

AVANCES EN EL ANÁLISIS DE ADN MITOCONDRIAL CON FINES FORENSES: LA COLABORACIÓN GHEP-EMPOP

Grupo de Habla Española y Portuguesa de la Sociedad Internacional de Genética Forense (GHEP-ISFG)¹

EDNAP mitochondrial DNA Population database (EMPOP)²

1.- INTRODUCCIÓN

En los últimos años, el genoma mitocondrial ha sido de interés en diversos campos como la clínica, la genética evolutiva y poblacional, y desde principios de los 90 también de la genética forense. Se trata de un genoma peculiar que consiste en una molécula circular de doble hebra y de pequeño tamaño, con sólo 16569 pares de bases aproximadamente (en contraposición a los cerca de 3000 millones del ADN nuclear) [1, 2]. Contiene información para unos cuantos genes, todos ellos relacionados con los procesos bioquímicos que ocurren en las mitocondrias y sólo una pequeña parte de la molécula parece ser no codificante, pero está involucrada en los procesos de replicación (la Región Control o D-Loop). El primer estudio que se realizó respecto a la composición de nucleótidos de este ADN fue llevado a cabo por un grupo de científicos de la universidad de Cambridge [1] y hoy en día a esta secuencia se le llama Secuencia de Referencia de Cambridge (CRS, Cambridge Reference Sequence). Sus nucleótidos están numerados del 1 al 16569 y la nomenclatura actualmente utilizada para nombrar las secuencias obtenidas en las muestras forenses tiene en cuenta sólo las diferencias con la versión corregida de la CRS, la llamada rCRS [2]. La zona más variable y más útil desde el punto de vista de la identificación es la Región Control (CR, Control Region) que se extiende desde el nucleótido 16024 al 576.

El número de copias de ADN mitocondrial (ADNmt) por célula varía entre los tipos celulares y el momento funcional en el que se encuentren éstos, pero es muy superior al número de copias de ADN nuclear [3, 4]; además, al ser una molécula circular se encuentra fuera del alcance de las enzimas exonucleasas. Estos son algunos de los motivos por los que el análisis de ADN mitocondrial tiene interés forense. Al disponerse de tantas copias, el porcentaje de éxito alcanzado en el análisis de ADNmt en muestras biológicas con escasa o ninguna cantidad de ADN nuclear (pelos telogénicos, fragmentos de pelos) y en muestras muy degradadas (cadáveres en muy mal estado de conservación), es muy superior al alcanzado con el análisis de ADN nuclear.

La herencia del ADNmt es también especial; se hereda íntegramente de madres a hijos sin sufrir recombinación, en forma de haplotipo (secuencia o conjunto de polimorfismos que se transmiten juntos, en bloque). Las hijas transmitirán su ADNmt a la siguiente generación, mientras que los hijos no lo transmiten. Por ello se ha utilizado en el campo forense para rastrear linajes maternos en casos de parentesco.

¹ Persona de contacto (en español, portugués o inglés): Lourdes Prieto, lourditasmt@gmail.com

² Contact person (German and English): Walther Parson, walther.parson@i-med.ac.at

Sin embargo, este tipo de herencia no nos permite distinguir individuos relacionados matrilinealmente. Es más, un mismo linaje mitocondrial puede haberse propagado en el pasado de forma más o menos amplia y hoy aparecernos intacto en personas que consideramos no relacionadas familiarmente en la actualidad. Por tanto, el poder de discriminación que se alcanza con este tipo de ADN es muy inferior al obtenido con polimorfismos autosómicos tipo STR, si bien, en la casuística forense nos encontramos con relativa frecuencia que no tenemos otra elección que el análisis mitocondrial.

La distribución de los haplotipos mitocondriales en las distintas áreas geográficas es muy diferente, y esta característica influye enormemente a la hora de realizar una valoración estadística de los resultados de este tipo de prueba. Los haplotipos próximos, que tienen un origen filogenético común, se pueden agrupar en lo que llamamos haplogrupos (conjunto de haplotipos que comparten ciertos polimorfismos debido a que proceden de un antecesor común). Así, hay ciertos polimorfismos que son típicos y característicos de ciertos haplogrupos y que no se encuentran en otros, los llamados "haplogroup motifs". Esta característica nos permite realizar un control de calidad a los haplotipos que obtenemos en las muestras forenses. Si tras el análisis tratamos de asignar el haplotipo obtenido a un haplogrupo y vemos que en nuestra muestra hay ciertos polimorfismos que "no encajan" o no han sido descritos con anterioridad en ese haplogrupo, podemos sospechar que quizá estamos cometiendo un error en el análisis [5].

2.- LOS PROBLEMAS EN EL ANÁLISIS DE ADN MITOCONDRIAL

2.1.- *Problemas debidos a la biología de la molécula*

La tasa de mutación del genoma mitocondrial es mayor que la del nuclear, entre otros motivos porque el ADNmt carece del efecto protector de las histonas y además se encuentra en un medio reactivo, donde ocurren muchos procesos metabólicos. Este hecho acarrea múltiples consecuencias, pero quizá la más llamativa sea la frecuencia con la que nos encontramos heteroplasmas (convivencia de dos tipos de ADNmt, muy parecidos pero no exactamente iguales, en un mismo individuo). Se han descrito dos tipos de heteroplasmas: las puntuales o de secuencia y las de longitud. Es imposible abordar aquí con detenimiento las características de las heteroplasmas, pero sirva como referencia decir que no aparecen con igual probabilidad a lo largo de toda la molécula mitocondrial, existiendo puntos calientes (hot spots) de aparición, como los trectos de poli-C (18184 a 16193, 303 a 315, 515 a 524 y 568 a 573) en el caso de las heteroplasmas de longitud, o las posiciones 16093 o 152 en el caso de las de secuencia.

Esto se debe a que la tasa de mutación en la molécula mitocondrial varía enormemente de unas regiones a otras, de unos nucleótidos a otros, e incluso varía según el tipo de tejido y la edad del individuo. Todo ello complica la interpretación de resultados de la prueba mitocondrial pues podemos encontrarnos muestras del mismo individuo procedente de distintos tejidos o muestras de individuos relacionados vía materna que no sean exactamente idénticas en su ADNmt.

Finalmente, el elevado número de copias de ADNmt, si bien es una ventaja para el éxito del análisis, también es un inconveniente por ser la causa de un alto riesgo de contaminación,

pues un mínimo contacto de una muestra crítica con material biológico de buena calidad producirá un resultado no concluyente.

2.2.- Problemas debidos a la analítica

El análisis de ADNmt presenta más dificultades que el análisis de los tradicionales polimorfismos STR (Short Tandem Repeats) nucleares. La gran cantidad de pasos que requiere el análisis, la ausencia de kits comerciales que faciliten los estudios a realizar y la ausencia de automatización se traducen en un incremento del riesgo de cometer errores y así se ha descrito en la literatura [6, entre otros]. Por ello, la revisión cuidadosa de los datos mitocondriales antes de ser informados es crucial, tanto en la casuística forense como en la realización de bases de datos que permitan conocer la frecuencia de los haplotipos en las poblaciones.

3.- EL PROYECTO GHEP-EMPOP

3.1.- Justificación

Uno de los desafíos mayores a la hora de aplicar el estudio de la molécula mitocondrial a la casuística real es la interpretación de los resultados que se obtienen. En los casos en los que la evidencia biológica (por ejemplo, fragmento de pelo hallado en la mano de una víctima) presenta un mismo tipo mitocondrial que la muestra de referencia (por ejemplo, muestra de mucosa bucal tomada al sospechoso) no se puede excluir que los dos haplotipos mitocondriales provengan de la misma fuente. En este caso es esencial el uso de bases de datos de haplotipos, de colecciones de datos mitocondriales de diferentes partes del mundo, con el fin de evaluar la rareza del haplotipo encontrado.

El tipo de herencia de este tipo de ADN (a través de la línea materna) hace necesario acumular un número muy elevado de muestras para conocer realmente la frecuencia de un haplotipo mitocondrial. Debemos pensar en el ADNmt como si fuera un único marcador genético (realmente lo es, pues se hereda en bloque, sin recombinación). En los marcadores tipo STR que normalmente se utilizan en genética forense, localizados en los cromosomas autosómicos, es suficiente el análisis de una pequeña muestra de la población para saber con qué frecuencia aparece cada alelo. De hecho, con sólo unos 200 individuos aparecen todos los alelos que están presentes en la población, salvo algunos alelos excepcionales. Pero no ocurre así con el ADNmt, donde el número de haplotipos es infinitamente mayor y el estudio de un grupo pequeño de individuos sólo nos da una visión parcial de la frecuencia con que aparecen en la población de estudio.

Uno de los proyectos más completos y ambiciosos de colección de haplotipos mitocondriales se está llevando a cabo en el Instituto de Medicina Legal de la Universidad de Innsbruck (Austria). Este proyecto, llamado base de datos EMPOP (EDNAP mitochondrial DNA Population database), presenta las ventajas de que todos sus datos han sido sometidos a exhaustivos controles de calidad y de que es posible realizar búsquedas de haplotipos procedentes de todo el mundo en sólo unos segundos [7], lo cual facilita la estimación de la frecuencia del haplotipo en estudio.

El Grupo de Habla Española y Portuguesa de la International Society for Forensic Genetics (GHEP-ISFG) propuso a EMPOP un trabajo de colaboración a finales del año 2008 con el fin de aportar un importante número de haplotipos a la base de datos [8]. El análisis de ADNmt de un elevado número de muestras es tedioso para un solo laboratorio y por ello se hace necesario realizar un esfuerzo conjunto entre varios laboratorios. En el GHEP hay un número bastante elevado de laboratorios forenses que realizan análisis mitocondrial, por lo que la unión de todos ellos permite realizar estudios poblacionales a gran escala.

Por otro lado, si bien EMPOP contenía en aquella época haplotipos procedentes de donantes de muchas partes del mundo, no existían muestras de poblaciones Nativo-Americanas (EMPOP Release 1: 4476 haplotipos europeos, 162 del este de Asia, 187 del sureste de Asia y 348 haplotipos de África). Se hacía por tanto necesario, realizar un esfuerzo para completar esta base de datos con el fin de que fuera útil para valorar la evidencia mitocondrial en nuestras poblaciones.

3.2.- Laboratorios participantes, muestras y requerimientos

Con el fin de obtener datos de la mayor calidad posible se fijaron algunos requerimientos a cumplir por los laboratorios participantes: (i) superación del ejercicio colaborativo del GHEP 2008 en el apartado de ADN mitocondrial; (ii) envío de datos procedentes de al menos 50 donantes no relacionados por vía materna de los cuales se conociera el origen geográfico (región/ciudad/población); (iii) secuenciación de al menos las regiones HV1 (16024-16365) y HV2 (73-340) y (iv) doble lectura (ambas hebras) de cada secuencia.

En la Tabla 1 se muestran los laboratorios participantes y el número de haplotipos aportados por cada uno de ellos.

3.3.- El proceso de revisión

Desde que se obtiene el resultado del análisis de ADNmt del secuenciador, hasta que dicho resultado se informa, hay aún un proceso que cumplimentar con el fin de evitar errores indeseables. El proceso de revisión de un haplotipo individual puede resumirse en las siguientes etapas:

- a) Comprobación de que todos los polimorfismos se detectan en ambas hebras, muy importante en el caso de muestras que presentan heteroplasmas de longitud y de secuencia.
- b) Lectura de las secuencias por parte de dos expertos diferentes en diferentes momentos.
- c) Determinación del rango de lectura dentro de la región analizada. Este rango se asemeja a lo que sería el nombre de un marcador STR en ADNn y sin él, el haplotipo no es válido.
- d) Comprobación de que todos los polimorfismos detectados están descritos con anterioridad (y si lo están en el mismo haplogrupo al que pertenece la muestra en estudio).
- e) En casos de heteroplasma de secuencia, chequear si la posición en donde aparece tiene una tasa de mutación elevada [9].

- f) Comprobar que el haplotipo obtenido tiene sentido desde el punto de vista filogenético [10].
- g) Realizar búsquedas en bases de datos con el fin de comprobar si el haplotipo completo se ha descrito con anterioridad.
- h) Si el haplotipo obtenido en el secuenciador no se puede transferir electrónicamente al informe final, realizar un doble chequeo por parte de dos personas distintas, pues los errores tipográficos son muy frecuentes.

Este proceso de revisión se hace tedioso y complicado en el caso de tener que chequear una gran cantidad de datos como el caso que nos ocupa. Por ello, en esta colaboración GHEP-EMPOP hemos evaluado además los haplotipos remitidos por los laboratorios con herramientas (software) basados en análisis filogenético (Networks, [11]), con el fin de evitar las siguientes fuentes de error (ver Figura 1):

- a) Desvíos respecto a la secuencia de referencia (reference bias)
- b) Mutaciones fantasma (polimorfismos no reales, producto de artefactos)
- c) Error en el nucleótido informado (Base mis-scoring)
- d) Errores de nomenclatura
- e) Errores en los alineamientos con la secuencia de referencia
- f) Errores de escritura

También hemos realizado un proceso de estandarización con el fin de dar uniformidad a los datos, relativa a la asignación de haplogrupos ([11], phylotree versión 10), al alineamiento y notación de las variantes de longitud, a la confirmación de las heteroplasmias puntuales y a la revisión de la filiación de las muestras; y se ha intentado además llegar a la mayor uniformidad en el rango de análisis.

El proceso de recolección y organización de los haplotipos procedentes de los laboratorios participantes, así como la revisión de los mismos se realizó en la Comisaría General de Policía Científica de Madrid y en el Instituto de Medicina Legal de Innsbruck. Todos los polimorfismos se cruzaron con una colección de las mutaciones fantasma más comunes [12] y los polimorfismos aparentemente no descritos con anterioridad se evaluaron mediante búsquedas en la literatura y/o realizando búsquedas directas en Internet [13].

Mención especial merece el hecho de que el correcto alineamiento de los haplotipos obtenidos en el laboratorio con respecto a la secuencia de referencia (rCRS) es de gran importancia para nombrar el haplotipo de forma adecuada. Existen situaciones (fundamentalmente cuando una secuencia muestra polimorfismos cercanos a los trectos de poli-C) en las que hay varios alineamientos posibles, lo cual resulta en diferentes nomenclaturas de un mismo haplotipo (ver Figura 2A). En este trabajo hemos unificado los criterios de alineamiento siguiendo las recomendaciones descritas en [14] (máxima parsimonia y sentido filogenético). EMPOP dispone además de un sistema de alineamiento (algoritmo SAM) [15] que evita inconsistencias en la nomenclatura, ya que aunque dos secuencias idénticas se introduzcan en la base de datos con nomenclaturas diferentes, el sistema SAM reconocerá que se trata del mismo haplotipo (ver Figura 2B). La gran utilidad de este algoritmo radica en que cuando se compara un haplotipo con la base de datos para evaluar su

frecuencia, se evitan falsas exclusiones en la comparación de haplotipos y errores en la valoración estadística de los resultados de las pericias basadas en ADNmt.

3.4.- Resultados

Hemos evaluado un total de 1019 haplotipos procedentes de 9 poblaciones. La tasa de error encontrada en estos haplotipos resultó ser del 8,5% (ver Tabla 2). El tipo de error más frecuente consistió en informar un nucleótido erróneo en una posición correcta (base mis-scoring). Normalmente este error consiste en informar el nucleótido correspondiente a la secuencia de referencia (rCRS) en lugar de informar el nucleótido real que aparece en la muestra que estamos analizando. Esta fuente de error tiene su origen en la transferencia manual de los datos analíticos al informe final, por una pobre revisión en la última fase del proceso. Se puede evitar fácilmente usando la propia base de datos EMPOP, pues si se introduce un haplotipo que contenga algún nucleótido de la rCRS nos aparecerá un aviso alertándonos de tal error (ver Figura 3).

4.- CONCLUSIONES

La transmisión del significado de la coincidencia entre una muestra desconocida y una muestra de referencia tras el análisis de ADN mitocondrial es una de las partes más delicadas y difíciles del análisis genético forense. Esto se debe a que el público en general y los profesionales del derecho están más o menos habituados al elevado poder de discriminación del ADN nuclear, que es el tipo de ADN que se usa de rutina en la mayoría de los casos forenses.

Por otro lado, a pesar de que hay numerosas publicaciones sobre estudios poblacionales, su calidad es dudosa, lo cual hace que estos estudios no se puedan utilizar para evaluar la evidencia mitocondrial. Y esta es una de las principales razones por las que iniciativas como EMPOP deben extenderse. Debido a la gran diversidad de poblaciones que hay en el GHEP y con el fin de aunar esfuerzos y poner las bases de datos individuales disponibles para toda la comunidad forense, hemos creído conveniente realizar este proyecto con EMPOP. Nuestro proceso de revisión confirma que la mayoría de los errores del análisis mitocondrial se concentran en el proceso manual de documentación [7, 16], debido a la ausencia de revisión en profundidad. El análisis filogenético resulta de gran utilidad para evitar errores y nos permite un mejor entendimiento de la distribución de haplotipos mitocondriales en la población mundial.

El impacto que esta colaboración ha tenido en los laboratorios del GHEP ha sido tremendamente positivo. En el ejercicio colaborativo del GHEP del año 2008, la tasa de error en el análisis de ADNmt se situaba alrededor del 13%; tras el proyecto GHEP-EMPOP, la tasa de error se redujo al 5% en el año 2010. Este hecho nos da una idea de los beneficios que aportan este tipo de colaboraciones.

Finalmente cabe destacar, que esta colaboración tiene continuidad y no está ni mucho menos cerrada a los laboratorios del GHEP. En la actualidad cualquier laboratorio puede enviar sus datos poblacionales al GHEP o directamente a EMPOP para que, tras ser sometidos a un control de calidad confidencial, puedan ponerse a disposición de toda la comunidad forense a

través de la página web EMPOP. Con ello, los laboratorios consiguen mejorar la calidad de sus análisis y adquieren un conocimiento profundo del comportamiento de esta molécula; y la comunidad forense en general se beneficia del incremento en el número de datos disponibles para poder realizar valoraciones de la evidencia mitocondrial más precisas.

AGRADECIMIENTOS

Este proyecto ha sido posible gracias al trabajo y a la generosidad de los siguientes científicos: L. Prieto, B. Zimmermann, A. Goios, A. Rodríguez-Monge, GG. Paneto, C. Alves, A. Alonso, C. Fridman, S. Cardoso, G. Lima, MJ. Anjos, MR. Whittle, M. Montesino, RMB. Cicarelli, AM. Rocha, C. Albarrán, MM. de Pancorbo, MF. Pinheiro, M. Carvalho, DR. Sumita, W. Parson. Antonio Amorim (IPATIMUP) puso todos los medios para la primera reunión del GHEP y EMPOP y nos enriqueció con sus comentarios. Arne Dür (Facultad de Matemáticas, Innsbruck) y Alexander Röck (Instituto de Medicina Legal, Innsbruck) nos proporcionaron muchas de las herramientas informáticas para el análisis de los datos (programación de las Networks de EMPOP y herramienta de búsqueda SAM).

BIBLIOGRAFÍA

- [1] S. Anderson, AT. Bankier, BG. Barrell, MH. de Bruijn, AR. Coulson, J. Drouin, IC. Eperon, DP. Nierlich, BA. Roe, F. Sanger, PH. Schreier, AJ. Smith, R. Staden, IG. Young, Sequence and organization of the human mitochondrial genome, *Nature* 290 (5806) (1981) 457–65.
- [2] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, *Nat. Genet.* 32 (1999) 147.
- [3] ED. Robin, R. Wong, Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells, *J. Cellular Physiology* 136 (3) (1988) 507–513.
- [4] M. Satoh, T. Kuroiwa, Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell, *Experimental Cell Res.* 196 (1) (1991) 137–140.
- [5] A. Salas, H.-J. Bandelt, V. Macaulay, M.B. Richards, Phylogeographic investigations: the role of trees in forensic genetics, *Forensic Sci. Int.* 168 (1) (2007) 1–13.
- [6] HJ. Bandelt, P. Lahermo, M. Richards, V. Macaulay, Detecting errors in mtDNA data by phylogenetic analyses, *Int. J. Legal Med.* 115 (2001) 64–69.
- [7] W. Parson, A.Dür, EMPOP—a forensic mtDNA database, *Forensic Sci. Int. Genet.* 1 (2007) 88–92.
- [8] L. Prieto, B. Zimmermann, A. Goios, A. Rodríguez-Monge, GG. Paneto, C. Alves, A. Alonso, C. Fridman, S. Cardoso, G. Lima, MJ. Anjos, MR. Whittle, M. Montesino, RMB. Cicarelli, AM. Rocha, C. Albarrán, MM. de Pancorbo, MF. Pinheiro, M. Carvalho, DR. Sumita, W. Parson, The GHEP–EMPOP collaboration on mtDNA population data—A new resource for forensic casework, *Forensic Sci. Int.: Genetics* 5 (2011) 146–151.
- [9] JA. Irwin, JL. Saunier, H. Niederstätter, KM. Strouss, KA. Sturk, TM. Diegoli, A. Brandstätter, W. Parson W, TJ. Parsons (2009), Investigation of Heteroplasmy in the Human Mitochondrial DNA Control Region: A Synthesis of Observations from More than 5000 Global Population Samples, *J. Mol. Evol.* 68 (5) (2009) 516–527.

- [10] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, *Hum. Mutat.* 30 (2009) E386–E394.
- [11] A. Brandstätter, R. Klein, N. Duftner, P. Wiegand, W. Parson, Application of a quasi-median network analysis for the visualization of character conflicts to a population sample of mitochondrial DNA control region sequences from southern Germany (Ulm), *Int. J. Legal Med.* 120 (2006) 310–314.
- [12] A. Brandstätter, T. Sängler, S. Lutz-Bonengel, W. Parson, E. Béraud-Colomb, B. Wen, QP. Kong, CM. Bravi, HJ. Bandelt, Phantom mutation hotspots in human mitochondrial DNA, *Electrophoresis* 26 (2005) 3414–3429.
- [13] HJ. Bandelt, A. Salas, CM. Bravi, What is a ‘novel’ mtDNA mutation – and does ‘novelty’ really matter?, *J. Hum. Genet.* 51 (2006) 1073–1082.
- [14] HJ. Bandelt and W. Parson, Consistent treatment of length variants in the human mtDNA control region: a reappraisal, *Int. J. Leg. Med.* 122 (2008) 11–21.
- [15] A. Röck, J. Irwin, A. Dur, T. Parsons, W. Parson, SAM: String-based sequence search algorithm for mitochondrial DNA database queries, *Forensic Sci. Int. Genetics* 5 (2011) 126–132.
- [16] W. Parson, A. Brandstätter, A. Alonso, N. Brandt, B. Brinkmann, A. Carracedo, D. Corach, O. Froment, I. Furac, T. Grzybowski, K. Hedberg, C. Keyser-Tracqui, T. Kupiec, S. Lutz-Bonengel, B. Mevag, R. Ploski, H. Schmitter, P. Schneider, D. Syndercombe-Court, E. Sørensen, H. Thew, G. Tully, R. Scheithauer, The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives, *Forensic Sci. Int.* 139 (2004) 215–226.
- [17] HJ. Bandelt, A. Dür, Translating DNA data tables into quasi-median networks for parsimony analysis and error detection, *Mol. Phylogenes and Evol.* 42 (2007) 256–271.

Tabla 1.- Listado de Laboratorios participantes en el proyecto de colaboración GHEP-EMPOP

Laboratorio	Muestras	Rango de análisis
Comisaría General de Policía Científica (Madrid, España)	249	Variable, pero al menos 16024-16365 y 72-340
Instituto Nacional de Toxicología y Ciencias Forenses, INTCF (Madrid, España)	154	16024-16365 y 73-340
Laboratorio de Paternidade, UNESP, Univ. Estadual Paulista (São Paulo, Brasil)	142	16024-576
Instituto de patología Molecular e Inmunología de la Universidad de Porto, IPATIMUP (Porto, Portugal)	132	16024-576
Departamento de Medicina Legal, Bioética y Medicina Ocupacional, Facultad de Medicina, Universidad de São Paulo (Brasil)	102	16024-576
BIOMICS Research Group. Centro de Investigación y Estudios Avanzados “Lucio Lascaray”. Universidad del País Vasco (Vitoria-Gasteiz, España)	84	16024-16383 y 66-370
Instituto Nacional de Medicina Legal. Delegación Norte (Porto, Portugal)	55	16024-16391 30-408; 10 SNPs R. cod + 1 SNP R. no codificante
Instituto Nacional de Medicina Legal, Delegación Centro (Coimbra, Portugal)	53	16024-16365 y 72-340
Genomic Engenharia Molecular (São Paulo, Brasil)	48	16024-16365 y 73-340
TOTAL	1,019	

Tabla 2.- Clasificación de las ambigüedades tras la revisión haplotipos y del dato bruto

Tabla 2a

DESVÍOS DE LA SECUENCIA DE REFERENCIA		
Polimorfismo	Veces	
72C	1	
73G	2	
210G	1	
315.1C	1	
16355T	1	
16360T	1	
16390A	1	Total = 8

Tabla 2b

MUTACIONES FANTASMA		
Posición	Veces	
16293M	1	
527G	1	Total = 2

Tabla 2c

ERROR EN EL NUCLEÓTIDO INFORMADO			
Erróneo	Correcto	Veces	
114G	114A	1	
146T	146C	1	
150C	150T	2	
150G	150T	1	
152T	152C	2	
195T	195C	1	
16278G	16278T	1	
16356T	16356C	3	Total = 12

Tabla 2d

NOMENCLATURA		
Posición	Veces	
309.2C sin 309.1C	8	Total = 8

Tabla 2e

ERRORES DE ALINEAMIENTO			
Posición		Veces	
523.1C 524.1A en lugar de 524.1A 524.2C		3	Total = 3

Tabla 2f

ERRORES DE ESCRITURA			
Erróneo	Correcto	Veces	
163G	263G	1	
315C	315.1C	2	
1620G	16207G	1	
16218C	16182C	1	
16223	16223T	1	
16278C	16288C	1	
19294T	16294T	1	Total = 8

Figura 1 (procedente de la figura 4 de [17]).- Networks: representaciones gráficas de un conjunto de datos de ADNmt. Los datos de buena calidad dan como resultado gráficas sencillas y con forma de estrella (izquierda y derecha). Los datos de mala calidad resultan en gráficas complejas y con formas irregulares (centro).

H.-J. Bandelt, A. Dür / Molecular Phylogenetics and Evolution 42 (2007) 256–271

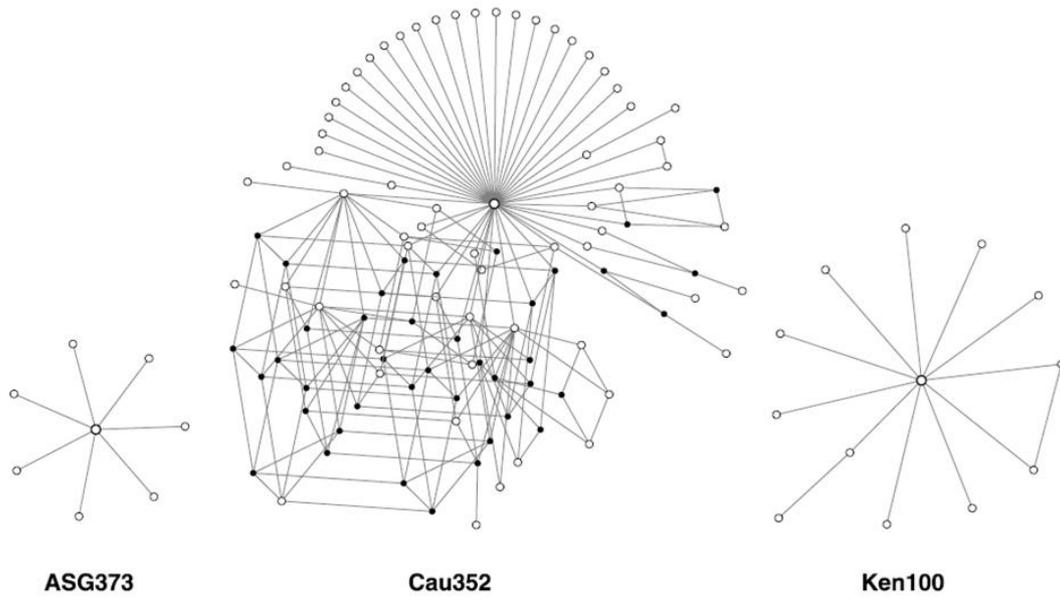


Figura 2A.- Dos posibles alineamientos de un mismo haplotipo. Según [14], el alineamiento más apropiado es el 2, ya que en la filogenia se han observado secuencias con transición C568T sin inserciones posteriores de Cs y no se han observado ejemplos de inserciones de T tras 567 (⁵⁶⁷ATCCCCCA⁵⁷⁴ está descrita y ⁵⁶⁷ATCCCCCA⁵⁷⁴ no lo está). Además, las inserciones de Cs son muy abundantes en este tracto (de 1 a 6 inserciones), por lo que es lógico pensar que se insertaran 5Cs.

rCRS	⁵⁶⁷ ACCCCCA ⁵⁷⁴
Muestra	⁵⁶⁷ ATCCCCCCCCCA ⁵⁷⁴

	5 6 7		5 6 8	5 6 9	5 7 0	5 7 1	5 7 2	5 7 3						5 7 4	HAPLOTIPO
rCRS	A	-	C	C	C	C	C	C	-	-	-	-	-	A	
ALINEAMIENTO 1	A	T	C	C	C	C	C	C	C	C	C	C	-	A	567.1T573.1C573.2C 573.3C573.4C
ALINEAMIENTO 2	A	-	T	C	C	C	C	C	C	C	C	C	C	A	568T573.1C573.2C 573.3C573.4C573.5C

Figura 2B.- Alineamiento con sistema SAM en EMPOP. Resultados de la búsqueda en EMPOP del haplotipo anterior introduciendo los alineamientos 1 y 2 de la figura 2A. El resultado que se obtiene en cuanto a coincidencias (hits) es el mismo en ambos casos.

Sample Info: FIGURE 2B
Type: string-based search: haplotype as differences to rCRS

Query: 567-576 **-567.1T -573.1C -573.2C -573.3C -573.4C**

Geographic affiliation: All | Metapopulation: All

DIFFERENCES TO QUERY PROFILE	NUMBER OF HAPLOTYPES	CUMULATIVE NUMBER OF HAPLOTYPES
0	3	3
1	8626	8629
2	25	8654
3	1	8655
4+	4	8659

Query: 567-576 **C568T -573.1C -573.2C -573.3C -573.4C -573.5C**

Geographic affiliation: All | Metapopulation: All

DIFFERENCES TO QUERY PROFILE	NUMBER OF HAPLOTYPES	CUMULATIVE NUMBER OF HAPLOTYPES
0	3	3
1	8626	8629
2	25	8654
3	1	8655
4+	4	8659

NUMBER OF HAPLOTYPES

Figura 3.- Ejemplo de aviso al introducir un dato erróneo en EMPOP (<http://empop.org/>). Se ha introducido el polimorfismo 73A (idéntico a rCRS) en lugar del real (73G).

The screenshot shows the EMPOP web interface. On the left is a navigation menu with 'Query' selected, containing sub-items 'Haplotype Publications' and 'Overview'. Below the menu are 'Tools' and 'Account' sections. The main content area is titled 'Input' and contains a red error box with the text: 'Profile 2: Mutation 1: nucleotide given is identical to rCRS.' Below this, the 'Type' is set to 'haplotvpe as differences to rCRS'. The 'Sample Info' is 'Ejemplo 1'. The 'Query' section shows two entries: one for range '16024-16365' with profile '16126C 16294T 16304C', and another for range '73-340' with profile '73A 263G 315.1C'. The second profile is highlighted in red, indicating the error.

Range	Profile
16024-16365	16126C 16294T 16304C
73-340	73A 263G 315.1C