

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 629 400**

51 Int. Cl.:

C12Q 1/68

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **16.02.2010 PCT/US2010/000426**

87 Fecha y número de publicación internacional: **30.09.2010 WO10110835**

96 Fecha de presentación y número de la solicitud europea: **16.02.2010 E 10708427 (9)**

97 Fecha y número de publicación de la concesión europea: **05.04.2017 EP 2411532**

54 Título: **Métodos de PCR para caracterizar la región 5' no traducida de los genes FMR1 y FMR2**

30 Prioridad:

24.03.2009 US 162977 P

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

09.08.2017

73 Titular/es:

**ASURAGEN, INC. (100.0%)
2150 Woodward St
Austin, TX 78744, US**

72 Inventor/es:

**LATHAM, GARY, J.;
CHEN, LIANGJING y
SAH, SACHIN**

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 629 400 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Métodos de PCR para caracterizar la región 5' no traducida de los genes FMR1 y FMR2

5 Campo de la invención

Esta invención se encuentra en los campos de la síntesis y análisis de ADN, particularmente, con relación a los moldes y productos ricos en GC.

10 Descripción General

Desde el primer aislamiento de una ADN polimerasa y la determinación de las condiciones bajo las que el ADN puede sintetizarse in vitro, las reacciones de síntesis de ADN se han usado ampliamente para propósitos preparativos y analíticos en aplicaciones biotecnológicas, médicas y de investigación. La reacción en cadena de la polimerasa, o PCR, es un tipo de reacción de síntesis de ADN mediante la cual una secuencia de ADN puede amplificarse rápida y exponencialmente. Al igual que otras reacciones de síntesis en ciclos, implica copiar repetidamente la secuencia objetivo de una manera cíclica. Una implementación típica de PCR implica proporcionar iniciadores complementarios a los extremos de la secuencia a amplificar, un regulador adecuado, una sal de magnesio, trifosfatos de desoxinucleótidos (dNTP) y una ADN polimerasa termofílica. El molde o ADN objetivo contenido, por ejemplo, dentro de una muestra de ADN genómico, se expone a estos componentes en solución acuosa. La mezcla se somete a ciclos a través de etapas a diferentes temperaturas que favorecen la desnaturalización del molde, hibridación de los iniciadores al molde, y después la extensión de los iniciadores por la polimerasa, creando un producto mayor. Como el producto de cada ciclo está disponible como molde en las reacciones posteriores, la cantidad de producto aumenta exponencialmente hasta aproximadamente que se agotan otros componentes de la reacción (inicialmente presentes en exceso). Ver, por ejemplo, la Patente de los Estados Unidos núm. 4.683.202; M. J. McPherson & S. G. Moller, PCR: The Basics (2da Ed., Taylor & Francis) (2006).

La PCR, junto con otras formas de reacciones de síntesis de ácidos nucleicos en ciclos, es una herramienta estándar en biología molecular, biotecnología y, cada vez más, en medicina. Las ventajas claves de la PCR y las técnicas relacionadas son la rapidez, bajo costo, sensibilidad, facilidad para el análisis de alto rendimiento y versatilidad. Las amplificaciones requieren sólo unas pocas horas o menos, pequeñas reacciones individuales pueden consumir mucho menos que el valor de un dólar americano en reactivos, la cantidad de molde requerida está típicamente en el intervalo de nanogramos, la automatización puede resultar en miles de corridas de reacciones por día por robótica, y los iniciadores pueden diseñarse para amplificar casi cualquier secuencia.

La PCR y técnicas relacionadas se adoptan ampliamente tanto para aplicaciones analíticas como preparativas. Una aplicación preparativa típica de PCR es amplificar una secuencia de manera que pueda clonarse en un vector heterólogo. Las aplicaciones analíticas notables de la PCR incluyen diagnósticos de afecciones o determinaciones de genotipos que implican los loci genéticos con polimorfismos de tamaño.

Un ejemplo de un locus que exhibe un polimorfismo de tamaño relevante desde el punto de vista médico es la región 5' no traducida (UTR) del gen FMR1 humano en el cromosoma X. Los individuos normales típicamente tienen 5-44 repeticiones de CGG en esta región. Por el contrario, los alelos de este locus que contienen 200 a 2000 o más repeticiones de CGG son indicativos del síndrome de X frágil (FXS). Tales alelos se denominan alelos de mutación completa. Estos alelos son genéticamente inestables. Los individuos con FXS pueden tener varias combinaciones de síntomas tales como ataxia, fallo ovárico prematuro, problemas de aprendizaje y otras afecciones cognitivas/conductuales, que incluyen síntomas similares al autismo.

Una desafortunada excepción a la versatilidad de la PCR se encuentra en la dificultad de amplificar largas corridas de secuencias altamente ricas en GC, que incluyen los alelos de mutación completa de 5' UTR de FMR1. Los intentos de optimizar la PCR de FMR1 han incluido modificaciones a las condiciones convencionales del ensayo de PCR. Ver Genome Res. 6(7):633-8 (1996); Nucleic Acids Res. 25(19):3957-8 (1997); J. Mol. Diagn 8:544-550 (2006); Am. J. Med. Genet. 51(4):527-34 (1994). Después, más aún de 15 años del desarrollo del ensayo de PCR de FMR1, un estudio de tamizaje piloto publicado tan sólo en 2008 (Genet. Med. 10(10):714-9 (2008)) para detectar X Frágil en los recién nacidos reportó que "dos métodos de análisis de la reacción en cadena de la polimerasa (PCR) cuantitativo... usados en el proceso interno de validación para determinar el número de repetición de FMR1 en las mujeres **fracasó** para producir resultados fiables y reproducibles", (énfasis añadido), y, además, que "un segundo fracaso [PCR] ya sea de aislamiento primario o secundario fue altamente sugerente de un FMR1 de tamaño anormal de repeticiones de CGG." Así, los expertos en la técnica siguen considerando la amplificación por PCR reproducible de los alelos de X frágil de mutación completa como un problema no resuelto.

Se ha observado que la detección de las regiones de repetición de tripletes CGG que contienen más de aproximadamente 100 repeticiones por PCR se hace progresivamente más débil con el aumento del número de repetición. J. Mol. Diagn. 7:605-12 (2005). Esta dificultad, combinada con la naturaleza heterogénea de los síntomas de FXS, ha contribuido al uso de procedimientos tales como la transferencia de Southern para detectar los alelos de

mutación completa. *Id.* La transferencia de Southern es generalmente costosa y lleva mucho tiempo, y es mucho menos fácil para implementaciones de alto rendimiento, que la PCR.

Una publicación reciente de Tassone y otros. (J. Mol. Diagn. 10:43-49 (2008); "Tassone 2008") describe un ensayo para probar la presencia de alelos CGG más largos sin amplificación de longitud completa del alelo. *Ver además el documento de patente núm. WO2008/011170.* El método utiliza un iniciador quimérico de PCR que se hibrida con sitios dentro de la región CGG expandida, de manera que la presencia de una amplia mancha de productos de PCR representa un resultado positivo para un alelo expandido. *Id.* en 46.

La estrategia de hibridación aleatoria puede resultar en la amplificación no específica y la pérdida de resolución. Las condiciones básicas de la PCR usadas en este documento se han probado ampliamente en muchos laboratorios diferentes y parece que el número máximo de repeticiones de CGG que pueden amplificarse con éxito es de aproximadamente 350 repeticiones de CGG. *Ver, por ejemplo,* Saluto y otros, J. Mol. Diagn. 7:605-612 (2005). La amplificación no específica es evidente en Tassone 2008. La mancha en el gel de agarosa que se muestra por Tassone 2008 como la Figura 1 parece que contiene productos más largos que la longitud máxima esperada de 350 repeticiones que contienen los amplicones y, así puede ser que muchas de las manchas no representan productos específicos de repeticiones de CGG (ver Figura 1 de Tassone 2008; la mancha en el carril 5 parece extenderse hasta la proximidad del pocillo de carga). Tassone y otros indican en la leyenda de la Figura 3 que usaron un patrón Alto-Bajo (Bionexus, Oakland, CA) como marcador de peso molecular, y el marcador en la Figura 1 parece ser el mismo que en la Figura 3. La banda más grande del patrón Alto-Bajo es de aproximadamente 10 kb, y la mancha de la Figura 1 se extiende por encima de esa banda. Esta longitud es además, más larga que el producto de longitud completa esperado de la muestra usada en el ensayo. El tamaño del alelo más largo se enumera como 615 repeticiones de CGG. Por lo tanto, se espera que los productos de longitud completa sean de aproximadamente 2 kb, estimados mediante la adición de aproximadamente 200 nt de secuencia flanqueante a 1845 nt de repeticiones. Los productos no específicos pueden reducir tanto la precisión como la confianza de las determinaciones basadas en la amplificación con respecto al estado del alelo asociado al X frágil.

Además, algunos ensayos convencionales de PCR para la detección de un X Frágil basado en diseños de iniciadores puramente específicos a genes, tienen limitaciones. Por ejemplo, tales ensayos proporcionan una estimación del número de repetición CGG basado en la movilidad del amplicón de PCR. Para permitir la cuantificación exacta del número de repetición, la movilidad del amplicón se mide generalmente con relación a calibradores externos, por ejemplo, un conjunto de controles de tamaño adecuado. Es conveniente permitir una cuantificación precisa de CGG sin depender de calibradores externos.

Además, los alelos de FMR1 pueden contener secuencias AGG que se intercalan entre las repeticiones de CGG, normalmente en la región 5' del segmento repetido. El conocimiento de los elementos de la secuencia AGG caracteriza al alelo en un aspecto. Pueden usarse elementos de secuencia AGG en la toma de decisiones clínicas. Por ejemplo, los casos de expansión del alelo FMR1 de una madre a un alelo de mutación completo en su hijo se han observado para un alelo con tan solo 59 repeticiones, y se conoce que este alelo carece de elementos AGG. *Ver Nolin y otros, Am. J. Hum. Genet. 72:454-464 (2003).* De hecho, las mutaciones completas rara vez parecen contener elementos AGG más allá de las primeras 20 repeticiones de CGG, y los estudios biofísicos han sugerido que los moldes con secuencias AGG "de interrupción" entre el segmento de repetición CGG son más propensas a adoptar estructuras de ADN más convencionales y son más estable y más susceptible de replicación precisa. *Ver, por ejemplo,* Weisman-Shomer y otros, Nucleic Acids Res. 28:1535-41 (2000); Zhong y otros, Am. J. Med. Genet. 64:261-5 (1996); Larsen y otros, Am. J. Med. Genet. 93:99-106 (2000); y Dombrowski y otros, Hum. Mol. Genet. 11:371-78 (2002). Así, existe una necesidad de una tecnología para mapear los elementos interruptores, tales como elementos AGG, dentro del gen FMR1. El mapeo de los elementos interruptores tiene por tanto, aplicaciones de investigación y diagnóstico relacionadas con el Síndrome del X Frágil y otras enfermedades de repetición asociadas.

Los métodos existentes para mapear elementos AGG incluyen la secuenciación y mapeo de restricción. La tecnología de secuenciación del ADN es relativamente laboriosa, particularmente porque generalmente requiere enriquecimiento o aislamiento (ya sea in vitro o in silico) de la secuencia de interés, y no se realiza rutinariamente en pruebas de diagnóstico de X Frágil. Los ensayos basados en el mapeo de restricción han usado la enzima *MnI*, que reconoce la secuencia GAGG y corta 7 pares de bases 5' a esa secuencia, para inferir las posiciones AGG, basadas en el tamaño de los fragmentos resultantes de los productos de PCR digeridos que comprenden la región de repetición CGG de la 5' UTR de FMR1. *Ver, por ejemplo,* Eichler y otros, Nat. Genet. 8:88-94 (1994); Zhong y otros, Am. J. Hum. Genet. 57:351-361 (1995). El ensayo de Eichler y otros implicó la amplificación por PCR de la región de repetición, purificación de los productos, digestión durante toda la noche, electroforesis durante 7 horas y transferencia de tipo Southern. En el ensayo de Zhong y otros, "el producto de PCR se extrajo una vez con fenol/cloroformo, se precipitó con etanol y se digirió parcialmente en 10 litros con 5 unidades de *MnI* a 37°C durante 50-70 min." *Id.* a 353. Esto se siguió por electroforesis y transferencia de tipo Southern. Ambos ensayos implicaron, múltiples etapas que incluyen purificación/limpieza, digestión por restricción y transferencia de tipo Southern además de PCR y electroforesis. Chiu y otros, Journal of Genetics 87:275-277 (2008) reportan un estudio que evalúa las repeticiones de CGG y el patrón de interrupción AGG del locus FMR1 entre la población taiwanesa. Bodega y otros, Human Reproduction 21:952-957 (2006) reportan sobre expansiones de FMR1 en pacientes con manifestación prematura de fallo ovárico.

En la presente descripción se proporcionan métodos para cuantificar las repeticiones de CGG e identificar, cuantificar y revelar el contexto secuencial de las secuencias de interrupción en la 5' UTR de los genes FMR1 y FMR2. Las aplicaciones de los productos potenciales de la invención incluyen aplicaciones clínicas para la prueba de X Frágil.

- 5 En algunas modalidades, la invención se refiere a un método para analizar en una muestra si una secuencia de interrupción está presente en al menos una región rica en CGG compuesta por al menos un molde, que comprende:
- a) proporcionar al menos dos iniciadores diferentes, incluyendo un primer iniciador que comprende repeticiones CGG, CCG, GCG, CGC, GCC o GGC, y un segundo iniciador que se hibrida a una posición fuera de la región rica en CGG;
 - 10 b) realizar la PCR con al menos dos iniciadores diferentes y al menos un molde que comprende al menos una región rica en CGG, en donde la PCR produce un conjunto de productos;
 - c) resolver el conjunto de productos con una técnica de alta resolución para producir una representación del tamaño y la abundancia del producto; y
 - 15 d) obtener información acerca de si una secuencia de interrupción está presente en al menos una región rica en CGG o donde dentro de al menos una región rica en CGG se localiza una secuencia de interrupción.

- En algunas modalidades, la invención se refiere a un método para analizar en una muestra si una secuencia de interrupción está presente en al menos una región rica en CGG compuesta por al menos un molde, que comprende:
- 20 a) proporcionar al menos tres iniciadores diferentes, incluyendo un primer iniciador que comprende repeticiones CGG, CCG, GCG, CGC, GCC, o GGC y una solapa 5', un segundo iniciador que se hibrida a una posición fuera de la región rica en CGG y un tercer iniciador que tiene una secuencia comprendida por la solapa 5' del primer iniciador, en donde el primer iniciador se proporciona a una concentración menor que el tercer iniciador;
 - 25 b) realizar la PCR con al menos tres iniciadores diferentes y al menos un molde, en donde la PCR produce un conjunto de productos;
 - c) resolver el conjunto de productos con una técnica de alta resolución para producir una representación del tamaño y la abundancia del producto; y
 - 30 d) obtener información acerca de si una secuencia de interrupción está presente en al menos una región rica en CGG o donde dentro de al menos una región rica en CGG se localiza una secuencia de interrupción a partir de dicha representación.

Además se describe en la presente descripción un método para analizar en una muestra al menos una región rica en CGG comprendida por al menos un molde, que comprende:

- 35 a) proporcionar al menos dos iniciadores diferentes, en donde el primer iniciador comprende repeticiones de CGG, CCG, GCG, CGC, GCC o GGC y el segundo iniciador se hibrida a una posición fuera de la región rica en CGG;
- b) realizar la PCR con al menos dos iniciadores diferentes y un molde que comprende la región rica en CGG, en donde la PCR produce un conjunto de productos;
- c) resolver el conjunto de productos con una técnica de alta resolución para producir una representación del tamaño y la abundancia del producto; y
- 40 d) obtener la información acerca del número de repetición CGG de dicha representación.

Además se describe en la presente descripción un método para analizar en una muestra al menos una región rica en CGG comprendida por al menos un molde, que comprende:

- 45 a) proporcionar tres iniciadores diferentes, en donde el primer iniciador comprende repeticiones de CGG, CCG, GCG, CGC, GCC o GGC y una solapa 5', el segundo iniciador se hibrida a una posición fuera de la región rica en CGG, el tercer iniciador tiene la misma secuencia de solapa 5' del primer iniciador, y el primer iniciador se proporciona a una concentración inferior que el tercer iniciador;
- b) realizar la PCR con los tres iniciadores diferentes y un molde que comprende la región rica en CGG, en donde la PCR produce un conjunto de productos;
- 50 c) resolver el conjunto de productos con una técnica de alta resolución para producir una representación del tamaño y la abundancia del producto; y
- d) obtener la información acerca del número de repetición CGG de dicha representación.

Además se describe en la presente descripción un oligonucleótido que comprende una secuencia seleccionada a partir de la sec. con núm. de ident.: 44 y la sec. con núm. de ident.: 45.

Se debe entender que la descripción general anterior y la siguiente descripción detallada son solamente ilustrativas y explicativas y no son restrictivas de la invención, según se reivindica.

60 Descripción detallada

La invención se refiere a reacciones de amplificación que amplifican toda o parte de una región de repetición CGG. En algunas modalidades, la región de repetición CGG está compuesta por la 5' UTR de FMR1, o la 5' UTR de FMR2.

65 La invención se refiere a reacciones de amplificación que usan un iniciador que se hibrida fuera de una región de repetición CGG y un iniciador que se hibrida a las secuencias de repetición CGG, permutaciones de secuencia o

complementos inversos de las secuencias (GCG, CCG, CGC, GCC, o GGC). Los iniciadores que pueden hibridar fuera de (corriente arriba o corriente abajo) de la región de repetición CGG pueden ser iniciadores directos o inversos. Los iniciadores pueden hibridarse a secuencias que flanquean la región de repetición CGG. Ejemplos de tales iniciadores directos incluyen CGG TGG AGG GCC GCC TCT GAG C (sec. con núm. de ident.: 1), CAG GCG CTC AGC TCC GTT TCG GTT T (sec. con núm. de ident.: 2), CAG TCA GGC GCT CAG CTC CGT TTC G (sec. con núm. de ident.: 3), TCC GGT GGA GGG CCG CCT CTG AGC (sec. con núm. de ident.: 4), GGT TCG GCC TCA GTC AGG CGC TCA GCT CCG TTT CG (sec. con núm. de ident.: 5), GGG TTC GGC CTC AGT CAG GCG CTC AGC TCC GTT TCG (sec. con núm. de ident.: 6), GCG GGC CGG GGG TTC GGC CTC AGT CA (sec. con núm. de ident.: 7), CAG CGG GCC GGG GGT TCG GCC TCA G (sec. con núm. de ident.: 8), GCA GCG GGC CGG GGG TTC GGC CTC A (sec. con núm. de ident.: 9), GGG CCG GGG GTT CGG CCT CAG TCA G (sec. con núm. de ident.: 10), GGG GTT CGG CCT CAG TCA GGC GCT CA (sec. con núm. de ident.: 11), GGG GTT CGG CCT CAG TCA GGC GCT CAG (sec. con núm. de ident.: 12), GGC GCT CAG CTC CGT TTC GGT TTC ACT TCC (sec. con núm. de ident.: 13), TCA GGC GCT CAG CTC CGT TTC GGT TTC A (sec. con núm. de ident.: 14), CAC TTC CGG TGG AGG GCC GCC TCT GA (sec. con núm. de ident.: 15), TTC TCG TGG AGG GCC GCC TCT GAG C (sec. con núm. de ident.: 16), and TCA GGC GCT CAG CTC CGT TTC GGT TTC ACG GCG GCG GCG GCG GA (sec. con núm. de ident.: 44). Ejemplos de tales iniciadores inversos incluyen CGC ACT TCC ACC ACC AGC TCC TCC A (sec. con núm. de ident.: 17), GGA GCC CGC CCC CGA GAG GTG (sec. con núm. de ident.: 18), GGG AGC CCG CCC CCG AGA GGT (sec. con núm. de ident.: 19), CGC ACT TCC ACC ACC AGC TCC TCC AT (sec. con núm. de ident.: 20), CGG GAG CCC GCC CCC GAG AGG TG (sec. con núm. de ident.: 21), CCG GGA GCC CGC CCC CGA GAG GT (sec. con núm. de ident.: 22), CCG GGA GCC CGC CCC CGA GAG GTG (sec. con núm. de ident.: 23), CGC CGG GAG CCC GCC CCC GAG AGG TG (sec. con núm. de ident.: 24), GCG CCG GGA GCC CGC CCC CGA GAG GT (sec. con núm. de ident.: 25), CGC CGG GAG CCC GCC CCC GAG AGG T (sec. con núm. de ident.: 26), GCG CCA TTG GAG CCC CGC ACT TCC ACC A (sec. con núm. de ident.: 27), GCG CCA TTG GAG CCC CGC ACT TCC A (sec. con núm. de ident.: 28), AGC GCC ATT GGA GCC CCG CAC TTC C (sec. con núm. de ident.: 29), CGC CAT TGG AGC CCC GCA CTT CCA C (sec. con núm. de ident.: 30), TTG GAG CCC CGC ACT TCC ACC ACC A (sec. con núm. de ident.: 31), AGC CCC GCA CTT CCA CCA CCA GCT CCT C (sec. con núm. de ident.: 32), GAG CCC CGC ACT TCC ACC ACC AGC TCC T (sec. con núm. de ident.: 33), CAT TGG AGC CCC GCA CTT CCA CCA CCA G (sec. con núm. de ident.: 34), CCC GCA CTT CCA CCA CCA GCT CCT CCA TCT (sec. con núm. de ident.: 35), TAG AAA GCG CCA TTG GAG CCC CGC ACT TCC (sec. con núm. de ident.: 36), AAG CGC CAT TGG AGC CCC GCA CTT CC (sec. con núm. de ident.: 37), AAG CGC CAT TGG AGC CCC GCA CTT CCC CGC CGC CGC CGC CG (sec. con núm. de ident.: 43), and AAG CGC CAT TGG AGC CCC GCA CTT CCC CGC CGC CGC CGC CT (sec. con núm. de ident.: 45).

La invención se refiere además al uso y métodos que comprenden proporcionar los iniciadores TCAGGCGCTCAGCTCCGTTTCGGTTTCACTTCCGGT (sec. con núm. de ident.: 38), AGCGTCTACTGTCTCGGCACT TGCCCGCCGCCCG (sec. con núm. de ident.: 39), TCA GGC GCT CAG CTC CGT TTC GGT TTC A (sec. con núm. de ident.: 40), y TCAGGCGCTCAGCTCCGTTTTCGGTTTCA CGGCGCGGCGGCGG (sec. con núm. de ident.: 41). La invención adicionalmente se refiere a iniciadores que comprenden la secuencia de cualquiera de las sec. con núms. de ident.: 1-38 o 40 con repeticiones de CGG o las permutaciones y complementos inversos de estas unidas al extremo 3'. La invención se refiere además, al uso de y métodos que comprenden proporcionar un iniciador que contiene un número de repeticiones trinucleotídicas de la secuencia CGG o las permutaciones y complementos inversos de estas. En algunas modalidades, el número de repeticiones de CGG en el iniciador es de cuatro o cinco. En algunas modalidades, el iniciador contiene 12-15 nucleótidos de secuencia de repetición trinucleotídica. En algunas modalidades, el iniciador contiene un número de repeticiones en el intervalo de 3 a 10. El iniciador puede contener 3, 4, 5, 6, 7, 8, 9 o 10 repeticiones, y opcionalmente una repetición parcial adicional de 1 o 2 residuos de C y/o G. Pueden proporcionarse iniciadores adicionales, por ejemplo, para asegurar la unión en un sitio polimórfico, o para amplificar una región de tamaño conocido que sirve como un control interno.

En algunas modalidades, el iniciador que se hibrida con secuencias de repetición CGG tiene una actividad de unión preferencial para sitios en la región rica en CGG que comprende un elemento de interrupción. La unión preferencial del iniciador que se hibrida a las secuencias de repetición CGG con al menos un sitio que comprende un elemento de interrupción puede resultar en una amplificación selectiva de al menos un producto que comprende el elemento de interrupción, por ejemplo, usando el iniciador en una reacción de PCR con un segundo iniciador de orientación opuesta que se une fuera de la región rica en CGG, como se describió anteriormente. La actividad de unión preferencial puede ser específica, por ejemplo, para los sitios que comprenden elementos CGG y AGG, o las permutaciones y/o complementos inversos de estos, tal como un sitio que comprende (1) un elemento AGG o una parte de un elemento AGG que comprende un A, y (2) tres, cuatro, cinco o seis elementos CGG y opcionalmente un elemento CGG parcial adicional.

El grado de unión preferencial, expresado en términos de una relación de la abundancia de producto amplificado selectivamente a productos de fondo a partir de una reacción de amplificación usando el iniciador con un segundo iniciador de orientación opuesta que se une fuera de la región rica en CGG, puede ser al menos 3 veces, 4 veces, 5 veces o 6 veces, o puede variar de 3 veces a 12 veces, 3 veces a 10 veces, 3 veces a 8 veces, 4 veces a 12 veces, 4 veces a 10 veces, 4 veces a 8 veces, 3 veces a 7 veces, 4 veces a 7 veces, 3 veces a 6 veces o 4 veces a 6 veces. Al menos un producto amplificado selectivamente tiene generalmente una longitud correspondiente a la distancia a lo largo del molde desde el extremo 5' del primer iniciador, cuando se une preferentemente a un sitio que comprende un

elemento de interrupción, hasta el extremo 5' del segundo iniciador, cuando se unen a su sitio fuera de la región rica en CGG.

5 En algunas modalidades, el iniciador que se hibrida a secuencias de repetición CGG y se une preferentemente a un sitio o sitios en la región rica en CGG que comprende un elemento de interrupción puede comprender un residuo A, T o U dentro o al final de la parte del iniciador que se hibrida a secuencias de repetición CGG, o en otras palabras, entre o en el extremo de las repeticiones de CGG, CCG, GCG, CGC, GCC o GGC; ver, por ejemplo, las sec. con núms. de ident.: 44 y 45 anteriormente. El residuo A, T o U puede producirse en el extremo 3' del iniciador. Cuando el residuo A, T o U se produce en el extremo de las repeticiones de CGG, CCG, GCG, CGC, o GCC, GGC puede o no haber una repetición parcial CGG, CCG, GCG, CGC, GCC, o GGC entre el residuo A, T o U y la última repetición completa de CGG, CCG, GCG, CGC, GGC, o GGC. Es posible sustituir residuos de nucleótidos no naturales, discutidos detalladamente más abajo, que preferentemente el par de bases con residuos T/U o A con relación a otros residuos nucleotídicos naturales para el residuo A, T o U. Similarmente, además es posible sustituir uno o más residuos de nucleótidos no naturales que preferentemente un par de bases con residuos C o G con relación a otros residuos nucleotídicos naturales para uno o más residuos G y/o C que forman las repeticiones de CGG, CCG, GCG, CGC, GCC, o GGC. La presencia de uno o más de dichos residuos no naturales dentro de una secuencia de cualquier otra forma constituida por repeticiones de CGG, CCG, GCG, CGC, GCC, o GGC (opcionalmente con un A, T, U o el residuo no natural correspondiente como se discutió anteriormente) no niega la identidad de dicha secuencia dentro del contexto de la presente descripción como una secuencia de repeticiones de CGG, CCG, GCG, CGC, GCC o GGC.

20 En un ensayo no anclado, se proporciona un primer iniciador que tiene una actividad de unión preferencial para sitios en la región rica de CGG que no comprenden elementos interruptores. La presencia de un elemento de interrupción puede señalarse en los resultados de un ensayo no anclado por un nivel relativamente bajo de productos cuya síntesis implica la extensión del primer iniciador unido a sitios que comprenden el elemento de interrupción. Estos niveles bajos pueden aparecer como una interrupción o conjunto de picos bajos rodeados por picos más altos en un electroferograma.

30 En un ensayo de anclaje, se proporciona un primer iniciador que tiene una actividad de unión preferencial para sitios en la región rica en CGG que comprenden elementos interruptores. Se debe señalar que un iniciador que tiene una actividad de unión preferencial para sitios en la región rica en CGG que comprende elementos interruptores puede proporcionarse en reacciones en las que al menos un molde comprende al menos una región rica en CGG que comprende cero, uno o una pluralidad de elementos interruptores y la repetición de un iniciador con "una actividad de unión preferente para sitios en la región rica de CGG que comprenden elementos interruptores" no implica, por ejemplo, que la región rica en CGG comprenda necesariamente una pluralidad de elementos interruptores. La presencia de un elemento de interrupción se señala en un ensayo de anclaje por un nivel relativamente alto de productos cuya síntesis implica la extensión del primer iniciador unido a sitios que comprenden el elemento de interrupción. El nivel alto puede aparecer como un pico rodeado por picos inferiores y/o señal de línea de base en un electroferograma.

40 En algunas modalidades, el primer iniciador usado en un ensayo de anclaje comprende un residuo A, T o U entre o en el extremo 3' de las repeticiones de CGG. En algunas modalidades, el primer iniciador usado en un ensayo de anclaje comprende un residuo de nucleótido no natural que preferentemente se basa en pares con residuos A o T/U con relación a otros residuos nucleotídicos naturales. Los residuos de nucleótidos no naturales son residuos nucleotídicos que comprenden una nucleobase distinta de adenina, timina, guanina, citosina y uracilo (A, T, G, C y U, respectivamente). Ejemplos de residuos de nucleótidos no naturales que incluyen preferentemente un par de bases con residuos A o T/U, sin limitación, aductos de residuos T, U o A que preferentemente un par de bases con residuos A o T/U con relación a otros residuos naturales (por ejemplo, análogos de uracilo 5-sustituido); y residuos que comprenden nucleobases tales como, por ejemplo, pseudouracilo y diaminopurina.

50 Dos iniciadores se consideran orientados de manera opuesta cuando se unen a cadenas opuestas de un molde de ácido nucleico bicatenario.

55 Como se usa en la presente descripción, una secuencia es "corriente arriba" de una región rica en CGG cuando se produce 5' de la región rica en CGG a lo largo de la cadena que comprende repeticiones de CGG. Como se usa en la presente descripción, una secuencia es "corriente abajo" de una región rica en CGG cuando se produce 3' de la región rica en CGG a lo largo de la cadena que comprende repeticiones de CGG.

60 Los métodos de la invención pueden referirse a reacciones de amplificación que comprenden proporcionar al menos dos o al menos tres iniciadores diferentes. En algunas modalidades, se proporcionan al menos tres iniciadores diferentes y uno de los iniciadores es una subsecuencia de otro iniciador. En algunas modalidades, un iniciador es un iniciador quimérico que comprende repeticiones de CGG y una secuencia solapa 5', y otro iniciador tiene la secuencia de la secuencia solapa 5' del iniciador quimérico. Se debe señalar que el iniciador que tiene la secuencia de la secuencia solapa 5' del iniciador quimérico puede, pero no necesariamente, tener la secuencia entera no repetitiva del iniciador quimérico. En otras palabras, la secuencia de parte o de todo un iniciador puede estar comprendida por la secuencia de otro iniciador; por ejemplo, el iniciador quimérico comprende una secuencia solapa 5', y otro iniciador puede comprender la secuencia de parte o de toda la secuencia solapa 5'. En algunas modalidades, el iniciador contiene 12-15 nucleótidos de secuencia de repetición CGG. La secuencia solapa 5' puede corresponder a una secuencia adyacente o cercana a la región de repetición CGG, o puede no estar relacionada con secuencias en y alrededor de la región de repetición CGG.

En algunas modalidades, la longitud del iniciador quimérico puede ser de aproximadamente 35, 40, 45, 50 ó 55 nt. En algunas modalidades, uno o más de los iniciadores tiene una Tm en el intervalo de 60°C a 75°C, por ejemplo, aproximadamente 60°C, 65°C, 70°C, o 75°C.

5 En algunas modalidades, se proporcionan al menos tres iniciadores diferentes y se proporciona un iniciador a una concentración inferior que la concentración de otro iniciador. Por ejemplo, el iniciador quimérico se proporciona opcionalmente a una concentración menor que el iniciador con la secuencia de la secuencia solapa 5' del iniciador quimérico. La relación de las concentraciones, expresada como diferencia de veces, puede encontrarse en el intervalo de 2 a 10.000 o más, por ejemplo 10, 20, 50, 100, 200, 500, 1.000, 2.000, 5.000, o 10.000. En tales modalidades, el
10 iniciador presente a una concentración inferior puede agotarse en las primeras rondas de la reacción de amplificación, de manera que la extensión es generalmente toda o casi toda, a partir de los iniciadores todavía presentes (que estaban inicialmente presentes a concentraciones relativamente superiores).

15 La invención se refiere a reacciones que amplifican ácidos nucleicos. Ejemplos de reacciones de amplificación incluyen, sin limitación, PCR, NASBA (amplificación basada en secuencias de ácidos nucleicos) y SDA (amplificación de desplazamiento de cadenas). *Ver, por ejemplo, la Patente de los Estados Unidos 4.683.202 (PCR); Patente de los Estados Unidos 6.326.173; y J. of Virol. Methods 151:283-293 (2008) (NASBA); y la Patente de los Estados Unidos 5.648.211 (SDA).* El técnico con experiencia comprenderá qué reactivos son apropiados para proporcionar. Cada uno de estos métodos implica la síntesis de ADN y, como tal, implica el uso de ADN polimerasas, nucleótidos y cationes
20 divalentes (suministrados como una sal), particularmente magnesio, en una solución que conduce a la polimerización del ADN y en la que está presente el molde. Los métodos varían en términos de proporcionar actividades catalíticas adicionales, el uso del termociclado o la incubación isotérmica, y el uso y estructura de los iniciadores. Además se proporciona típicamente un regulador a un pH adecuado tales como entre 7 y 8, entre 6,5 y 8,5; entre 6 y 9, o aproximadamente 7,4 o 7,5.

25 En la PCR de conformidad con la invención, se proporciona al menos un par de iniciadores que se unen en cada extremo o dentro de una región objetivo, sobre cadenas opuestas, de manera que cada uno inicia la síntesis hacia el otro iniciador. La reacción se somete a termociclado para conducir la desnaturalización del sustrato en una etapa a alta temperatura, hibridación de los iniciadores a una etapa de temperatura inferior y extensión a una temperatura que puede ser pero no necesariamente superior que la de la etapa de hibridación. La amplificación se produce porque los
30 productos de un ciclo pueden servir como moldes en el siguiente ciclo.

En NASBA, se proporciona una ARN polimerasa (ARNP) además de la ADN polimerasa, que puede ser además una transcriptasa inversa (RT) (por ejemplo, una enzima que puede catalizar la síntesis de ADN usando o un molde de ARN o ADN). Se proporcionan iniciadores que son similares a los usados en la PCR excepto que al menos un iniciador comprende adicionalmente una secuencia promotora que se reconoce por la ARNP. Así, el producto de la RT sirve como molde para la ARNP, que sintetiza el ARN que sirve como molde para la RT, dando lugar a la amplificación. En algunas formas de NASBA, se proporciona RNasa H para producir el ADN monocatenario después de la síntesis de un híbrido de ARN-ADN por RT. La amplificación se produce por la acción combinada de la RT y ARNP, en ausencia de la
35 desnaturalización térmica repetida.

SDA es una técnica en la que el ADN se amplifica de forma isotérmica y asincrónica, lo que significa que la desnaturalización térmica cíclica no se usa para separar las cadenas; en su lugar, el desplazamiento de la cadena ocurre a través de la propia síntesis de ADN, en donde la extensión de un 3' OH provoca el desplazamiento de la
45 cadena corriente abajo. El 3' OH se proporciona inicialmente por un iniciador exterior y posteriormente por una reacción de corte. Se proporcionan dos pares de iniciadores. Un par "interior" se une alrededor del amplicón y comprende adicionalmente solapas 5' que contienen un sitio de restricción. El otro par "exterior" se sitúa distalmente, es decir, más alejado de la región objetivo. Un iniciador interior puede unirse al molde, extenderse y después desplazarse mediante síntesis a partir del correspondiente iniciador exterior. Posteriormente, el ADN desplazado se hace bicatenario, por
50 ejemplo, mediante síntesis de la segunda cadena. La siguiente etapa consiste en cortar una cadena de la molécula bicatenaria, que puede hacerse usando nucleótidos modificados y un sitio de restricción en donde el sitio de escisión se inactiva en una cadena (pero no en la otra) por el nucleótido modificado. La enzima de restricción correspondiente a este sitio se proporciona en la reacción y genera el corte. El 3' OH en el corte resultante se extiende después por la ADN polimerasa, desplazando una cadena (que puede servir de nuevo como molde ¹) y la molécula bicatenaria regenerada es de nuevo un sustrato para el corte seguido de extensión y desplazamiento, dando lugar a la
55 amplificación. No es necesaria la desnaturalización térmica repetida.

En algunas modalidades, los métodos de la invención comprenden proporcionar los dNTP en una relación GC/AT mayor que uno, y a una concentración total de dNTP que conduce a la síntesis de ADN usando moldes ricos en GC. La relación GC/AT puede ser de aproximadamente 1,1; 1,2; 1,4; 1,6; 2; 2,5; 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, o superior. La relación GC/AT puede estar entre 1,1 y 20, 1,1 y 15, 1,1 y 10, 1,1 y 8, 1 y 15, 1,1 y 7, 1,1 y 6, 1,1 y 5, 1,2 y 25, 1,4 y 25, 1,6 y 25, 2 y 25, 3 y 25, 4 y 25, 5 y 25, 2 y 15, 2,5 y 10, o 4 y 10. La concentración total de dNTP puede ser de aproximadamente 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1, 1,2; 1,5; 2, o 3 mM. La concentración de dNTP puede estar entre 0,4 y 3 mM, 0,5 y 3 mM, 0,6 y 3 mM, 0,7 y 3 mM, 0,8 y 3 mM, 0,9 y 3 mM, 1 y 3 mM, 0,4 y 2 mM, 0,4 y 1,5 mM, 0,4 y 1,2 mM, 0,4 y 1 mM, 0,4 y 0,9 mM, 0,4 y 0,8 mM, 0,4 y 0,7 mM, 0,5 y 2 mM, 0,5 y 1 mM, o 0,6 y 0,9 mM. "Relación GC/AT " significa la relación de la concentración de la suma de dCTP, dGTP y todos sus análogos de
60

nucleótidos, a la concentración de la suma de dATP, dTTP, dUTP y todos sus análogos de nucleótidos, en una solución dada o mezcla. "dNTP" significa desoxinucleótido trifosfato y se refiere a dATP, dCTP, dGTP, dTTP, dUTP y sus análogos. Los "análogos de nucleótidos" son moléculas o iones que comprenden una porción base distinta de las bases naturales adenina (A), citosina (C), guanina (G), timina (T) uracilo (U), una porción azúcar idéntica o similar a desoxirribosa, y al menos una porción fosfato o fosfato múltiple (por ejemplo, difosfato o trifosfato). El análogo de nucleótidos es un análogo de un nucleótido específico, particularmente dATP, dCTP, dGTP, dTTP o dUTP, cuando comprende un trifosfato y una porción de azúcar, cuya estructura y configuración son adecuadas para la incorporación en un ácido nucleico por una polimerasa y una base cuyas propiedades de apareamiento de bases en un ácido nucleico bicatenario y los loci de incorporación por ADN polimerasas en un ácido nucleico bicatenario son más similares a uno de los cinco nucleótidos enumerados anteriormente, con la excepción de que los análogos de dTTP serán generalmente además análogos de dUTP y viceversa. El término "análogo" usado junto con los términos que incluyen pero no se limitan a "nucleósido", "base", "nucleobase", o "residuo" debe interpretarse de la misma manera que si se usaran junto con "nucleótido."

¹ Tenga en cuenta que algunas cadenas desplazadas no serán inicialmente de longitud completa, sino que carecerán del complemento de la porción distal de la solapa interior del iniciador, como consecuencia del corte. Esto no perjudica la unión del iniciador (recuerde que la porción sin solapa del iniciador tiene una longitud suficiente para hibridar establemente) y, tras la unión del iniciador, se genera un saliente 5' que la polimerasa es capaz de llenar.

En algunas modalidades, se pueden proporcionar potenciadores. Los potenciadores contribuyen al éxito de las reacciones que generan productos ricos en GC. Puede incluirse generalmente, una variedad de potenciadores en las reacciones de PCR para aumentar el rendimiento, la especificidad y la consistencia, y pueden operar disminuyendo la T_m del ADN molde. Los potenciadores pueden funcionar mediante la desestabilización de la hélice, la neutralización de los inhibidores de la reacción u otros mecanismos, que incluyen mecanismos desconocidos. Los potenciadores incluyen, sin limitación, betaína, análogos de betaína, glicerol, albúmina de suero bovino (BSA), polietilenglicol, cloruro de tetrametilamonio, 7-deaza-GTP, detergentes neutros, dimetilsulfóxido (DMSO), metanol, etanol, isopropanol, formamida, acetona, acetamida, N-metilformamida, N,N-dimetilformamida, acetona, acetimida, N-metilacetimida, N,N-dimetilacetimida, 2-pirrolidona, N-metilpirrolidona, propionamida e isobutiramida. Los detergentes neutros incluyen, sin limitación, TWEEN-20, β-octil-glucósido, Octil-β-tio-glucopiranosido, Tritón X-100, Tritón X-114, NP-40, Brij-35, Brij-58, Tween-80, Pluronic F-68, Pluronic F-127, Deoxi Big CHAP, CHAPS, CHES, nonil fenoxipolietoxietanol (Tergitol-tipo NP-40) y octil fenoxipolietoxietanol (Igepal CA-630). Los análogos de betaína incluyen, sin limitación, homodeanol betaína, deanol betaína, propio betaína, homoglicerol betaína, dietanol homobetaína, trietanol homobetaína, hidroxipropil homobetaína, sal interna de N-metil-N-(2-carboxietil)morfolinio, sal interna de N-metil-N-(2-carboxietil)piperidinio, sal interna de N-Metil-N-(2-carboxietil)pirrolidinio, sal interna de N,N dimetil-N-(2-hidroxi-etil)-N-(2-sulfoetil)amonio, sal interna de N,N-dimetil-N-(2-hidroxi-etil)-N-(3-sulfopropil)amonio, sal interna de N,N-dihidroxi-etil-N-metil-N-(3-sulfopropil)amonio, sal interna de N,N-dimetil-N-(2-hidroxi-etil)-N-(4-sulfo-butil)amonio, sal interna de N-metil-N-(3-sulfopropil)morfolinio, y sal interna de N-metil-N-(3-sulfopropil)piperidio.

Se pueden proporcionar betaína, análogos de betaína y/o los otros potenciadores a concentraciones molares entre 0,01 y 5 M, 0,01 y 4 M, 0,01 y 3 M, 0,01 y 2,5 M, 0,02 y 5 M, 0,03 y 5 M, 0,04 y 5 M, 0,05 y 5 M, 0,07 y 5 M, 0,1 y 5 M, 0,2 y 5 M, 0,3 y 5 M, 0,4 y 5 M, 0,5 y 5 M, 0,7 y 5 M, 1 y 5 M, 1,5 y 5 M, 0,1 y 4 M, 0,5 y 3 M, 1 y 2,5 M, o 1,5 y 2,5 M, por ejemplo, aproximadamente 0,01; 0,02; 0,05; 0,1; 0,2; 0,5; 0,75; 1; 1,25; 1,5; 1,6; 1,7; 1,8; 1,9; 2,0; 2,1; 2,2; 2,3; 2,4; 2,5; 3; 3,5; 4; 4,5 o 5 M. Alternativamente, se pueden proporcionar potenciadores a concentraciones en porcentaje de p/v o v/v de entre 0,1 y 50%, 0,2 y 50%, 0,5 y 50%, 1 y 50%, 2 y 50%, 5 y 50%, 0,1 y 40%, 0,1 y 30%, 0,1 y 20%, 0,5 y 40% 1 y 30%, o 2 y 20%, por ejemplo, aproximadamente 0,1; 0,2; 0,5; 1; 2; 5; 10; 15; 20; 25; 30; 35; 40; 45 o 50% en volumen. Los detergentes neutros pueden proporcionarse entre 0,0001 y 10% en volumen, 0,0002 y 10%, 0,0005 y 10%, 0,001 y 10%, 0,002 y 10%, 0,005 y 10%, 0,01 y 10%, 0,02 y 10%, 0,05 y 10%, 0,0001 y 5%, 0,0001 y 2%, 0,0001 y 1%, 0,0005 y 1%, o 0,001 y 1%, por ejemplo, aproximadamente 0,0001; 0,0002; 0,0003; 0,0004; 0,0005; 0,0006; 0,0007; 0,0008; 0,0009; 0,001; 0,002; 0,003; 0,004; 0,005; 0,006; 0,007; 0,008; 0,009; 0,01; 0,02; 0,03; 0,04; 0,05; 0,06; 0,07; 0,08; 0,09; 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1; 2; 3; 4; 5; 6; 7; 8; 9; o 10% en volumen. Los expertos en la técnica reconocerán las concentraciones apropiadas para diversos potenciadores.

La invención se refiere a los métodos que comprenden proporcionar reguladores para reacciones de amplificación. Los reguladores pueden comprender, por ejemplo y sin limitación, tris(hidroximetil)aminometano (Tris), bis-tris propano, bicarbonato, fosfato, glicina, histidina, ácido 4-(2-hidroxi-etil)-1-piperazinoetanosulfónico (HEPES), ácido 3-(N-morfolino)propanosulfónico (MOPS) y diversas bases/ácidos conjugadas y sus sales.

La invención se refiere a los métodos que comprenden proporcionar al menos una ADN polimerasa para sintetizar ADN a partir de los dNTP de una manera dependiente del molde. La ADN polimerasa puede comprender una polimerasa silvestre, modificada, termofílica, quimérica, manipulada y/o una mezcla de más de una. La ADN polimerasa puede comprender Polimerasa Exacta (5 PRIME GmbH), AccuSure™ ADN Polimerasa (Bioline), Phusion™ AccuPrime™ Pfx (Invitrogen), Platinum Taq ADN Polimerasa de Alta Fidelidad (Invitrogen), Phire™ Hot Start ADN Polimerasa (New England Biolabs), Phusion® Hot Start ADN Polimerasa de Alta Fidelidad (New England Biolabs), JumpStart™ REDTaq™ ADN Polimerasa (Sigma-Aldrich), PfuUltra™ Hotstart ADN Polimerasa (Stratagene), PfuTurbo® Cx Hotstart ADN Polimerasa (Stratagene), PrimeSTART™ HS ADN Polimerasa (Takara), PCR Enzima de Alta Fidelidad Extensora (ABgene), ACCUZYME™ ADN Polimerasa (Bioline), SAHARA™ ADN Polimerasa (Bioline), VELOCITY ADN Polimerasa

(Bioline), GeneChoice® ADN Polimerasa AccuPOL™ GeneChoice, Inc.), GeneChoice® UniPOL™ ADN Polimerasa (GeneChoice, Inc.), Mezcla enzima Elongasa (Invitrogen), Pfx50™ ADN Polimerasa (Invitrogen), Phusion ADN Polimerasa (New England Biolabs), KOD HiFi ADN Polimerasa (Novagen), KOD XL ADN Polimerasa (Novagen), Expand 20 kb PLUS mezcla de ADN Polimerasa Termostable (Roche Applied Science), mezcla de ADN Polimerasa Termostable de alta fidelidad PLUS para expandirse (Roche Applied Science), mezcla de ADN Polimerasa Termostable de alta fidelidad para expandirse (Roche Applied Science), mezcla de ADN Polimerasa Termostable de alta fidelidad de molde largo para expandirse (Roche Applied Science), Easy-A™ Enzima para el clonaje de alta fidelidad (Stratagene), EXL™ ADN Polimerasa (Stratagene), Herculase® Enhanced ADN Polimerasa (Stratagene), Herculase® II Fusion ADN Polimerasa (Stratagene), Kapa LongRange™ ADN Polimerasa (Kapa Biosystems), Kapa HiFi™ ADN Polimerasa (Kapa Biosystems), Kapa2G™ ADN Polimerasa Robusta (Kapa Biosystems), Kapa2G™ HotStart ADN Polimerasa Robusta (Kapa Biosystems), Kapa2G™ ADN Polimerasa rápida (Kapa Biosystems), Kapa2G™ HotStart ADN Polimerasa rápida (Kapa Biosystems), LA TAQ ADN Polimerasa (Takara), Optimasa ADN Polimerasa (Transgenomic, Inc.), Exo- Pfu ADN Polimerasa (Stratagene), HotMaster Taq ADN Polimerasa (5 PRIME GmbH), HotTaq ADN Polimerasa (Abnova Corporation), AmpliTaq Gold® ADN Polimerasa (Applied Biosystems), Bst ADN Polimerasa Lg Frag (New England Biolabs), MasterAmp™ Tfl ADN Polimerasa (EPICENTRE Biotechnologies), Red Hot ADN Polimerasa (ABGene), Thermoprime Plus ADN Polimerasa (ABGene), Taq-red ADN Polimerasa (AppliChem GmbH), BIO-X-ACT™ Long ADN Polimerasa (Bioline), BIO-X-ACT™ Short ADN Polimerasa (Bioline), Bioline HybriPol™ ADN Polimerasa (Bioline), BioTherm Taq ADN Polimerasa (eEnzyme LLC), EU-Taq ADN Polimerasa (eEnzyme LLC), Synergy Taq ADN Polimerasa (eEnzyme LLC), GeneChoice® RedPOL™ ADN Polimerasa (GeneChoice, Inc.), AccuPrime™ GC-Rich ADN Polimerasa (Invitrogen), PyroPhage® 3173 ADN Polimerasa, Exo Minus (Lucigen), 9 Degrees North (Modified) ADN Polimerasa (New England Biolabs), Therminator ADN Polimerasa (New England Biolabs), Pwo ADN Polimerasa (Roche Applied Science), Paq5000™ ADN Polimerasa (Stratagene), YieldAce™ ADN Polimerasa (Stratagene), e2TAK™ ADN Polimerasa (Takara), o ADN polimerasas de origen natural de *P. kodakaraensis*, *P. furiosus*, *T. gorgonarius*, *T. zilligii*, *T. litoralis* "Vent™", *P. GB-D* "Deep Vent", *T. 9N-7*, *T. aggregans*, *T. barossii*, *T. fumicolans*, *T. celer*, *Pyrococcus sp. cepa ST700*, *T. pacificus*, *P. abyssi*, *T. profundus*, *T. siculi*, *T. hydrothermalis*, *Thermococcus sp. cepa GE8*, *T. thioreducens*, *P. horikoshii* o *T. onnurineus NA1*, *Thermococcus sp. 9°N-7*, *Thermococcus sp. GI-J*, *Thermococcus sp. MAR-13*, *Thermococcus sp. GB-C*, *Thermococcus sp. GI-H*, *Thermus aquaticus*, *Thermus thermophilus*, *Thermus caldophilus*, *Thermus filiformis*, *Thermus flavus*, *Thermotoga maritima*, *Bacillus stearothermophilus*, o *Bacillus caldotenax*.

Los datos obtenidos a través de la invención, por ejemplo, los resultados de las pruebas, pueden usarse en el proceso de diagnóstico de la presencia o ausencia de una afección o enfermedad. Los datos obtenidos a través del uso de la invención pueden usarse en la determinación del genotipo de un individuo. Los datos obtenidos a través de la invención pueden usarse para detectar genotipos asociados con síndrome de X frágil, síndrome de ataxia de temblor asociado con X frágil e insuficiencia ovárica primaria asociada con X frágil. Los loci genéticos asociados con estas afecciones son conocidos en la técnica e incluyen sin limitación FMR1, FMR2, el 5' UTR de FMR1, el 5' UTR de FMR2, las repeticiones de CGG dentro del 5' UTR de FMR1 y repeticiones de CGG dentro del 5' UTR de FMR2. En una modalidad adicional, los datos obtenidos a través de la invención pueden usarse para detectar genotipos asociados con trastornos de repetición de trinucleótidos ricos en GC y/o con elementos de interrupción, tales como síndrome de X frágil, síndrome de temblor/ataxia asociado a X frágil y la insuficiencia ovárica primaria asociada a X frágil, distrofia miotónica, enfermedad de Huntington, atrofia muscular espinobulbar, atrofia de atrofia dentato-rubro-pálido-luisiana, y/o ataxia espinocerebelosa. Un elemento de interrupción, como el nombre sugiere, interrumpe una serie de repeticiones. Una región rica en CGG puede comprender o no elementos de interrupción. Así, una región rica en CGG puede comprender series de repeticiones trinucleotídicas y al menos un elemento de interrupción entre las series. Los loci genéticos asociados con estas afecciones son conocidos en la técnica e incluyen sin limitación FMR1, FMR2, DMPK, ZNF9, HTT, AR, ATN1, ATXN1-3, ATXN7, ATXN10, CACNA1A, SCA8, PPP2R2B, y TBP. Ver, por ejemplo, Nat. Genet. 13(1):105-8 (1996); Nat. Genet. 13(1):109-13 (1996).

Los métodos pueden comprender la determinación de la presencia o ausencia de elementos de interrupción cerca de cualquiera de los extremos de una región rica en CGG comprendida por al menos un alelo comprendido en la muestra, por ejemplo, dentro de 60, 90, 120, 150, 180, 210, 240, 270 o 300 pb de cada extremo. Determinar la presencia o ausencia de elementos de interrupción dentro de una distancia dada de cada extremo de una región rica en CGG comprendida por al menos un alelo comprendido por la muestra se entiende que incluye determinar la presencia o ausencia de interrupción en regiones ricas en CGG con tamaños de manera que toda la región rica en CGG está dentro de la distancia dada de cada extremo.

En algunas modalidades, los métodos comprenden analizar al menos una región rica en CGG independientemente de los ensayos auxiliares o reflejos que comprenden procedimientos tales como transferencia de Southern, digestión por restricción o secuenciación, por ejemplo, secuenciación de Sanger, secuenciación de Maxam-Gilbert, secuenciación por ligación y secuenciación por síntesis de una sola molécula (además conocida como secuenciación de segunda generación), que incluye secuenciación terminadora reversible y pirosecuenciación. Sin embargo, los métodos pueden comprender la determinación de información sobre al menos una región rica en CGG tal como el contenido o la posición del elemento de longitud y/o de interrupción como se describe en la presente descripción. En algunas modalidades, los métodos comprenden obtener información que es necesaria y suficiente para la determinación de información acerca de al menos una región rica en CGG, tal como el contenido o la posición del elemento de longitud y/o interrupción como se describe en la presente descripción a partir de la representación del tamaño del producto y la abundancia producida por una técnica de alta resolución a partir de una reacción de amplificación como se describe en la presente descripción.

La información acerca de si una secuencia de interrupción está presente en al menos una región rica en CGG o donde dentro de al menos una región rica en CGG se localiza una secuencia de interrupción puede comprender información tal como si al menos una región rica en CGG comprendida por al menos un molde en una muestra comprende una secuencia de interrupción; en los casos donde más de un molde diferente está presente, si cada molde en una muestra comprende una secuencia de interrupción; una estimación de límite inferior del número de secuencias de interrupción presentes en al menos una región rica en CGG; y/o al menos una posición de al menos un elemento de interrupción (que incluyen la determinación dentro de un nivel dado de precisión, como se discute más abajo). La lista anterior de tipos específicos de información no incluye otros tipos de información que puedan determinarse a través de los métodos de la invención descritos explícita o implícitamente en la presente descripción, teniendo en cuenta el conocimiento de un experto en la técnica.

En algunas modalidades, los métodos comprenden realizar al menos dos ensayos, en donde que cada ensayo comprende (a) proporcionar iniciadores, en donde los iniciadores de al menos dos ensayos no son idénticos; (b) realizar una reacción de amplificación para producir un conjunto de productos; (c) resolver el conjunto de productos, produciendo al menos dos representaciones del tamaño del producto y la abundancia a partir de al menos dos ensayos; y (d) obtener información acerca de si una secuencia de interrupción está presente en al menos una región rica en CGG o donde dentro de al menos una región rica en CGG se localiza una secuencia de interrupción. En algunas modalidades, la información obtenida comprende información que no se pudo obtener si solamente se había realizado un ensayo, tal como, en los casos en que la muestra comprende al menos dos moldes diferentes que comprenden regiones ricas en CGG y al menos un elemento de interrupción está presente en al menos una de las regiones ricas en CGG, cuyas al menos dos regiones ricas en CGG comprenden al menos un elemento de interrupción. Por ejemplo, una situación en la que esta información puede ser ambigua a partir de los resultados de un ensayo se ilustra en la Figura 10. Otras posibles ambigüedades, y ejemplos de cómo resolverlas, se discuten más abajo.

En algunas modalidades, al menos dos ensayos comprenden un ensayo de anclaje y un ensayo sin anclaje. En algunas modalidades, al menos dos ensayos usan iniciadores orientados opuestamente que comprenden repeticiones de CGG, CCG, GCG, CGC, GCC o GGC. Es decir, un primer ensayo puede comprender proporcionar iniciadores que comprenden un iniciador que comprende repeticiones que se orienta corriente arriba, y el segundo ensayo puede comprender proporcionar iniciadores que comprenden un iniciador que comprende repeticiones que se orienta corriente abajo, o viceversa.

En algunas modalidades, los métodos comprenden al menos tres ensayos, que comprenden al menos dos ensayos sin anclaje y al menos un ensayo de anclaje, o viceversa. Al menos dos ensayos sin anclaje (o de anclaje) pueden usar iniciadores orientados opuestamente que comprenden repeticiones de CGG, CCG, GCG, CGC, GCC o GGC, como se discutió en el párrafo anterior.

Se debería señalar que en algunas situaciones, un "conjunto de productos" puede no comprender una pluralidad de productos principales o detectables, tales como ensayos de anclajes con muestras que comprenden como máximo una región rica en CGG que comprende como máximo una interrupción. Los iniciadores se consideran "no idénticos" entre ensayos si al menos uno de los iniciadores usados en uno de los ensayos difiere de los iniciadores usados en los otros ensayos; en otras palabras, algunos de los iniciadores pueden ser idénticos.

Los métodos pueden comprender determinar al menos un detalle de un genotipo heterocigoto, aneuploide o mosaico, realizando al menos dos ensayos como se describió anteriormente, en donde las reacciones de amplificación usan conjuntos de iniciadores no idénticos y al menos un detalle se determina comparando los resultados de al menos dos reacciones de amplificación.

El genotipo de una muestra se refiere al genotipo de la fuente a partir de la cual se obtuvo una muestra; este término puede abarcar múltiples genotipos individuales en el caso de muestras mixtas.

Una muestra que tiene un genotipo heterocigoto comprende material genético de una célula que comprende dos alelos del locus que comprende la región rica en CGG.

Una muestra que tiene un genotipo de mosaico comprende al menos dos alelos del locus que comprende la región rica en CGG, siendo las células de las que se obtuvo la muestra genéticamente diferentes en al menos un alelo del locus que comprende la región rica en CGG. Al menos dos alelos comprendidos por una muestra que tiene un genotipo de mosaico pueden clasificarse como alelos mayor o menor dependiendo de su abundancia. El alelo con mayor abundancia se considera un alelo principal y el alelo con la menor abundancia se considera un alelo menor; cuando están presentes más de dos alelos, los alelos se clasifican como mayores o menores basados en si su abundancia relativa, expresada como un porcentaje, es aritméticamente más próxima al alelo con la abundancia más alta o más baja. Por ejemplo, en una muestra que comprende alelos primero, segundo y tercero con abundancias relativas del 60%, 31% y 9%, respectivamente, el segundo alelo se considera un alelo menor. La presencia de alelos mayores y menores además puede resultar de la aneuploidía, por ejemplo, si un genoma comprende tres copias de una región rica en CGG, de las cuales dos son iguales (el alelo principal) y uno es diferente (el alelo menor). La neuploidía se discute en detalle más abajo.

En algunas modalidades, los métodos comprenden detectar si una muestra comprende un alelo menor con una abundancia relativa de al menos 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, o 25%. Se entiende que la región rica en CGG del alelo menor puede ser diferente de la región rica en CGG del alelo principal. Por ejemplo, el alelo menor puede tener una región rica en CGG con un número diferente de repeticiones y/o una distribución diferente de elementos AGG. En algunas modalidades, los métodos comprenden detectar si una muestra comprende un alelo menor que comprende un elemento de interrupción. En algunas modalidades, los métodos comprenden detectar la localización de un elemento de interrupción en un alelo menor, que puede estar dentro de un nivel de precisión como se describe más abajo.

Una muestra que tiene un genotipo aneuploide comprende un número de alelos del locus que comprende la región rica en CGG distinta del número euploide. El número euploide es uno para los loci cromosómicos somáticos en los gametos y células germinales que han sufrido una división meiótica reductora, excepto los loci cromosómicos sexuales en gametos masculinos y células germinales masculinas que han sufrido una división meiótica reductora, donde puede ser 0 o 1 para células individuales y promedios hasta aproximadamente 0,5 para una población de esta (sujeto a variación dependiendo del número de células que portan cromosomas X e Y presentes). El número euploide es además uno para los loci del cromosoma X en las células somáticas masculinas y las células germinales masculinas que aún no han sufrido una división meiótica reductora. El número euploide es de 2 para los loci en cromosomas somáticos de células somáticas y en células germinales que aún no han sufrido una división meiótica reductora y para los loci de cromosomas X en células femeninas somáticas y células germinales femeninas que aún no han sufrido una división meiótica reductora. Es posible que una muestra tenga más de una de las condiciones de heterocigosidad, aneuploidía y mosaicidad.

Los genotipos de mosaico pueden producirse, por ejemplo, en muestras de individuos que comprenden células derivadas de diferentes células progenitoras (cigotos, blastómeros, células madre, etc.) o individuos que comprenden células que son genéticamente diferentes debido a mutación somática, que incluyen alteración del número de repetición, tal como la expansión de la repetición. La aneuploidía está presente en ciertos síndromes, por ejemplo los de Down, Turner y Klinefelter, y además puede surgir somáticamente a través de eventos de no disyunción cromosómica; tales eventos pueden dar lugar a individuos que pueden proporcionar muestras aneuploides en mosaico, que pueden además ser o no heterocigotos.

Las muestras que comprendan al menos dos alelos diferentes del locus que comprende la región rica en CGG (debido a ser al menos uno de heterocigoto, aneuploide o mosaico para ese locus) pueden analizarse para determinar al menos un detalle con respecto al genotipo.

Al menos un detalle puede comprender si uno, dos, ninguno o más de dos (si corresponde) de al menos los dos alelos diferentes comprenden al menos un elemento de interrupción.

Al menos un detalle puede comprender el número mínimo de elementos de interrupción comprendidos por al menos uno de al menos dos alelos diferentes. Por ejemplo, se pudo determinar que un alelo comprende al menos un elemento de interrupción. Al menos un detalle puede comprender además que otro alelo comprende, por ejemplo, al menos dos elementos de interrupción.

Al menos un detalle puede comprender la localización de al menos un elemento de interrupción en uno de los alelos.

En algunas modalidades, al menos un detalle comprende al menos un detalle que no puede determinarse sin ambigüedad acerca del genotipo de una muestra heterocigótica, aneuploide y/o mosaica a partir de los resultados de un único ensayo. Lo siguiente es una lista no exclusiva de detalles que pueden ser ambiguos a partir de los resultados de un único ensayo:

1. Los resultados de un solo ensayo pueden indicar que al menos un elemento de interrupción está presente, pero dichos resultados pueden ser ambiguos en cuanto a qué el alelo comprende al menos un elemento de interrupción.
2. Los resultados de un único ensayo pueden indicar que al menos dos elementos de interrupción están presentes, pero dichos resultados pueden ser ambiguos en cuanto a si los elementos de interrupción están comprendidos por el mismo alelo o alelos diferentes.
3. Los resultados de un único ensayo pueden indicar que al menos tres elementos de interrupción están presentes, pero dichos resultados pueden ser ambiguos en cuanto a cómo al menos los tres elementos de interrupción se distribuyen entre al menos los dos alelos diferentes.
4. Los resultados de un único ensayo pueden ser ambiguos en cuanto a si un elemento de interrupción está presente en una posición en un primer alelo que está oscurecido en la representación del tamaño del producto y la abundancia por la señal, tal como un pico grande, que corresponde a un segundo alelo.

En algunas modalidades, los métodos de la invención resultan en la síntesis de productos que no contienen, o están libres de, una fracción sustancial de material amplificado inespecíficamente. En algunas modalidades, los métodos resultan en la síntesis de productos que no contienen, o están libres de una fracción sustancial de material amplificado inespecíficamente cuando el molde es una de las muestras descritas en los Ejemplos 1-3 más abajo. En algunas modalidades, los métodos de la invención resultan en la síntesis de productos que no contienen o están libres de una fracción sustancial de material amplificado inespecíficamente que es mayor que un tamaño esperado, calculándose

dicho tamaño esperado añadiendo la longitud de la región rica de CGG a un ajuste basado en la posición en la cual hibrida el segundo iniciador. En algunas modalidades, los productos contienen menos del 20%, 15%, 10%, 5%, 2% o 1% de material que es mayor que el tamaño esperado descrito en la oración anterior, determinado por densitometría o integración de una representación del tamaño y la cantidad del producto. En algunas modalidades, esta representación es un electroferograma. Como se define en la presente descripción, los productos son "esencialmente libres" de material amplificado inespecíficamente cuando contienen menos de aproximadamente 10% de material que es mayor que el tamaño esperado, como se determinó anteriormente.

La invención se refiere a métodos que comprenden el análisis de un molde que comprende secuencias repetidas ricas en GC, tales como, por ejemplo, trinucleótidos CGG o CCG. Dicho análisis puede comprender realizar una reacción de amplificación y resolver los productos a alta resolución. La alta resolución puede ser una resolución suficiente para distinguir productos que contienen, por ejemplo, 20 frente a 21, o 20 frente a 22, 20 frente a 23, 20 frente a 24, o 20 frente a 25 repeticiones de trinucleótidos, o en algunas modalidades, productos que difieren en longitud por 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 o 18 nucleótidos o pares de bases. Ejemplos de técnicas que pueden emplearse en los métodos de la invención para llevar a cabo esta etapa de resolución incluyen, sin limitación, electroforesis capilar y electroforesis en gel de poliacrilamida (PAGE; algunas modalidades de PAGE se denominan comúnmente en la técnica como "correr un gel de secuenciación"), o sistemas de electroforesis/microfluidos del tipo "lab-on-a-chip". La instrumentación para realizar la electroforesis capilar puede obtenerse, por ejemplo, de vendedores tales como Applied Biosystems (ABI), Beckman, Agilent y Qiagen, y los instrumentos adecuados incluyen sin limitación ABI 310, 3100, 3130/3130xl o 3730/3730xl; Beckman P/ACE MDQ; Agilent 2100 Bioanalyzer; y Qiagen QIAxcel. El proceso de resolver los productos mediante electroforesis puede implicar el uso de polímeros líquidos, que incluyen, por ejemplo, POP-7, POP-6, POP-5 o POP-4, todos los cuales se venden por Applied Biosystems. En algunas modalidades, se usa un polímero que puede resolver productos que contienen aproximadamente 250, 300, 350, 400 o más repeticiones de trinucleótidos de conformidad con precisamente el tamaño, tal como, por ejemplo, POP-4, para resolver los productos. La resolución del conjunto de productos a alta resolución puede realizarse usando una máquina, por ejemplo, elegida a partir de máquinas que comprenden una fuente de voltaje (por ejemplo, una fuente de alimentación de DC), máquinas que comprenden una fuente de presión (por ejemplo, una bomba), y máquinas que comprenden una columna o capilar adecuado para realizar separaciones químicas. Por supuesto, la máquina puede comprender más de uno de los componentes anteriores.

La resolución de los productos a alta resolución conduce a la producción de una representación del tamaño y la abundancia del producto. Esta representación puede ser una imagen o gráfica que un experto en la técnica pueda interpretar visualmente o con ayuda de una instrumentación tal como, por ejemplo, un ordenador con programa apropiado o un densitómetro, para comprender los tamaños y las cantidades de los productos de la reacción. En algunas modalidades, la representación es un electroferograma, fotografía, gráfico, diagrama o autorradiograma. La representación puede obtenerse o registrarse a partir de fotones o partículas beta emitidas por los productos o moléculas de colorantes unidas a los productos; éstos pueden detectarse, por ejemplo, fotográficamente o electrónicamente, y procesarse o desarrollarse para generar la representación.

La invención se refiere a métodos que comprenden obtener información acerca del número de repetición CGG (es decir, cuántas repeticiones de CGG, o repeticiones de trinucleótidos generalmente, están comprendidas por el molde) a partir de la representación del tamaño y la abundancia del producto. Tal obtención puede comprender contar el número de especies observables en la representación para determinar el número de repetición (a partir del número de repeticiones comprendidas por el producto más pequeño, que puede ser, por ejemplo, 4 o 5) o estimar el número de repeticiones basado en la posición del producto más grande observable en la representación.

En algunas modalidades, la región rica en CGG está compuesta por la 5' UTR de FMR1 o la 5' UTR de FMR2.

En algunas modalidades, los métodos comprenden detectar si hay elementos de interrupción presentes dentro de la región rica en CGG. En algunas modalidades, los elementos de interrupción son trinucleótidos AGG. La detección de si están presentes puede lograrse, por ejemplo, determinando posiciones en el molde donde la unión del primer iniciador se redujo esencialmente o determinando un conjunto de longitudes en las que la cantidad de producto se reduce esencialmente en comparación con longitudes cercanas. Por ejemplo, si el 10^{mo} trinucleótido fuera AGG en una región con al menos 14 repeticiones totales de trinucleótidos y se realizó una reacción de amplificación que implica el uso de un iniciador que comprendió 4 repeticiones de CGG, puede esperarse que los productos con 10, 11, 12 y 13 repeticiones de CGG están presentes en una cantidad esencialmente reducida con relación a los productos cercanos con 9 y 14 repeticiones. El grado en que se reduce la cantidad puede variar de 25% a 95% o más, por ejemplo, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, o 95%. El grado de reducción dependerá generalmente de la cigosidad de la muestra; por ejemplo, una muestra en donde la región rica en CGG es de un heterocigoto individual para un alelo que comprende la región rica en CGG, o para una repetición AGG específica, puede mostrar una reducción en la cantidad de los productos correspondientes en el intervalo de 25% y 75%. Una muestra en donde la región rica en CGG es de un hemocigoto o homocigótico individual para un alelo que comprende la región rica en CGG, o para una repetición AGG específica, puede mostrar una reducción en la cantidad de los productos correspondientes en el intervalo de 50% a 95% o más.

En algunas modalidades, los métodos no requieren el uso de estándares externos o calibradores tales como, por ejemplo, patrones de referencia o patrones moleculares. Esta independencia puede lograrse porque en algunas modalidades pueden usarse picos correspondientes a productos que difieren en longitud por una repetición trinucleotídica para contar hasta el tamaño del producto más grande, en donde el conteo puede indicar el número de repeticiones en el molde (sujeto a cualquier ajuste necesario de conformidad con el número de repeticiones comprendidas por el iniciador). En algunas modalidades, la separación entre picos en parte de la representación del tamaño del producto y la cantidad puede usarse para estimar el tamaño del producto más grande, por ejemplo, por extrapolación; esto puede ser útil cuando la resolución de los picos correspondientes a números de repetición más grandes es insuficiente para contar todos los picos hasta e incluir el del producto más grande.

En algunas modalidades, los métodos pueden determinar el número de repeticiones en la región rica en CGG hasta dentro de 100, 50, 20, 10, 5, 4, 3, 2, 1, o 0 repeticiones. Determinación de una cantidad "dentro de 0 unidades" significa que se determina su valor preciso. En algunas modalidades, los métodos comprenden determinar el número de repeticiones en la región rica en CGG hasta dentro de las repeticiones de 100, 50, 20, 10, 5, 4, 3, 2, 1, o 0 repeticiones. Por ejemplo, los métodos pueden comprender determinar el número de repeticiones en una región rica en CGG que comprende menos de 70 repeticiones de CGG hasta dentro de 5, 4, 3, 2, 1 o 0 repeticiones; determinar el número de repeticiones en una región rica en CGG que comprende de 70 a 120 repeticiones de CGG hasta dentro de 10, 5, 4 o 3 repeticiones; o determinar el número de repeticiones en una región rica en CGG que comprende más de 120 repeticiones de CGG, hasta aproximadamente 200 repeticiones de CGG, hasta dentro de 20, 10 o 5 repeticiones. La determinación de los niveles de exactitud para regiones repetidas CGG aún más grandes puede ser difícil debido a la falta de estándares conocidos fiables.

En algunas modalidades, los métodos pueden determinar la posición de un elemento de interrupción dentro de los nucleótidos de 60, 45, 30, 15, 12, 9, 6, 5, 4, 3, 2, 1 o 0. La posición del elemento de interrupción puede determinarse, por ejemplo, en términos de distancia con relación a cualquiera de los extremos de la región rica en CGG.

En algunas modalidades, al menos uno de los iniciadores comprende una porción radiológicamente o electromagnéticamente detectable. Las porciones radiológicamente detectables incluyen isótopos radiactivos que emiten partículas detectables, tales como partículas beta o gamma, por ejemplo, ¹⁴C, ³H, ³²P, ³³P, ³⁵S, y ¹²⁵I. Las porciones electromagnéticamente detectables incluyen entidades químicas que interactúan con radiación electromagnética (incluyendo absorbancia, emisión o ambas) de una manera detectable, tales como cromóforos y fluoróforos, por ejemplo, fluoresceína, FAM, tintes de cianina, tintes de rodamina, etc.

Los aspectos adicionales de la invención se expondrán en parte en la descripción que sigue, y en parte serán evidentes a partir de la descripción, o pueden aprenderse al llevar a la práctica la invención. Los aspectos de la invención se realizarán y conseguirán por medio de los elementos y combinaciones particularmente destacadas en las reivindicaciones anexas.

Breve descripción de los dibujos

Los dibujos adjuntos ilustran una o más modalidades de la invención y, junto con la descripción, sirven para explicar los principios de la invención.

La Figura 1 describe la amplificación de un molde rico en CGG que comprende los trinucleótidos AGG. Las secuencias de iniciadores son

FAM_FX-F	5'-FAM-TCAGGCGCTCAGCTCCGTTTCGGTTTCACTTCCGGT (sec. con núm. de ident.: 38)
Tag-(CCG) ₄	AGCGTCTACTGTCTCGGCACTTGCCCGCCGCCGCG (sec. con núm. de ident.: 39)

Un molde que comprende (CGG)₁₀AGG(CGG)₉AGG(CGG)₉ (sec. con núm. de ident.: 42) se muestra. Representa una posible región de repetición CGG en 5' UTR' de FMR1. El iniciador Tag-(GCC)₄ puede unirse internamente en múltiples posiciones en la región de repetición; con el iniciador directo marcado con FAM (FAM-FX-F), que se hibrida corriente arriba de la región de repetición CGG, puede amplificar una pluralidad de productos de PCR. El amplicón CGG más corto tendrá 4 repeticiones de CGG y el amplicón CGG más largo comprenderá la longitud completa de la sec. con núm. de ident.: 40. Todos los productos que son significativamente más largos que los productos de longitud completa se consideran productos no específicos.

Las Figuras 2A y 2B muestran electroferogramas de los resultados de la PCR a partir de 5 muestras clínicas (AFM104, AFB107, ABB001, AFB011 y AMB12, como se indica en las áreas de trazado en la Figura 2A y el panel superior izquierdo de la Figura 2B) y tres patrones de Coriell (31/46 CGG, 31/54 CGG, y 30/75 CGG, como se indica en las áreas de trazado de la Figura 2B diferentes de la parte superior izquierda). Para las muestras clínicas, la secuencia de las regiones ricas en CGG según se determina por la secuenciación se ordena en las áreas de trazado. Para los patrones de Coriell, el contenido de repetición CGG de los dos alelos 5' UTR' de FMR1 presentes en la muestra como se indicó

5 por el electroferograma se clasifica como "Confirmado". Los picos seleccionados se etiquetan con el número de trinucleótidos CGG comprendidos por el producto correspondiente al pico. Las unidades en el eje vertical representan la intensidad de fluorescencia y son arbitrarias. Las unidades en el eje horizontal representan el tamaño estimado en nucleótidos. Estos tamaños estimados se determinaron sobre la base del número de exploración (relacionado con el tiempo de retención) del instrumento de la CE.

10 Las Figuras 3A y 3B muestran electroferogramas de los resultados de PCR de ocho patrones de Coriell. En cada uno, el pico correspondiente al número máximo detectado de trinucleótidos CGG se indica cuando fue soluble. En los casos en que el número de trinucleótidos difiere de la designación Coriell o cuando se detectan trinucleótidos AGG, se indica la secuencia deducida del electroferograma en el área de trazado. En el panel superior derecho de la Figura 3B, no pudieron resolverse todos los picos correspondientes a especies de productos individuales, pero la posición del pico más a la derecha correspondió a un contenido estimado de 122 trinucleótidos CGG. En los dos paneles inferiores, no se detectó la amplificación de los productos cercanos a la longitud total de los alelos más largos. Las etiquetas de los ejes horizontal y vertical son como en la Figura 2A.

15 La Figura 4 representa un ensayo usando tres iniciadores. FMR1R-FAM se hibrida corriente abajo de la región de repetición CGG y comprende un fluoróforo FAM, para uso en la detección de productos. El iniciador quimérico comprende repeticiones de CGG y además, una aleta 5' que coincide con la secuencia del tercer iniciador, en este caso FMR1-F, que se hibrida con el molde original como se indica, pero hibrida también con los productos que comprenden la secuencia iniciadora quimérica. El iniciador quimérico se proporciona a una concentración menor tal que se consume rápidamente para limitar la reducción progresiva en el tamaño del producto de ciclo a ciclo.

20 La Figura 5 muestra electroferogramas de PCR de un ensayo de tres iniciadores de los patrones de Coriell heterocigóticos indicados en cada área de trazado. Las etiquetas de los ejes horizontal y vertical son como en la Figura 2A. En cada panel, el pico más alto corresponde al menor de los dos alelos presentes en la muestra y se indica de conformidad con el número de repeticiones de CGG determinadas por este ensayo. Se contaron picos hasta el pico 190, que se muestra como una ampliación de la inserción.

25 La Figura 6 muestra los electroferogramas de los resultados de PCR a partir de cinco muestras de ADN genómico (paneles A, B, C, D y E). Los paneles A, B y C se derivan de muestras de ADN genómico masculino. Los paneles D y E se derivan de muestras de ADN genómico femenino. Los picos seleccionados se etiquetan con el número de trinucleótidos CGG comprendidos por el producto correspondiente al pico. Las secuencias previstas de los alelos genómicos se enumeran en las áreas de trazado. Las unidades en el eje vertical representan la intensidad de fluorescencia y son arbitrarias. Las unidades en el eje horizontal representan el tamaño estimado en nucleótidos. Como se describe en los Ejemplos, se determinó que las muestras que se usaron para generar los perfiles de los paneles A-E tienen regiones de repetición CGG con las siguientes secuencias. Panel A: 5'-(CGG)₁₀AGG(CGG)₉AGG(CGG)₉-3' (sec. con núm. de ident.:46). Panel B: 5'-(CGG)₉AGG(CGG)₉AGG(CGG)₂₇-3' (sec. con núm. de ident.:47). Panel C: 5'-(CGG)₆₁-3' (sec. con núm. de ident.:48). Panel D: 5'-(CGG)₂₀-3' (sec. con núm. de ident.:49) y 5'-(CGG)₁₀AGG(CGG)₉AGG(CGG)₁₀-3' (sec. con núm. de ident.:50). Panel E: 5'-(CGG)₉AGG(CGG)₉AGG(CGG)₂₆-3' (sec. con núm. de ident.:51) y 5'-(CGG)₁₀AGG(CGG)₈₆-3' (sec. con núm. de ident.:46).

30 La Figura 7 muestra los electroferogramas de la CE previstos (paneles superior e inferior lateral izquierdo) para dos asignaciones de repetición AGG diferentes para diferentes alelos (panel superior, 5'-(CGG)₉AGG(CGG)₁₀-3' (sec. con núm. de ident.:53) y 5'-(CGG)₁₀AGG(CGG)₂₀-3' (sec. con núm. de ident.: 54); panel inferior, 5'-(CGG)₂₀-3' (sec. con núm. de ident.:49) y 5'-(CGG)₁₀AGG(CGG)₉AGG(CGG)₁₀-3' (sec. con núm. de ident.:50)) y un electroferograma de CE de simulación para productos de amplificación de ambos alelos (panel inferior derecho). El electroferograma de CE observado se muestra en el panel superior derecho.

35 La Figura 8 representa ocho ensayos de amplificación que pueden usarse en los métodos de la invención. FMR1_F hibrida corriente arriba de la región de repetición CGG. FMR1_R hibrida corriente abajo de la región de repetición CGG. FMR1_F_FAM y FMR1_R_FAM se corresponden con los mismos dos iniciadores respectivamente, pero que comprenden un fluoróforo FAM, para uso en la detección de productos. FMR1_F_(CGG)_n es un iniciador quimérico que comprende repeticiones de CGG y una solapa 5' que coincide con la secuencia de FMR1_F, que se hibrida con el molde original como se indica e hibrida también a productos que comprenden la secuencia iniciadora quimérica. El iniciador quimérico puede proporcionarse a una concentración menor de modo que se consume rápidamente, para limitar la reducción progresiva en el tamaño del producto de ciclo a ciclo. FMR1_R_(CGG)_n es un iniciador quimérico que comprende repeticiones CCG y una solapa 5' que coincide con la secuencia de FMR1_R, que se hibrida con el molde original como se indica e hibrida también a productos que comprenden la secuencia iniciadora quimérica. FMR1_F_(CGG)_nA es un iniciador quimérico correspondiente a FMR1_F_(CGG)_n pero que tiene además un dA terminal 3'. El dA terminal 3' se hibrida al T en la cadena de ADN en la región de complementariedad con la secuencia AGG presente en algunas regiones de repetición CGG. FMR1_R_(CGG)_nCCT es un iniciador quimérico con la misma secuencia que FMR1_R_(CGG)_n pero también tiene un dCdCdT terminal 3'. El dCdCdT terminal se hibrida con la secuencia de AGG presente en la cadena de ADN complementaria en algunas regiones de repetición CGG. Las secuencias de iniciadores usadas para estos ensayos pueden ser, por ejemplo:

65

	FMR1_F	TCAGGCGCTCAGCTCCGTTTCGGTTTCA (sec. con núm. de ident.:14)
5	FMR1_F_FAM	5'-FAM-TCAGGCGCTCAGCTCCGTTTCGGTTTCA (sec. con núm. de ident.:14)
	FMR1_F_(CGG) ₅	TCAGGCGCTCAGCTCCGTTTCGGTTTCACGGCGGGCGGCGGCGG (sec. con núm. de ident.: 41).
	n=5	
10	FMR1_R	AAGCGCCATTGGAGCCCCGCACTTCC (sec. con núm. de ident.:37)
	FMR1_R_FAM	5'-FAM-AAGCGCCATTGGAGCCCCGCACTTCC (sec. con núm. de ident.:37)
15	FMR1_R_(CCG) ₅	AAGCGCCATTGGAGCCCCGCACTTCCCCGCCGCCGCCGCCG (sec. con núm. de ident.:43)
	n=5	
	FMR1_F_(CGG) ₅ A	TCAGGCGCTCAGCTCCGTTTCGGTTTCACGGCGGGCGGCGGCGGA (sec. con núm de ident.: 44)
20	n=5	
	FMR1_R_(CCG) ₄ CCT	AAGCGCCATTGGAGCCCCGCACTTCCCCGCCGCCGCCGCCCT (sec. con núm. de ident.: 45)
	n=4	

25

Las secuencias FMR_F y FMR_R mostradas pueden substituirse por cualquier otra secuencia de iniciador adecuada de las regiones 5 'y 3' que flanquean las repeticiones de CGG, respectivamente.

30

La Figura 9 describe la amplificación de un molde rico en CGG que comprende trinucleótidos AGG, de conformidad con el esquema de PCR mostrado en la Figura 8H. Las porciones superior e inferior de la figura representan dos posibles configuraciones de alelos presentes en la muestra de ADN genómico cuyo electroferograma de CE real se muestra en la Figura 6D. La porción superior describe la amplificación de dos alelos incluyendo un alelo de 31 repeticiones que tiene una secuencia AGG presente después de la 10^{ma} repetición CGG (5'-(CGG)₁₀AGG(CGG)₂₀3' (sec. con núm. de ident.:54)) y un alelo de 20 repeticiones que tiene un AGG presente después de la 9^{na} repetición de CGG (5'-(CGG)₉AGG(CGG)₁₀-3' (sec. con núm. de ident.:53)). La porción inferior describe la amplificación de dos alelos que incluyen un alelo de 20 repeticiones que no tiene secuencia AGG dentro de las repeticiones de CGG (5'-(CGG)₂₀-3' (sec. con núm. de ident.: 49)) y un alelo de 31 repeticiones que tiene un AGG presente en la posición de repetición 11 y un AGG en la posición de repetición 21 (5'-(CGG)₁₀AGG(CGG)₉AGG(CGG)₁₀-3' (sec. con núm. de ident.: 50)). El panel de la derecha muestra un electroferograma de CE real de productos de amplificación de la muestra de ADN genómico, que demuestra la presencia de productos de amplificación que representan la configuración del alelo en la porción inferior de la figura.

40

45

La Figura 10 muestra electroferogramas simulados de la CE de productos de amplificación de alelos que se generan mediante el esquema de PCR mostrado en la Figura 8B a partir de la muestra de ADN genómico cuyo electroferograma de CE real se muestra en la Figura 6E y que contiene alelos que tienen 46 y 97 repeticiones de CGG. Después de la amplificación de conformidad con el esquema de PCR de la Figura 8B, dos configuraciones de alelos posibles (paneles de la izquierda, 5'-(CGG)₉AGG(CGG)₉AGG(CGG)₂₆-3' (sec. con núm. de ident.: 51) y 5'-(CGG)₁₀AGG(CGG)₈₆-3' (sec. con núm. de ident.: 52); paneles de la derecha, 5'-(CGG)₄₆-3' (sec. con núm. de ident.: 55) y 5'-(CGG)₁₀AGG(CGG)₄₉AGG(CGG)₉AGG(CGG)₂₆-3' (sec. con núm. de ident.: 56)) pueden generar electroferogramas de la CE idénticos como se muestra en los paneles inferiores izquierdo y derecho y en la Figura 6E.

50

55

La Figura 11 muestra electroferogramas simulados de la CE de productos de amplificación de alelos que pueden generarse mediante el esquema de PCR mostrado en la Figura 8D a partir de la muestra de ADN genómico cuyo electroferograma de CE real se muestra en la Figura 6E y que contiene alelos que tienen 46 y 97 repeticiones de CGG. Esta figura simula los productos de amplificación a partir de las dos posibles configuraciones de alelo (paneles de la izquierda, 5'-(CGG)₉AGG(CGG)₉AGG(CGG)₂₆-3' (sec. con núm. de ident.: 51) y 5'-(CGG)₁₀AGG(CGG)₈₆-3' (sec. con núm. de ident.: 52); paneles de la derecha, 5'-(CGG)₄₆-3' (sec. con núm. de ident.: 55) y 5'-(CGG)₁₀AGG(CGG)₄₉AGG(CGG)₉AGG(CGG)₂₆-3' (sec. con núm. de ident.: 56)), seguido la amplificación por el esquema de ensayo de PCR mostrado en la Figura 8D, en el que se invierte la direccionalidad del iniciador de repetición (FMR_R_(CCG)_n). Los dos alelos pueden generar escenarios de diferentes perfiles de la CE (Figura 11, paneles inferiores).

60

65

La Figura 12 muestra un electroferograma de CE real de los productos de amplificación de alelos de la misma muestra de ADN genómico descrita en las descripciones de las Figuras 10 y 11 anteriores (que contienen el alelo de 46/97 repeticiones CGG). Los electroferogramas de la CE se generaron a partir de muestras amplificadas de conformidad con el esquema de ensayo de PCR en la Figura 8B (panel superior) o Figura 8F (panel inferior).

La Figura 13 muestra esquemáticamente los productos de amplificación esperados de la misma muestra de ADN genómico descrita en las descripciones de las Figuras 10, 11 y 12 anteriores (que contienen el alelo de 46/97 repeticiones CGG) (5'-(CGG)₉AGG(CGG)₉AGG(CGG)₂₆-3' (sec. con núm. de ident.: 51), 5'-(CGG)₁₀AGG(CGG)₈₆-3' (sec. con núm. de ident.: 52), 5'-(CGG)₄₆-3' (sec. con núm. de ident.: 55), y 5'-(CGG)₁₀AGG(CGG)₄₉AGG(CGG)₉AGG(CGG)₂₆-3' (sec. con núm. de ident.: 56)), después de la amplificación por el esquema de ensayo de PCR mostrado en la Figura 8F.

La Figura 14 muestra esquemáticamente los productos de amplificación que pueden esperarse usando el esquema de ensayo de PCR de la Figura 8H, después de la amplificación de genómica ADN (descrito en las descripciones de las Figuras 10-13 anteriores y que contiene el alelo de 46/97 repeticiones de CGG) con dos alelos en escenarios diferentes (parte superior de la figura: 5'-(CGG)₉AGG(CGG)₉AGG(CGG)₂₆-3' (sec. con núm. de ident.: 51) y 5'-(CGG)₁₀AGG(CGG)₈₆-3' (sec. con núm. de ident.: 52); porción inferior de la figura: 5'-(CGG)₄₆-3' (sec. con núm. de ident.: 55) y 5'-(CGG)₁₀AGG(CGG)₄₉AGG(CGG)₉AGG(CGG)₂₆-3' (sec. con núm. de ident.: 56)). El ensayo de PCR de la Figura 8H utiliza un iniciador dT anclado de orientación inversa (FMR1_R_(CCG)nCCT) (Figura 8H). Este ensayo debe producir un electroferograma de CE con picos de 14, 15 y 24 CGG de tamaños equivalentes para el escenario de alelos representado en la parte superior de la figura, en el que el alelo de 46 repeticiones tiene dos AGG y el alelo de 97 repeticiones tiene un AGG, o el ensayo debe producir un electroferograma de CE con picos de 15, 65 y 75 CGG de tamaños equivalentes para el escenario de alelos representado en la porción inferior de la figura, en el que el alelo de 46 repeticiones no tiene AGG y el alelo de 97 repeticiones tiene las tres AGG. La figura muestra también el electroferograma de CE real (porción superior derecha de la figura) después de la amplificación de esta muestra de ADN genómico. El electroferograma es consistente con el alelo en el escenario mostrado en la parte superior izquierda de la figura.

La Figura 15 representa electroferogramas de la CE después de la amplificación de muestras de ADN genómico y muestras de ADN genómico mixto, de conformidad con el método de ensayo de la invención mostrado en la Figura 8C. Se analizaron por separado dos muestras de ADN diferentes, una con un alelo de 30 repeticiones de CGG (panel A) y otra con un alelo de 645 repeticiones de CGG (panel B). También se analizó una combinación artificial de ADN cromosómico de estas dos muestras (panel C) para demostrar que el método era capaz de mapear la presencia o ausencia de trinucleótidos AGG cerca del extremo 5' de repeticiones de CGG largas en presencia de un alelo más corto que tiene insertos del trinucleótido AGG.

La Figura 16 muestra electroferogramas de la CE después de la amplificación del ADN cromosómico de la muestra 20 (Tabla 4). La muestra tiene dos alelos principales, uno de los cuales es un alelo de mutación completa y dos alelos menores. Se cree que la configuración del alelo se deriva de una población de células en mosaico dentro de la muestra. Panel A, productos de amplificación después de la amplificación por PCR mediante el esquema mostrado en la Figura 8A. Panel B, productos de amplificación después de la amplificación por PCR mediante el esquema mostrado en la Figura 8H. Panel C, productos de amplificación después de la amplificación por PCR mediante el esquema mostrado en la Figura 8D. Además de demostrar las capacidades de mapeo de AGG/CGG de los ensayos, los datos demuestran también la presencia de un trinucleótido AGG en un alelo de mutación completa.

La Figura 17 muestra electroferogramas de la CE después de la amplificación del ADN cromosómico de la muestra 27 (Tabla 4). Panel A, productos de amplificación después de la amplificación por PCR mediante el esquema mostrado en la Figura 8A. Panel B, productos de amplificación después de la amplificación por PCR mediante el esquema mostrado en la Figura 8H. Panel C, productos de amplificación después de la amplificación por PCR mediante el esquema mostrado en la Figura 8C. Tenga en cuenta que el Panel C se muestra a una escala más cercana que el Panel A, de modo que la región donde podría aparecer el pico de longitud completa no está visible.

La Figura 18 muestra electroferogramas de la CE después de la amplificación del ADN cromosómico de la muestra 6 y 20 (Tabla 4). Los paneles A y B muestran los productos de amplificación de las muestras 6 y 20 respectivamente, después de la amplificación por PCR mediante el esquema mostrado en la Figura 8F. Como se describe en los Ejemplos, las muestras que se usaron para generar los perfiles de los paneles A-E se determinaron que tienen regiones de repetición CGG en mosaico, compuestas de las siguientes secuencias de repetición CGG. Panel A: 5'-(CGG)₉AGG(CGG)₉AGG(CGG)₉-3' (sec. con núm. de ident.:57), 5'-(CGG)₉AGG(CGG)₇AGG(CGG)₂₄-3' (sec. con núm. de ident.:58) (alelo menor), y 5'-(CGG)₉AGG(CGG)₇AGG(CGG)₄₂-3' (sec. con núm. de ident.:59). Panel B: 5'-(CGG)₉AGG(CGG)₉AGG(CGG)₉-3' (sec. con núm. de ident.:57), 5'-(CGG)₉AGG(CGG)₄₄-3' (sec. con núm. de ident.: 60) y 5'-(CGG)₉AGG(CGG)₆₀-3' (sec. con núm. de ident.: 61).

La Figura 19 representa un flujo de trabajo para mapear trinucleótidos AGG dentro de repeticiones de CGG en la región 5' no traducida de los genes FMR1 y FMR2. Los ensayos A, C, G y F se refieren a ensayos de PCR correspondientes mostrados en la Figura 8.

Ejemplos

Se hará referencia en detalle a modalidades de la invención, cuyos aspectos y resultados se ilustran en los dibujos que se acompañan. A los efectos de una mayor claridad y continuidad, se proporcionan a continuación varios segmentos de

la discusión e interpretación de los métodos y resultados de ciertos ejemplos; la presentación de ejemplos se resume siguiendo estos segmentos.

5 Ejemplo 1: Determinación del número de repetición CGG y la posición AGG en el promotor FMR1 para los alelos de premutación normales y bajos mediante un ensayo de PCR con iniciador de repetición y electroforesis capilar de alta resolución

10 Ocho muestras de ADN genómico que contienen números de premutación normal a baja de repeticiones de CGG (5 muestras clínicas: AFM104, AFB107, ABB001, AFB011, y AMB12; y tres patrones Coriell: 31/46 CGG, 31/54 CGG y 30/75 CGG) se evaluaron de la siguiente manera. Se usaron iniciadores de sec. con núms. de ident.: 38-39. Las condiciones de reacción de PCR que se usaron se basaron en un protocolo publicado (Saluto y otros, J. Mol. Diagn. 7: 605-12 (2005)) con ligeras modificaciones. 15 a 20 ng de ADN genómico se amplificaron en un regulador de reacción que contiene regulador de PCR 2 para molde de ampliación Roche (núm. de catálogo Roche 11681834001) más 2,2 M de betaína (núm. de catálogo Sigma B0300-1VL), 250 µM de cada dNTP (Roche, grado GMP núm. de catálogo G 04631129103, C 04631072103, A 04631056103, T 04631137103), 1,5 µM de cada iniciador, y 1,25 U de Taq recombinante de ADN polimerasa Roche GMA (Roche, núm. de catálogo 03734935001), en un volumen de reacción de 15 µl. Las condiciones de ciclación por PCR fueron 95°C durante 5 min; después 10 ciclos de 97°C durante 35 seg - 62°C durante 35 seg - 68°C durante 4 min; después 20 ciclos de 97°C durante 35 seg - 62°C durante 35 seg - 68°C durante 4 min con 20 seg de auto-extensión por ciclo. 1 µl de productos de PCR se mezclaron con 2 µl de patrón ROX 1007 (preparado de conformidad con DeWoody y otros, Biotechniques 37:348, 350, 352 (2004)) en 12 µl de formamida Hi-Di™ (Applied Biosystems (ABI) núm. de parte 4311320) y se desnaturalizó por calor a 95°C durante 2 min antes de la electroforesis capilar en un instrumento ABI 3130xl con una longitud capilar de 36 cm usando el polímero líquido POP7 (ABI núm. de parte 4352759). Los electroferogramas resultantes se muestran en las Figuras 2A y 2B.

25 Los picos en los electroferogramas se numeraron comenzando con 4, el contenido mínimo de producto CGG posible. Una reducción severa en la intensidad máxima desde el pico n a n+1, por ejemplo, desde el pico 10 a 11, indicó la presencia de un trinucleótido AGG en la posición correspondiente al pico n+1. Un trinucleótido resultó cuatro picos de baja intensidad, que se cree que sean porque el trinucleótido AGG redujo la afinidad del iniciador que contiene CGG para las cuatro posiciones de unión que abarcan ese trinucleótido (recuerde que el iniciador con la secuencia de sec. con núm. de ident.: 39, contiene cuatro repeticiones de CGG). El número total de trinucleótidos se determinó contando el número total de picos, siendo el primero numerado como 4 se describió anteriormente. Se cree que el pico pequeño a la derecha de cada panel de la Figura 2A y en el panel superior izquierdo de la Figura 2B resulta de sucesos de hibridación en donde, sólo se hibridan 3 repeticiones de CGG del iniciador y molde y por lo tanto no refleja la longitud real del molde. Se confirmaron mediante técnicas estándar el contenido total de repetición y las localizaciones de repetición de trinucleótidos AGG secuenciando las cinco muestras clínicas (ver Tabla 1). Para las cinco muestras clínicas, los resultados obtenidos a través del método descrito coincidieron con los resultados obtenidos a través de la secuenciación.

Tabla 1

40

ID de la muestra	Fuente	# de CGG	posición AGG determinada por secuenciación	posición AGG determinada por repetición de PCR	# de CGG determinado por repetición de PCR
45 AFM104	Lavado bucal donante#4 Asuragen	30/30	11, 21	11, 21	30
AFB107	donante de sangre Asuragen#7	30/30	11, 21	11, 21	30
50 AMB001	donante de sangre Asuragen#1	29	10, 20	10, 20	29
AMB011	donante de sangre Asuragen#11	29	10, 20	10, 20	29
55 AMB012	donante de sangre Asuragen#12	37	10, 18, 28	10, 18, 28	37
NA20234	Coriell	31	N/A		31
		46	N/A		46
60 NA20236	Coriell	31	N/A		31
		54	N/A		54
NA20242	Coriell	30	N/A		30
65		73	N/A		75

Ejemplo 2: Dificultad para determinar el número y posición de la repetición en algunos alelos largos de muestras femeninas con el ensayo de PCR con iniciación de dos iniciadores de repetición

5 Para evaluar este ensayo con muestras que comprenden repeticiones de CGG desde el intervalo normal hasta la completa mutación, otro conjunto de ocho muestras, a saber, dos muestras clínicas de sangre completa (ID de las muestras 00100 y 00065, correspondientes a los paneles de la Figura 3 marcados con CGG 69 y CGG 87) y seis muestras de ADN genómico de línea celular Coriell (con designaciones Coriell que indican CGG 56, 76, 96, 118, 20/183-193 y 28/336). Las reacciones de PCR se realizaron como se describe en el Ejemplo 1. Los electroferogramas resultantes se muestran en las Figuras 3A-3B. En estos electroferogramas, el pico final a contar se determinó mediante inspección visual. Para las seis muestras masculinas (todos los paneles de la Figura 3A y los dos paneles superiores de la Figura 3B), este ensayo fue capaz de determinar tanto el tamaño (número de repetición) como la posición de los trinucleótidos AGG (ver tabla 2). Los números repetidos coincidieron con las designaciones de Coriell o el número obtenido mediante secuenciación hasta un 4%. El ensayo no pudo detectar el número total de repeticiones presentes en los alelos más largos de las muestras femeninas (paneles inferiores de la Figura 3B).

15 Tabla 2

ID de la muestra	Fuente	# de CGG	posición AGG determinada por secuenciación	posición AGG determinada por repetición de PCR	# de CGG determinado por repetición de PCR
CD00014	Línea celular Coriell	56	11,21	11,21	57
00100	Sangre	69	N/A	núm de AGG	69
NA20231	Línea celular Coriell	76	11	11	78
00065	Sangre	87	N/A	núm de AGG	87
NA06906	Línea celular Coriell	96	11	11	100
NA06891	Línea celular Coriell	118	11	11	122
NA20239	Línea celular Coriell	20/182-193	N/A	11 (20 CGG)	20/>60
NA07537	Línea celular Coriell	28/336	N/A	10, 20 (29 CGG)	29>50

Ejemplo 3: Determinación del número de repetición CGG y de la posición AGG en alelos de premutación y mutación completa usando el sistema modificado de tres iniciadores para la PCR con iniciador de repetición

45 Para aumentar el número de repeticiones que se pueden detectar, el procedimiento se modificó como se indica en la Figura 4. Se usaron tres iniciadores. El primer iniciador fue un iniciador quimérico que contenía 5 repeticiones de CGG en el extremo 3' y tenía la secuencia de sec. con núm. de ident.: 41. El segundo iniciador fue un iniciador inverso con relación al primer iniciador; este segundo iniciador tenía la secuencia de sec. con núm. de ident.: 37. El tercer iniciador se orientó hacia delante con respecto al iniciador quimérico y tenía la misma secuencia que el iniciador quimérico pero sin las 5 repeticiones de CGG. El iniciador quimérico se proporcionó a una concentración, 1,5 nM, aproximadamente 1000 veces por debajo de las del segundo y tercer iniciador (cada uno a 1,5 μ M), de tal manera que el iniciador quimérico puede agotarse dentro de los primeros ciclos de la PCR. Las otras condiciones fueron como en el Ejemplo 1. Este procedimiento, realizado con moldes como se indica en la Figura 5, resultó en productos que se obtienen y que tuvieron aproximadamente 196-199 repeticiones en el experimento representado en el panel superior. Este valor fue cercano, pero ligeramente superior, a la designación Coriell de 183-193 repeticiones de CGG. En el experimento representado en el panel inferior, se observó un pico muy a la derecha del electroferograma, marcado 336 repeticiones de CGG de conformidad con la designación de Coriell. El número de repetición aparente fue entre 250-300 repeticiones, pero este tamaño aparente no se consideró preciso porque se cree que el tamaño de este producto no está dentro del intervalo en donde la CE a base del polímero POP-7 puede resolver con precisión los productos que contienen la repetición CGG de acuerdo con el tamaño. Con la CE a base del polímero POP-7, los productos que contienen aproximadamente más de 200 repeticiones de CGG tienden a tener un tamaño aparente artificialmente acortado. Así, este resultado indicó que el pico representaba un producto con un número de repetición no menor de 250. En ambos paneles, se observaron picos en la posición aproximada esperada para un producto con el número de repetición indicado por los alelos más cortos enumerados en las designaciones Coriell, siendo estas designaciones 20 para el panel superior y 28 para el panel inferior (el pico grande mostrado en este panel corresponde a un número de repetición de 29). En cada panel, el pico correspondiente al alelo más largo (números de repetición designados por Coriell 183-193

y 336) fue más amplio que los picos para los productos que contienen números de repetición individuales. El ancho del pico 183-193 de CGG en el panel superior de la Figura 5 parece ser aproximadamente 5-10 repeticiones de CGG o 15-30 pb. Obsérvese que el pico más a la izquierda en cada electroferograma de la Figura 5 corresponde a 5 CGG, debido al iniciador quimérico que contiene 5 repeticiones de CGG.

5

Ejemplo 4: Determinación de la localización AGG dentro de las repeticiones de CGG del promotor de FMR1 en los alelos masculinos mediante un sistema modificado de tres iniciadores para la PCR con iniciador de repetición

10

15

20

25

30

35

40

45

50

55

60

65

Se evaluaron cinco muestras de ADN genómico que contienen alelos con un número de repeticiones de CGG en el intervalo de pre-mutación normal a bajo (30 CGG, 47 CGG, 61 CGG, 20/31 CGG y 46/97 CGG). Para mayor brevedad, los números de repeticiones de CGG enumeradas reflejan el número total de trinucleótidos, es decir, la suma del número de trinucleótidos CGG y el número de trinucleótidos AGG de interrupción. Las muestras se amplificaron por PCR mediante la preparación de una mezcla maestra que contiene 11,45 µl de regulador de AMP rico en GC (Asuragen núm. de catálogo #49387), 1,5 µl de iniciadores de FMR1 marcados con FAM (Asuragen núm. de catálogo #49386; FMR1_F (sec. con núm. de ident.: 14), FMR1_R_FAM (sec. con núm. de ident.: 37 que tiene un 5'FAM)), 0,5 µl FMR1_F_(CGG)_n (sec. con núm. de ident.: 41) (Asuragen núm. de catálogo #49393), 0,5 µl agua libre de nucleasa, y 0,05 µl mezcla de polimerasa rica en GC (Asuragen núm. de catálogo #49388) de Asuragen Inc. (Austin, TX, EE.UU.). La mezcla principal de PCR se agitó en vórtice antes de dispensarse en una placa de microtitulación (placas de 96 ó 384 pocillos, Phenix Research Products, Candler, NC, EE.UU.). Las concentraciones finales de reacción de FMR_F y FMR_R_FAM fueron 1,3 µM, y la concentración final de reacción de FMR1_F_(CGG)_n fue 1,3 nM. Alícuotas de las muestras de ADN genómico, típicamente 1 µl a 20 ng/µl, se transfirieron a cada pocillo de la placa de microvaloración. Se utilizaron hojas de película de aluminio ABgene (Thermo Fisher Scientific) para sellar las placas. Las placas selladas se sometieron a vórtice, se centrifugaron y se transfirieron a un termociclador (Sistema de PCR GeneAmp® 9700, Applied Biosystems™, Foster City, CA, EE.UU.). Las muestras se amplificaron con una etapa de desnaturalización térmica inicial de 95°C durante 5 min, seguido por 10 ciclos de 97°C durante 35 segundos, 62°C durante 35 segundos, 68°C durante 4 min y luego 20 ciclos de 97°C durante 35 segundos, 62°C durante 35 segundos y 68°C durante 4 min con una extensión automática de 20 segundos en cada ciclo. La etapa de extensión final fue 72°C durante 10 min. Este sistema de tres iniciadores para ensayar repeticiones de CGG se representa esquemáticamente en la Figura 4 y la Figura 8A.

Después de la PCR, las muestras se almacenaron a -15 hasta -30°C (protegidas de la luz antes del análisis) o se usaron inmediatamente para el análisis del producto de amplificación por electroforesis capilar (CE). Para la CE, los productos de PCR (1 µl) se mezclaron con 2 µl de patrón ROX 1007 (preparado de conformidad con DeWoody y otros, Biotechniques 37:348, 350, 352 (2004)) en 12 µl de formamida Hi-Di™ (Applied Biosystems (ABI) núm. de parte 4311320) y se desnaturalizó por calor a 95°C durante 2 min antes de la electroforesis capilar en un instrumento ABI 3130xl con una longitud capilar de 36 cm usando el polímero líquido POP7 (ABI núm. de parte 4352759). Los electroferogramas resultantes se muestran en la Figura 6.

Los picos en los electroferogramas se numeraron comenzando con 5, el contenido de repetición CGG mínimo posible de productos de PCR, basado en el diseño del iniciador quimérico. Una reducción severa en la intensidad máxima desde el pico n a $n+1$, (por ejemplo, en la Figura 6 panel A desde el pico 9 a 10,) indicó la presencia de un trinucleótido AGG en la posición correspondiente al pico $n+1$. Un trinucleótido AGG resultó de cinco picos de baja intensidad, que se cree sean porque el trinucleótido AGG redujo la afinidad del iniciador que contiene CGG para las cinco posiciones de unión que abarcan ese trinucleótido (recuerde que el iniciador con la secuencia de sec. con núm. de ident.: 41, contiene cinco repeticiones de CGG). El número total de trinucleótidos CGG/AGG se determinó contando el número total de picos, siendo el primero numerado como 5 se describió anteriormente. Por ejemplo, el electroferograma de la Figura 6, panel A, perfila productos amplificados a partir de un molde hemicigota masculino normal con 30 repeticiones de CGG/AGG. Tres conjuntos de picos bien definidos fueron evidentes en el lado izquierdo del trazado, correspondientes a los productos del iniciador de repetición CGG. El pico más a la izquierda se correspondió a productos con 5 CGG ya que el iniciador quimérico está compuesto por 5 repeticiones de CGG complementarias. La interrupción entre los primeros 5 picos y el siguiente conjunto de 5 picos más intensos refleja la interferencia de una secuencia de intervención AGG. Esta interrupción fue equivalente a 5 repeticiones de CGG, es decir, la amplitud del iniciador quimérico cuando interroga cada posición posible para la hibridación (que está, a su vez, comprometida con cada unidad de repetición en el iniciador por desajustes entre la "C" del iniciador de repetición CGG y la "T" dentro del complemento inverso (CCT) de la secuencia del interruptor AGG). El siguiente conjunto de 5 picos reflejó un marcado incremento en la intensidad de la señal cuando el iniciador quimérico se unió a sitios más allá de la secuencia de interrupción AGG. Entonces se encontró un segundo elemento AGG, y después de la extensión del iniciador a través del último conjunto de secuencias CGG, la señal del producto CGG se perdió por completo. Los datos de la CE indicaron la presencia de 30 repeticiones de CGG/AGG en este alelo. Después de los amplicones repetidos, se observó una interrupción en el electroferograma y el pico más a la derecha fue una banda de producto muy intensa correspondiente al amplicón de longitud completa, producido a partir de los iniciadores FMR1-F y FMR1 R-FAM, que abarcó el tramo de 30 CGG/AGG. Este perfil de la CE correspondió a una secuencia 5'-(CGG)₁₀AGG(CGG)₉AGG(CGG)₉-3' (sec. con núm. de ident.:46).

Se obtuvieron cuatro ejemplos adicionales de perfiles de producto de PCR a partir de los alelos FMR1 (Figura 6, paneles B-E). Se representan muestras masculinas en los paneles B y C y muestras femeninas en los paneles D y E. En cada caso, el pico específico del gen se midió en comparación al tamaño con un ADN de referencia (es decir, el

patrón ROX). Este tamaño estuvo de acuerdo con el método libre de calibración del conteo de picos de producto CGG descrito anteriormente. La precisión de la cuantificación de CGG usando este enfoque también se correlacionó bien con los resultados de Wilson y otros, (J Mol Diagn 10:2-12, 2008) usando materiales consensuales de X frágiles publicados. Usando la misma estrategia de análisis, las muestras B y C se descodificaron como (CGG)₉AGG(CGG)₉AGG(CGG)₂₇ (sec. con núm. de ident.:47) y (CGG)₆₁ (sec. con núm. de ident.:48) (no AGG), respectivamente.

Discusión de las posibles interpretaciones de los resultados en las Figuras 6D-E, y ensayos reflejos para distinguirlos

Los resultados con muestras de ADN genómico de mujeres pueden ser más complejos de interpretar. Por ejemplo, los resultados de la muestra presentados en la Figura 6, paneles D y E revelaron una combinación de picos de intensidad media y alta. Este patrón de producto reflejó la superposición de dos poblaciones de productos de repetición CGG, una de cada alelo. Las posiciones de AGG pueden determinarse a partir de la posición de las dos señales "caídas" (es decir, reducciones en la intensidad de la señal del pico) en el perfil heterocigótico. La única diferencia con los perfiles de los alelos hemicigóticos masculinos es que la intensidad de la señal en la "caída" nunca puede reducirse hasta la línea base en la muestra femenina, porque en estos casos, los dos alelos femeninos no tuvieron trinucleótidos AGG en la misma posición con relación al extremo 3' del amplicón. Para la muestra mostrada en el panel D (alelo 20/31 CGG), la primera "caída" de la intensidad de la señal AGG, entre los picos 10 y 16, nunca se redujo hasta la línea de base, porque sólo un alelo tuvo un trinucleótido AGG en esa posición, mientras que otros tuvieron un trinucleótido CGG. Por el contrario, la segunda caída de la señal regresó a la línea de base porque el alelo 20 CGG termina en la posición 20 y la segunda caída de AGG junto con los picos CGG corrió abajo en el electroferograma pertenecieron únicamente al alelo 31 CGG. La única ambigüedad en la presente descripción era cuál de los dos alelos tuvo el trinucleótido AGG identificado por la primera caída de la señal. Basándose en la incidencia de haplotipos AGG conocidos, el caso más probable es que el alelo muy corto (20 CGG) no contiene el AGG mientras que el alelo 31 CGG tiene dos AGG (como se muestra por las secuencias en el panel D). Sin embargo, el ensayo estándar de tres iniciadores con iniciador de repetición CGG, solo, no puede determinar definitivamente qué alelo contribuye a la primera caída de AGG.

El análisis de la siguiente muestra (Figura 6, panel E) fue más complejo. La traza de la CE mostró tres caídas de AGG aparentes (flechas no marcadas). Usando el escenario más probable basado en las estadísticas de haplotipos, las dos caídas de AGG en el lado izquierdo de la traza de la CE se pueden asignar al alelo de 46 repeticiones y la tercera caída de AGG (flecha no marcada más a la derecha) puede asignarse al alelo de 97 repeticiones. Como resultado, las secuencias alelo pueden ser (CGG)₉AGG(CGG)₉AGG(CGG)₂₆ y (CGG)₁₀AGG(CGG)₈₆. Sin embargo, existen dos incertidumbres en esta secuencia. En primer lugar, cada uno de los trinucleótidos AGG identificados por las caídas de señal, en teoría, pueden asignarse a cualquiera de los dos alelos. Si los tres trinucleótidos AGG ocurren en el alelo de 97 repeticiones, y ninguno en el alelo de 46 repeticiones de CGG, las secuencias deducidas del trazado del electroferograma pueden ser (CGG)₄₆ y (CGG)₉AGG(CGG)₆₀AGG(CGG)₉AGG(CGG)₂₆. Esta combinación de secuencia puede generar el perfil de la CE idéntico al descrito anteriormente. En segundo lugar, el pico específico del gen completo del alelo de 46 repeticiones (el pico de intensidad muy alta cerca del centro de la traza de la CE) puede superponerse y oscurecer cualquier caída de AGG en el alelo de 97 repeticiones. Así, puede haber una caída adicional de la señal de AGG dentro del alelo de 97 repeticiones que migra en la misma posición en el electroferograma de CE que el alelo de 46 genes específicos de longitud completa.

Se desarrollaron métodos adicionales para diferenciar las posibilidades de mapeo de AGG específicas para los alelos 20/31 de CGG y 46/97 de CGG mostrados en el Ejemplo 4 (Figura 6 paneles D y E). El ejemplo 5 más abajo demuestra la utilidad de los métodos de la invención para resolver con precisión estas incertidumbres.

Para la muestra de alelo 20/31 (Figura 6 panel D), dos asignaciones posibles de AGG a los alelos de 20 y 31 repeticiones se muestran en la Figura 7 (paneles de la izquierda). Se prevé que ambos escenarios producirán electroferogramas de la CE idénticos cuando los productos de amplificación por PCR se separen por la CE (simulada en la Figura 7, panel inferior derecho). El perfil simulado de la CE de las dos configuraciones de alelos posibles (Figura 7, panel inferior derecho) estuvo de acuerdo con los resultados empíricos de la CE (Figura 7, panel superior derecho).

Aunque el ensayo estándar de tres iniciadores con iniciador de repetición CGG (Figura 4, Figura 8A) no pudo distinguir el estado AGG de los dos alelos, se desarrolló un ensayo reflejo, representado esquemáticamente en la Figura 8H, que puede distinguir entre el dos posibles escenarios de alelos, como se muestra en la Figura 9. Usando el ensayo de PCR mostrado en la Figura 8H, un electroferograma de CE para el escenario del primer alelo (es decir, cada alelo tiene un AGG, Figura 9, parte superior) pueden generarse dos picos adyacentes con equivalentes de tamaño de 14 CGG y 15 CGG. Un electroferograma de CE para el segundo alelo, en el que el alelo más corto de 20 repeticiones no tiene CGG y el alelo de 31 repeticiones tiene dos AGG (Figura 9, parte inferior), puede generar picos con equivalentes de tamaño de 15 CGG y 25 CGG. En otro ejemplo (Figura 6, panel E), la muestra contuvo dos alelos que tenían 46 y 97 repeticiones de CGG. Como se mencionó anteriormente, el ensayo estándar de tres iniciadores con iniciador de repetición CGG (Figura 4, Figura 8A) no puede diferenciar el contenido de AGG de los alelos específicos. Los electroferogramas simulados de la CE de los productos de amplificación de alelos para esta muestra se muestran en la Figura 10. Los paneles inferiores izquierdo y derecho de la Figura 10 revelan que el ensayo de PCR de tres iniciadores con iniciador de repetición CGG es incapaz de resolver las dos posibles configuraciones de alelos. Sin embargo, cuando se usa un ensayo de PCR (Figura 8C) en el que se invierte la direccionalidad del iniciador de repetición (FMR1_R_(CGG)_n; sec.

con núm. de ident.: 43), se predice que los dos escenarios de alelos generan diferentes perfiles de la CE como se muestra en Figura 11, paneles inferiores.

5 Ejemplo 5: Ensayos de PCR para determinar la localización de AGG dentro de las repeticiones de CGG del promotor FMR1 en los alelos complejos femeninos

10 Se usaron las condiciones de PCR y CE en un ensayo reflejo de la muestra 20/31 de CGG de la Figura 6D que fueron idénticos a los descritos en el Ejemplo 4 anterior excepto que se usaron diferentes iniciadores. La secuencia de iniciador directo fue FMR1_F_FAM (sec. con núm. de ident.: 14) y el iniciador inverso fue un iniciador anclado que tiene un dT 3' terminal (Figura 8H, FMR_R_(CCG)_nCCT, donde n=4; sec. con núm. de ident.: 45). El perfil de la CE real para la muestra contuvo 15 y 25 picos de CGG (Figura 9, traza de la CE) confirmando que las dos secuencias de alelo fueron (CGG)₂₀ (sec. con núm. de ident.: 49) y (CGG)₁₀AGG(CGG)₉AGG(CGG)₁₀ (sec. con núm. de ident.: 50).

15 Cuando se realizó un ensayo de PCR en la muestra de 46/97 CGG de la Figura 6E con un diseño de dos iniciadores con iniciador de repetición, representado esquemáticamente en la Figura 8B, solamente tres caídas de AGG estuvieron presentes en el perfil de la CE (Figura 12, parte superior del panel B, flechas sin marcar). Estos datos excluyeron la posibilidad de una 4^{ta} caída de AGG que se podía superponer por el pico del gen específico del alelo de 46 repeticiones (Figura 6E). También, cuando se realizó el ensayo de PCR anclado representado esquemáticamente en la Figura 8F, se observaron tres picos de alelos CGG distintos, en las posiciones de repetición CGG de 32, 42 y 92 (Figura 12, panel inferior F). Este resultado confirmó claramente que sólo 3 AGG estuvieron presentes en los dos alelos.

Discusión

25 Sin embargo, ninguno de estos dos ensayos (Figura 8B, 8F) puede asignar definitivamente el número de elementos AGG presentes en cada alelo. La Figura 13 explica por qué esto es así, el ensayo anclado del iniciador A (FMR1_F_(CGG)_nA; sec. con núm. de ident.: 44), mostrado en la Figura 8F y la Figura 13, podría generar resultados de la CE idénticos para el escenario alélico (1) en el que el alelo de 46 repeticiones tiene dos AGG y el alelo de 97 repeticiones tiene un AGG (Figura 13, esquema superior) y para el escenario alélico (2) en el que el alelo de 46 repeticiones no tiene AGG y el alelo de 97 repeticiones tiene las tres AGG (Figura 13, esquema inferior). Sin embargo, cada AGG se puede asignar con precisión al alelo apropiado utilizando un ensayo en el que la direccionalidad del iniciador CGG se invierta. Como se muestra en la Figura 14, un ensayo de PCR que utiliza un iniciador dT anclado de orientación inversa (FMR1_R_(CCG)_nCct; sec. con núm. de ident.: 45); (Figura 8H, Figura 14) debe producir un electroferograma de CE con picos equivalentes de tamaño de CGG 14, 24, y 15 para el escenario alélico (1) o un electroferograma de CE con picos equivalentes de tamaño de CGG 15, 65 y 75 para el escenario alélico (2). La Tabla 3 enumera todas las distribuciones AGG posibles en los dos alelos, junto con el tamaño de pico CGG esperado después de la separación de la CE de los productos de amplificación por PCR de este ensayo de iniciador dT anclado (Figura 8H).

Tabla 3

40

Picos de la CE (Equivalentes de repetición de CGG)	posición AGG en el alelo 46	posición AGG en el alelo 97
CGG 15, 14, 24	(CGG) ₉ AGG(CGG) ₉ AGG(CGG) ₂₆ (sec. con núm. de ident.: 51)	(CGG) ₁₀ AGG(CGG) ₈₆ (sec. con núm. de ident.: 52)
CGG 15, 65, 24	(CGG) ₁₉ AGG(CGG) ₂₆ (sec. con núm. de ident.: 68)	(CGG) ₁₀ AGG(CGG) ₄₉ AGG(CGG) ₃₆ (sec. con núm. de ident.: 69)
CGG 15, 14, 75	(CGG) ₉ AGG(CGG) ₃₆ (sec. con núm. de ident.: 70)	(CGG) ₁₀ AGG(CGG) ₅₉ AGG(CGG) ₂₆ (sec. con núm. de ident.: 71)
CGG 15, 65, 75	(CGG) ₄₆ (sec. con núm. de ident.: 55)	(CGG) ₁₀ AGG(CGG) ₄₉ AGG(CGG) ₉ AGG(CGG) ₂₆ (sec. con núm. de ident.: 56)

55

Ejemplo 6: Determinación de la distribución AGG usando un ensayo anclado con un iniciador de CGG orientado de forma opuesta

60 Se realizó un ensayo de PCR como esquematizado en la Figura 8H sobre la muestra de 46/97 CGG de la Figura 6E con un iniciador dT anclado (FMR1_R_(CCG)_nCCT; sec. con núm. de ident.: 45). El perfil de la CE resultante (Figura 14) confirmó que el alelo de 46 repeticiones CGG tuvo dos AGG en las posiciones de repetición 10 y 20; mientras que el alelo de 97 repeticiones CGG tiene un AGG en la posición de repetición 11 (contando de 5' a 3' a partir de la primera CGG, según el consenso de la literatura). Así, pueden usarse los ensayos de mapeo de la invención para resolver patrones complejos de repetición alélica de CGG/AGG en genes de FMR cromosómicos.

65

Ejemplo 7: Determinación del número de repetición de CGG y la posición de AGG en alelos de mutación normal y completa a partir de muestras femeninas utilizando un ensayo de PCR de orientación inversa con iniciador de repetición CCG

5 El número de repeticiones CGG y la presencia y localización de trinucleótidos AGG, se analizaron para 29 muestras clínicas de ADN cromosómico usando el ensayo de PCR de tres iniciadores con iniciador de repetición CGG (Figura 8A) y el ensayo de PCR inversa dT anclado (Figura 8H). Los métodos de PCR fueron como los descritos en los Ejemplos 4 y 5, excepto que se usaron iniciadores apropiados. Los resultados se muestran en la Tabla 4. Los números de la posición AGG siguen el consenso de la literatura para la numeración. NA indica que no se detectaron trinucleótidos de AGG en ese alelo. Se cree que las muestras masculinas tienen más de un alelo (1, 4, 17, 19, 25 y 26) y las muestras femeninas con más de dos alelos (6 y 20) provienen de una población de células en mosaico dentro de la muestra.

Tabla 4

ID de la muestra	Sexo	Picos alelos de CE equivalentes de repetición de CGG	posición de AGG
1	M	>200	11
		170	11
2	F	20	NA
		31	11,21
3	M	47	10,20
4	M	154	NA
		174	NA
		>200	NA
5	M	61	NA
6	F	29	10,20
		60	10,18
		42**	10,18
7	M	51	11
8	F	31	11,21
		47	12
9	F	30	11,21
		50	10
10	M	46	10,20
11	F	30	11,21
		49	10
12	M	54	10
13	M	>200	11
14	F	19	NA
		57	10,20
15	M	57	10,20
16	F	41	11,21,32
		57	10,20
17	M	53	NA
		>200	NA
		152	NA

5	18	F	30	11,21
			60	10,20
10	19	M	>200	12
			61	NA
15	20	F	29	10,20
			>200**	10
			54**	10
20	21	M	90**	10
			50	10
25	22	F	32	10,23
			53	10
30	23	F	46	10,20
			97	11
35	24	M	46	10,20
			64	NA
40	25	M	>200	11
			108	NA
45	26	M	>200	NA
			30	11,21
50	27	F	>200**	0
			58	10,20
55	28	M	29	10,20
			59	10,20

Después de la amplificación por PCR con estos dos ensayos (Figura 8A, 8H) y el análisis de la CE, el número de repeticiones CGG y el estado AGG de los alelos se determinaron positivamente para 26 de las 29 muestras. Estos dos formatos de ensayo no pudieron determinar positivamente el estado AGG para sólo cinco alelos de este grupo. Estos alelos se designan con "*" en la Tabla 4 (alelo 42 CGG en la muestra 6; alelos 54 CGG, 90 CGG y >200 CGG en la muestra 20; alelo >200 CGG en la muestra 27). Se usaron métodos de ensayo adicionales de la invención (Figura 8) para resolver el número de repetición CGG y el estado de AGG para esos alelos.

El análisis de CE de los productos de PCR reveló que dos muestras (20 y 27) tuvieron trinucleótidos AGG en uno de los alelos normales. Por ejemplo, en la muestra 20, los trinucleótidos AGG están presentes en las posiciones 10 y 20 en el alelo 29 CGG. Es posible que otros alelos (*por ejemplo*, >200 CGG) que tienen trinucleótidos AGG en las mismas posiciones exactas (10, 20) no puedan detectarse por el ensayo con iniciador de repetición CGG (Figuras 4, 8A) debido a que las caídas de AGG que son más que ~100 nucleótidos lejos del extremo 3' pueden ser difíciles de detectar debido a la baja intensidad de pico. Además, en esta situación, el ensayo dT anclado (Figura 8H) puede generar picos idénticos a los del alelo más corto, si la localización AGG con respecto al extremo 5' de la región de repetición CGG es idéntica en ambos alelos. La muestra 27 presenta una situación similar en que los trinucleótidos AGG se identificaron en las posiciones 11 y 21 para el alelo 30 CGG, pero la presencia de un trinucleótido AGG en el alelo >200-CGG no puede determinarse sin ambigüedad.

Para resolver estos problemas, se diseñó otro ensayo reflejo (Figura 8C). En una prueba de principio del experimento, se analizaron tres moldes de ADN genómico artificial (Instituto Coriell para la Investigación Médica, Camden, NJ, EE.UU.) que tienen 30 repeticiones de CGG (NA07174), 645 repeticiones de CGG (NA04025) y 30/645 repeticiones de CGG (50% NA07174 y 50% NA04025). El ensayo de iniciador de repetición CCG (Figura 8C) se realizó en la dirección inversa con los iniciadores FMR1_F_FAM, FMR1_R_(CCG)₅ y FMR1_R (sec. con núms. de ident.: 14, 43, 37). La Figura 15, panel A muestra un electroferograma de CE producido a partir del alelo 30 CGG (NA07174). Se observaron dos trinucleótidos AGG, uno en la posición 10 y otro en la posición 20. El panel B de la Figura 15 muestra que no hubo trinucleótido AGG en el alelo 645 CGG (NA04025). Cuando las muestras cromosómicas que tenían estos dos alelos se combinaron para imitar una muestra femenina de alelo 30/645, se observaron dos caídas de AGG, pero la señal no se redujo a la línea de base (Figura 15, panel C). Este resultado confirmó que el alelo CGG 30 tuvo dos trinucleótidos AGG

en las posiciones 10 y 20, mientras que el alelo 645 CGG no contuvo ningún de los interruptores AGG en esas posiciones. Por lo tanto, este ensayo de PCR fue capaz de interrogar el extremo 5' de repeticiones largas, tales como el alelo 645 CGG, para determinar la presencia o ausencia de interruptores AGG.

5 Este ensayo se utilizó para analizar muestras clínicas 20 y 27, que tienen alelos de mutación completa que contienen >200 repeticiones de CGG (Tabla 4). La Figura 16 muestra los resultados de la muestra 20. El panel A de la Figura 16 muestra un electroferograma de CE de los productos de PCR después de la amplificación con el ensayo de iniciación CGG estándar (Figura 8A) con el iniciador de repetición orientado en la dirección hacia delante. Se observaron dos AGG en las posiciones 10 y 20 (contando a partir del extremo 3' de las repeticiones CGG). Basándose en la incidencia de haplotipos, los AGG se sitúan normalmente en las posiciones 10, 11, 20 y 21, contando desde el extremo 5' de la región de repetición CGG. *ver, por ejemplo*, Zhong y otros, Am. J. Hum. Genet. 57: 351-61 (1995); Kunst y otros, Am. J. Hum. Genet. 58:513-22 (1996); y Eichler y otros, Hum. Mol. Genet. 4:2199-208 (1995). La incidencia del haplotipo sugiere, por lo tanto, que los interruptores AGG están presentes en el alelo 29 CGG. El panel B de la Figura 16 muestra un electroferograma de CE de los productos de PCR después de la amplificación de las mismas muestras usando el ensayo de dT anclado con dos iniciadores orientado en la dirección inversa (Figura 8H). Estos resultados confirmaron que los dos trinucleótidos AGG se localizaron en las posiciones 10 y 20 contando a partir del extremo 5' de la región de repetición CGG. El análisis de los resultados en la Figura 16, paneles A y B reveló que dos trinucleótidos AGG se localizaron dentro del alelo 29 CGG, cuya región de repetición CGG tenía por lo tanto la secuencia (CGG)₉(AGG)(CGG)₉AGG(CGG)₉ (sec. con núm. de ident.: 57). Una incertidumbre restante con esta muestra es que el alelo de mutación completo (> 200 CGG) puede tener trinucleótidos AGG en las mismas posiciones (10, 20) a partir del extremo 5'. Si es así, estas secuencias de AGG no pueden detectarse mediante un ensayo T anclado con orientación inversa ni diferenciarse de las AGG en el alelo corto (debido a la distancia neta al iniciador directo y la correspondiente reducción en la intensidad de señal observada para mapear los elementos de la secuencia 5'). Por lo tanto, se usó el ensayo de PCR mostrado en la Figura 8D, un ensayo de PCR de tres iniciadores con el iniciador de repetición CCG en la dirección inversa, para resolver este problema. Los resultados del ensayo (Figura 16, panel C) confirmaron que el alelo CGG >200 tuvo un AGG en la posición 10 desde el extremo 5' de las repeticiones CGG, ya que la señal cayó casi completamente hasta la línea base en esa posición. Así, tanto el alelo 29 CGG como el alelo 200 CGG tuvieron trinucleótidos AGG en la posición 10. No hubo AGG presente en la posición 20 del alelo de mutación completa (>200 CGG) debido a que la señal en ese lugar no se redujo hasta cerca de la línea de base (Figura 16 panel C).

30

Discusión

Estos análisis demostraron la aparición de los trinucleótidos AGG en alelos de mutación completa. Se cree que esto contrasta con la posición establecida de múltiples expertos en el campo de que los interruptores AGG no se producen en los alelos de mutación completa. Además, los métodos y ensayos de la invención son capaces de detectar interruptores de trinucleótidos AGG cerca del extremo 5' de la región de repetición CGG.

35

Ejemplo 8: Mapeo de los elementos del interruptor en la muestra 27

40 La Figura 17 muestra los resultados del mapeo de AGG obtenidos para la muestra 27 (Tabla 4). El panel A de la Figura 17 muestra un electroferograma de CE después de la amplificación con el ensayo estándar de iniciación con CGG en la dirección directa (Figura 8A), que demostró que estuvieron presentes dos trinucleótidos AGG. El patrón de pico CGG reveló dos AGG, uno en la posición 10 y otro en la posición 20 (desde el extremo 3' de las repeticiones CGG). Basándose en el conocimiento de los haplotipos comunes, las interrupciones AGG de las repeticiones CGG se colocan característicamente en las posiciones 10, 11, 20, y/o 21 contando desde el extremo 5' de las repeticiones CGG, lo que sugiere que las repeticiones AGG se presentan en el alelo 30 CGG en esta muestra. El panel B de la Figura 17 presenta los resultados del ensayo de PCR de dos iniciadores dT anclado, orientado en la dirección inversa (Figura 8H). Este ensayo confirmó que los dos trinucleótidos AGG estuvieron en las posiciones 11 y 21, contando desde el extremo 5' de las repeticiones CGG - (CGG)₁₀AGG(CGG)₉AGG(CGG)₉ (sec. con núm. de ident.: 46). Una incertidumbre restante con esta muestra es que el alelo de mutación completa (> 200 CGG) puede tener trinucleótidos AGG en las mismas posiciones que las del alelo de repetición 30 CGG. El panel C de la Figura 17 ofrece los resultados del ensayo de PCR de tres iniciadores con iniciador CCG (Figura 8D) orientado en la dirección inversa. Este ensayo confirmó que el alelo >200 CGG no tiene AGG en la posición 11 ó 21, porque la caída de AGG no se redujo hasta cerca de la línea de base en estas dos posiciones.

55

Ejemplo 9: Resolución de las posiciones de AGG para alelos de baja abundancia en muestras de mosaico usando un ensayo de PCR con anclaje A de orientación directa

60 El análisis del ADN cromosómico de algunas muestras en la Tabla 4 reveló la presencia de alelos de baja abundancia en las muestras 6 y 20. Se cree que estos son alelos derivados de una población de células en mosaico presentes en esas muestras. El ensayo de PCR de la invención, dA anclado orientado hacia delante (Figura 8F) fue suficientemente sensible para detectar trinucleótidos AGG en la secuencia de estos alelos menores. Los resultados del ensayo dA anclado de dos iniciadores para las muestras 6 y 20 se muestran en la Figura 18. La amplificación del ADN cromosómico de la muestra 6, usando el ensayo de PCR mostrado en la Figura 8F, y el posterior análisis de CE de los productos de PCR reveló cuatro picos principales, con longitudes de 15, 25, 48 y 56 repeticiones de CGG y dos picos menores, con longitudes de 30 y 38 repeticiones de CGG (Figura 18 panel A). Estas longitudes incluyen las cinco

65

repeticiones de CGG en el iniciador, el interruptor AGG y el número de repeticiones entre el interruptor AGG y el extremo 3' de la región de repetición CGG. Los cuatro picos principales confirmaron las posiciones AGG para los dos alelos principales (posiciones 10 y 20 para el alelo 29 CGG y posiciones 10 y 18 para el alelo 60 CGG - todos a partir del extremo 5' de las repeticiones CGG). Los dos picos menores confirmaron que el alelo menor (42 CGG) tiene dos AGG, uno en la posición 10 y uno en la posición 18 (a partir del extremo 5' de las repeticiones CGG). A pesar de que tanto los alelos 42 CGG como los 60 CGG tienen trinucleótidos AGG en las posiciones 10 y 18 a partir del extremo 5', ya que el ensayo dA anclado se sitúa en la dirección directa, el tamaño de los amplicones del PCR de iniciador dA anclado se determinó por el número de repeticiones CGG contadas desde el extremo 3' para cualquier AGG que pueda estar presente (*por ejemplo*, 24 para el alelo 42 CGG, y 42 para el alelo 60 CGG). Estos productos de PCR se separaron bien por CE (30, 38 frente a 48, 56 unidades repetidas).

La amplificación del ADN cromosómico de la muestra 20 usando el ensayo de PCR mostrado en la Figura 8F y el análisis de CE subsiguiente de los productos de PCR reveló dos picos primarios (Figura 18B) para el alelo principal (previamente asignado como alelo 29 CGG basado en los datos mostrados en la Figura 16). Sin embargo, el ensayo reveló también, dos picos de alelos menores en la posición 50 CGG y 86 CGG indicando que había dos AGG en estos dos alelos menores, ambos en la posición 10 cuando se cuenta a partir del extremo 5' (*es decir*, las posiciones 45 y 81 si se cuenta desde el extremo 3').

En conclusión, las combinaciones apropiadas de los cuatro ensayos descritos anteriormente permitieron el mapeo de los interruptores trinucleotídicos AGG en las regiones de repetición CGG de cada alelo de las 29 muestras clínicas mostradas en la Tabla 4.

Ejemplo 10: Flujoograma de trabajo de la muestra

Un ejemplo de un flujoograma de mapeo de AGG y recuento de CGG usando los métodos de la invención se muestra en la Figura 19 y combina algunos de los formatos de ensayo mostrados en la Figura 8. El flujoograma de trabajo que se muestra en la presente descripción es sólo un ejemplo. Las designaciones de letras del ensayo se refieren a las mostradas en la Figura 8A-H. Algunos ensayos usados en este flujoograma de trabajo pueden reemplazarse por otros formatos de ensayo para lograr el propósito de la muestra.

Las modalidades dentro de la descripción proporcionan una ilustración de las modalidades de la invención y no deben interpretarse como limitantes del alcance de la invención. El experto en la técnica reconoce fácilmente que muchas otras modalidades se incluyen en la invención.

A menos que se indique de cualquier otra forma, todos los números que expresan cantidades de ingredientes, condiciones de reacción, y así sucesivamente usados en la descripción, que incluyen las reivindicaciones, se entenderán que pueden modificarse en todos los casos por el término "aproximadamente,". En consecuencia, a menos que se indique de cualquier otra forma, los parámetros numéricos son aproximaciones que pueden variar en dependencia de las propiedades que se buscan obtener por la presente invención. Por lo menos, y no como un intento de limitar la aplicación de la doctrina de los equivalentes al alcance de las reivindicaciones, cada parámetro numérico debe al menos interpretarse a la luz del número de dígitos significativos informados y aplicando técnicas de redondeo ordinarias.

A menos que se indique de cualquier otra forma, el término "al menos" que precede a una serie de elementos debe entenderse que se refiere a cada elemento de la serie. Aquellos con experiencia en la materia reconocerán, o serán capaces de determinar usando no más que la experimentación de rutina, muchos equivalentes para las modalidades específicas descritas específicamente en la presente descripción. Se pretende que esos equivalentes se incluyan en las reivindicaciones siguientes.

Cuando se reciten en una reivindicación reacciones de amplificación múltiple (por ejemplo, PCR), debe entenderse que al referirse a las reacciones como "primero", "segundo", etc., no se refiere al orden cronológico en el que se realizan las reacciones, y que tales reivindicaciones abarcan métodos en los que las reacciones descritas se realizan en cualquier orden o simultáneamente, incluyendo, por ejemplo, realizar la "segunda" reacción antes, al mismo tiempo o después de la "primera" reacción.

Lista de secuencias

<110> Asuragen, Inc.

<120> Métodos de PCR para caracterizar la región 5' no traducida de los genes FMR1 y FMR2

<130> 10256.31-304

<150> US 61/162.977

<151> 2009-03-24

<160> 71

<170> PatentIn versión 3.5

5 <210> 1
<211> 22
<212> ADN
<213> Secuencia artificial

10 <220>
<223> iniciador

15 <400> 1
cgggtggaggg ccgcctctga gc 22

20 <210> 2
<211> 25
<212> ADN
<213> Secuencia artificial

25 <220>
<223> iniciador

30 <400> 2
caggcgctca gctcggttc ggttt 25

35 <210> 3
<211> 25
<212> ADN
<213> Secuencia artificial

40 <220>
<223> iniciador

45 <400> 3
cagtcaggcg ctcagctccg ttctg 25

50 <210> 4
<211> 24
<212> ADN
<213> Secuencia artificial

55 <220>
<223> iniciador

60 <400> 4
tccggtggag ggccgcctct gagc 24

65 <210> 5
<211> 35
<212> ADN
<213> Secuencia artificial

70 <220>
<223> iniciador

75 <400> 5
ggttcggcct cagtcaggcg ctcagctccg ttctg 35

80 <210> 6
<211> 36
<212> ADN
<213> Secuencia artificial

85 <220>
<223> iniciador

ES 2 629 400 T3

<400> 6
 gggttcggcc tcagtcaggc gctcagctcc gtttcg 36

5
 <210> 7
 <211> 26
 <212> ADN
 <213> Secuencia artificial

10
 <220>
 <223> iniciador

<400> 7
 gcgggcccggg gggtcgcct cagtca 26

15
 <210> 8
 <211> 25
 <212> ADN
 <213> Secuencia artificial

20
 <220>
 <223> iniciador

<400> 8
 cagcgggccc ggggttcggc ctca 25

25
 <210> 9
 <211> 25
 <212> ADN
 <213> Secuencia artificial

30
 <220>
 <223> iniciador

<400> 9
 gcagcgggcc ggggttcgg cctca 25

35
 <210> 10
 <211> 25
 <212> ADN
 <213> Secuencia artificial

40
 <220>
 <223> iniciador

<400> 10
 gggccggggg ttcggcctca gtcag 25

45
 <210> 11
 <211> 26
 <212> ADN
 <213> Secuencia artificial

50
 <220>
 <223> iniciador

<400> 11
 ggggttcggc tcagtcagg cgctca 26

55
 <210> 12
 <211> 27
 <212> ADN
 <213> Secuencia artificial

60
 <220>
 <223> iniciador

65
 <400> 12

5 ggggttcggc ctca^gtcagg cgctcag 27
 <210> 13
 <211> 30
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> iniciador
 10 <400> 13
 ggcgctcagc tccgttcgg ttcacttc 30
 15 <210> 14
 <211> 28
 <212> ADN
 <213> Secuencia artificial
 20 <220>
 <223> iniciador
 <400> 14
 tcaggcgctc agctccg^tt cggttca 28
 25 <210> 15
 <211> 26
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> iniciador
 <400> 15
 cacttccggt ggagggccgc ctctga 26
 35 <210> 16
 <211> 25
 <212> ADN
 <213> Secuencia artificial
 40 <220>
 <223> iniciador
 <400> 16
 ttccggtgga gggccgctc tgagc 25
 45 <210> 17
 <211> 25
 <212> ADN
 <213> Secuencia artificial
 50 <220>
 <223> iniciador
 <400> 17
 cgacttcca ccaccagctc ctcca 25
 55 <210> 18
 <211> 21
 <212> ADN
 <213> Secuencia artificial
 60 <220>
 <223> iniciador
 <400> 18
 ggagcccgcc cccgagaggt g 21
 65

ES 2 629 400 T3

5
 <210> 19
 <211> 21
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> iniciador

10
 <400> 19
 gggagcccg ccccgagagg t 21

15
 <210> 20
 <211> 26
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> iniciador

20
 <400> 20
 cgacttcca ccaccagctc ctccat 26

25
 <210> 21
 <211> 23
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> iniciador

30
 <400> 21
 cgggagcccg ccccgagag gtg 23

35
 <210> 22
 <211> 23
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> iniciador

40
 <400> 22
 ccgggagccc gcccccgaga ggt 23

45
 <210> 23
 <211> 24
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> iniciador

50
 <400> 23
 ccgggagccc gcccccgaga ggtg 24

55
 <210> 24
 <211> 26
 <212> ADN
 <213> Secuencia artificial

60
 <220>
 <223> iniciador

65
 <400> 24
 cgccgggagc ccgccccga gaggtg 26

<210> 25

ES 2 629 400 T3

<211> 26
 <212> ADN
 <213> Secuencia artificial

5

<220>
 <223> iniciador

<400> 25
 gcgccgggag cccgccccg agaggt 26

10

<210> 26
 <211> 25
 <212> ADN
 <213> Secuencia artificial

15

<220>
 <223> iniciador

<400> 26
 cgccgggagc ccgccccga gaggt 25

20

<210> 27
 <211> 28
 <212> ADN
 <213> Secuencia artificial

25

<220>
 <223> iniciador

<400> 27
 gcgccattgg agccccgcac ttccacca 28

30

<210> 28
 <211> 25
 <212> ADN
 <213> Secuencia artificial

35

<220>
 <223> iniciador

<400> 28
 gcgccattgg agccccgcac ttcca 25

40

<210> 29
 <211> 25
 <212> ADN
 <213> Secuencia artificial

45

<220>
 <223> iniciador

<400> 29
 agcgccattg gagccccgca ctcc 25

50

<210> 30
 <211> 25
 <212> ADN
 <213> Secuencia artificial

55

<220>
 <223> iniciador

<400> 30
 cgccattgga gccccgact tccac 25

60

<210> 31
 <211> 25

65

ES 2 629 400 T3

	<212> ADN	
	<213> Secuencia artificial	
5	<220>	
	<223> iniciador	
	<400> 31	
	ttggagcccc gcactccac cacca	25
10	<210> 32	
	<211> 28	
	<212> ADN	
	<213> Secuencia artificial	
15	<220>	
	<223> iniciador	
	<400> 32	
20	agccccgcac ttcaccacc agctcctc	28
	<210> 33	
	<211> 28	
	<212> ADN	
25	<213> Secuencia artificial	
	<220>	
	<223> iniciador	
	<400> 33	
30	gagccccgca ctccaccac cagctcct	28
	<210> 34	
	<211> 28	
35	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> iniciador	
	<400> 34	
40	cattggagcc ccgcacttc accaccag	28
	<210> 35	
	<211> 30	
45	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> iniciador	
	<400> 35	
50	cccgcacttc caccaccagc tctccatct	30
	<210> 36	
	<211> 30	
55	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> iniciador	
60	<400> 36	
	tagaaagcgc cattggagcc ccgcacttc	30
	<210> 37	
65	<211> 26	
	<212> ADN	

ES 2 629 400 T3

<213> Secuencia artificial
 <220>
 <223> iniciador
 5
 <400> 37
 aagcgccatt ggagccccgc acttcc 26
 <210> 38
 <211> 36
 <212> ADN
 <213> Secuencia artificial
 10
 <220>
 <223> iniciador
 15
 <400> 38
 tcaggcgctc agctccgttt cggtttact tccggt 36
 <210> 39
 <211> 36
 <212> ADN
 <213> Secuencia artificial
 20
 <220>
 <223> iniciador
 25
 <400> 39
 agcgtctact gtctcggcac tgcccgccg ccgccg 36
 <210> 40
 <211> 28
 <212> ADN
 <213> Secuencia artificial
 30
 <220>
 <223> iniciador
 <400> 40
 tcaggcgctc agctccgttt cggtttca 28
 <210> 41
 <211> 43
 <212> ADN
 <213> Secuencia artificial
 35
 <220>
 <223> iniciador
 <400> 41
 tcaggcgctc agctccgttt cggtttcacg gcggcggcgg cgg 43
 <210> 42
 <211> 90
 <212> ADN
 <213> Homo sapiens
 40
 <400> 42
 cggcggcggc ggcggcggcgc gcggcggcgg aggcggcggc ggcggcggcgc gcggcggcgg 60
 aggcggcggc ggcggcggcgc gcggcggcgg 90
 45
 <210> 43
 <211> 41
 <212> ADN
 <213> Secuencia artificial
 50
 <210> 43
 <211> 41
 <212> ADN
 <213> Secuencia artificial
 55
 <210> 43
 <211> 41
 <212> ADN
 <213> Secuencia artificial
 60

ES 2 629 400 T3

<220>
 <223> iniciador

5 <400> 43
 aagcgccatt ggagccccgc acttccccgc cgccgcccgc g 41

<210> 44
 <211> 44
 <212> ADN
 <213> Secuencia artificial

10

<220>
 <223> iniciador

15 <400> 44
 tcaggcgctc agctccggtt cggttcacg gcggcgccgg cgga 44

<210> 45
 <211> 41
 <212> ADN
 <213> Secuencia artificial

20

<220>
 <223> iniciador

25 <400> 45
 aagcgccatt ggagccccgc acttccccgc cgccgcccgc t 41

<210> 46
 <211> 90
 <212> ADN
 <213> Homo sapiens

30

<400> 46
 cggcggcggc ggcggcggcg gcggcggcgg aggcggcggc ggcggcggcg gcggcggcgg 60

35 aggcggcggc ggcggcggcg gcggcggcgg 90
 <210> 47
 <211> 141
 <212> ADN
 <213> Homo sapiens

40

<400> 47
 cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg 60
 cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg 120
 cggcggcggc ggcggcggcg g 141

<210> 48
 <211> 183
 <212> ADN
 <213> Homo sapiens

45

<400> 48
 cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg 60
 cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg 120
 cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg 180

50 cgg 183

<210> 49
 <211> 60
 <212> ADN
 <213> Homo sapiens

55

ES 2 629 400 T3

	<400> 49		
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	60	
5	<210> 50 <211> 93 <212> ADN <213> Homo sapiens		
10	<400> 50		
	cggcggcggc ggcggcggcg gcggcggcgg aggcggcggc ggcggcggcg gcggcggcgg	60	
	aggcggcggc ggcggcggcg gcggcggcgg cgg	93	
15	<210> 51 <211> 138 <212> ADN <213> Homo sapiens		
20	<400> 51		
	cggcggcggc ggcggcggcg gcggcggcgg aggcggcggc ggcggcggcg gcggcggcgg	60	
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	120	
25	cggcggcggc ggcggcggcg	138	
30	<210> 52 <211> 291 <212> ADN <213> Homo sapiens		
	<400> 52		
35	cggcggcggc ggcggcggcg gcggcggcgg aggcggcggc ggcggcggcg gcggcggcgg	60	
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	120	
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	180	
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	240	
40	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg g	291	
45	<210> 53 <211> 60 <212> ADN <213> Homo sapiens		
50	<400> 53		
	cggcggcggc ggcggcggcg gcggcggcgg aggcggcggc ggcggcggcg gcggcggcgg	60	
55	<210> 54 <211> 93 <212> ADN <213> Homo sapiens		
60	<400> 54		
	cggcggcggc ggcggcggcg gcggcggcgg aggcggcggc ggcggcggcg gcggcggcgg	60	
	cggcggcggc ggcggcggcg gcggcggcgg cgg	93	
65	<210> 55 <211> 138 <212> ADN <213> Homo sapiens		

ES 2 629 400 T3

	<400> 55	
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	60
5	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	120
	cggcggcggc ggcggcgg	138
10	<210> 56 <211> 291 <212> ADN <213> Homo sapiens	
15	<400> 56	
	cggcggcggc ggcggcggcg gcggcggcgg aggcggcggc ggcggcggcg gcggcggcgg	60
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	120
20	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	180
	aggcggcggc ggcggcggcg gcggcggcgg aggcggcggc ggcggcggcg gcggcggcgg	240
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg g	291
25	<210> 57 <211> 87 <212> ADN <213> Homo sapiens	
30	<400> 57	
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	60
	cggcggcggc ggcggcggcg gcggcggcgg	87
35	<211> 126 <212> ADN <213> Homo sapiens	
40	<400> 58	
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gaggcggcgg	60
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	120
	cggcgg	126
45	<210> 59 <211> 180 <212> ADN <213> Homo sapiens	
50	<400> 59	
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gaggcggcgg	60
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	120
55	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	180
60	<210> 60 <211> 162 <212> ADN <213> Homo sapiens	
65	<400> 60	

ES 2 629 400 T3

	cggcggcggc ggcggcggcg gcggcggagg cggcggcggc ggcggcggcg gcggcggcgg	60
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	120
5	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc gg	162
	<210> 61	
	<211> 270	
	<212> ADN	
10	<213> Homo sapiens	
	<400> 61	
	cggcggcggc ggcggcggcg gcggcggagg cggcggcggc ggcggcggcg gcggcggcgg	60
15	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	120
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	180
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	240
20	cggcggcggc ggcggcggcg gcggcggcgg	270
	<210> 62	
	<211> 111	
	<212> ADN	
25	<213> Homo sapiens	
	<400> 62	
	cggcggcggc ggcggcggcg gcggcggagg cggcggcggc ggcggcggcg gaggcggcgg	60
30	cggcggcggc ggcggcggcg gaggcggcgg cggcggcggc ggcggcggcg g	111
	<210> 63	
	<211> 171	
	<212> ADN	
	<213> Homo sapiens	
	<400> 63	
	cggcggcggc ggcggcggcg gcggcggagg cggcggcggc ggcggcggcg gcggcggagg	60
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	120
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg g	171
45		
	<210> 64	
	<211> 234	
	<212> ADN	
50	<213> Homo sapiens	
	<400> 64	
	cggcggcggc ggcggcggcg gcggcggcgg aggcggcggc ggcggcggcg gcggcggcgg	60
55	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	120
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	180
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcgg	234
60		
	<210> 65	
	<211> 300	
	<212> ADN	
	<213> Homo sapiens	
65		
	<400> 65	

ES 2 629 400 T3

	cggcggcggc ggcggcggcg ggcggcggcg aggcggcggc ggcggcggcg ggcggcggcg	60
	cggcggcggc ggcggcggcg ggcggcggcg cggcggcggc ggcggcggcg ggcggcggcg	120
5	cggcggcggc ggcggcggcg ggcggcggcg cggcggcggc ggcggcggcg ggcggcggcg	180
	cggcggcggc ggcggcggcg ggcggcggcg cggcggcggc ggcggcggcg ggcggcggcg	240
	cggcggcggc ggcggcggcg ggcggcggcg cggcggcggc ggcggcggcg ggcggcggcg	300
10	<210> 66	
	<211> 366	
	<212> ADN	
	<213> Homo sapiens	
15	<400> 66	
	cggcggcggc ggcggcggcg ggcggcggcg aggcggcggc ggcggcggcg ggcggcggcg	60
20	cggcggcggc ggcggcggcg ggcggcggcg cggcggcggc ggcggcggcg ggcggcggcg	120
	cggcggcggc ggcggcggcg ggcggcggcg cggcggcggc ggcggcggcg ggcggcggcg	180
	cggcggcggc ggcggcggcg ggcggcggcg cggcggcggc ggcggcggcg ggcggcggcg	240
25	cggcggcggc ggcggcggcg ggcggcggcg cggcggcggc ggcggcggcg ggcggcggcg	300
	cggcggcggc ggcggcggcg ggcggcggcg cggcggcggc ggcggcggcg ggcggcggcg	360
	cggcgg	366
30	<210> 67	
	<211> 60	
	<212> ADN	
	<213> Homo sapiens	
35	<400> 67	
	cggcggcggc ggcggcggcg ggcggcggcg aggcggcggc ggcggcggcg ggcggcggcg	60
40	<210> 68	
	<211> 138	
	<212> ADN	
	<213> Homo sapiens	
45	<400> 68	
	cggcggcggc ggcggcggcg ggcggcggcg cggcggcggc ggcggcggcg ggcggcggcg	60
	cggcggcggc ggcggcggcg ggcggcggcg cggcggcggc ggcggcggcg ggcggcggcg	120
	cggcggcggc ggcggcggcg	138
50	<210> 69	
	<211> 291	
	<212> ADN	
	<213> Homo sapiens	
55	<400> 69	
	cggcggcggc ggcggcggcg ggcggcggcg aggcggcggc ggcggcggcg ggcggcggcg	60
60	cggcggcggc ggcggcggcg ggcggcggcg cggcggcggc ggcggcggcg ggcggcggcg	120
	cggcggcggc ggcggcggcg ggcggcggcg cggcggcggc ggcggcggcg ggcggcggcg	180
	aggcggcggc ggcggcggcg ggcggcggcg cggcggcggc ggcggcggcg ggcggcggcg	240
65	cggcggcggc ggcggcggcg ggcggcggcg cggcggcggc ggcggcggcg g	291

ES 2 629 400 T3

	<210> 70	
	<211> 138	
	<212> ADN	
	<213> Homo sapiens	
5	<400> 70	
	cggcggcggc ggcggcggcg gcggcggagg cggcggcggc ggcggcggcg gcggcggcgg	60
10	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	120
	cggcggcggc ggcggcgg	138
15	<210> 71	
	<211> 291	
	<212> ADN	
	<213> Homo sapiens	
20	<400> 71	
	cggcggcggc ggcggcggcg gcggcggcgg aggcggcggc ggcggcggcg gcggcggcgg	60
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	120
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg gcggcggcgg	180
25	cggcggcggc ggcggcggcg gcggcggcgg aggcggcggc ggcggcggcg gcggcggcgg	240
	cggcggcggc ggcggcggcg gcggcggcgg cggcggcggc ggcggcggcg g	291
30		

Reivindicaciones

1. Un método para analizar si una secuencia interruptor está presente en al menos una región rica en CGG comprendida por al menos un molde en una muestra, que comprende:
 - a) proporcionar al menos dos iniciadores diferentes, incluyendo un primer iniciador que comprende repeticiones CGG, CCG, GCG, CGC, GCC o GGC, y un segundo iniciador que se hibrida a una posición fuera de la región rica en CGG;
 - b) realizar la PCR con al menos dos iniciadores diferentes y al menos un molde que comprende al menos una región rica en CGG, en donde la PCR produce un conjunto de productos;
 - c) resolver el conjunto de productos con una técnica de alta resolución para producir una representación del tamaño y la abundancia del producto; y
 - d) derivar la información acerca de si una secuencia de interruptor está presente en al menos una región rica en CGG o dónde al menos dentro de una región rica en CGG se encuentra una secuencia de interruptor a partir de dicha representación.
2. Un método para analizar si una secuencia interruptor está presente en al menos una región rica en CGG comprendida por al menos un molde en una muestra, que comprende:
 - a) proporcionar al menos tres iniciadores diferentes, incluyendo un primer iniciador que comprende repeticiones CGG, CCG, GCG, CGC, GCC, o GGC y una solapa 5', un segundo iniciador que se hibrida a una posición fuera de la región rica en CGG y un tercer iniciador que tiene una secuencia comprendida por la solapa 5' del primer iniciador, en donde el primer iniciador se proporciona a una concentración menor que el tercer iniciador;
 - b) realizar la PCR con al menos tres iniciadores diferentes y al menos un molde, en donde la PCR produce un conjunto de productos;
 - c) resolver el conjunto de productos con una técnica de alta resolución para producir una representación del tamaño y la abundancia del producto; y
 - d) derivar la información acerca de si una secuencia de interruptor está presente en al menos una región rica en CGG o dónde al menos dentro de una región rica en CGG se encuentra una secuencia de interruptor a partir de dicha representación.
3. El método de conformidad con cualquiera de las reivindicaciones 1 ó 2, en donde la secuencia de interruptor es un elemento AGG.
4. El método de conformidad con cualquiera de las reivindicaciones 1 ó 2, que comprende además derivar la información sobre el número de repeticiones CGG de dicha representación que determina si la región de repetición rica en CGG comprende más o menos de 200 repeticiones CGG y opcionalmente determina el número de repeticiones CGG presentes en la repetición región rica en CGG.
5. El método de conformidad con cualquiera de las reivindicaciones 1 ó 2, con la condición de que no se utilice un patrón o calibrador externo en la derivación de la información sobre si una secuencia de interrupción está presente en la región rica en CGG o cuándo se localiza dentro de la región rica en CGG la secuencia de interrupción a partir de dicha representación.
6. El método de conformidad con cualquiera de las reivindicaciones 1 ó 2, en donde derivar la información acerca de si una secuencia de interrupción está presente en la región rica en CGG o dónde se localiza una secuencia de interrupción dentro de la región rica en CGG a partir de dicha representación comprende determinar una o más longitudes del producto en el que se reduce la cantidad de producto en al menos 50% en comparación con la cantidad de productos de longitud vecinos, o en donde derivar información acerca de si una secuencia de interrupción está presente en la región rica en CGG o dónde se localiza una secuencia de interrupción dentro de la región rica en CGG a partir de dicha representación, comprende determinar una o más longitudes de producto en las que la cantidad de producto se reduce en al menos 25% en comparación con la cantidad de productos de longitud vecinos, en donde la región rica en CGG es a partir de un individuo heterocigótico para el alelo que comprende la región rica en CGG.
7. El método de cualquiera de las reivindicaciones 1 ó 2, en donde el método es un ensayo anclado y el primer iniciador comprende una selección subsiguiente de A, T, AG, CT, AGG y CCT entre o en el extremo 3' de las repeticiones de CGG, CCG, GCG, CGC, GCC o GGC.
8. El método de conformidad con cualquiera de las reivindicaciones 1 ó 2, que comprende proporcionar además al menos un primer iniciador adicional y opcionalmente un segundo iniciador adicional, el primer iniciador adicional que comprende repeticiones de CGG, CCG, GCG, CGC, GCC o GGC; realizar una segunda PCR con al menos el primer iniciador adicional, un iniciador seleccionado a partir del segundo iniciador de la etapa (a) y el segundo iniciador adicional, y la al menos una plantilla, en donde la segunda PCR produce un segundo conjunto de productos; Y resolver el segundo conjunto de productos con una técnica de alta resolución para producir una segunda representación del tamaño y la abundancia del producto; en donde el primer iniciador de la etapa (a) tiene una actividad de unión preferencial para sitios en la región rica de CGG que no comprende un elemento de

interrupción y en donde el primer iniciador adicional tiene una actividad de unión preferencial para sitios en la región rica en CGG que comprenden un elemento de interrupción.

- 5 9. El método de la reivindicación 8, que comprende además, determinar al menos una longitud de al menos una región rica en CGG, en donde la muestra comprende el material genético a partir de células que tienen una ploidía de al menos 2 con respecto a la región CGG y el método comprende determinar al menos dos longitudes de al menos dos regiones ricas en CGG.
- 10 10. El método de la reivindicación 8, que comprende además, determinar si la muestra comprende alelos mayores y menores con elementos de interrupción posicionados de forma diferente.
- 15 11. El método de la reivindicación 8, que comprende además, proporcionar al menos un tercer iniciador adicional y opcionalmente un cuarto iniciador adicional, el tercer iniciador adicional que comprende repeticiones de CGG, CCG, GCG, CGC, GCC o GGC; realizar una tercera PCR con al menos el tercer iniciador adicional, un iniciador seleccionado a partir del segundo iniciador adicional y el cuarto iniciador adicional, y al menos un molde, en donde la tercera PCR produce un tercer conjunto de productos; y resolver el tercer conjunto de productos con una técnica de alta resolución para producir una tercera representación del tamaño y la abundancia del producto; en donde el tercer iniciador adicional está orientado de manera opuesta con relación al primer iniciador de la etapa (a) y es diferente del primer iniciador adicional.
- 20 12. El método de la reivindicación 11, que comprende además, determinar la presencia o ausencia de elementos de interrupción dentro de 150 pb de cada extremo de al menos un alelo comprendido por la muestra.
- 25 13. El método de conformidad con cualquiera de las reivindicaciones 1 ó 2, que comprende además, proporcionar además al menos un primer iniciador adicional y un segundo iniciador adicional, el primer iniciador adicional que comprende repeticiones de CGG, CCG, GCG, CGC, GCC o GGC; realizar una segunda PCR con al menos el primer iniciador adicional y el segundo iniciador adicional, y al menos una plantilla, en donde la segunda PCR produce un segundo conjunto de productos; y resolver el segundo conjunto de productos con una técnica de alta resolución para producir una segunda representación del tamaño y la abundancia del producto; en donde el primer iniciador adicional está orientado de manera opuesta con relación al primer iniciador de la etapa (a).
- 30 14. El método de la reivindicación 13, en donde el primer iniciador tiene una actividad de unión preferencial para sitios en la región rica de CGG que no comprende elementos de interrupción, el primer iniciador adicional tiene una actividad de unión preferencial para sitios en la región rica de CGG que comprende elementos de interrupción, y la muestra comprende al menos dos alelos que comprenden regiones ricas en CGG de diferentes longitudes, el método además que comprende determinar las longitudes de al menos dos alelos y detectar al menos un elemento de interrupción y determinar la longitud del alelo mediante el cual al menos está comprendido un elemento de interrupción.
- 35 15. El método de conformidad con cualquiera de las reivindicaciones 1 ó 2, en donde la región rica en CGG está compuesta por un 5' UTR de FMR1 o 5' UTR de FMR2.
- 40

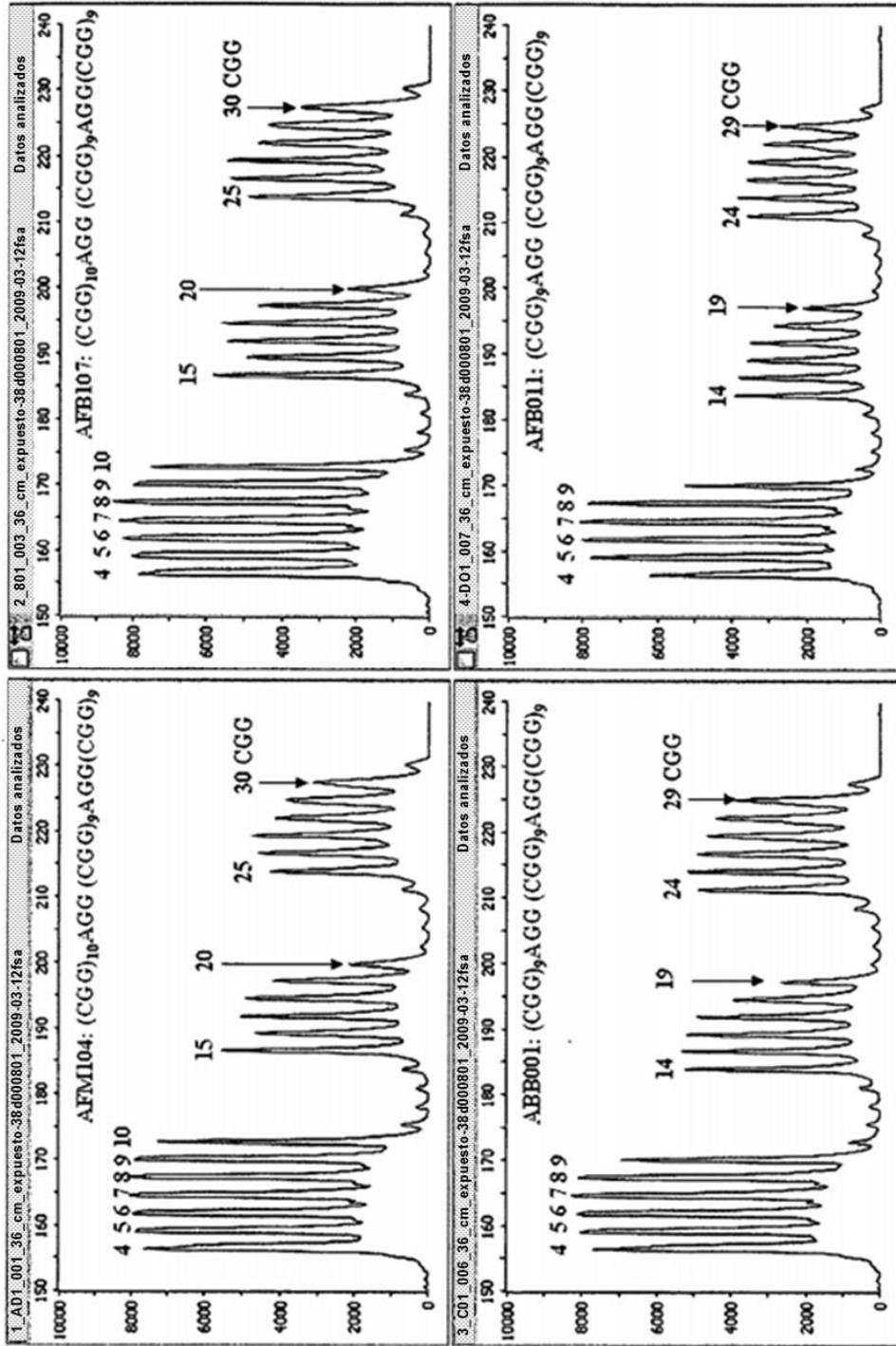


Figura 2A

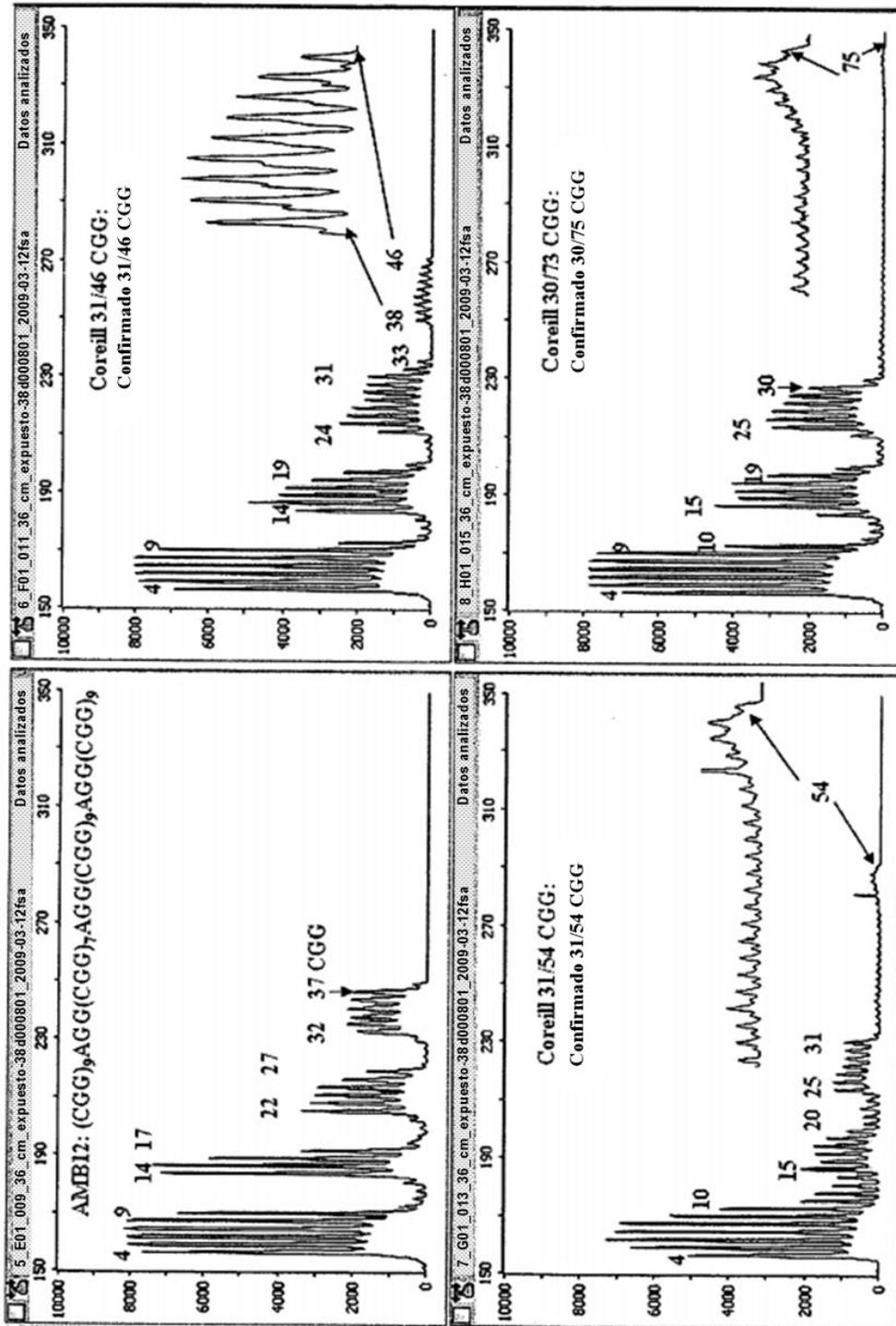


Figura 2B

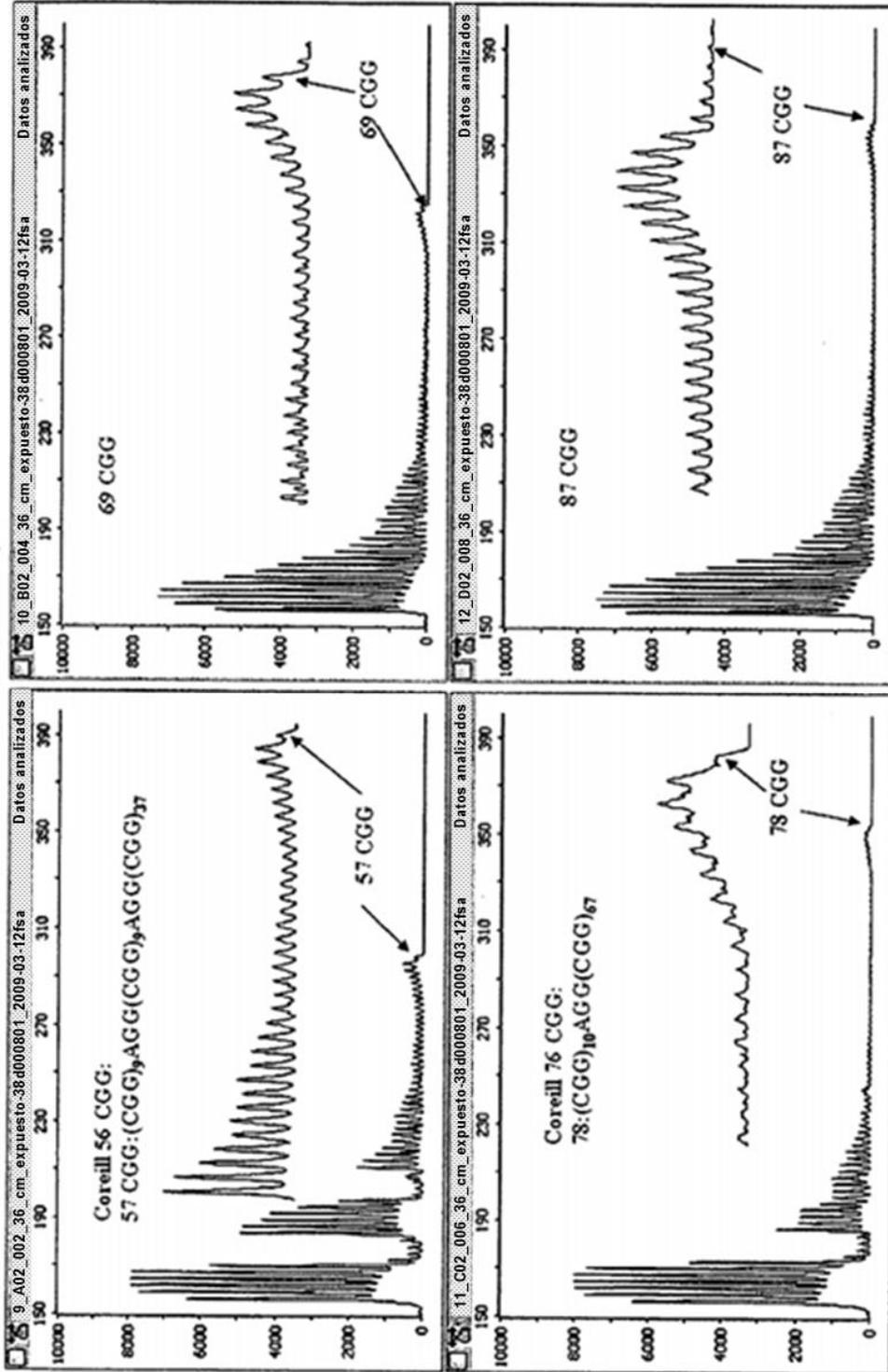


Figura 3A

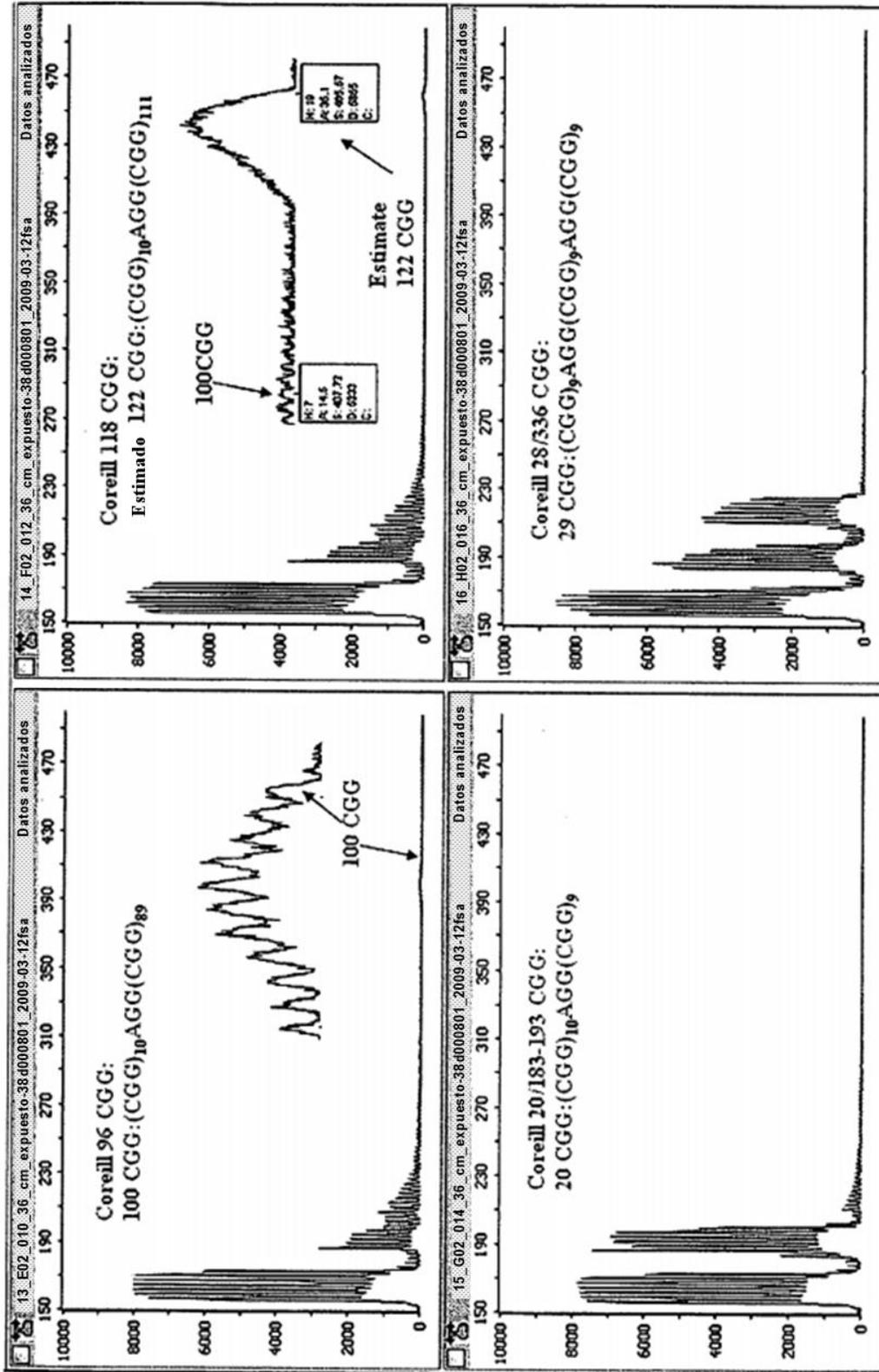
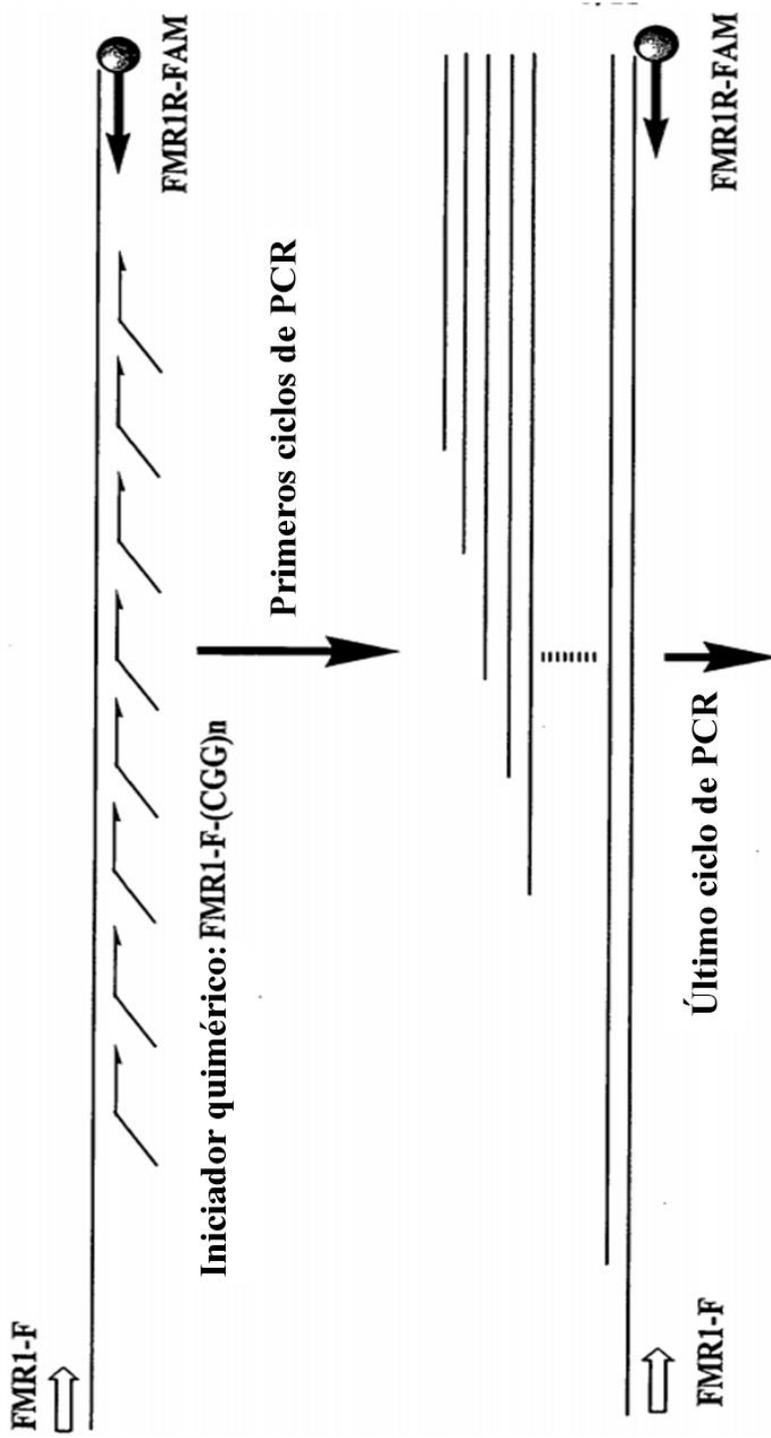


Figura 3B



Los patrones CGG (cortos) compiten con productos de longitud completa para formar un espectro amplio de productos

Figura 4

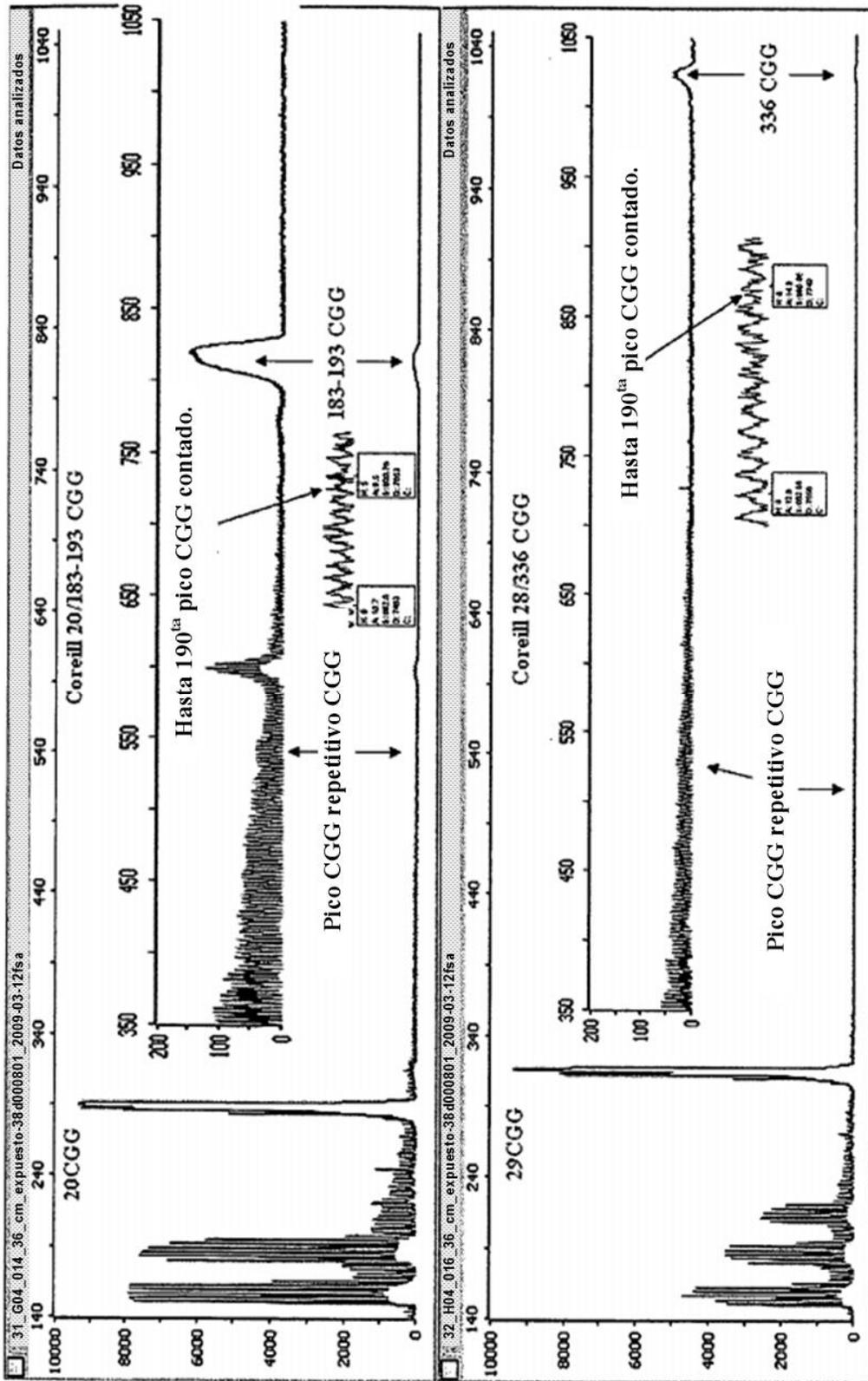


Figura 5

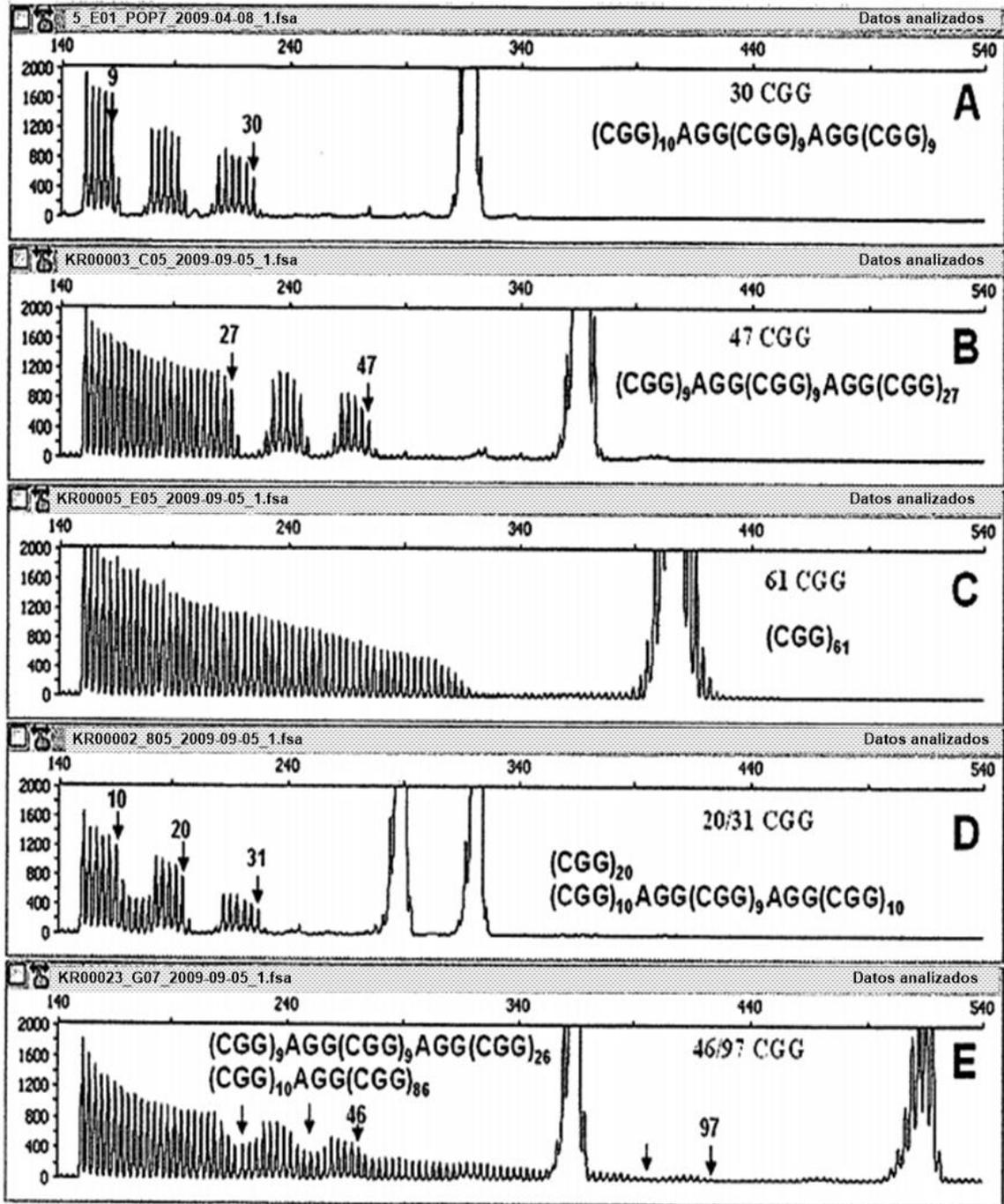


Figura 6

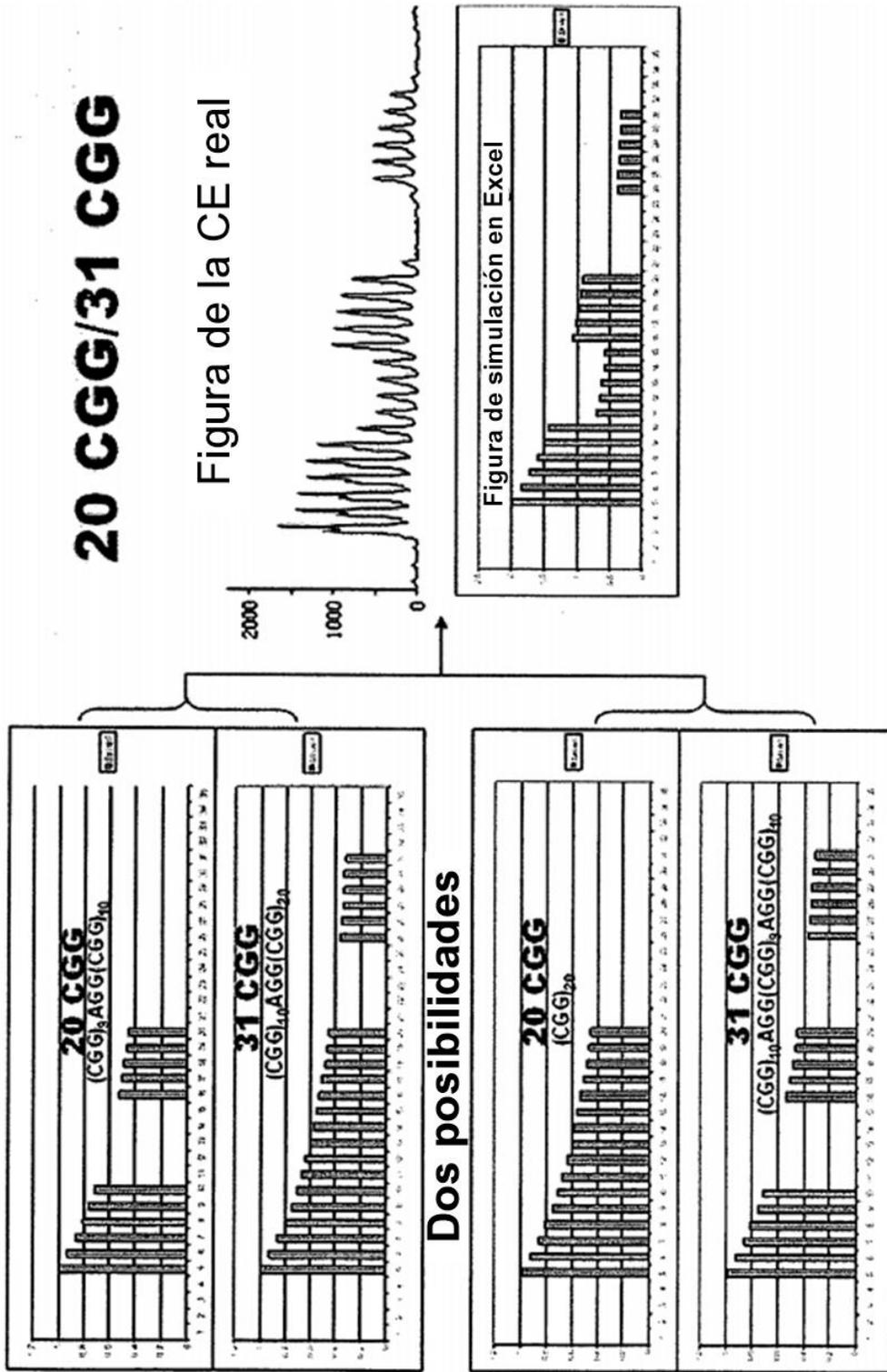


Figura 7

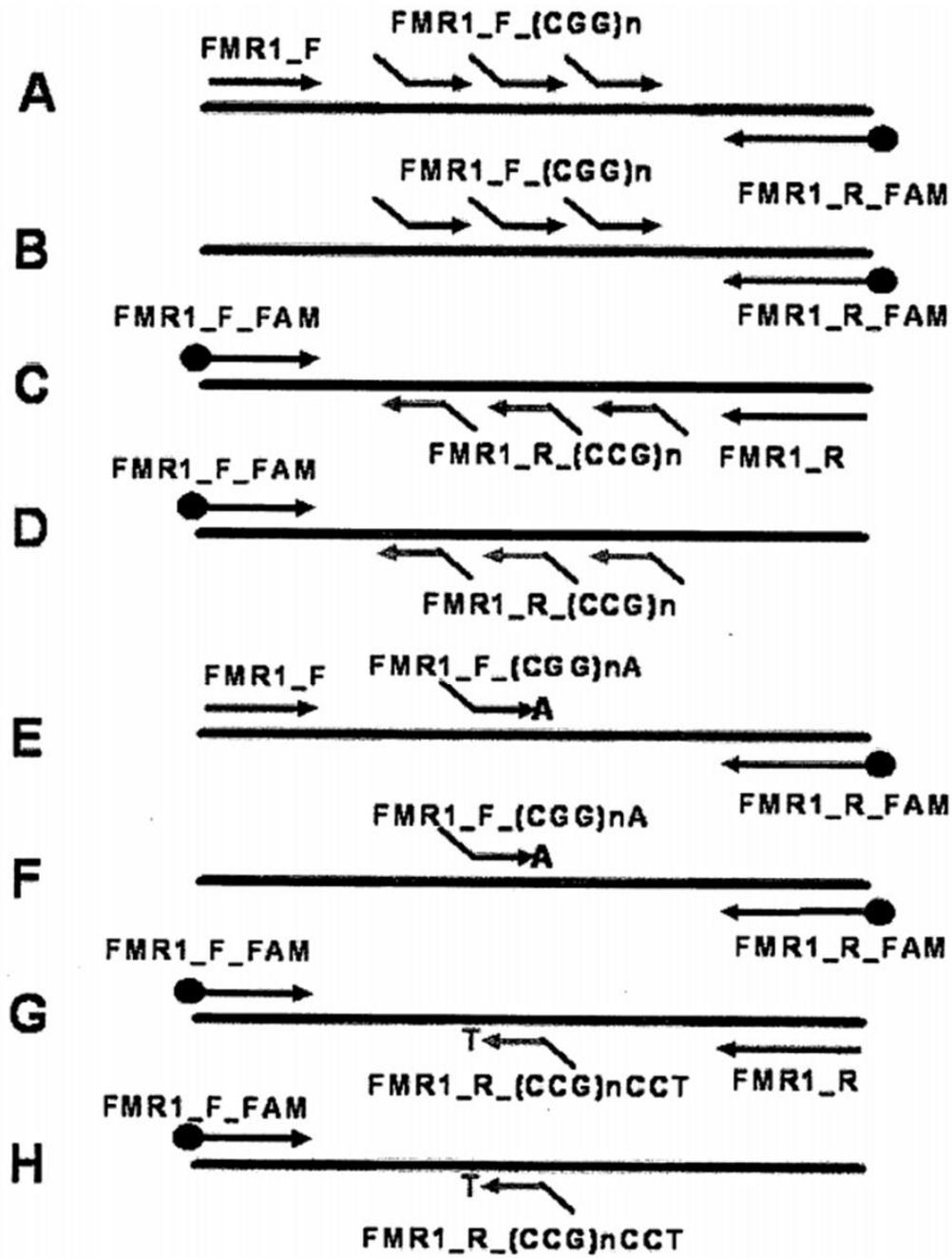


Figura 8

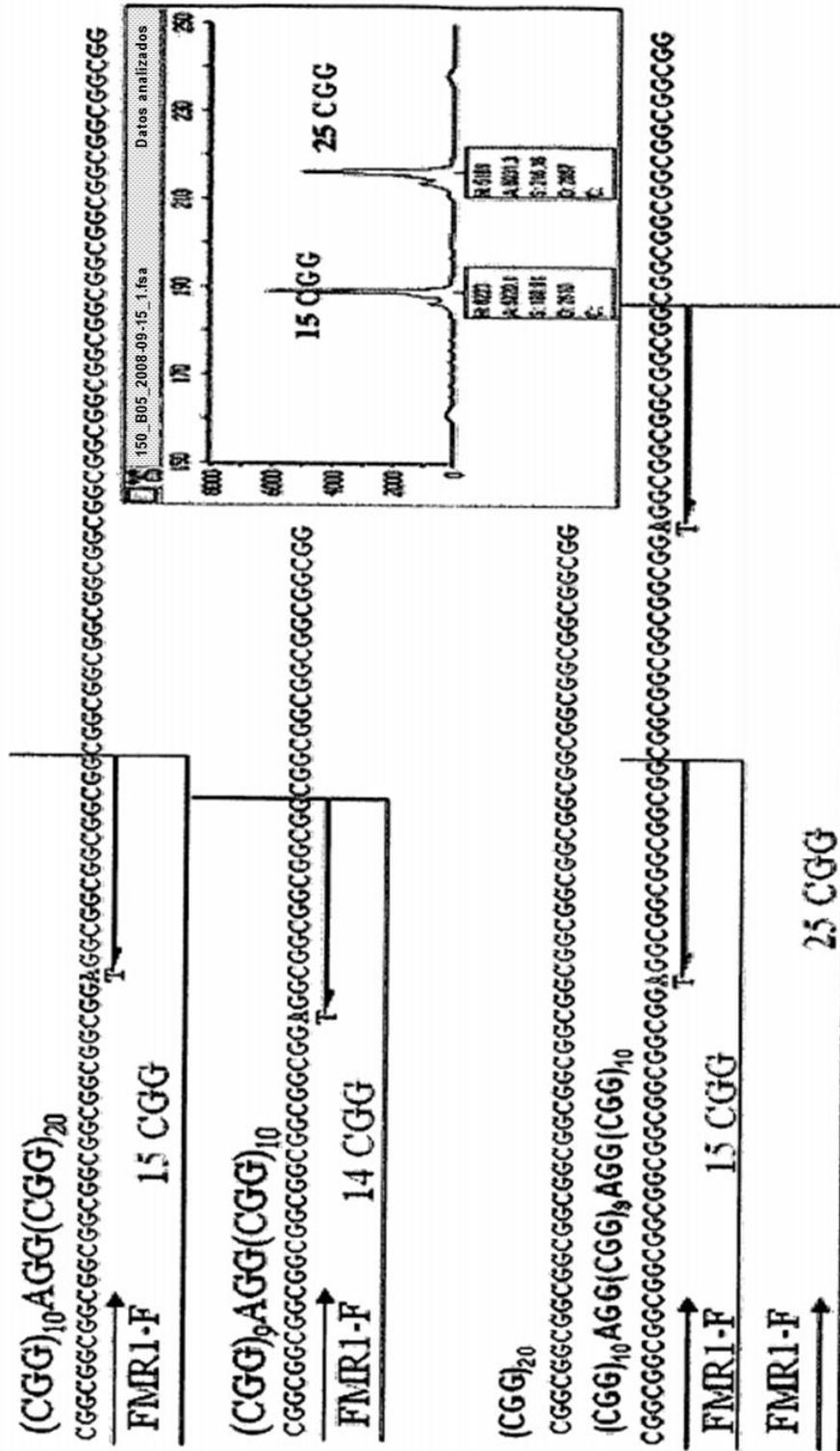


Figura 9

Los tres AGG se asignan al alelo 97

El alelo 46 tiene dos y el 97 tiene un AGG

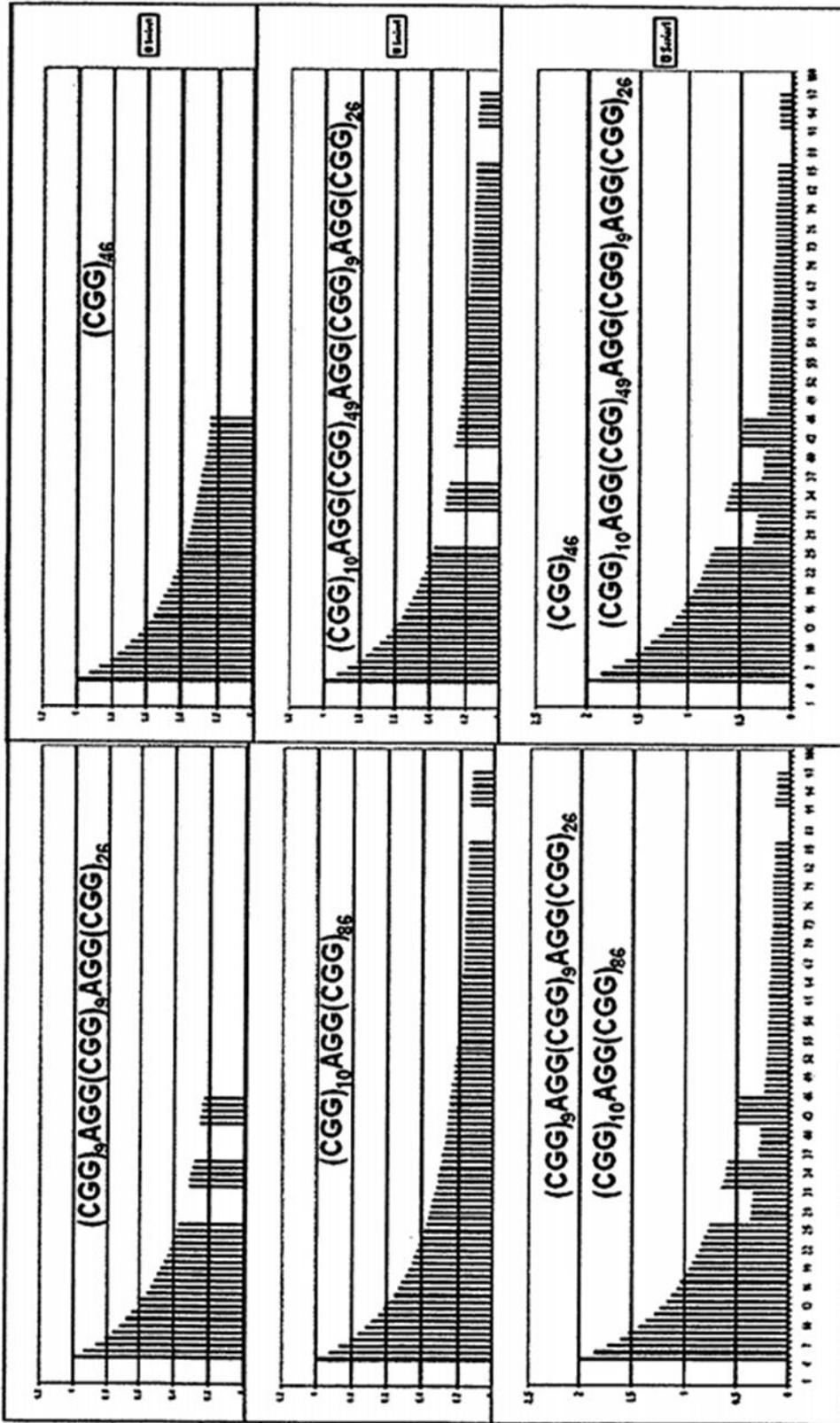


Figura 10

El alelo 46 tiene dos y el 97 tiene un AGG Los tres AGG se asignan al alelo 97

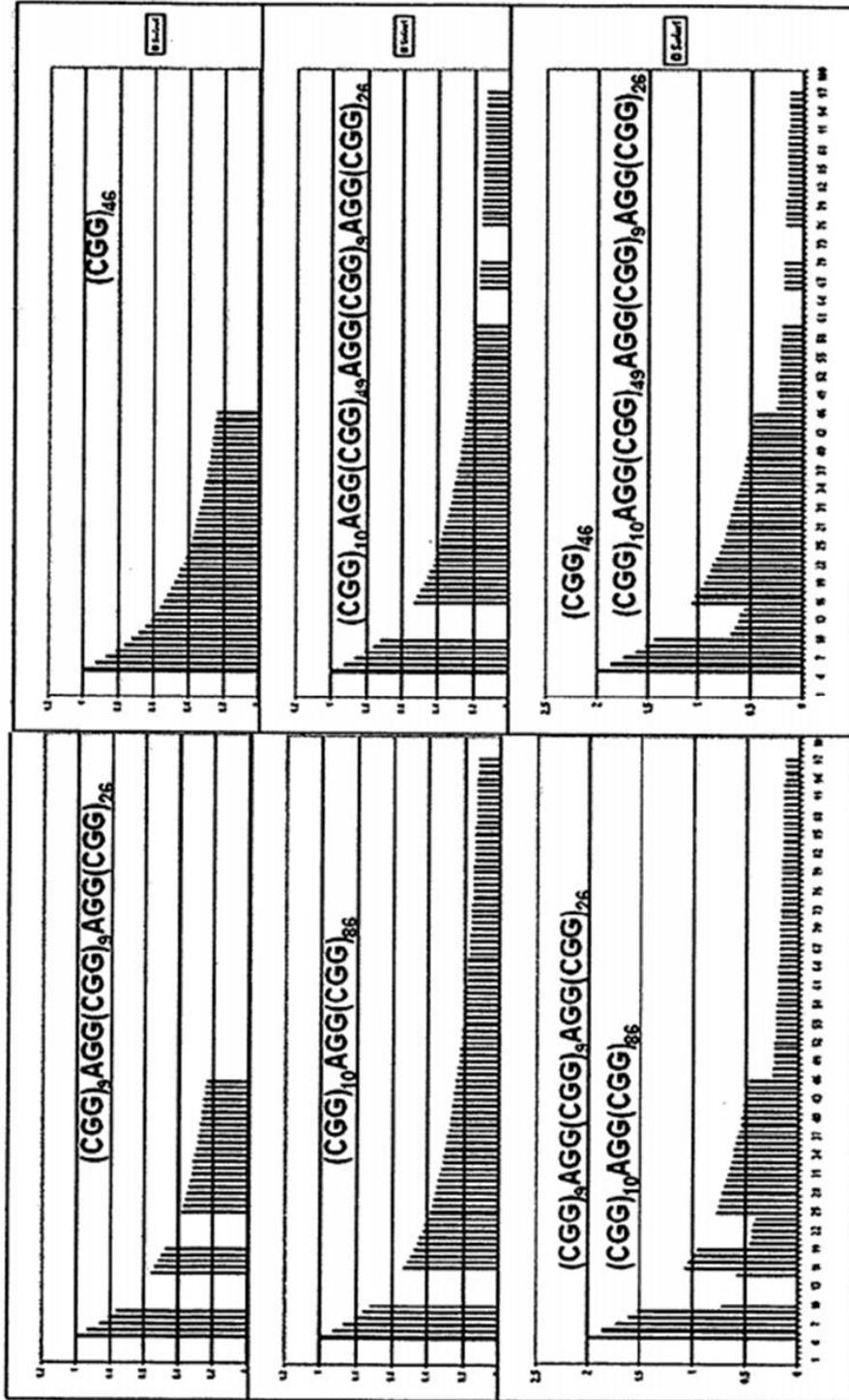


Figura 11

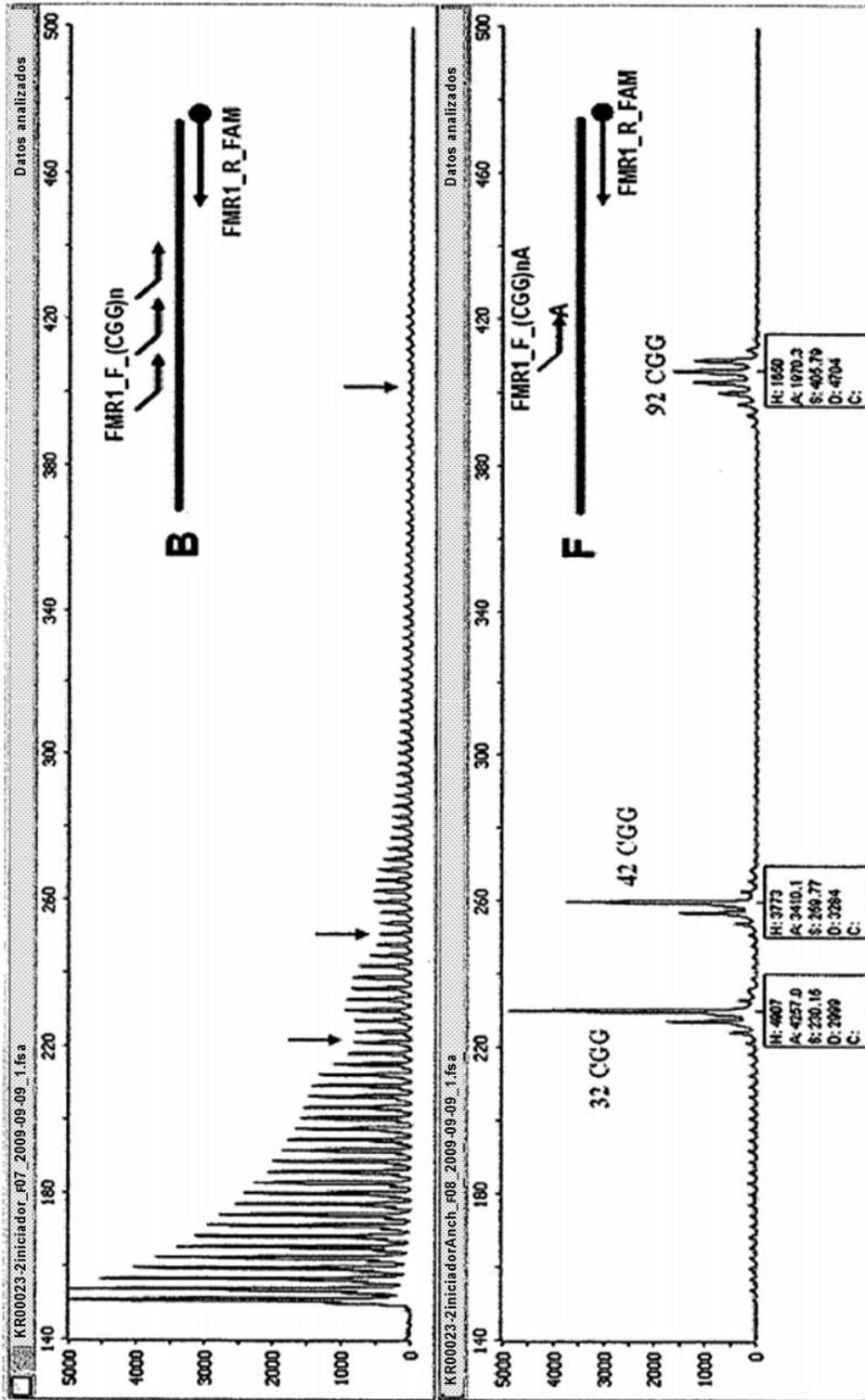


Figura 12

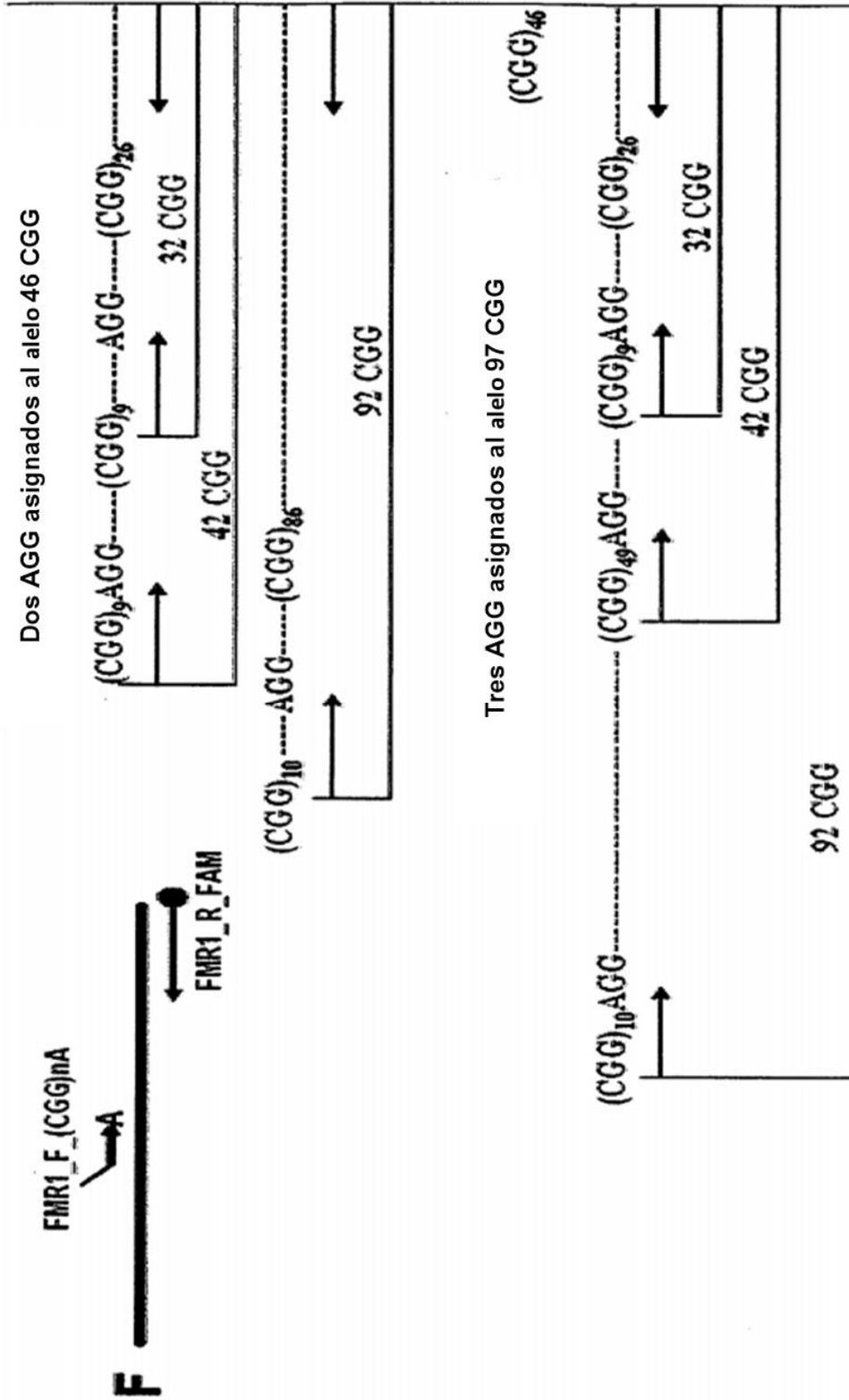


Figura 13

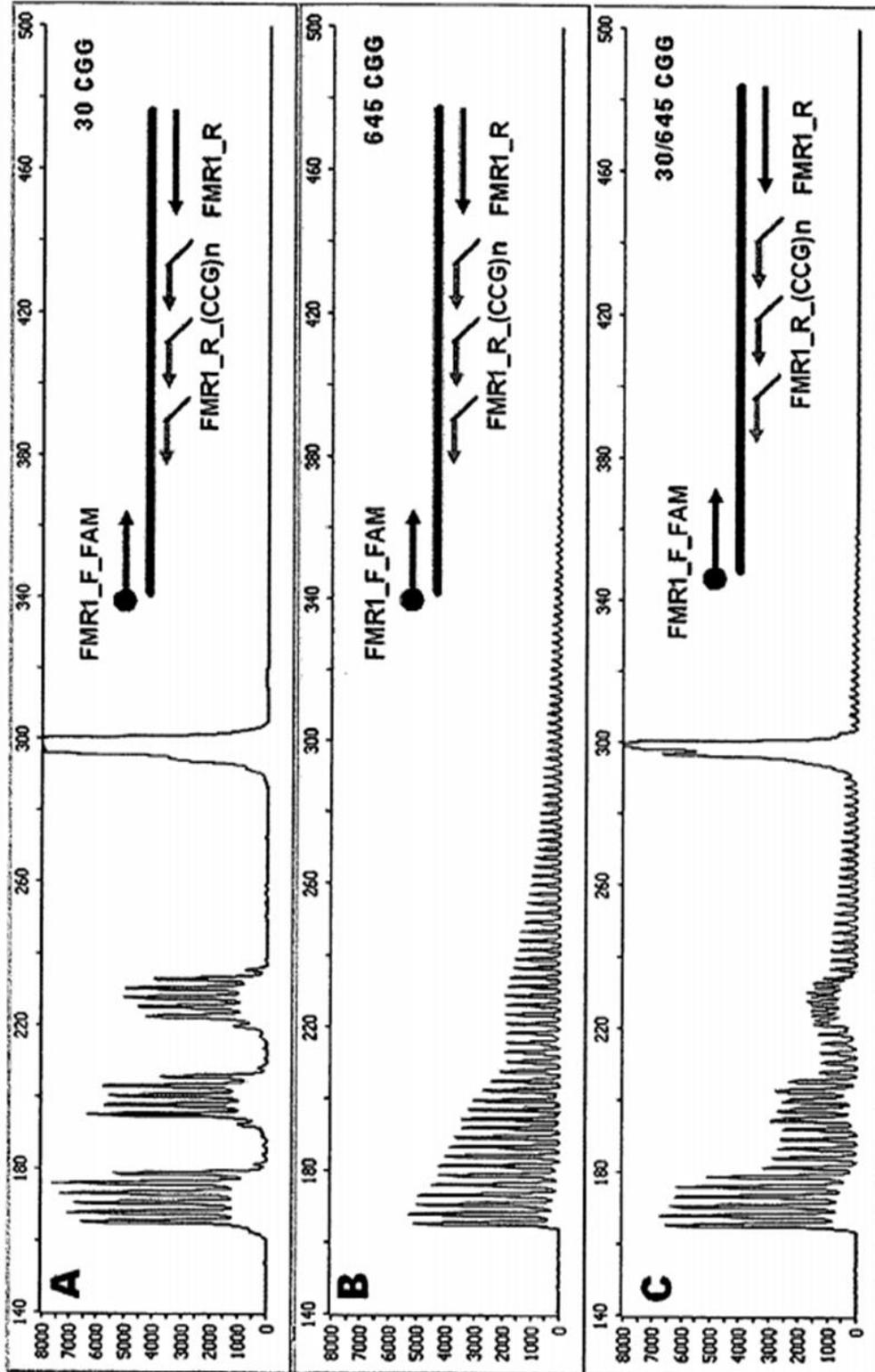


Figura 15

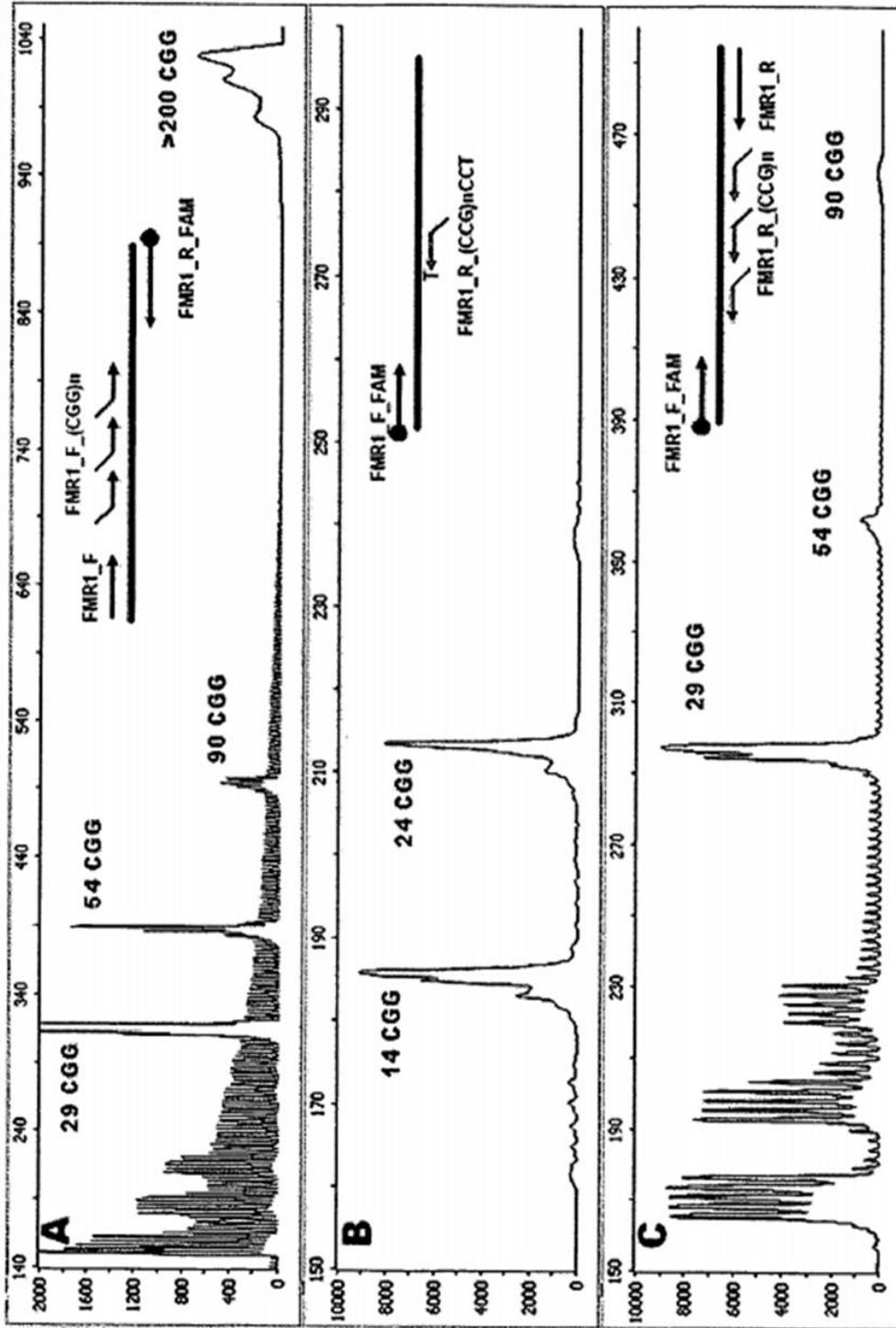


Figura 16

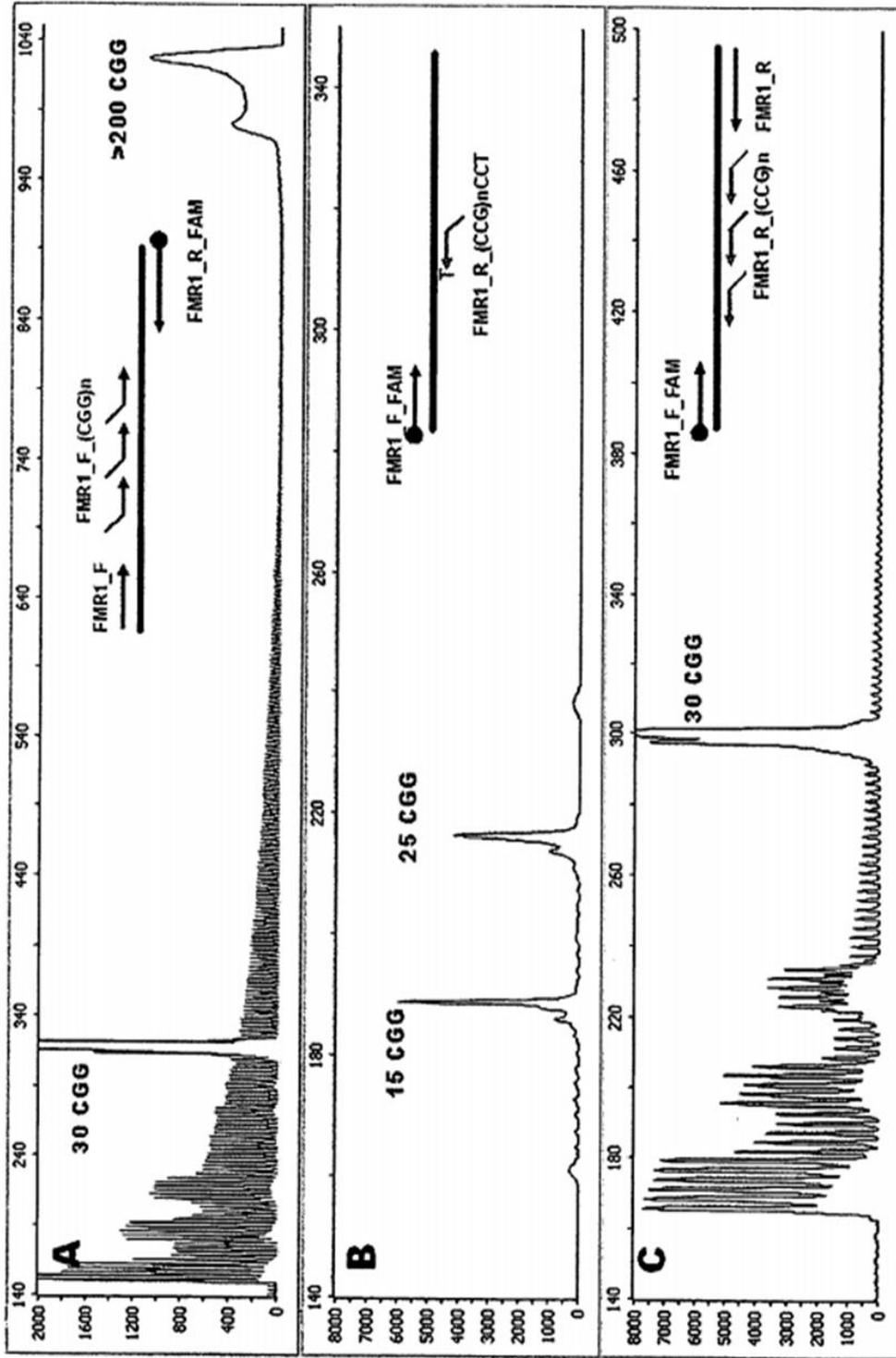


Figura 17

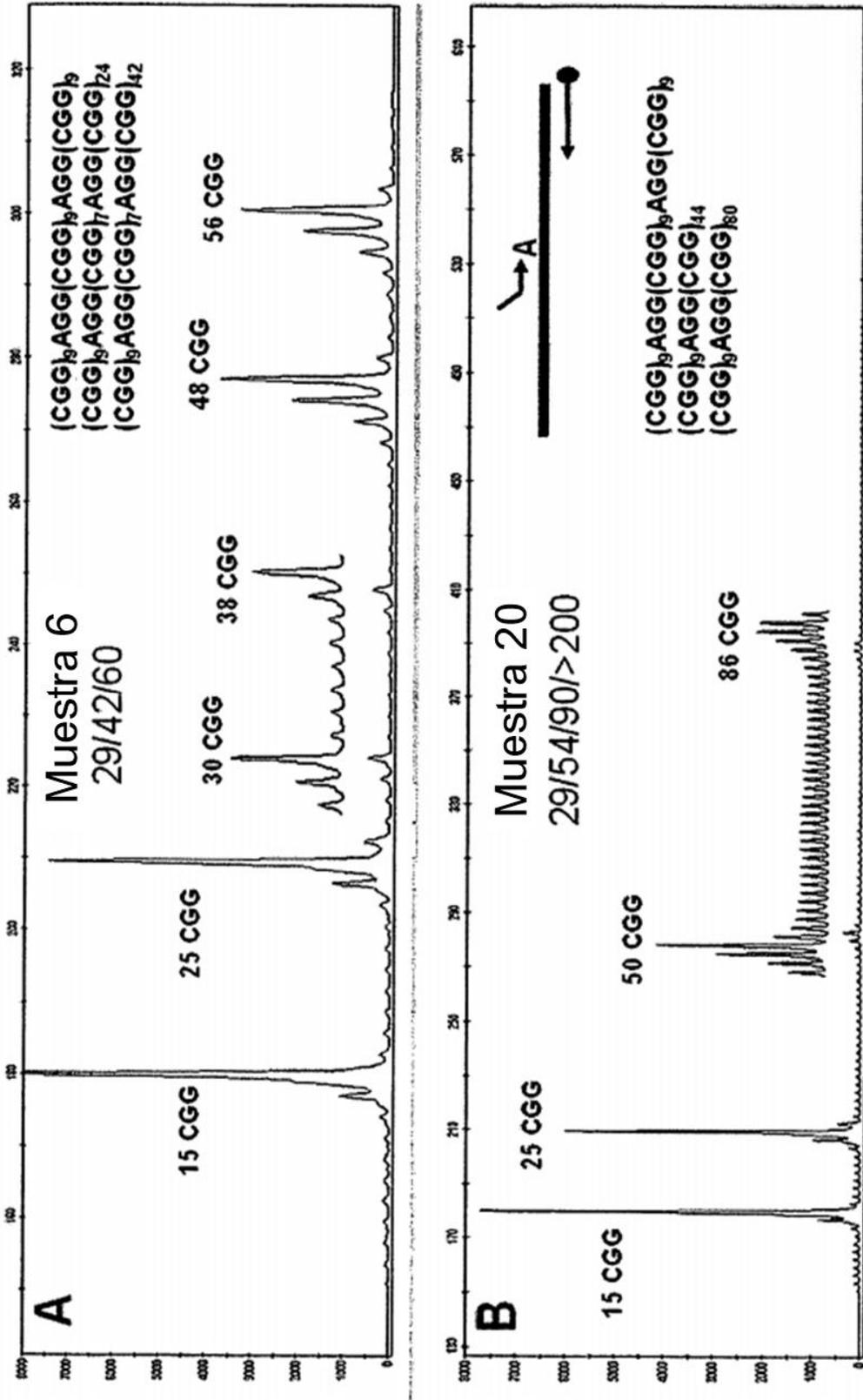


Figura 18

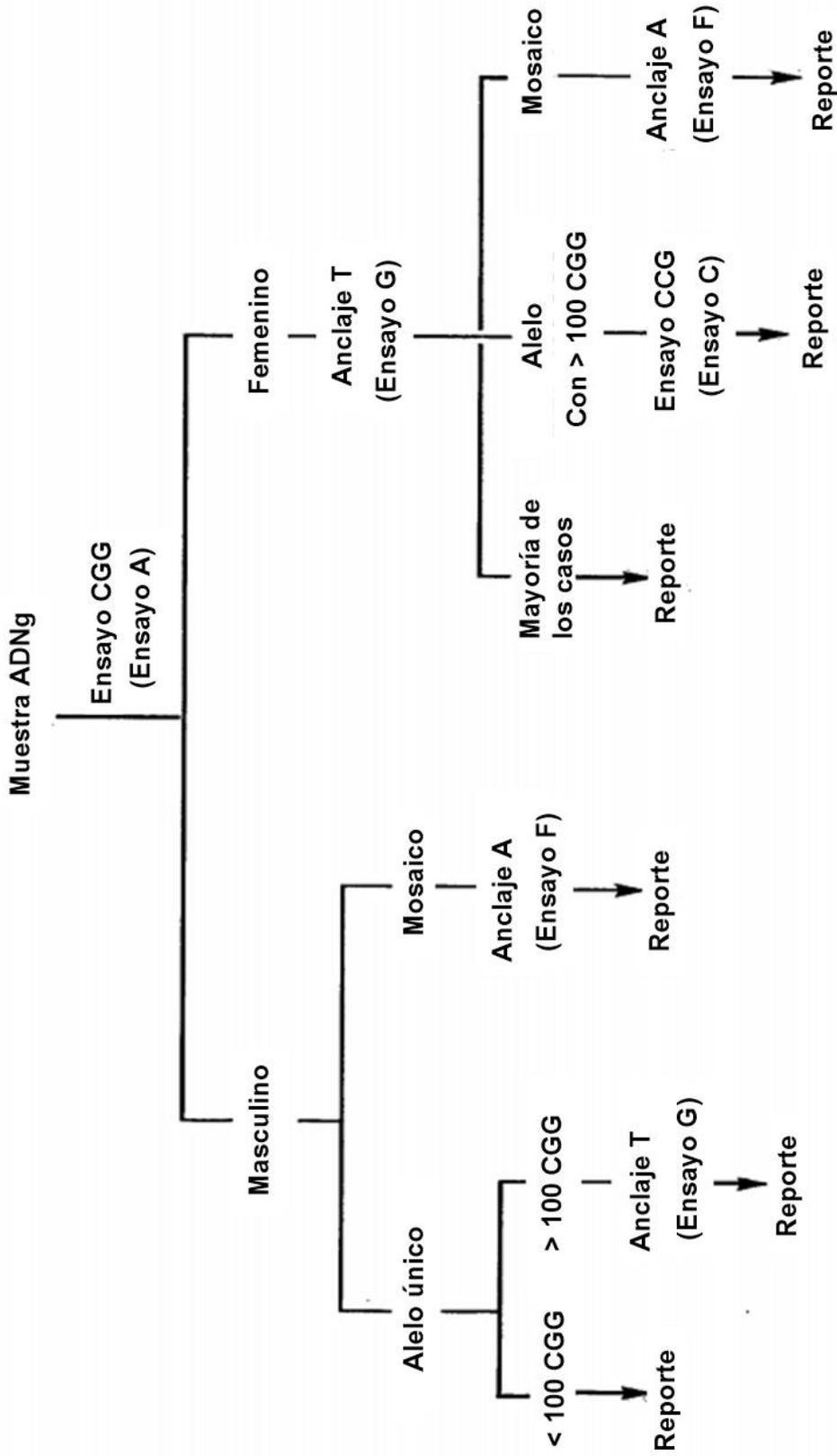


Figura 19