCTT CTGAAAGAGG  
 251 ATGGAGGTTC ACTGGATTTT GTTTTTGAGA AAAGCAGTGT TAGAACCAGT  
 301 TATTCTTCCC AGGCTATAGC ATCAATTCTC AATGGATGTG AGAACTACAA  
 351 CTCCCAGAAG GCCTCTGGCT GGCCAGCTCC TTAATATTCC TGGAATTACC  
 401 TTCATTTTATT CATTCAACCA CATTGATCTG AATGAAGAAG TGAAACATGG  
 451 GACCTTCTCC CCCATGCCTA GATATCGTGA CCCCCAATGG ACTGAATGTG  
 501 AAGAAGTTCT CTGCCATGCA TGAGTTCCAG AACCTCCATG CTCAGAGCAA  
 551 GGCCCGAATC CAGGAGTTTG TGCGTGGCCA TTTTATGGG TATGTGGGCC  
 601 AGATACCCAG GTCTTGAGAG AGGTGGGGGT TGGGTGCCCA GACTCCCGGG  
 651 TCTAAGGGAG GGGACAGCTG GGGGCTCAGA CTCTTGAGTT CCTGAGCGTC  
 701 AGAACAAATGG CTGAGTATTG CAGGAGATCC CCGTTTAAGG AGAAAAGAGG  
 751 CACAAGTCCC ATGAGTCCCA TGATGCCTCT GTACTGTCTA AGATGTTCTT  
 801 TGCCCCTGTG GCTTCTGGGG TTTGTAGITT TGAGCCAGGG CTAGAGGGTG  
 851 AGGATCTTGG CTTACACCTG TCTTGTGGCT TTTTAG (SEC ID NO:6)

Figura 5.

1 MPLNRTLMS SLPGLDDWED EFDLENTVLF EVAWEVANKV GGIYTVLQTK  
 51 AKVTGDEWGD NYLVGPTYE QGVRTQVELL EPPTPALKRT LDSMNSKGCK  
 101 VYFGRWLIEG GPLVVLLDVG ASAWALERWK GELWDTCNIG VPWYDREAND  
 151 AVLFGFLTTW FLGEFLAQSE EKPHVVAHFH EWLAGIGLCL CRARRLPVAT  
 201 IFTTHATLLG RYLCAGAVDF YNNLENFNVD KEAGERQIYH RYCMERAAAH  
 251 CTHVFTTVSQ ITAIEAQHLL KRKPDIVTPN GLNVKKFSAM HEFQNLHAQS  
 301 KARIQEFVRG HPYGHLDENL DKTLYFFIAG RYEFNSKPAD VFLEALARLN  
 351 YLLRVNGSEQ TVVAFFIMPA RTNNFNVELL KGQAVRKQLW DTANTVKEKF  
 401 GRKLYESLLV GSLPDMNKML DKEDFTMMKR AIFATORQSF PPVCTHNMLD  
 451 DSSDPILTTI RRIGLFNSSA DRVKVIFHPE FLSSTSPLLP VDYEEFVRGC  
 501 HLGVFPSYYE PWGYTPAECT VMGIPSISTN LSGFGCFMEE HIADPSAYGI  
 551 YILDRRFRSL DDSCSQLTSF LYSFCQOSRR QRIIQNRNTE RLSDLLDWKY  
 601 LGRYYMSARH MALAKAFPEH FTYEPREADA TQGYRYPRPA SVPPSPSLSR  
 651 HSSPHQSEDE EEPRDVPPDE DSERYDEDEE AAKDRRNIRA PEWPRRASCT  
 701 SSTSGSKRGS VDTGPSSSL S TPSEPLSPAS SLGEERN-