



**UNIVERSIDAD AUTÓNOMA DE
QUERÉTARO**

FACULTAD DE QUÍMICA

**“ANÁLISIS DE LA CARGA DE DELECCIONES EN EL
GENOMA MITOCONDRIAL DEL CEREBRO EMBRIONARIO
DE RATÓN”**

TESIS INDIVIDUAL

QUE PARA OBTENER EL TÍTULO DE

INGENIERO EN BIOTECNOLOGÍA

PRESENTA

FERNANDO LARRIVA SÁNCHEZ

DIRIGIDA POR

Dr. ALFREDO VARELA ECHAVARRÍA

SANTIAGO DE QUERÉTARO, QUERÉTARO, 2019



Dirección General de Bibliotecas y Servicios Digitales
de Información



ANÁLISIS DE LA CARGA DE DELECCIONES EN EL
GENOMA MITOCONDRIAL DEL CEREBRO
EMBRIONARIO DE RATÓN

por

Fernando Larriva Sánchez

se distribuye bajo una [Licencia Creative Commons
Atribución-NoComercial-SinDerivadas 4.0 Internacional](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Clave RI: FQLIN-237266



**UNIVERSIDAD AUTÓNOMA DE
QUERÉTARO**

FACULTAD DE QUÍMICA

**“ANÁLISIS DE LA CARGA DE DELECCIONES
EN EL GENOMA MITOCONDRIAL DEL
CEREBRO EMBRIONARIO DE RATÓN”**

TESIS INDIVIDUAL

**QUE PARA OBTENER EL TÍTULO DE
INGENIERO EN BIOTECNOLOGÍA**

PRESENTA

FERNANDO LARRIVA SÁNCHEZ

DIRIGIDA POR

Dr. ALFREDO VARELA ECHAVARRÍA

SINODALES

Dr. ALFREDO VARELA ECHAVARRÍA
DIRECTOR Y ASESOR EXTERNO

Dra. BLANCA ESTELA GARCÍA ALMENDRÁREZ
SINODAL INTERNO

Dr. VIJAYKUMAR MULEY
SINODAL

M. en I.M. DAVID GUSTAVO GARCÍA GUTIÉRREZ
SINODAL

ÍNDICE GENERAL

Contenido	Página
ÍNDICE GENERAL	i
ÍNDICE DE CUADROS	iii
ÍNDICE DE FIGURAS	iv
AGRADECIMIENTOS	
RESUMEN	
ANTECEDENTES	1
1.1 La mitocondria	1
1.1.1 Origen evolutivo de la mitocondria	1
1.1.2 Rol celular de la mitocondria	2
1.1.3 Funciones metabólicas de la mitocondria	3
1.1.4 Otras funciones de la mitocondria	6
1.1.5 Plasticidad mitocondrial	8
1.2 El genoma mitocondrial	10
1.2.1 El genoma mitocondrial de mamíferos	11
1.2.2 Transcripción y replicación del mtDNA de mamíferos	12
1.2.3 Sistemas de reparación del mtDNA en mamíferos	14
1.2.4 El mtDNA del ratón	18
1.3 Mutaciones y enfermedades mitocondriales	18
2. HIPÓTESIS	21
3. OBJETIVOS	22
4. METODOLOGÍA	23
4.1 Parte experimental	23
4.1.1 Animales	23
4.1.2 Clonación y re-secuenciación del genoma mitocondrial de	

ratón	23
4.1.3 Extracción y secuenciación del DNA mitocondrial de ratón	24
4.2 Análisis bioinformáticos	25
4.2.1 Control de calidad de las lecturas	25
4.2.2 Mapeo de las lecturas	25
4.2.3 Análisis estadístico de las deleciones	26
4.3 Diagrama de flujo	27
4.4 Diseño estadístico	28
5. RESULTADOS	29
5.1 Procesamiento y filtrado de los datos de secuenciación	29
5.2 Puntajes de calidad del mapeo en una resolución por base a lo largo del genoma mitocondrial ensamblado	32
5.3 Las deleciones de gran longitud son un evento muy raro	36
5.4 Número superior de eventos de deleción en muestras biológicas comparado con el grupo control	40
5.5 Los genomas mitocondriales son altamente resistentes a Eventos de deleción	43
5.6 Razones posibles para deleciones artefactuales	44
6. DISCUSIÓN	48
7. CONCLUSIONES	52
8. REFERENCIAS	53

ÍNDICE DE CUADROS

Cuadro	Página
1	32
2	43

ÍNDICE DE FIGURAS

Figura	Página
1	7
2	27
3	29
4	30
5	30
6	31
7	34
8	35
9	36
10	38
11	39
12	41
13	42
14	45
15	46
16	47

AGRADECIMIENTOS

Este trabajo se llevó a cabo con fondos del proyecto “CONACYT FOMIX QRO-2018-01-01-88344”.

Agradecimientos Institucionales

Universidad Nacional Autónoma de México e Instituto de Neurobiología

Laboratorio Nacional de Visualización Científica Avanzada (LAVIS), UNAM

Luis Alberto Aguilar Bautista

Alejandro de León Cuevas

Carlos Sair Flores Bautista

Jair García Sotelo

Instituto de Neurobiología, UNAM

Laboratorio de Diferenciación Neural y Axogénesis

Alfredo Varela Echavarría

Vijaykumar Muley

Carlos Lozano Flores

Laura González Sánchez

Ofelia Mora Izaguirre

Unidad de Proteogenómica

Adriana González Gallardo

Anaid Antaramian Salas

Michael Conrad Jeziorski

Unidad de Cómputo

Alberto Lara Rubalcaba

Omar González Hernández

Ramón Martínez Olvera

Bioterio

José Martín García Servín

Alejandra Castilla León

Agradecimientos personales

Le agradezco de manera muy especial a Alfredo Varela Echavarría por la oportunidad de llevar a cabo mi Tesis en su laboratorio, por su apoyo constante y por su flexibilidad con mi proyecto de investigación.

Le agradezco a Vijaykumar Muley por haber sido mi mentor a lo largo de esta tesis. También por sus importantes contribuciones a los mapeos con “Rsubread” y a todo el análisis bioinformático.

Le agradezco a Carlos Lozano Flores por la obtención y la preparación de todas las muestras genómicas.

Le agradezco a Jorge Tonatiuh Ayala Sumuano por el aporte científico a esta línea de investigación. Agradezco a Brenda Farías Serratos por sus enseñanzas de Bioinformática. También, le agradezco Fernando Soto Barragán, Bertha Rueda Zarazúa, Leopoldo González Santos y Carlos Javier López Victorio por sus aportaciones a scripts.

RESUMEN

La mitocondria es un organelo de gran importancia en células eucariontes al ser clave para diversos procesos como metabolismo energético, protección ante estrés e incluso para desencadenar muerte celular apoptótica. En las últimas tres décadas ha adquirido mucha mayor atención estudiar la mitocondria y su genoma. La acumulación de mutaciones en el genoma mitocondrial ha sido relacionada con fenotipos de envejecimiento y con diversas enfermedades neuromusculares. Los métodos actuales no han logrado determinar a fondo la dinámica de estas deleciones, sobre todo en individuos sanos. Existen hipótesis contrastantes de si el origen exacto de estas deleciones se puede explicar con la rápida amplificación del mtDNA durante la embriogénesis o si más bien estas deleciones se acumulan en los tejidos somáticos. Hemos investigado este fenómeno secuenciando genomas mitocondriales del tejido cerebral de embriones de ratón y fragmentos del genoma mitocondrial clonados y amplificados en *E. coli*. Un análisis bioinformático de los datos de secuenciación obtenidos a partir de las muestras de estos grupos reveló que el tamaño de deleción es inversamente proporcional a su abundancia. En las muestras biológicas hallamos deleciones en muy bajas frecuencias a lo largo del genoma mitocondrial. En contraste con las muestras del grupo control, asumimos que estos eventos de deleción corresponden a copias defectuosas del genoma mitocondrial que surgieron de manera aleatoria y sin una preferencia posicional. Asumimos que estas pequeñas poblaciones de genomas mitocondriales defectuosos se encuentran en números extremadamente pequeños y no tienen consecuencias funcionales importantes en la célula. De un total de las 4,500 deleciones que hallamos no hubo ninguna deleción diferencial. Por lo mismo, descartamos la posibilidad de deleciones mitocondriales recurrentes con consecuencias funcionales, al menos en tejido cerebral en desarrollo.

1. ANTECEDENTES.

1.1. La mitocondria

1.1.1. Origen evolutivo de la mitocondria

La mitocondria es un organelo eucarionte responsable de numerosas funciones vitales a nivel celular. Las mitocondrias, cuyo volumen ocupa una gran porción del volumen citoplasmático, han desempeñado un papel esencial en la evolución de animales complejos. La presencia de las mitocondrias en sus primeras manifestaciones, importante evento la evolución del dominio eucarionte, data a hace aproximadamente 2 mil millones de años (Gabaldón y Huynen, 2004; Martin y Mentel, 2010).

La teoría endosimbiótica (Margulis, 1967) es, incluso hoy en día, una de las más ampliamente aceptadas al explicar el origen evolutivo de ciertos organelos hoy presentes en eucariontes. La teoría explica, apoyándose del uso de evidencias de biología celular, geología y filogenia, el origen de los organismos eucariontes. Lynn Margulis basa esta teoría en dividir la historia de la vida en dos eras geológicas: en una anaeróbica y otra aeróbica. Margulis propone que en la primera fase (anaeróbica) surgieron ciertas cianobacterias capaces de llevar a cabo fotosíntesis. Posteriormente, propone que durante la siguiente y prolongada fase oxigenada surgieron los eucariontes mediante varias interacciones de endosimbiosis, en las cuales cianobacterias y proteobacterias fueron fagocitadas por bacterias de mayor tamaño. Es así que Lynn Margulis justifica la aparición de los plástidos fotosintéticos, las mitocondrias y los cuerpos basales (9+2) de los flagelos. A pesar de que este artículo fue una primera aproximación y hace ciertas suposiciones que posteriormente fueron refutadas, tuvo un gran impacto en la biología de su tiempo. Transformó la endosimbiosis en un evento central en el proceso evolutivo, además de englobar casi un siglo de descubrimientos científicos en el tema (Lane, 2017).

En los años posteriores, con el descubrimiento del dominio arquea (Woese y Fox, 1977) y los cambios mayores que se hicieron en árbol filogenético universal

(Woese y col., 1990; Eme y col., 2017), se hizo una modificación importante a la propuesta de Margulis, planteando ahora que lo más probable es que los hospederos de las interacciones endosimbióticas hayan sido arqueas. Gracias al surgimiento de la genómica y la bioenergética, también se han incluido otros cambios en la teoría endosimbiótica original. Margulis subestimó la cantidad de cambios que ocurrieron a nivel celular después de los eventos endosimbióticos. Los aportes más novedosos apuntan a que hubo una integración del genoma mitocondrial al nuclear, sobre todo en las regiones que codifican proteínas de las membranas bioenergéticas. Sorprendentemente, este inesperado evento fue la clave para la evolución de los eucariontes y el aumento gradual de su complejidad morfológica. Otras regiones del genoma mitocondrial se han perdido a lo largo del proceso evolutivo. Así mismo, lo más probable es que la evolución de los eucariontes haya comenzado incluso antes de la formación de una atmósfera aeróbica (Friedman y Nunnari, 2014; Knoll y col., 2016; Lane, 2017).

1.1.2. Estructura funcional de la mitocondria

A lo largo de la evolución, las mitocondrias han adquirido cada vez más funciones a nivel celular. La primera, y probablemente la más importante, fue completar el metabolismo de azúcares. Gracias a la presencia de mitocondrias es que se puede llevar a cabo una respiración aeróbica. Esta importante porción del metabolismo es responsable de una producción de ATP 15 veces mayor que la glucólisis (Karp, 2009).

Cabe destacar que muchas funciones de la mitocondria son posibles gracias a su particular estructura. Cada una de las mitocondrias se encuentra recubierta por dos membranas altamente especializadas, que forman dos compartimentos dentro de la mitocondria, el más céntrico denominado matriz mitocondrial interna y el más periférico conocido como espacio intermembranal. Las dos membranas que forman parte de la estructura mitocondrial presentan a su vez composiciones y grados de permeabilidad muy diferentes. La membrana externa está compuesta en gran proporción de lípidos (cerca de un 50%) y contiene proteínas como las porinas y

enzimas involucradas en metabolismo energético. Posee un grado de permeabilidad relativamente alto al permitir el paso de moléculas de hasta 5000 Da. La bicapa mitocondrial interna es rica en el fosfolípido cardiolipina, el cual le atribuye propiedades de permeabilidad más selectivas, impidiendo incluso el paso de iones. La membrana interna es responsable de la mayoría de las funciones metabólicas. Puede llevarlas a cabo debido a su gran área superficial generada por crestas al igual que debido a su alta proporción proteína/lípido (Alberts y col., 2008; Karp, 2009).

1.1.3. Funciones metabólicas de la mitocondria

La mitocondria participa de manera importante tanto en el metabolismo aeróbico de azúcares como en el de lípidos. La mitocondria puede utilizar como sustrato para el ciclo de Krebs ácidos grasos o piruvato, los cuales son transportados a la membrana mitocondrial interna. Dentro de la matriz mitocondrial, cualquiera de esos sustratos es convertido al intermediario metabólico acetil coenzima A, el cual termina por entrar al ciclo de Krebs. Con la entrada de cada molécula de acetil coenzima A al ciclo de Krebs y la presencia de una molécula de oxaloacetato saliente del ciclo previo, la enzima citrato sintasa cataliza una reacción de condensación y genera una molécula de citrato. El citrato posee 6 carbonos y entrará a una nueva vuelta del ciclo. Durante el ciclo de Krebs, mediante un proceso oxidativo de 8 pasos, dos de los carbonos provenientes del citrato son liberados en forma de dióxido de carbono. Otro producto del proceso oxidativo es agua. Las reacciones de oxidación consecutivas generan 5 electrones de alta energía por vuelta, los cuales son almacenados en moléculas acarreadoras (NADH y FADH₂). En cada ciclo de la reacción, un enlace fosfato de alta energía también es generado y transferido a una molécula acarreadora GDP, convirtiéndola así a su forma activa (GTP) (Alberts y col., 2008; Karp, 2009).

Las moléculas acarreadoras finalmente liberan sus electrones transfiriéndolos a proteínas de la membrana mitocondrial interna; de esta manera se desencadena la cadena transportadora de electrones (CTE), que está compuesta por cuatro complejos proteicos portadores de electrones y dos portadores más: citocromo C y

ubiquinona. Los electrones donados por NADH entran a la cadena mediante el complejo I y los donados por FADH₂ mediante el complejo II. Estos electrones son transferidos del complejo I o II a la ubiquinona. Después de que la ubiquinona pasa a su forma reducida (ubiquinol), los electrones son transferidos al complejo III y posteriormente al citocromo C y que los acarrea al complejo IV. Finalmente, esta serie de reacciones culmina en la formación de una molécula de agua a partir de una molécula de oxígeno molecular y cuatro protones de hidrógeno. Ya que los complejos I, III y IV translocan protones al espacio intermembranal se forma un gradiente electroquímico. Este gradiente da origen a la fuerza protón motriz (FPM), la cual es esencial para la producción de ATP. El gradiente electroquímico surge a consecuencia de diferenciales de pH y voltaje entre ambos lados de la membrana. Ejerce una fuerza protón motriz, la cual en una mitocondria en respiración equivale a valores de entre 180 y 190 mV (Alberts y col., 2008).

Aprovechando el gradiente formado entre ambos lados de la membrana mitocondrial interna, la ATP sintasa fosforila moléculas de ADP. La ATP sintasa es un elaborado complejo proteico multimérico (también conocido como el complejo V de la CTE), el cual por su estructura utiliza la FPM ejercida en la membrana mitocondrial interna por los protones previamente translocados. La consecuencia de esta fuerza es que la ATP sintasa fosforila moléculas de ADP convirtiéndolas en ATP. La ATP sintasa es una enzima funcionalmente reversible, ya que también puede hidrolizar ATP al bombear protones en contra del gradiente electroquímico. El mecanismo esencial mediante el cual el ATP es sintetizado, denominado fosforilación oxidativa, es compartido entre mitocondrias, cloroplastos y bacterias. La estructura básica de esta enzima está conformada por dos complejos: F₀ y F₁. La ATP sintasa genera ATP mediante un proceso llamado catálisis rotatoria. Este proceso es impulsado por el paso de protones a través de la membrana por el canal base F₀. El paso de los electrones a través del canal genera un torque, el cual es transferido F₁ gracias al eje asimétrico de la subunidad helicoidal γ . Las subunidades alfa y beta que forman un anillo dentro del complejo F₁ giran con respecto al tallo central. Dentro

del anillo, la rotación produce tres distintas conformaciones del sitio activo, las cuales varían en su afinidad de unión al ATP y logran su catálisis a partir de ADP y P_i.

Al ocurrir el transporte de electrones en organismos aerobios, entre el 1 y el 2% del oxígeno metabolizado no es convertido en agua de manera exitosa. Por ello, se generan superóxidos y otras moléculas que poseen oxígeno con un alto grado de reactividad, tanto en sus formas radicales como no radicales, que se denominan especies reactivas de oxígeno. La mayoría de ellos son altamente reactivos y dañinos al entorno celular y tisular. Las ROS en sus formas radicales (electrones desapareados en la capa de valencia) tienen un alto potencial reactivo ya que los productos de sus reacciones pueden ser a su vez más radicales. Presentes o no en sus formas de radicales, las ROS tienen un alto potencial dañino al poder reaccionar e interferir con la función de biomoléculas como los son lípidos, proteínas y hasta el DNA. A pesar de que en niveles normales las ROS son parte del estrés oxidativo, su producción inadecuada o excesiva está relacionada con muchas enfermedades crónicas degenerativas al igual que envejecimiento (Lodish, 2016).

En el tejido graso pardo se ha observado que las mitocondrias incluso participan en un proceso de termogénesis. La grasa parda recibe su nombre, de hecho, por el color que adquiere por la abundancia de mitocondrias. La membrana interna de estas mitocondrias contiene termogenina. La termogenina es una proteína transportadora de protones similar en estructura a los transportadores mitocondriales ATP/ADP. La termogenina, o UCP1 desacopla la FPM de la síntesis de ATP. Esto permite que los protones puedan fluir a través de la membrana mitocondrial a favor de su gradiente de concentración. El retorno de los electrones de una manera altamente controlada es esencial para que la termogenina lleve a cabo una reacción para la generación de calor. Este mecanismo es una importante fuente adicional de calor a partir de grasa en mamíferos y posee vital importancia en varias situaciones (Nicholls y col., 1987; Lodish, 2016; Couchani y col., 2017).

1.1.4. Otras funciones de la mitocondria

La mitocondria se encuentra fuertemente involucrada en las vías de señalización de calcio. La proteína uniporter mitocondrial de calcio (MCU), presente en la membrana mitocondrial interna, es fundamental para la captura de iones de Ca^{2+} en la matriz. Además de su papel protagónico en la señalización de calcio, la MCU también controla aspectos como el metabolismo aeróbico e incluso la apoptosis. A raíz del estudio de la señalización de calcio mitocondrial se sabe desde ya hace algunos años que los niveles de Ca^{2+} varían drásticamente entre células y tejidos, y que la apertura de canales depende fuertemente de las concentraciones de Ca^{2+} extramitocondrial. Debido a esta variabilidad tisular, al igual que a otro tipo de variaciones circunstanciales, la señalización de calcio mitocondrial es un proceso complejo y poco comprendido. En los últimos años ha surgido mayor interés en el estudio de la proteína MCU como parte de un complejo conformado por diversas moléculas. Un importante hallazgo que surgió a partir de estos esfuerzos es que además de ser altamente compleja, la señalización de Ca^{2+} mitocondrial es afectada por diversos genes (De Stefani y col., 2015).

La mitocondria está fuertemente involucrada en la vía apoptótica intrínseca. Esta vía, también conocida como apoptosis mitocondrial, es iniciada por agentes intracelulares como los son un daño irreparable al material genético, concentraciones demasiado altas de iones Ca^{2+} , estrés oxidativo grave. La vía de apoptosis extrínseca, en cambio es iniciada por una activación de los receptores de muerte, los más comunes son "Fas" y "TNF", lo que a su vez activa proteínas proapoptóticas de la familia Bcl-2, como Bax, que pueden trascolarse al interior de la mitocondria y aumentar la permeabilidad mitocondrial al unirse a su membrana interna (Wali y col., 2013).

La mitocondria es blanco de interacción de proteínas proapoptóticas y antiapoptóticas. Ambos grupos de proteínas afectan la permeabilidad mitocondrial, siendo las proteínas proapoptóticas las que promueven la apoptosis. A la par con un aumento en la permeabilidad, el citocromo C y otras proteínas proapoptóticas son

liberados del espacio intermembranal al citosol. La presencia de estos agentes en el citosol permite la formación de un complejo llamado apoptosoma. Una vez completada la formación del apoptosoma, la procaspasa 9 pasa a su forma activa como caspasa 9 o caspasa iniciadora que desencadena el programa apoptótico de la célula. En ciertas líneas celulares, la vía intrínseca activada por algún receptor también es capaz de hacer una interacción cruzada con la apoptosis mitocondrial. Después de comenzar la vía extrínseca, cuando la caspasa 8 se encuentra en su forma madura, corta a “Bid” en su extremo N-terminal. Al ocurrir este corte se produce un fragmento de menor tamaño denominado “tBid”, el cual se translocará a la mitocondria y desencadenará la liberación de citocromo C (Luo y col., 1998; Wang y Youle, 2009; Wang y Youle, 2016).

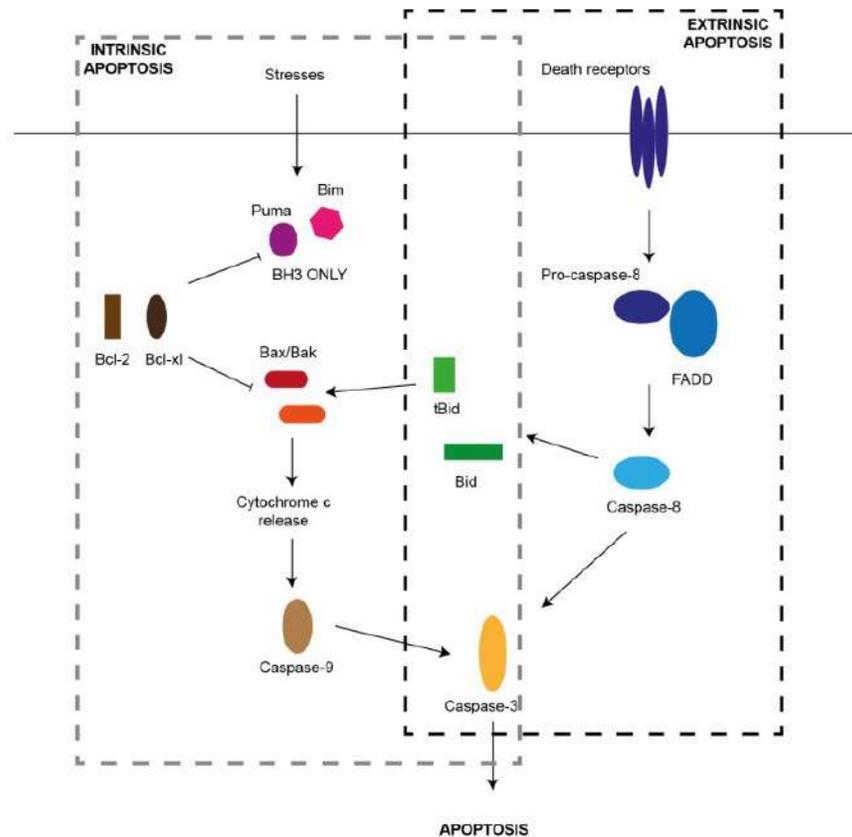


Figura 1. Las dos vías de apoptosis. La proteína Bid de dominio BH3 es esencial para la apoptosis mediada por el receptor de la muerte en las células beta, que proporciona una comunicación cruzada entre las dos vías apoptóticas (Wali y col., 2013).

1.1.5. Plasticidad mitocondrial

Al ser responsables de un tan extenso repertorio de procesos celulares, las mitocondrias requieren una gran adaptabilidad a los requerimientos de la célula.

Las mitocondrias pueden sentir las prioridades y adaptarse dentro de la célula de una manera autónoma. La organización espacial de las mitocondrias dentro de la célula se debe a una delicada regulación entre fusión y fisión (o fragmentación) mitocondrial. También poseen la capacidad de adaptarse estructuralmente desde su interior, mediante la remodelación de sus crestas. Esto se ha observado de manera mucho más evidente con la inmensa variación morfológica de las mitocondrias en condiciones tan distintas como desarrollo embrionario, respuesta a ayuno, apoptosis, y muchas otras (Pernas y Scorrano, 2016).

La actividad y la biogénesis mitocondrial son controladas a nivel transcripcional por la familia de coactivadores PGC-1 α . La actividad de este coactivador es regulada por una acetilación reversible. Está bien establecido que PGC-1 α es desacetilado y activado por Sirt1 y acetilado e inhibido por GCN5, sin embargo, el proceso regulatorio de actividad de estas dos enzimas es menos conocido. Los cambios en la actividad del coactivador PGC-1 α desencadenan una respuesta transcripcional, la cual aumenta la actividad mitocondrial en tiempos de requerimientos energéticos y la disminuye en momentos de baja demanda energética. Mientras estructuras mitocondriales alargadas y extensas son ideales para la producción energética, estructuras más fragmentadas y redondeadas son más adecuadas para los mecanismos de protección de la célula (Jeninga y col., 2010; Friedman y Nunnari, 2014).

La fusión mitocondrial es el proceso que favorece estructuras más alargadas e interconectadas. Es un proceso de gran importancia y ocurre en toda célula eucarionte que posea mitocondrias. Su inhibición resulta en letalidad embrionaria en ratones (Detmer y col., 2003). Una vez que las dos mitocondrias que se fusionan comienzan las interacciones proteicas. Las proteínas Mitofusina 1 y Mitofusina 2

forman complejos homodiméricos o heterodiméricos entre el par de mitocondrias. Estas dos proteínas median la fusión de la membrana mitocondrial externa siempre formando dímeros entre proteínas presentes en las dos mitocondrias que se fusionarán. La fusión de la membrana mitocondrial interna es realizada por la proteína OPA1. Tal proceso es dependiente del potencial de membrana y para que ocurra no son necesarios dímeros, sino solamente basta con que OPA1 se encuentre en una de las membranas mitocondriales internas (Pernas y Scorrano, 2016).

La fisión mitocondrial es un proceso un poco más complejo. Consiste en la división de una mitocondria para dar origen a una o más mitocondrias hijas. Una gran diversidad de elementos del citoplasma, citoesqueleto y organelos colaboran en sincronía para llevarla a cabo. De una manera muy general el proceso de fisión puede dividirse en tres pasos. En el primero ocurre el marcaje del sitio de fisión, este marcaje mediante el factor de fisión mitocondrial (MFF) sobre la membrana mitocondrial externa recluta con un alto grado de afinidad a la proteína DRP1 (proteína relacionada a la dinamina 1). En el siguiente paso del proceso ocurre un ensamblaje de complejos proteicos formados a base de monómeros de la proteína DRP1, los cuales van alineándose en forma de espirales alrededor del sitio de fisión. Para finalizar, en el tercer paso ocurre una hidrólisis de GTP simultánea a una constricción de los complejos de DRP1, lo cual provoca finalmente la fragmentación de la mitocondria. Hoy en día se sabe que los fenómenos de ensamblaje y constricción mediados por DRP1 son posibles gracias a interacciones de las mitocondrias con el retículo endoplásmico y algunas de sus proteínas (Friedman y col., 2011; Wang y Youle, 2016).

Las crestas mitocondriales también poseen gran adaptabilidad, pueden variar en: longitud, grosor, alineación lateral, rigidez y angularidad. Como consecuencia de la exposición a un sustrato no glucolítico, la abundancia de crestas aumenta. Este aumento va de la mano con un incremento en la cantidad de las proteínas que forman parte de la cadena respiratoria. En contraste, con la escasez de nutrientes disminuye el tamaño de crestas y disminuye su grosor. Cuando las crestas disminuyen en

grosor se favorece la dimerización de la ATPasa y su actividad aumenta. Una disminución del grosor de las crestas incluso puede afectar la apoptosis al disminuir la liberación de moléculas iniciadoras de la misma. El proceso de remodelación de las crestas es mediado por varias proteínas, pero la más importante y conocida es la OPA1. La proteína OPA-1 puede existir en ocho distintas isoformas: unas más largas y otras más cortas. A pesar de haber sido mencionada previamente, OPA1 participa en procesos independientes para la remodelación de crestas y para la fusión mitocondrial.

1.2. El genoma mitocondrial

La primera evidencia de la presencia de genes fuera del núcleo fue obtenida con el descubrimiento la herencia no mendeliana en plantas. En la herencia no mendeliana se observa una ausencia de segregación para las características parentales. Por lo tanto, es una clara evidencia de que no hubo una asociación entre el carácter de segregación y el huso meiótico (Lewin, 2000).

El DNA mitocondrial es una estructura circular de doble cadena. Sus dos cadenas son denominadas hebra ligera (L) y hebra pesada (H). La hebra H tiene un mayor contenido de citosinas y guaninas que la hebra L y su peso molecular es relativamente mayor. Por lo mismo, ambas hebras pueden ser separadas mediante una centrifugación por densidad en gradientes alcalinos de cloruro de cesio (Holt y Jacobs, 2000). A menudo se menciona que la mitocondria y su proteoma combinaciones de nuevos y viejos mecanismos y procesos moleculares Esto se refiere a que existen estructuras, sobre todo proteínas, altamente conservadas entre la mitocondria y las bacterias, al igual que otras más novedosas que han surgido por las interacciones endosimbióticas de la mitocondria con el resto la maquinaria celular (Gabaldón y Huynen, 2004; Friedman y col., 2011).

En la gran mayoría de los eucariontes que presentan reproducción sexual, los genomas citoplasmáticos, como el mitocondrial, son heredados de manera

uniparental. Es decir, todas las copias de su genoma son originadas solamente de un padre. En una pequeña cantidad de estos casos, se presenta una herencia paterna del DNA mitocondrial como ocurre algunas plantas como los pepinos. En cambio, la gran mayoría de los animales presentan una herencia mitocondrial materna con mecanismos que varían en los distintos organismos, lo que sugiere su importancia evolutiva. Este tipo de herencia probablemente surge como una medida preventiva, la cual limita el potencial replicativo de mutaciones “egoístas” o de replicación rápida dentro del genoma. Una segunda hipótesis le atribuye su aparición evolutiva a la herencia mitocondrial materna como un mecanismo para la protección contra la coexistencia de versiones mutadas y no mutadas del genoma, llamada heteroplasmia (Havey, 1997; Allen y Paula, 2013).

En el caso de los mamíferos, la herencia mitocondrial materna es lograda gracias a un proceso de ubiquitinación y degradación de las mitocondrias presentes en los espermatozoides (Sutovsky y col., 1999). Este proceso degradativo propicia a la presencia de copias genómicas idénticas (homoplasmia). La homoplasmia es promovida aún más debido a dos efectos de “cuello de botella” presentes a durante la reproducción sexual en mamíferos antes y después de la fertilización (Ladoukakis y Zouros, 2017).

1.2.1. El genoma mitocondrial de mamíferos

El genoma mitocondrial de animales presenta una estructura extremadamente compacta. Aunque existen diferencias entre filos de animales, se mantiene el común denominador de un genoma pequeño que codifica para un número restringido de funciones. En el caso de los mamíferos, la organización del mtDNA es aún más compacta. No existen intrones e incluso algunos de sus genes se traslapan; casi cada par de bases puede ser asignado a un gen. Las únicas excepciones de esto son dos regiones no codificantes. La más grande mide aproximadamente 900 pb y es denominada D-loop. Contiene los promotores de transcripción de ambas hebras (HSP y LSP) y el origen de replicación de la hebra ligera (denominado OH). La segunda región no codificante es mucho más corta al medir solamente 30 pb. Esta

región contiene únicamente el origen de replicación de la hebra ligera (denominado OL) (Perata y col., 2012; Lewin, 2000).

El mtDNA de mamíferos posee una longitud de aproximadamente 16.5 kpb y codifica únicamente para 37 genes. De estos, trece son proteínas que participan como subunidades del complejo OXPHOS, dos son rRNAs (12S y 16S) y 22 son tRNAs. Este genoma mitocondrial codifica para proteínas esenciales de los complejos I, III, IV y V de la CTE, pero las subunidades del complejo II son codificadas en el núcleo. Los genes presentes en el mtDNA de mamíferos se encuentran distribuidos de manera asimétrica entre ambas cadenas: la cadena pesada codifica para 12 de los 13 mRNAs, ambos rRNAs y 14 de los 22 tRNAs (Perata y col., 2012; Gustaffson y col., 2016).

1.2.2. Transcripción y replicación del mtDNA de mamíferos

La transcripción del genoma mitocondrial comienza ya sea en el promotor de la hebra ligera (LSP) o en el promotor de la hebra pesada (HSP). Esta transcripción produce transcritos policistrónicos con una longitud cercana a la del genoma completo. Se previene que la transcripción proceda hasta la región de control por eventos de terminación de la transcripción. Los transcritos primarios son rápidamente procesados para la liberación de moléculas individuales de RNA, ya que los transcritos no procesados son altamente inestables. En estudios de secuenciación del mtDNA de humanos y ratones se descubrió que las secuencias de tRNAs a menudo flanquean secuencias que codifican para proteínas y rRNA. Por ello, surgió en estos organismos el “modelo de puntuación de tRNA”. De acuerdo a este modelo las secuencias de tRNA son reconocidas específicamente y cortadas. En mamíferos, los transcritos policistrónicos son cortados en la terminación 5' por la RNasa P, mientras que en el extremo 3' son cortados por la RNasa Z. Al ocurrir esto, las secuencias de tRNAs, mRNAs y rRNAs son liberadas por separado y sufren modificaciones posteriores como parte de su maduración (Gustaffson y col., 2016).

La maquinaria involucrada en el mantenimiento y la expresión génica del mtDNA de mamíferos es muy distinta a la del núcleo (Falkenberg y col., 2007). Esto es consistente con la teoría endosimbiótica. Aunque muchos de los componentes de esta maquinaria tampoco coinciden con la de α -proteobacterias. En cambio, muchas proteínas se parecen a las presentes en el linaje “T-odd” de bacteriófagos (Shutt y Gray, 2006). Algunos ejemplos de estos factores son la RNA-polimerasa dirigida por mtDNA (POLYRMT), la subunidad catalítica de la mtDNA polimerasa (POL γ A) y la helicasa replicativa mitocondrial (TWINKLE). Este tipo de factores seguramente han ido reemplazando los originales de α -proteobacterias a lo largo del proceso de endosimbiosis.

Aunque el proceso de replicación del mtDNA en mamíferos es un proceso aún poco comprendido, es de gran importancia. Esta importancia surge ya que, aunque algunas mutaciones en el mtDNA en tejidos somáticos surgen por el envejecimiento, la mayoría de las mutaciones en el mtDNA surgen a raíz de errores replicativos sobre todo frecuentes durante la embriogénesis (Zheng y col., 2006; Ameer y col., 2011).

El replisoma mitocondrial está conformado en su núcleo por la DNA polimerasa γ (POL γ). Esta polimerasa es un trímero compuesto por una subunidad catalítica POL γ A y dos subunidades accesorias POL γ B. Para que la POL γ pueda llevar a cabo su función catalítica requiere de la helicasa TWINKLE (Korhonen y col., 2014). Esta helicasa se mueve a la par con la POL γ en la horquilla de replicación con una direccionalidad de 5' a 3'. La helicasa TWINKLE tiene un efecto de desenrollamiento en la doble hélice del mtDNA. En mamíferos, los cebadores necesarios para comenzar la síntesis del mtDNA probablemente son sintetizados por la RNA polimerasa mitocondrial (POLRMT). La proteína mitocondrial de unión a cadena simple desempeña múltiples funciones: estimula la actividad de la helicasa TWINKLE, estabiliza segmentos largos de DNA de cadena simple y estimula la síntesis del mtDNA al mejorar el reconocimiento de POLRMT de cebadores. Así mismo, es indispensable la presencia de topoisomerasas para que pueda ocurrir una replicación de DNA. La topoisomerasa mitocondrial I (TOP1mt) cataliza rupturas en

la doble hebra del mtDNA manera transitoria. También se ha descrito recientemente la presencia de dos topoisomerasas nucleares tipo IIA (TOP2 α y TOP2 β) en la mitocondria de mamíferos (Zhang, 2014). Una segunda polimerasa, llamada polimerasa primasa (PrimPol) se ha visto participando tanto en el núcleo como en la mitocondria, aunque no es esencial para la replicación del genoma mitocondrial (Ciesielski y col., 2016).

De acuerdo al modelo de desplazamiento de hebra, la replicación del mtDNA procede de manera continua en ambas hebras y no hay formación de fragmentos de Okazaki. La replicación comienza en OH y ocurre únicamente en la hebra pesada hasta cubrir dos tercios de su longitud. Mientras esto ocurre, mtSSB cubre la hebra ligera la cual está siendo desplazada constantemente y no es posible que la POLRMT inicie la transcripción. Cuando la maquinaria replicativa alcanza el OL, esta secuencia se vuelve una hebra simple y adquiere una estructura de tallo-bucle. El tallo obstaculiza la unión de mtSSB y permite finalmente la unión de POLRMT al bucle. Después de la síntesis de un cebador de alrededor de 25 nucleótidos, POLRMT es reemplazada por POL γ la cual lleva a cabo la síntesis de la hebra ligera. Aunque el modelo de desplazamiento de hebra es el más ampliamente aceptado existen otros dos modelos alternativos: el modelo de incorporación de ribonucleótidos a lo largo de la hebra retrasada (RITOLS) y el modelo de replicación de mtDNA acoplado a hebra (Perata y col., 2012).

1.2.3. Sistemas de reparación del mtDNA en mamíferos

El genoma mitocondrial se encuentra expuesto a estrés de manera constante y por lo tanto está en riesgo de sufrir mutaciones, ya sea por causas endógenas o exógenas. La principal fuente de mutaciones en el mtDNA son errores replicativos introducidos por la POL γ de manera espontánea. La introducción de este tipo de mutaciones es más probable si existen defectos en el dominio exonucleasa de la POL γ . El mtDNA también puede sufrir mutaciones por agentes químicos nocivos, sobre todo los que provienen de fuentes oxidantes. El daño oxidante del mtDNA surge a raíz de la formación de las ROS por la fuga de electrones de la CTE. Se sabe

que la mitocondria cuenta con varios mecanismos para la reparación de su genoma. Muchos de los mecanismos de reparación mitocondriales tienen gran similitud con los nucleares, aunque los mecanismos de reparación del mtDNA son mucho menos robustos y efectivos que los del núcleo (Schneider y Excoffier, 1999; Alexeyev y col., 2013; Copeland y Longley, 2014).

El primer mecanismo que ocurre en la mitocondria es el de reparación por escisión de base BER. Este sistema es activado a consecuencia de las glicosilasas de DNA, las cuales poseen capacidad para reconocer varios tipos de daño a la base. El mecanismo de escisión de base puede desencadenar dos procesos de reparación dependiendo del caso: una reparación BER de un solo nucleótido (SN-BER) o una reparación BER de parche largo (LP-BER). En ambos procesos de reparación una base o bases que presentan daño u oxidación son reconocidas y cortadas por una glicosilasa específica por lo que se crea un sitio AP (apurínico/apirimídico) el cual es cortado ahora por la AP endonucleasa para generar un pliegue 5' desoxirribosa fosfato (dRP). En la SN-BER la polimerasa mitocondrial POLy rellena el hueco y recorta la porción dRP antes de la ligación. En la LP-BER, la POLy también rellena el hueco, pero es necesario remover la hebra desplazada 5' ya que esta forma un pliegue. Esta hebra sobresaliente aparece a consecuencia de una síntesis por desplazamiento de hebra, y es removida mediante una coordinación entre DNA2 y FEN-1. Después de completarse cualquiera de estos procesos de reparación por BER, se concluye con una ligación catalizada por la DNA ligasa III la cual resana las muescas que causan discontinuidad en las hebras (Liu y Demple, 2010; Copeland y Longley, 2014).

Se ha demostrado que las mitocondrias carecen de mecanismo de reparación por escisión de nucleótido (NER). El primer experimento que sugirió esto fue uno que utilizó mutagénesis inducida por UV, en la cual se producen dímeros de pirimidinas (Clayton y col., 1975). Otros experimentos con daño por alquilación han concluido lo mismo. En este tipo de daños al mtDNA no se observó la intervención de un mecanismo de NER.

La reparación por “mismatch” (MMR) es esencial para la eliminación de bases que no presentan complementariedad o inserciones y deleciones cortas en el genoma nuclear. La presencia de un sistema MMR con homólogos de la MutS bacteriana ha sido descrita en *Saccharomyces cerevisiae* al igual que *Schizosaccharomyces pombe*, pero no en eucariontes más complejos (Copeland y Longley, 2014). No obstante, se ha descrito un proceso de reparación similar al MMR en las mitocondrias de hígado de rata, aunque este sistema es mucho menos potente. En humanos, se le ha atribuido esta función reparativa una proteína de unión al DNA específica para “mismatch” (YB-1), aunque aún se conoce poco acerca de su rol y mecanismo de acción exactos (de Souza-Pinto y col., 2009).

La reparación de rupturas de doble hélice (DSB) puede ocurrir por recombinación homóloga (HR) o por una unión de terminaciones no homólogas (NHEJ). Organismos como *Drosophila melanogaster* presentan mecanismos de reparación a rupturas de doble hélice altamente eficientes y muy frecuentes, a diferencia de los mamíferos. (Morel y col., 2008) La recombinación homóloga en mamíferos parece ser un evento poco frecuente, por lo que su inducción y observación en experimentos ha sido limitada. Sin embargo, un experimento utilizando cultivos celulares y tejido hepático de ratas pudo demostrar que ocurría una HR. Como consecuencia a rupturas de doble hebra hubo formación de deleciones, algunas de las cuales estaban flanqueadas en sus puntos de ruptura por fragmentos repetitivos (Bacman y col., 2009). En cuanto a la reparación mitocondrial por NHEJ, la vía es conocida solamente de manera parcial. No obstante, las mitocondrias de mamíferos poseen actividad de unión de terminaciones en su DNA en bajas frecuencias (Coffey y col., 1999). Los mamíferos presentan una NHEJ poco eficiente pero la cual podría tener implicaciones estructurales y de organización topológica importantes. La reparación por NHEJ también podría tener importante influencia sobre la aparición de deleciones de gran tamaño en el genoma mitocondrial (Alexeyev y col., 2013).

Recientemente se ha podido comprobar que la mitocondria posee un mecanismo de degradación de DNA como alternativa a la reparación. Varios estudios en mamíferos han demostrado que bajos niveles de rupturas de doble hélice en el mtDNA inducen una reparación mediante recombinación. En situaciones más graves, rupturas de doble hélice extensas o persistentes ocasionan una degradación de estas copias del mtDNA. Inicialmente se creía que la endonucleasa G (EndoG) desempeñaba este rol degradativo del mtDNA irreparable, aunque la disminución en sus niveles en mamíferos no afectó de manera observable la incidencia de mutaciones (Davies y col., 2003). Por esta razón, es probable que aún se desconozca la endonucleasa o exonucleasa responsable un proceso degradativo del mtDNA.

Desde los años 70 se describió la presencia de ribonucleótidos dentro del genoma mitocondrial (Grossman y col., 1973). Más recientemente se ha estimado que su presencia varía en un rango que va desde 10 hasta 30 por genoma mitocondrial (Yang y col., 2002). La fuente exacta de estos ribonucleótidos es controversial; algunas de las propuestas más convincentes postulan que son residuos de los cebadores que inician la replicación de DNA o que son incorporados erróneamente durante su síntesis. Igualmente se sabe que las relaciones entre ribonucleótidos y desoxirribonucleótidos afectan la tasa de incorporación y que existen mayores probabilidades de incorporación dependiendo de la identidad de la base. Como mecanismo de protección a este tipo de errores en el núcleo existe un sistema de reparación por escisión de ribonucleótidos (RER). Este proceso es realizado por las RNasas H1 y H2 en el núcleo, aunque también se ha hallado participación de la RNasa H1 en la mitocondria. La RNasa H1 logra remediar la incorporación de ribonucleótidos cortando largos fragmentos híbridos de RNA/DNA, los cuales frecuentemente son generados en la replicación a causa de la fuerte actividad de transcriptasa reversa que posee la POL γ . Aún no se conoce el mecanismo exacto de reparación mediado por la RNasa H1 en la mitocondria (Cerritelli y col., 2003; Copeland y Longley, 2014).

1.2.4. El mtDNA del ratón

El ratón común (*Mus musculus*) es uno de los organismos modelo más utilizados para experimentos de biología, genómica y ciencias biomédicas. Su mayor ventaja es que es un organismo modelo bien establecido, y esto implica una gran facilidad para encontrar homología entre genes y por lo tanto imitar o predecir enfermedades presentes en humanos (NIH, 2010). El genoma mitocondrial del ratón fue el segundo en ser secuenciado, en el mismo año que el del humano (Anderson y col., 1981; Bibb y col., 1981). Se ha hecho, la longitud de sus genomas mitocondriales es prácticamente igual: el del ratón mide 16,299 pb y el del humano mide 16,569 pb. Además, esta diferencia de aproximadamente 300 pb debe casi por completo a las diferencias de longitud en la región del “D-loop”. Así mismo, el resto de las diferencias entre las secuencias no tienen implicaciones funcionales obvias (Bayona-Bafaluy y col., 2003).

1.3. Mutaciones y enfermedades mitocondriales

Las enfermedades mitocondriales son un amplio y muy diverso grupo de desórdenes que se caracterizan por una ausencia de energía celular debido a anomalías en el complejo de fosforilación oxidativa (OXPHOS). Ahora se sabe que también pueden existir enfermedades mitocondriales en las cuales la mutación afecte genes necesarios para la traducción o el ensamblaje de componentes de los complejos OXPHOS (Schon y col., 2012).

Las enfermedades mitocondriales primarias pueden ocurrir por mutaciones puntuales o a causa de rearrreglos, los cuales pueden ser inserciones, deleciones o inversiones. Se han descrito rearrreglos en el mtDNA de longitudes muy variables, que van desde una sola base hasta varias kilobases (Ameur y col., 2011; Grady y col., 2013). Inicialmente se creía que las mutaciones del mtDNA en mamíferos solamente podían ser adquiridas por herencia materna (mutaciones de línea germinal). Ahora se sabe que también pueden surgir en cualquier punto posterior a

la fertilización (mutaciones somáticas). Es un hecho es que la gran mayoría de las enfermedades mitocondriales ocurren en pacientes con heteroplasmia. Además, casi todas estas mutaciones son de carácter recesivo, por lo que son necesarios niveles muy altos de ellas para que se pueda observar un fenotípico clínico. Hoy en día existe controversia acerca de los factores más influyentes en la inducción de mutaciones en el genoma mitocondrial. Las hipótesis más populares en el tema son: a) las mutaciones se originan por errores replicativos con la rápida amplificación del mtDNA durante la embriogénesis y b) las mutaciones ocurren de manera consistente a lo largo de la vida de los individuos y se van acumulando gradualmente (Larsson, 2010; Schon y col., 2012).

Es de trascendente importancia clínica estudiar más a fondo la progresión de enfermedades mitocondriales y averiguar con mayor contundencia los factores que las promueven o desencadenan. Si se obtiene más información acerca de la progresión de estas enfermedades los médicos podrán ofrecerles a sus pacientes mayor certeza en el diagnóstico y pronóstico (Phillips y col., 2014). Las principales fuentes de enfermedades mitocondriales pueden ser errores replicativos el genoma mitocondrial con consecuencias funcionales, daño del genoma mitocondrial por estrés oxidativo, irregularidades en el número de copias del genoma mitocondrial por célula y mutaciones o daño en genes mitocondriales del genoma nuclear (Tuppen y col, 2010). Las mutaciones mitocondriales, sobre todo de gran escala, se han relacionado fuertemente con un fenotipo de envejecimiento. Los estudios más ampliamente utilizados en mutaciones mitocondriales son los de deleciones, ya que estas son una de las causas más recurrentes de enfermedades mitocondriales. (Ling-Ping y col., 2003; Ameer y col., 2011; Grady y col., 2013).

La mayoría de los rearrreglos en el mtDNA son deleciones de gran escala que van de 1.3 a 8 kb y abarcan varios genes y obstaculizan el funcionamiento mitocondrial. Las mutaciones individuales del mtDNA podrían ocurrir de manera esporádica durante el desarrollo temprano, por lo que más adelante una misma mutación puede encontrarse distribuida en todo un tejido o en una línea celular específica. El surgimiento de mutaciones de longitudes variadas puede deberse a

mutaciones congénitas en genes nucleares cuyos productos participan en el mantenimiento, la replicación del mtDNA y en el metabolismo de nucleótidos de la mitocondria. La acumulación exponencial de mutaciones de tamaños variados ha sido también reportada en tejidos post-mitóticos viejos en enfermedades neurodegenerativas (Tuppen y col., 2010).

Algunas de las enfermedades mitocondriales más comunes incluyen enfermedades neurodegenerativas como algunos casos de Parkinson, Alzheimer y demencia, el síndrome de Leigh, MERRF y esclerosis múltiple. Otras enfermedades mitocondriales son más diversas en sus manifestaciones clínicas, como el síndrome de Kearns-Sayre, el síndrome de Pearson, la neuropatía óptica hereditaria de Leber, el síndrome MELAS, CPEO, NARP y síndromes de depleción. La mayoría de estas enfermedades mitocondriales no neurodegenerativas afectan sobre todo la función muscular, neuromuscular o la visión del paciente (Wallace, 1994; Tuppen y col. 2010; Grady y col., 2013).

Aún no se sabe claramente hasta qué grado la mitocondria puede resistir la carga mutacional. Existen publicaciones de investigaciones que analizaron el genoma mitocondrial de individuos sanos en busca de heteroplasmia debida a deleciones (Gadaleta y col., 1992; Elliott y col., 2008; Bosworth y col., 2017; He y col, 2002). Estos estudios afirman poder detectar deleciones con niveles de heteroplasmia muy por debajo del 1%. Sin embargo, no está claro hasta qué punto estas deleciones podrían ser resultado de mutaciones espontáneas en el genoma. La mayoría de los estudios utilizan individuos sanos contra enfermos, o individuos jóvenes contra de mayor edad, así que las deleciones detectadas por estos estudios ya se deben encontrar presentes de manera prominente en uno de los grupos. Por ello mismo, no debería ser difícil identificar estas deleciones. Permanece la pregunta de si las deleciones detectadas en estos estudios son consecuencia de una variación biológica entre las muestras y el genoma de referencia más que regiones deletadas. Con el fin de evaluar la carga de deleciones diseñamos una serie de experimentos húmedos e *in silico* utilizando muestras biológicas y las comparamos con muestras control que poseen un menor grado de variabilidad biológica.

2. HIPÓTESIS.

El genoma mitocondrial es altamente resistente a mutaciones, por lo que es improbable que acumule deleciones recurrentes en heteroplasmia.

3. OBJETIVOS.

3.1. General.

Analizar la distribución y carga mutacional de deleciones en el genoma mitocondrial de cerebro embrionario de ratón E12.5.

3.2. Específicos.

Evaluar la calidad de las lecturas obtenidas a partir de secuenciación NGS.

Identificar deleciones mediante un mapeo genómico.

Comparar las deleciones mapeadas en ambos grupos experimentales mediante análisis estadísticos.

4. METODOLOGÍA.

4.1. Parte experimental

Este proyecto es continuación de una serie de experimentos relacionados con la detección de mutaciones en el genoma mitocondrial murino. La selección de los animales, la obtención de su material genético, su clonación (grupo control) y secuenciación fue llevado a cabo previamente para su uso en varios proyectos de investigación. En el caso de esta tesis, se partirá de los datos de NGS en formato “.fastq”.

4.1.1. Animales

Se utilizaron ratones C57BL/6J (Stock 000664, The Jackson Laboratory, Bar Harbor, ME). Los ratones en el Instituto de Neurobiología (Universidad Nacional Autónoma de México) fueron manejados humanamente y sacrificados mediante dislocación cervical con el mínimo estrés posible por el personal de instituto de acuerdo a las legislaciones mexicanas correspondientes al uso de animales de laboratorio para fines de investigación (NOM-062-ZOO-1999) y siguiendo la “Guide for the Care and Use of Laboratory Animals” del “Institute of Laboratory Animal Resources Council” (ILAR, National Research Council, USA).

4.1.2. Clonación y re-secuenciación del genoma mitocondrial de ratón

El genoma mitocondrial de ratón fue clonado en seis fragmentos superpuestos de una longitud aproximada de 3.3 kb cada uno. Los pares de los cebadores utilizados fueron los diseñados por Mizutani y col. (2001). Los fragmentos fueron amplificados con la enzima GoTaq DNA polymerase (Promega Corporation, Madison, WI) y clonados con el kit pGEM-T (Promega Corporation). Sus identidades fueron confirmadas mediante secuenciación capilar de sus extremos. Se hicieron cinco mezclas independientes con cantidades equimolares de los seis plásmidos con sus respectivos insertos y fueron secuenciadas mediante la plataforma Illumina HiSeq 2000/2500. Estas secuencias proporcionaron un control y línea base para las deleciones introducidas por el método de secuenciación Illumina. De todas las

muestras de ratón, solamente se conservaron las deleciones ausentes en el DNA clonado o que tuvieran significativamente mayor abundancia en las muestras de ratones que en el DNA clonado. Como el DNA plasmídico también incluye deleciones generadas en su proceso de replicación en *Escherichia coli* (XL1-blue, Agilent Technologies, Inc., Santa Clara, CA) el método eliminó deleciones presentes en muestras de ratón también presentes en el mtDNA clonado. Debido a lo mismo, los cálculos que fueron hechos podrían subestimar o ignorar completamente algunas de las deleciones presentes en las muestras de ratón.

4.1.3. Extracción y secuenciación del DNA mitocondrial de ratón

El mtDNA fue extraído de cerebros embrionarios E12.5 con un método adaptado para DNA circular utilizando el kit Qiagen Maxi (Qiagen, Germantown, MD) y ciertas modificaciones. Los tejidos fueron homogenizados en buffer P1 utilizando un mortero para tejidos de vidrio y un pistilo de vidrio o PTFE. Después los homogenizados de tejido fueron lisados con buffer P2 en agitación lenta durante cinco minutos a temperatura ambiente. Los tejidos lisados fueron neutralizados con buffer P3 helado y se les hizo un mezclado suave pero constante, el cual fue seguido de una incubación en hielo de 10 minutos. Se obtuvo un sobrenadante claro después de centrifugar las muestras a 4000 x g durante 30 minutos y filtrarlas a través de una capa de toalla de papel de poca pelusa (Kimberly-Clark Professional, Rosswell, GA). En seguida las muestras fueron colocadas en columnas "Tip-500" pre-equilibradas. Después de dos lavados, las muestras fueron eluidas con solución de elución precalentada a 65 °C. La columna se mantuvo en incubación a esta misma temperatura durante la elución. El DNA fue precipitado de acuerdo a las instrucciones del kit y resuspendido en agua libre de nucleasas.

Para eliminar las moléculas lineares acarreadas por la extracción la preparación de DNA fue digerida utilizando la enzima ATP-dependent Plasmid-Safe exonuclease (Epicentre, Madison, WI) con la siguiente preparación: 40 µg de DNA fueron digeridos por 15 horas a 37 °C con 960 unidades de enzima sin diluir en 1440 µl de buffer de reacción completo 1X. Luego se añadió una mezcla fresca de enzima

(300 unidades en 600 µl de buffer de reacción completo) y se hizo una segunda incubación por 4 horas. Cien unidades de enzima concentrada fueron añadidas y se hizo una tercera incubación por tres horas. El DNA no digerido fue precipitado por centrifugación utilizando acetato de sodio y etanol absoluto. El precipitado fue resuspendido en 30 µl de buffer TE (10 mM Tris-base, 1 mM EDTA, pH 8.0) el cual contenía SDS al 1%. Se hizo una incubación de 10 minutos a 65 °C y se hizo una dilución añadiendo 500 µl más de TE. Posterior a dos extracciones con una mezcla de fenol-cloroformo-álcool isoamílico (25:24:1, Sigma) y una extracción con cloroformo, el DNA fue precipitado nuevamente con acetato de sodio y etanol absoluto, se le hizo un lavado con etanol al 70% y se resuspendió el DNA en agua libre de nucleasas. Una parte de esta preparación de DNA altamente enriquecida en mtDNA circular fue secuenciada directamente con la plataforma Illumina HiSeq2000 o HiSeq2500 (lecturas de hebra pareada de 50, 100, ó 125 nt de longitud, todas con un tamaño de inserto de 300 nucleótidos). Las secuenciaciones se hicieron en el "Genome Sequencing Shared Resource" parte del "Center for Genomic and Computational Biology" en "Duke University" (Durham, NC).

4.2. Análisis bioinformáticos

4.2.1. Control de calidad de las lecturas

Se utilizó el programa "FASTQC" versión 0.11.3 para evaluar la calidad de los dos archivos (lectura en sentido y lectura en anti-sentido) con terminación ".fastq" de cada una de las muestras. Este control de calidad fue hecho antes y después de filtrar secuencias cortas.

4.2.2. Mapeo de las lecturas

Para observar todo tipo de mutaciones se hizo un mapeo de cada una de las muestras contra un genoma de referencia. La secuencia de referencia utilizada para todos los mapeos fue el archivo "chrM.fa.gz", correspondiente al genoma mitocondrial de ratón C57BL/6J obtenido en diciembre del 2011 (ensamblaje

GRCm38/mm10) por “University of California Santa Cruz”. Esta secuencia de referencia puede hallarse en la base de datos de NCBI con el número de acceso NC_005089.1. Antes de utilizar el genoma, su archivo fue descargado y se descomprimió mediante el comando “gunzip” en la línea de comando de LINUX. Los mapeos y la detección de variaciones estructurales se hicieron primero con la función “align()” de la herramienta “Rsubread” en el paquete “Bioconductor” para el lenguaje de programación estadístico R (Liao y col, 2013) y también con el programa Segemehl 0.11 (Hoffmann y col, 2009). Para la función “align()” se utilizaron los parámetros “detectSV=True” e “indels=200”. Para el mapeo en Segemehl las opciones “-S” y “-D 0”.

La función align() genera un archivo .vsf por muestra, el cual contiene las variaciones estructurales incluyendo INDELS. Después se utilizó la función “vcf2bed” con el parámetro “--deletions” para extraer únicamente deleciones en el formato “.bed”. El programa Segemehl genera archivos “.bed” para las lecturas divididas, los cuales fueron procesados y resumidos utilizando el programa Haarz.x del paquete Segemehl.

4.2.3. Análisis estadístico de las deleciones

Los datos previamente generados fueron procesados y analizados utilizando un script de R hecho a la medida. Más adelante se hizo un análisis diferencial de deleciones con el paquete “EdgeR”, parte del repositorio de R “Bioconductor”.

4.3. Diagrama de flujo

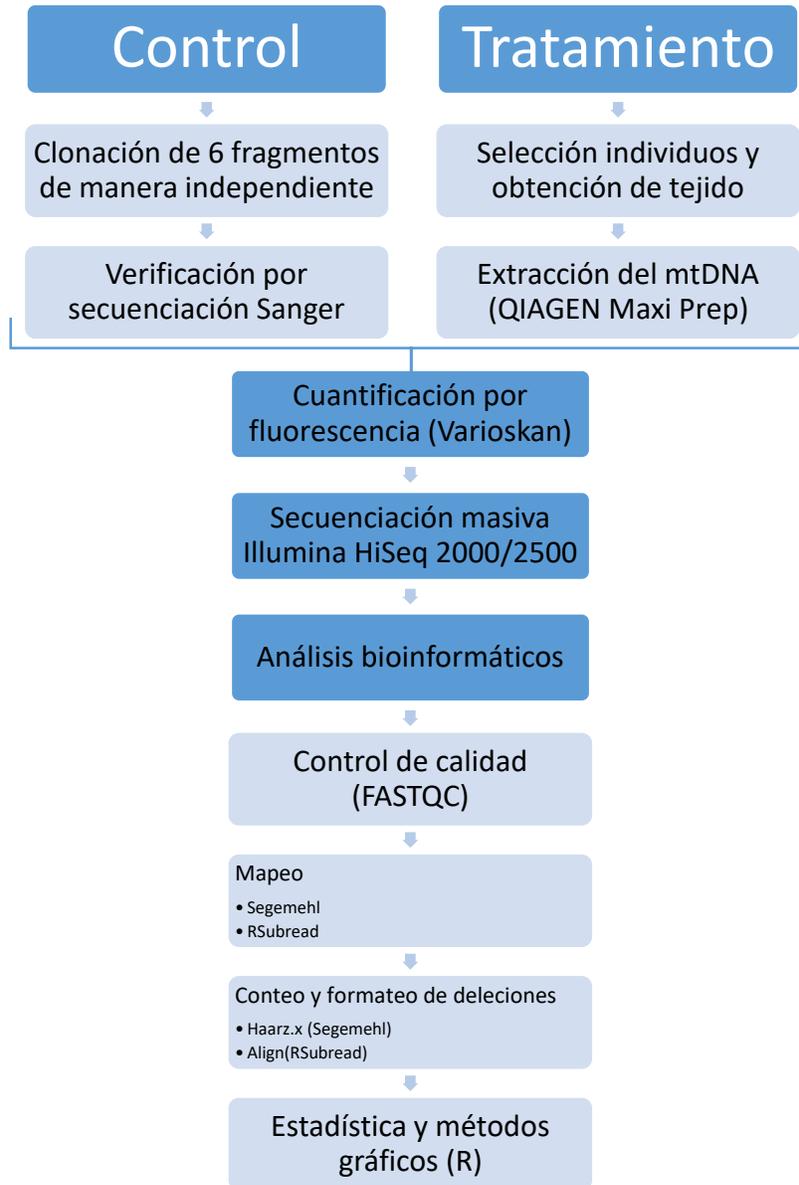


Figura 2. Diagrama de flujo de los métodos que se siguieron para ambos grupos experimentales. La parte superior del diagrama muestra la obtención de las muestras y la parte inferior muestra el chequeo de calidad de las muestras y los pasos posteriores de secuenciación y análisis.

4.4. Diseño estadístico

Para el grupo control se hizo una mezcla equimolar de los 6 fragmentos del mtDNA previamente clonados en plásmidos. A partir de esta mezcla equimolar, se obtuvieron 5 alícuotas y estas fueron secuenciadas de manera independiente.

El grupo experimental consistió de 5 muestras de cerebro de ratón. Cada una de estas muestras fue una mezcla del tejido cerebral de 25 embriones murinos E12.5. A partir de estas mezclas se hizo una extracción para mtDNA. Estas 5 muestras de mtDNA fueron secuenciadas de manera independiente.

5. RESULTADOS.

Para cada muestra del grupo del tratamiento los genomas mitocondriales fueron obtenidos a partir de los tejidos cerebrales de 5 distintos embriones de ratón en desarrollo. Como cada muestra proviene de distintos embriones, esperamos una inherente variación biológica en los genomas mitocondriales de este grupo. Para el grupo control se secuenciaron 5 muestras provenientes de una librería del mismo genoma mitocondrial previamente clonado en 6 fragmentos y amplificado en *E. coli*. Como todas estas copias del genoma fueron amplificadas a partir de los mismos fragmentos, para este grupo esperamos ausencia de la variabilidad biológica presente en las muestras obtenidas directamente del tejido murino.

5.1. Procesamiento y filtrado de los datos de secuenciación

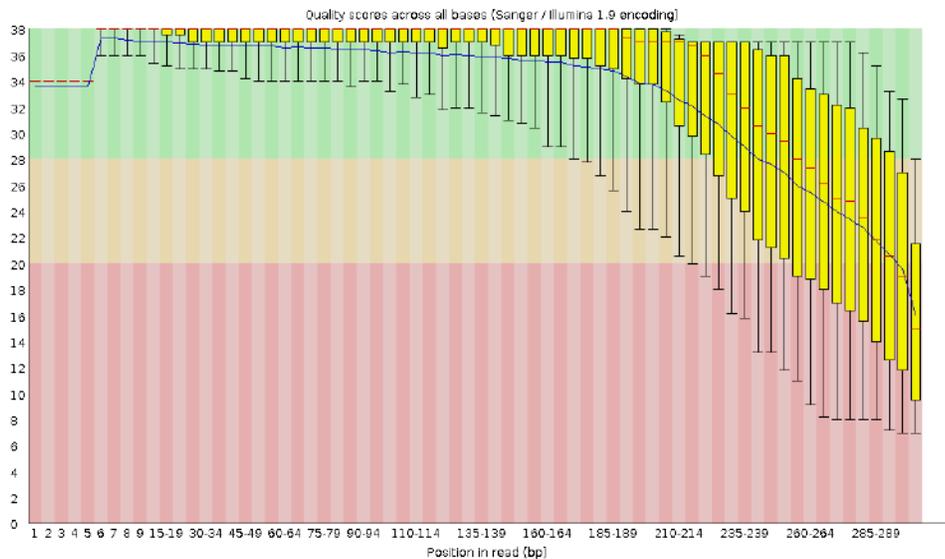


Figura 3. Diagrama de caja y bigote representando los puntajes de calidad por base (Phred 33) de las lecturas secuenciadas. Lecturas correspondientes a la muestra C1-R1. Esta gráfica corresponde a las lecturas en sentido positivo de la primera muestra del grupo control.

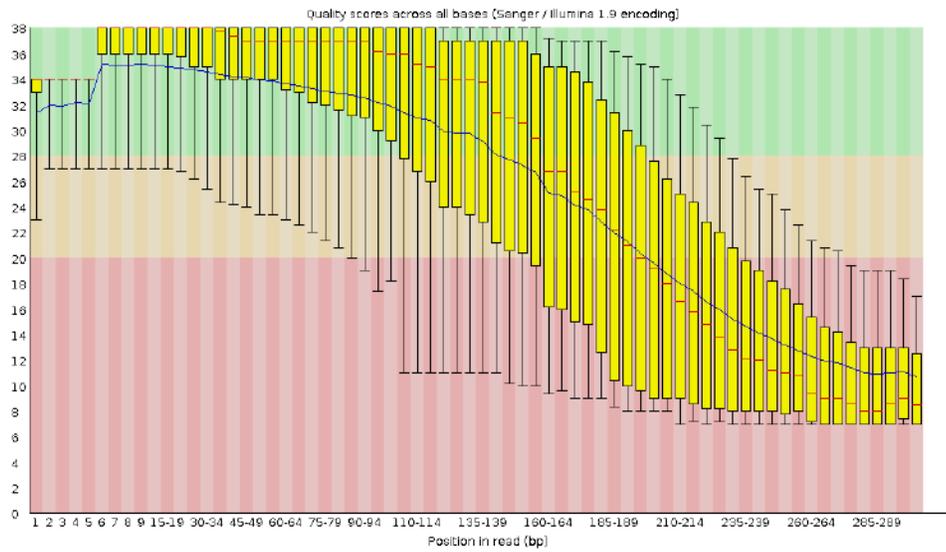


Figura 4. Diagrama de caja y bigote representando los puntajes de calidad por base (Phred 33) de las lecturas secuenciadas. Lecturas correspondientes a la muestra C1-R2. Esta gráfica corresponde a las lecturas en antisentido de la primera muestra del grupo control.

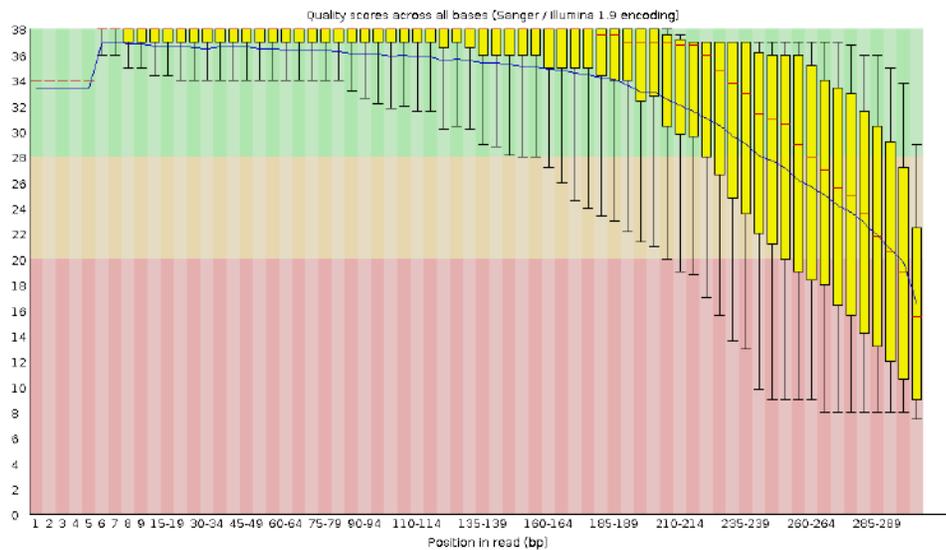


Figura 5. Diagrama de caja y bigote representando los puntajes de calidad por base (Phred 33) de las lecturas secuenciadas. Lecturas correspondientes a la muestra T1-R1. Esta gráfica corresponde a las lecturas en sentido positivo de la primera muestra del grupo del tratamiento.

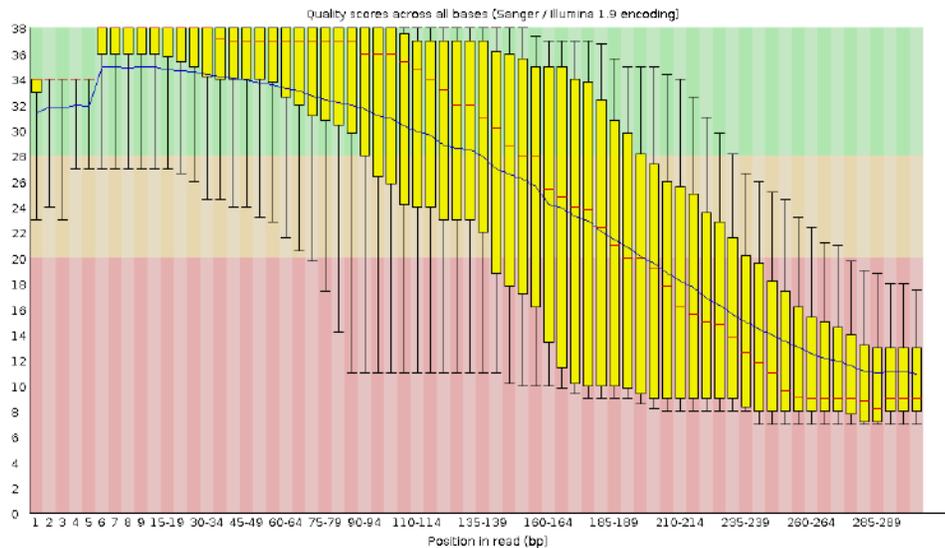


Figura 6. Diagrama de caja y bigote representando los puntajes de calidad por base (Phred 33) de las lecturas secuenciadas. Lecturas correspondientes a la muestra T1-R2. Esta gráfica corresponde a las lecturas en antisentido de la primera muestra del grupo del tratamiento.

Llevamos a cabo un chequeo de la calidad de las lecturas provenientes de la secuenciación. El puntaje de calidad de las lecturas en antisentido (Figuras 4 y 6) fue mucho más bajo que en las lecturas en sentido positivo (Figura 3 y 5). A pesar de la baja calidad al final de las lecturas en antisentido, los programas de mapeo utilizan “k-mers” así que la baja calidad no afectará la calidad del mapeo. Por ello mismo decidimos no recortar las secuencias para conservar la mayor información de las lecturas posible. Como se observa en el Cuadro 1, en promedio pudimos mapear un 54% de las lecturas por muestra al genoma de referencia.

Cuadro 1. Estadísticas de la reconstrucción del genoma mitocondrial con “Rsubread”. Conteos de las lecturas correspondientes a la secuenciación de cada una de las muestras (hebra pareada).

Muestra	Lecturas totales	Lecturas mapeadas	Proporción de lecturas mapeadas
C1	1426441	682806	0.478678
C2	1545897	711897	0.460507
C3	1381664	682712	0.494123
C4	1391208	647610	0.465502
C5	1301945	641714	0.492889
T1	1264894	765577	0.60525
T2	550371	394713	0.717176
T3	1841748	1160115	0.629899
T4	1431551	685831	0.479082
T5	1272493	733446	0.576385

5.2. Puntajes de calidad del mapeo en una resolución por base a lo largo del genoma mitocondrial ensamblado

Identificamos deleciones utilizando Segemehl y Rsubread. Solamente hallamos 32 eventos de deleción originados del mapeo en Segemehl, con 5 o más lecturas divididas para cada deleción, por lo que decidimos no proseguir con el análisis estadístico con estos datos. Un total de 4,500 deleciones fueron halladas en el mapeo hecho en el paquete “Rsubread”. Todas ellas fueron utilizadas para los análisis estadísticos y los métodos gráficos.

En la primera parte del análisis estadístico se evaluó la calidad de los mapeos en cada posición del genoma (puntaje de calidad reportado por base en una escala Phred-33). Por los objetivos del análisis nos fueron de interés las partes de lecturas mapeadas en dos diferentes posiciones en el genoma de referencia. Una lectura dividida es una lectura con dos alineaciones parciales en dos distintos sitios del genoma de referencia, donde en este caso la región entre ambos sitios sería la parte deletada en el genoma mitocondrial de la muestra. Para las lecturas divididas calculamos la calidad de mapeo en una resolución por base a lo largo del genoma mitocondrial de cada una de las muestras. Observamos un alto puntaje de calidad

en lecturas de secuenciación en la gran mayoría del genoma mitocondrial, pero hubo un número selecto de regiones en las que la calidad fue muy baja. En algunos de los casos la misma región tiene un puntaje de calidad baja en las muestras de ambos grupos experimentales. Después utilizamos los puntajes de calidad de mapeo máximos de cada grupo y los graficamos. Hicimos esto para darnos una idea general de las posiciones con bajos puntajes de una manera independiente a de qué muestra se trate. De manera consistente con las observaciones obtenidas de muestras individuales, existen regiones en el genoma que probablemente no fueron secuenciadas adecuadamente. Sin embargo, la mayoría del genoma fue cubierto adecuadamente y esto permitirá proseguir con un análisis estadístico en la mayoría de los sitios del genoma sin ningún problema.

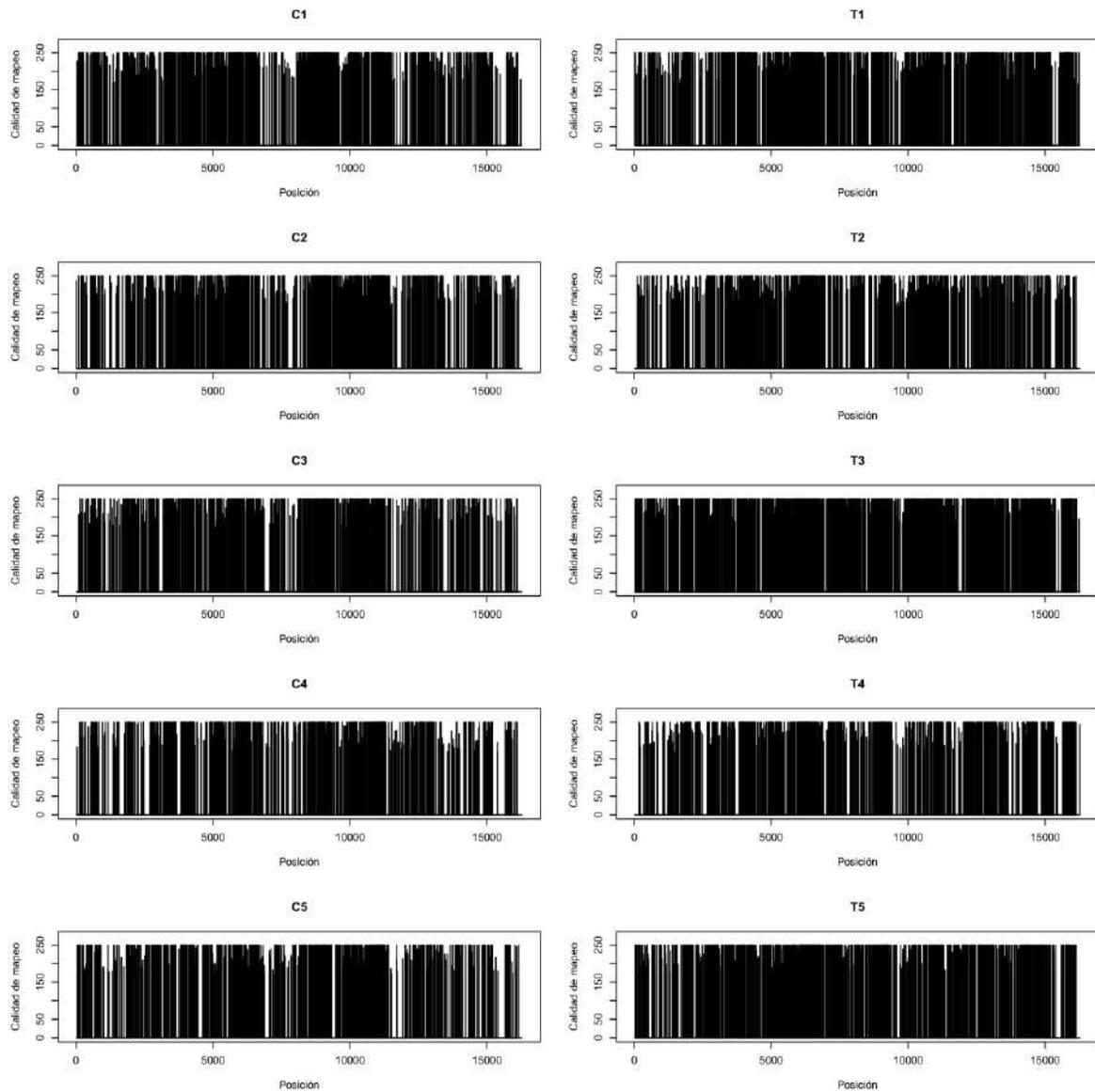


Figura 7. Calidad de mapeo en una resolución por base a lo largo del genoma mitocondrial. Las regiones blancas dentro de las gráficas corresponden a zonas del genoma donde hubo problemas en la secuenciación.

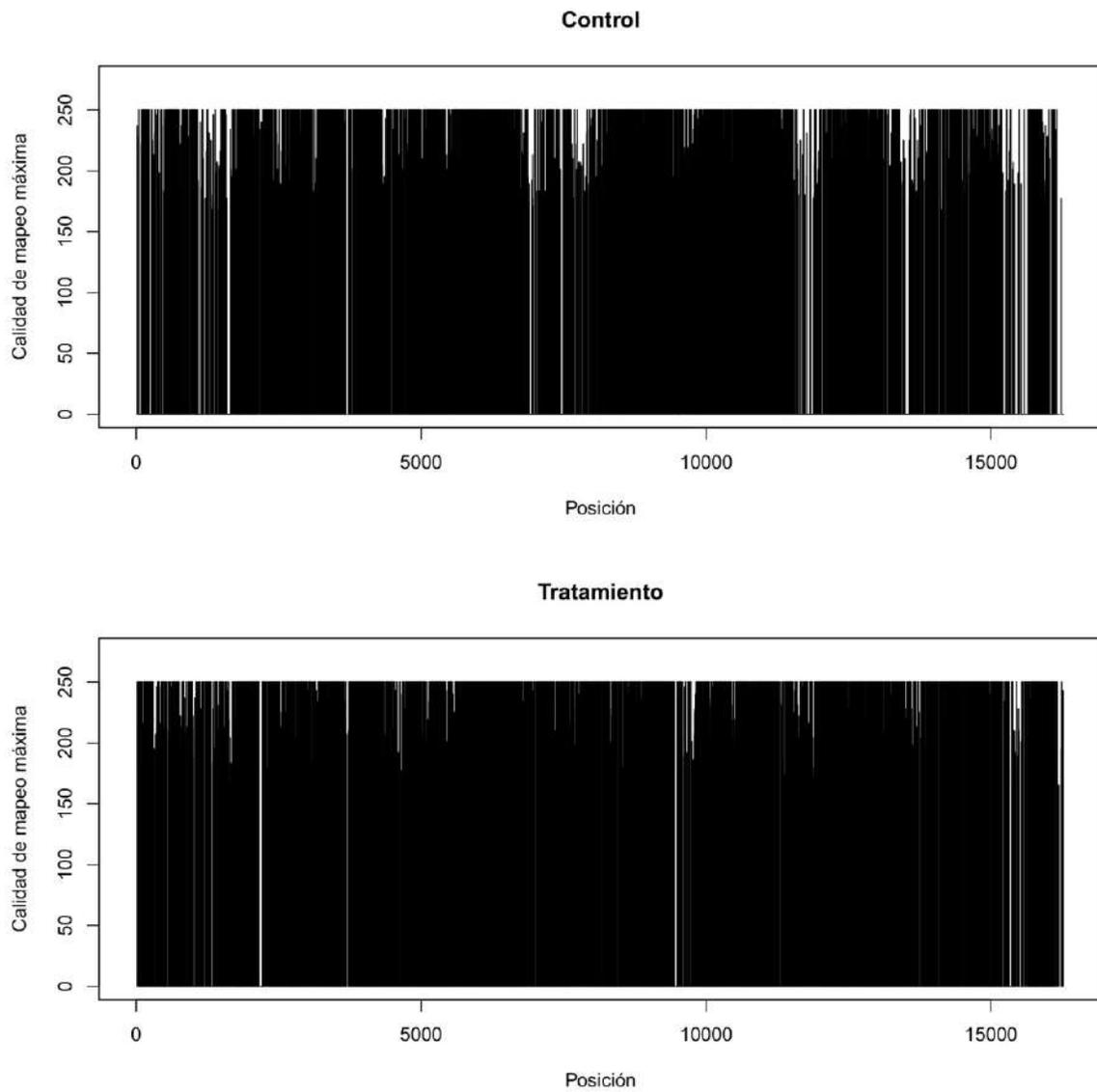


Figura 8. Calidad de mapeo máxima en una resolución por base a lo largo del genoma mitocondrial. Cada una de las gráficas incluye las 5 muestras dentro del grupo experimental.

5.3. Las deleciones de gran longitud son un evento muy raro

Lo primero que queríamos después del chequeo de calidad era la probabilidad de hallar deleciones de cierta longitud. Hallamos que la ocurrencia de deleciones de más de dos nucleótidos es muy baja (Figura 9). No observamos deleciones con una longitud superior a 15 nucleótidos, lo que sugiere que la mitocondria no puede tolerar deleciones más grandes en su genoma.

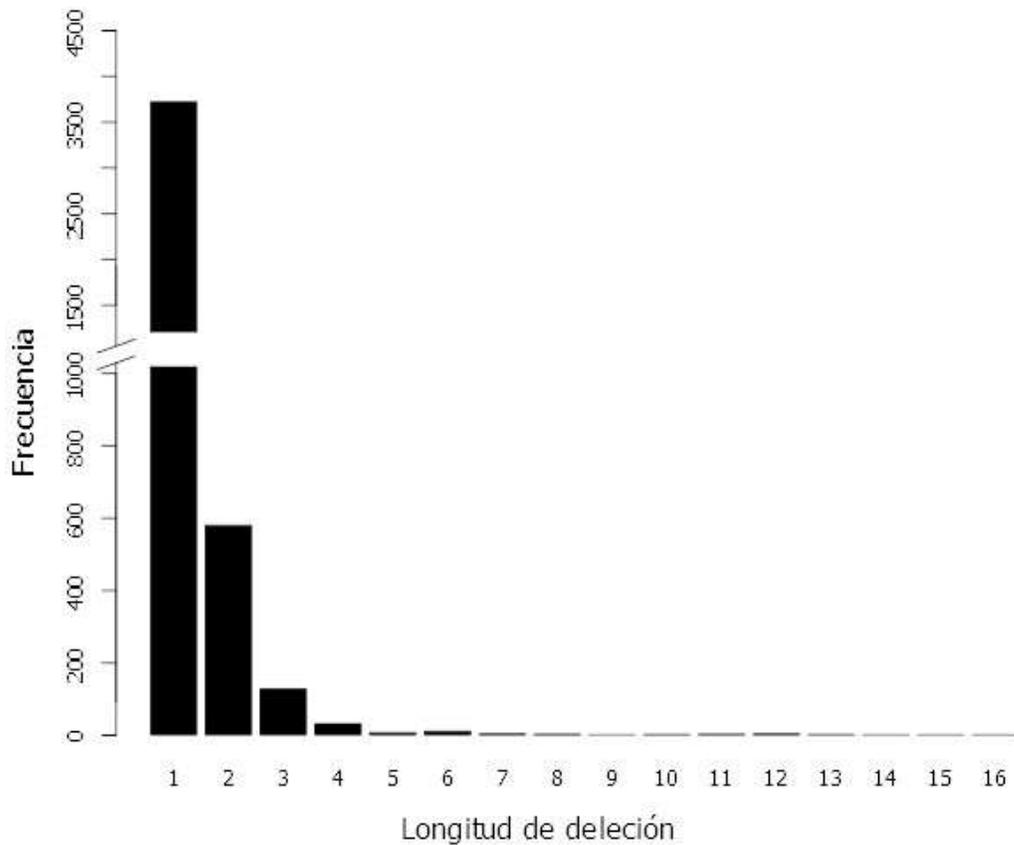


Figura 9. Distribución de las frecuencias en las distintas longitudes de todos los eventos de deleción.

El número de lecturas divididas obtenido en cada posición en el genoma de referencia podría variar entre muestras debido a diferencias en la profundidad de secuenciación. Fue por ello que normalizamos los conteos dividiéndolos por profundidad de secuenciación de cada muestra. La profundidad de secuenciación es el número total de lecturas presentes en cada muestra. Después de la normalización el número de lecturas es comparable entre muestras. Para llevar estos números de lecturas a la misma escala en cada muestra los multiplicamos por un factor de escalamiento de 1000. Esto significa que para cada muestra el número de lecturas divididas se encuentra en una escala por mil genomas. Por ejemplo, si se graficaran 50 lecturas divididas eventos de cierta longitud esto significaría que 50 de cada 1000 genomas en la muestra presentan esta delección. Esta normalización es similar a la de CPM (counts per million), pero en este caso la adaptamos los valores de nuestros datos.

Calculamos la longitud de las delecciones y el número total de lecturas divididas representándolas. Como era esperado, la longitud es inversamente proporcional al número de lecturas divididas (Figura 10). Esta tendencia es consistente con los fenómenos biológicos celulares ya que entre mayor sea una delección incrementará exponencialmente el riesgo de fatalidad. Esta es la razón por la que en nuestros datos no hallamos ninguna delección con una longitud superior a 16 nucleótidos (Figuras 9 y 10).

Como se muestra en la Figura 10, el número total de lecturas halladas en todas las muestras graficado contra la longitud de delección muestra la muy baja cobertura de lecturas divididas que difícilmente alcanza 25 lecturas divididas incluso para delecciones de un solo nucleótido. En nuestra escala de normalización, el número de lecturas divididas mostrado en la Figura 10 es por cada 10000 genomas. Como ninguno de los eventos de delección supera las 25 lecturas divididas, son artefacto. En un nivel de confianza de 95%, con un tamaño total de muestras de 10000, serían necesarios números de lecturas divididas superiores a 500 para ser considerados significativos. De otra manera, son lecturas originadas por probabilidad aleatoria.

Tampoco observamos números significativos de lecturas divididas en muestras individuales (Figura 11).

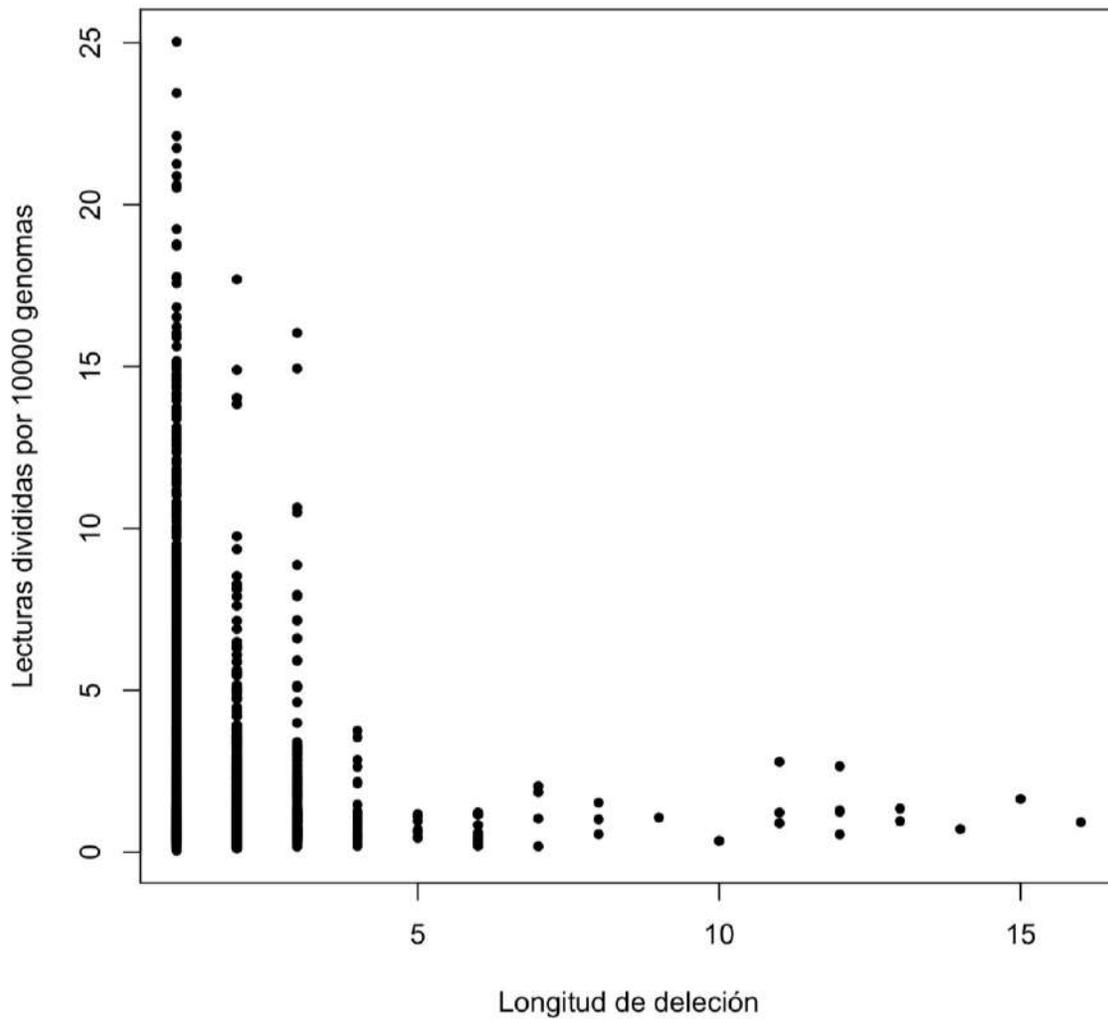


Figura 10. Número de lecturas divididas normalizado correspondiente a cada longitud de deleción. A longitudes de deleciones superiores a 5 nucleótidos el número de lecturas divididas es escaso.

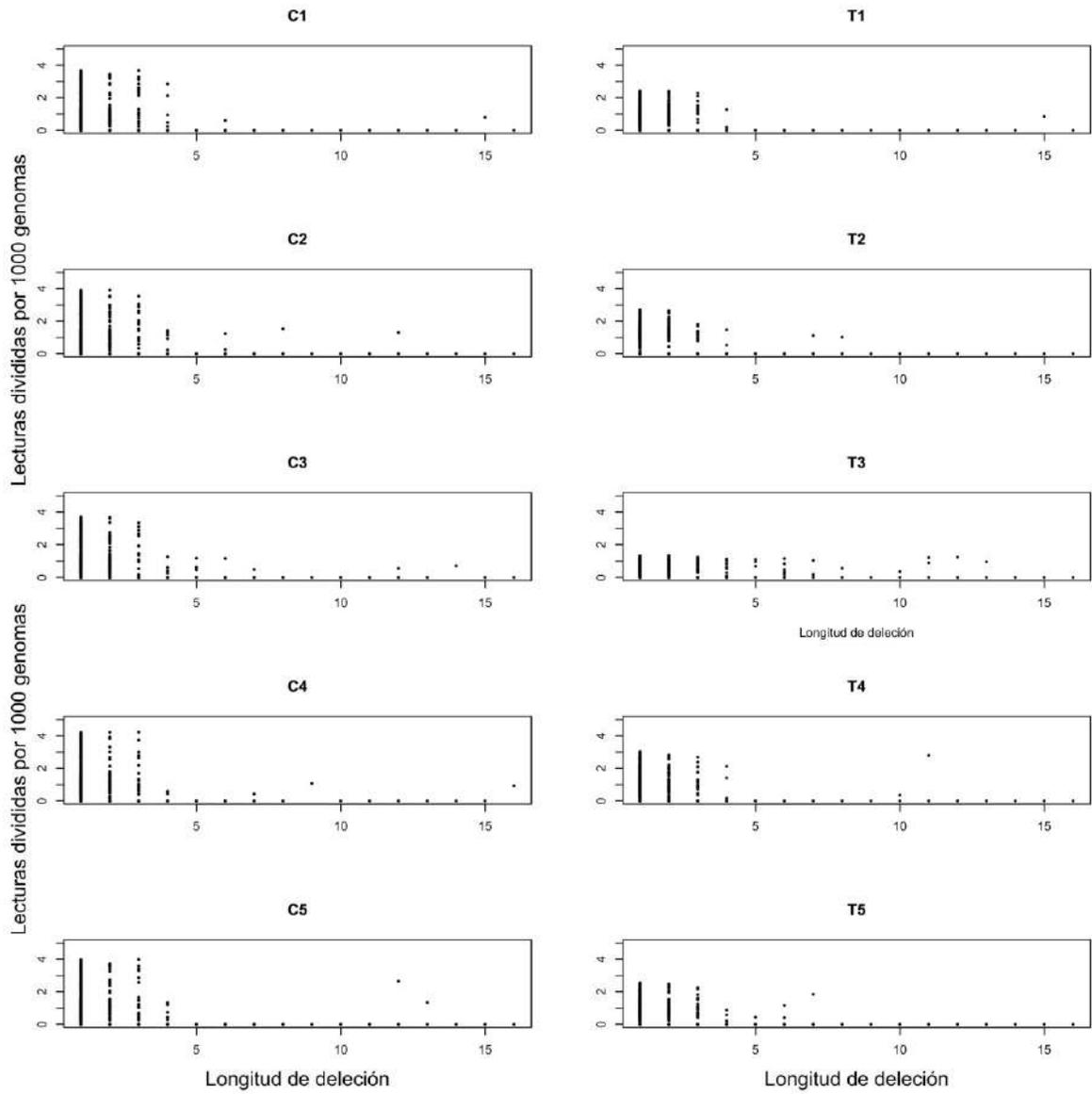


Figura 11. Número de lecturas divididas normalizado graficado contra longitud de deleción. Cada gráfica corresponde a una muestra.

5.4. Número superior de eventos de delección en muestras biológicas comparado con el grupo control

Calculamos el número de lecturas en una resolución por base para cada una de las muestras y lo graficamos. De manera general observamos un mayor número de deleciones a través del genoma en las muestras biológicas comparándolas con el control. El número total de eventos de delección en las muestras del grupo control fue de 2,683 y de 3200 en el grupo del tratamiento, de las cuales 1817 fueron eventos en común entre ambos grupos. Existe un aumento sustancial en el número de lecturas normalizado mapeadas en las muestras biológicas comparadas con el control (Figura 12). Después de observar los números de lecturas divididas a lo largo del genoma para cada una de las muestras, sumamos los números de lecturas a lo largo del genoma para cada grupo. Como lo esperábamos, hay un mayor número de lecturas en el grupo control posiblemente debido a su menor variabilidad biológica. En el grupo control las deleciones se encuentran distribuidas de una manera más consistente a lo largo del genoma que en el grupo control. Por el otro lado, en el control hallamos picos localizados en ciertas posiciones y una distribución de deleciones más heterogénea que en las muestras biológicas (Figura 13).

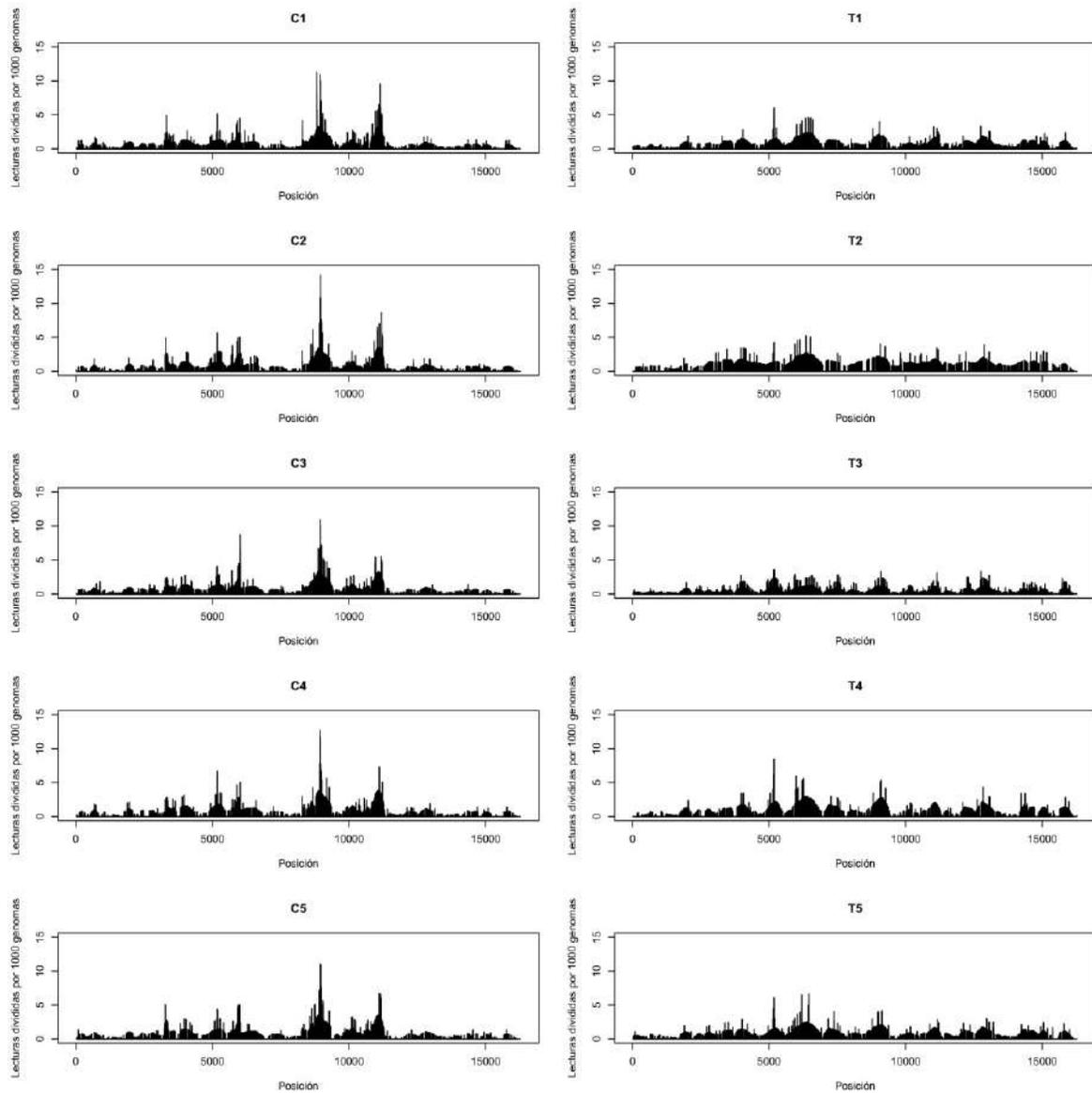


Figura 12. Número de lecturas divididas normalizado de cada una de las muestras graficado contra la posición dentro del genoma mitocondrial. El número de lecturas en cada posición fue obtenido en una resolución por base a partir de una matriz con posiciones iniciales y finales de deleción.

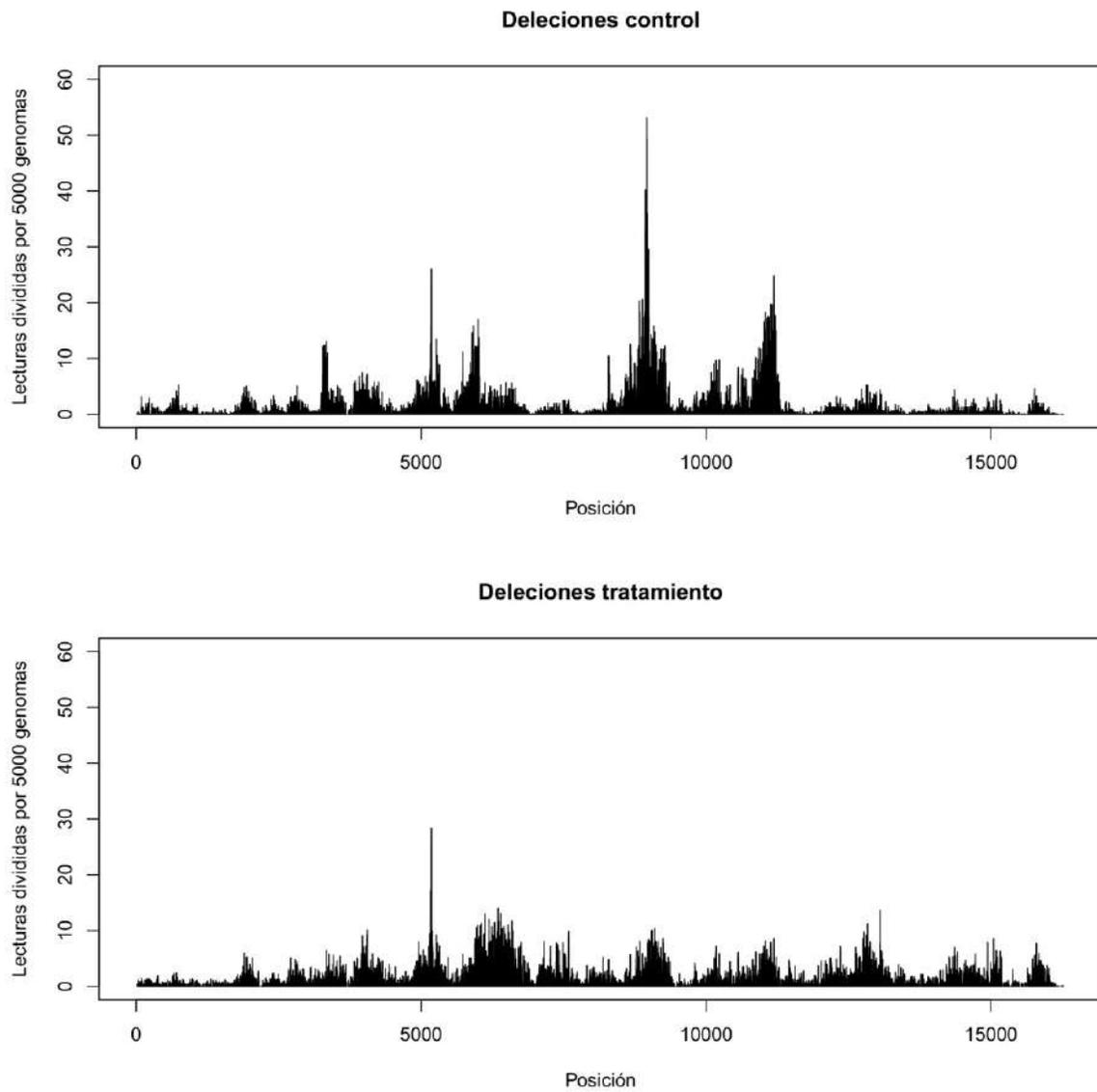


Figura 13. Total de lecturas divididas normalizado de cada grupo experimental graficado contra la posición en el genoma mitocondrial. El número de lecturas en cada posición fue obtenido en una resolución por base a partir de una matriz con posiciones iniciales y finales de deleción.

5.5. Los genomas mitocondriales son altamente resistentes a eventos de delección

Los datos fueron sometidos a un análisis de significancia en deleciones diferenciales entre grupos. En el Cuadro 2 se muestran las deleciones significativas con “p-value” menor o igual a 0.05 de manera previa a la corrección de la prueba de hipótesis múltiples. Este análisis no arrojó deleciones significativas después de una prueba de múltiples hipótesis. Esto sugiere que las deleciones que hallamos son un artefacto metodológico. De las 44 deleciones, 22 fueron halladas de manera diferencial en el grupo control y 22 en el grupo del tratamiento. De hecho, no esperábamos deleciones diferenciales en el grupo control por que este carece de variabilidad biológica.

Cuadro 2. Deleciones con más bajo “p-value” obtenidas del análisis en “EdgeR”

Delección	logFC	logCPM	PValue	FDR
8960-8962	-18.8228	10.83089	0.001014	0.99995
8956-8958	-18.8202	10.82826	0.001015	0.99995
8956-8959	-18.8166	10.82472	0.001017	0.99995
9073-9075	-18.4176	10.42568	0.001371	0.99995
8994-8996	-18.2979	10.30605	0.001498	0.99995
6594-6596	17.94802	9.956181	0.001914	0.99995
7396-7398	17.28415	9.29242	0.004006	0.99995
7587-7590	16.93018	8.938521	0.005553	0.99995
8980-8982	-18.6064	10.61453	0.006175	0.99995
8953-8955	-18.5655	10.57358	0.006368	0.99995
8960-8963	-18.4873	10.49543	0.006622	0.99995
745-747	-16.6128	8.621249	0.006909	0.99995
10960-10962	-17.9268	9.934989	0.009041	0.99995
4325-4327	16.48689	8.495366	0.010385	0.99995
5865-5867	-17.5689	9.577147	0.011557	0.99995
12789-12791	17.4808	9.489026	0.012394	0.99995
12769-12771	17.41206	9.420302	0.013261	0.99995
5191-5193	17.19645	9.204732	0.016604	0.99995
15833-15835	16.94221	8.95055	0.01918	0.99995
7266-7269	16.86315	8.871513	0.020299	0.99995
12185-12187	16.74262	8.751016	0.021707	0.99995
8950-8952	-18.2034	10.2115	0.025294	0.99995

8935-8937	-18.1238	10.13199	0.02635	0.99995
8971-8975	-18.0641	10.07224	0.026713	0.99995
11085-11087	-18.0641	10.07228	0.026833	0.99995
11122-11124	-18.0365	10.04462	0.027103	0.99995
8873-8875	-18.0354	10.04356	0.027542	0.99995
8985-8987	-17.9791	9.987279	0.028324	0.99995
11163-11165	-17.9373	9.945444	0.02849	0.99995
9033-9035	-17.7477	9.755842	0.031606	0.99995
7074-7076	16.43483	8.443318	0.035358	0.99995
9198-9200	-17.4864	9.494606	0.036773	0.99995
6417-6419	17.47861	9.486845	0.037035	0.99995
9127-9129	-17.4614	9.469584	0.037589	0.99995
15638-15640	16.37603	8.384543	0.037767	0.99995
6489-6491	17.36146	9.369714	0.040562	0.99995
6503-6505	17.36081	9.369065	0.040769	0.99995
6456-6458	17.31267	9.320933	0.04258	0.99995
6569-6571	17.31049	9.318752	0.042596	0.99995
6227-6230	17.29972	9.307988	0.042971	0.99995
11466-11468	16.25614	8.264702	0.045183	0.99995
6220-6222	17.19298	9.201263	0.046014	0.99995
6517-6519	17.23643	9.244705	0.046145	0.99995
12774-12776	17.07564	9.08395	0.048894	0.99995

5.6. Razones posibles para deleciones artefactuales

Los análisis estadísticos reportados en la parte superior indican que las deleciones identificadas son artefactuales y esto se podría deber a varias razones. Estas razones podrían ser la preparación de las muestras, la preparación de las librerías genómicas, amplificación de los fragmentos, secuenciación, errores de mapeo y cualquiera de estos errores reportados por nuestras herramientas como variaciones estructurales. La mayoría de estos potenciales problemas son difíciles de confirmar, por lo que visualizamos todas las lecturas divididas de las distintas muestras a lo largo del genoma mitocondrial utilizando el programa “IGV”. Hallamos una distribución de las lecturas divididas bastante uniforme a través del genoma mitocondrial, indicando una ausencia de especificidad posicional y por lo tanto ausencia de deleciones recurrentes (Figura 14). Así mismo, hicimos un acercamiento a la región que abarca las posiciones 8,953 y 8,962 en la cual hallamos

varias deleciones con bajos puntajes de “p-value”. Como se muestra en la Figura 15, la mayoría de las lecturas divididas se hallan en el control, en donde no eran esperadas. Por el otro lado, las muestras biológicas tienen muy pocas lecturas divididas de deleciones. La distribución de las lecturas divididas en esta región to sigue ningún patrón inherente, indicando una vez más que son artefacto. La razón más probable de que se hayan generado es la existencia de varios dímeros del mismo nucleótido en esta región (Figura 15).

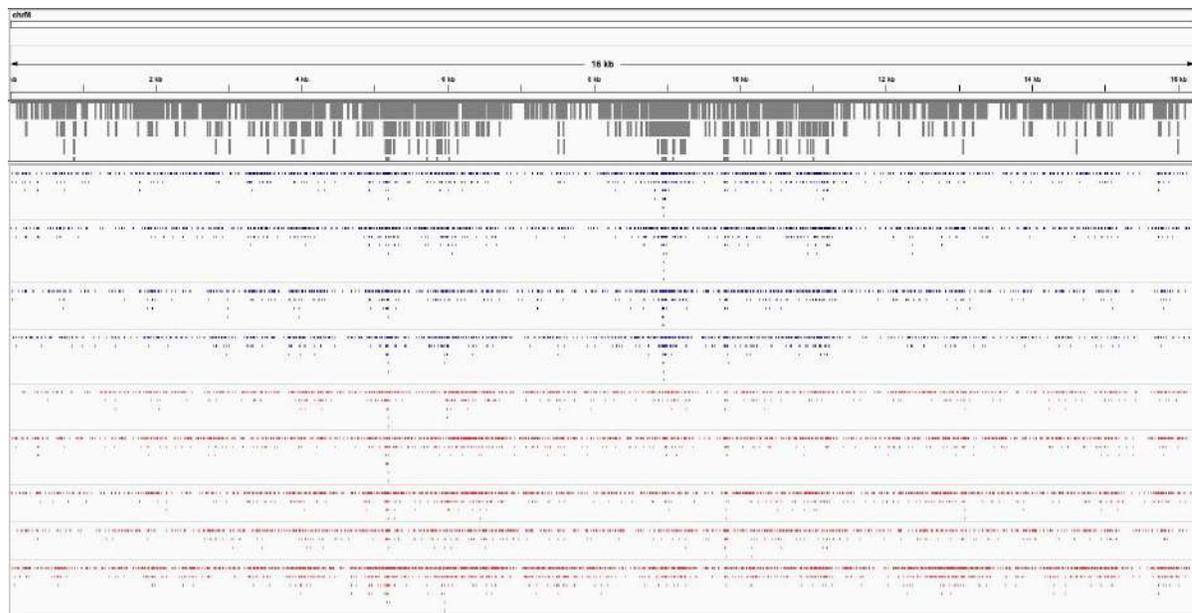


Figura 14. Visualización de las lecturas divididas mapeadas en IGV de todo el genoma. En la parte superior se encuentran las posiciones del genoma mitocondrial. Más abajo, una gráfica de cobertura de mapeo. En la parte inferior se encuentra una representación gráfica de las diferencias entre las lecturas de cada muestra contra el genoma de referencia.

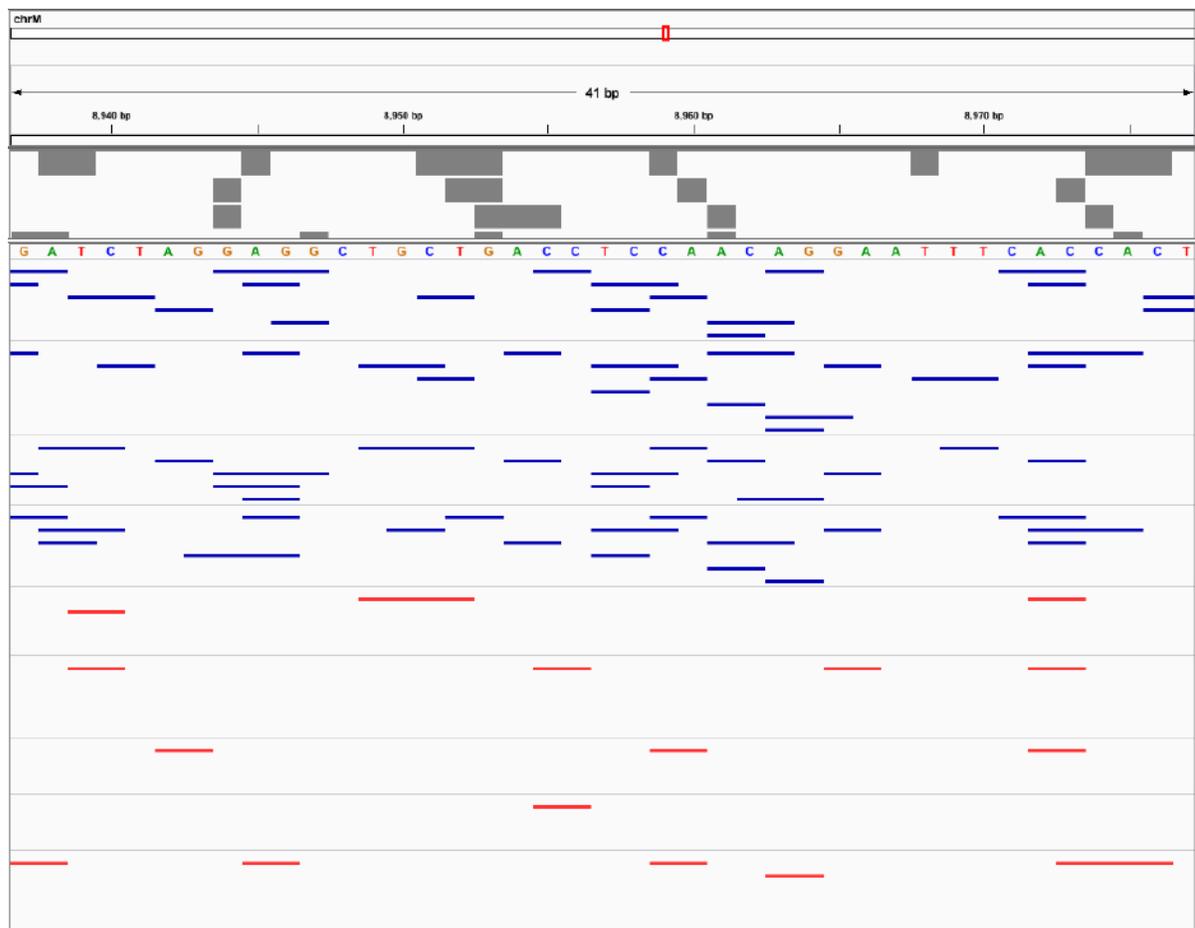


Figura 15. Visualización de las lecturas divididas mapeadas en IGV de la zona en la que se hallaron la gran mayoría de las deleciones con “p-value” bajos. En la parte superior se encuentran las posiciones del genoma mitocondrial. Más abajo, una gráfica de cobertura de mapeo y la secuencia del genoma de referencia. En la parte inferior se encuentra una representación gráfica de las diferencias entre las lecturas de cada muestra contra el genoma de referencia.

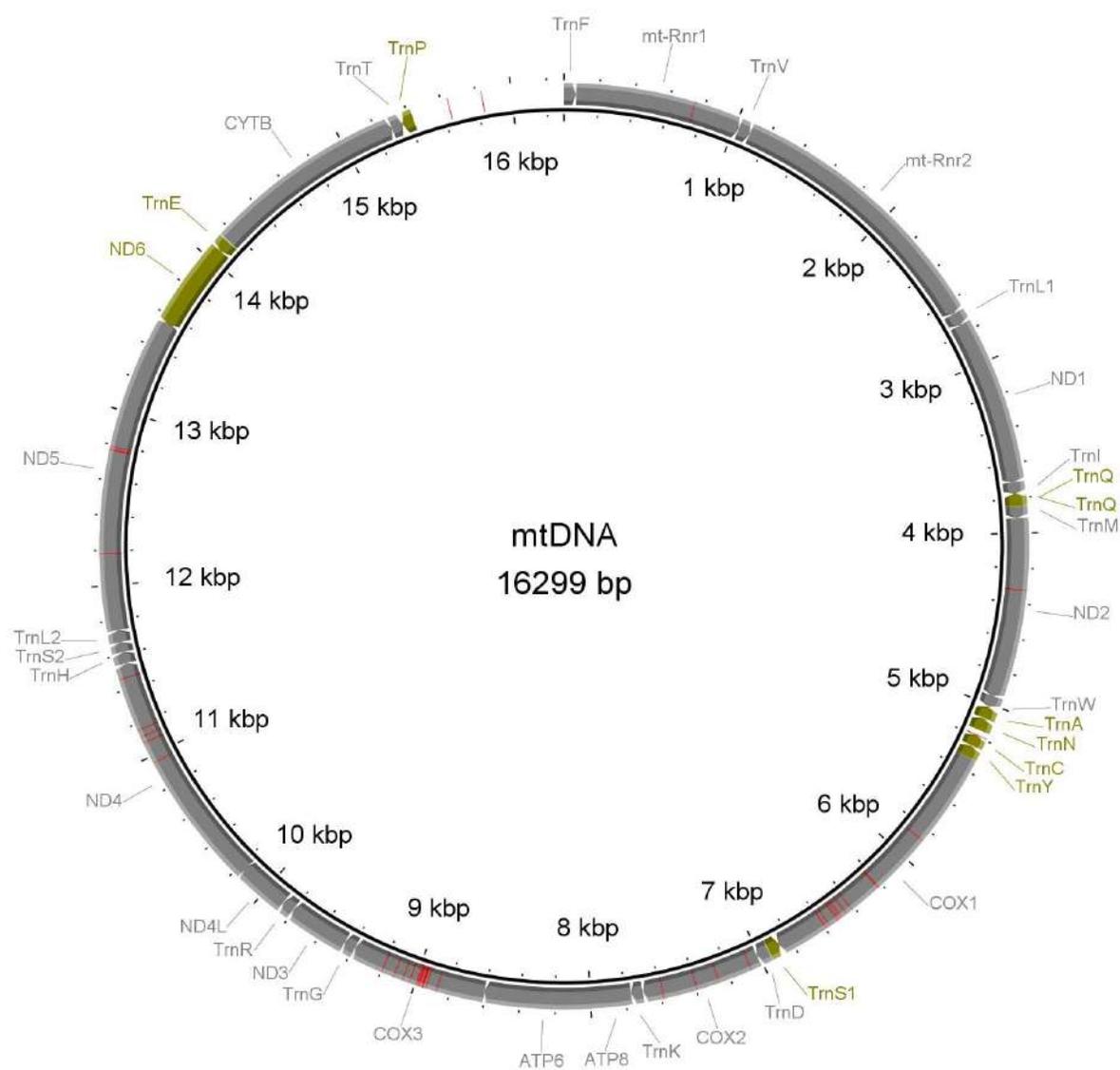


Figura 16. Representación gráfica de las deleciones (rojo) con los valores más bajos de “p-value” en el genoma mitocondrial de ratón. En color gris se encuentran los genes en sentido positivo y en color verde los genes en antisentido. La mayoría de las deleciones se encuentran en los genes COX1, COX3 y ND5.

6. DISCUSIÓN.

Las mitocondrias de todos los organismos eucariontes vivos se derivan de lo que una vez fue una alfa-proteobacteria la cual era endosimbionte de la célula hospedera y juntos fungieron como el último ancestro en común de los eucariontes hace aproximadamente 2 mil millones de años. Una alfa-proteobacteria actual tiene en promedio 3000 genes que fueron perdidos a lo largo de la evolución o transferidos al genoma nuclear, dejando solamente 37 genes en mamíferos codificados por un total de aproximadamente 16 kilobases de las cuales menos de 1 kilobase son regiones no codificantes. Además, la gran mayoría de estas regiones no codificantes las abarcan los orígenes de replicación. Esto indica que el genoma de la alfa-proteobacteria que permaneció en la mitocondria pasó por una muy fuerte presión selectiva a lo largo de la evolución de los eucariontes. No es sorprendente, pues las mitocondrias (endosimbiontes de alfa-proteobacterias) han proveído la energía requerida para el mantenimiento y la supervivencia de la célula en una atmósfera aeróbica. Considerando la importancia crítica de los genomas mitocondriales dentro de organismos aerobios y su organización con una alta densidad génica, es menos probable que pueda tolerar mutaciones. Sin embargo, existen varios reportes que sugieren heteroplasmia a nivel de secuencia en el genoma mitocondrial y sus consecuencias funcionales positivas, en una manera tejida específica y durante el desarrollo embrionario. Estos resultados son intrigantes considerando el hecho de las mitocondrias ya perdieron todas y cada una de las partes de sus genomas que no son esenciales y algunas de las que son esenciales fueron transferidas el genoma nuclear. Por ello, investigamos sistemáticamente la autenticidad de estas aseveraciones.

Utilizamos todas las lecturas divididas y su calidad de mapeo. La mayoría del genoma mostró una buena calidad de mapeo (representando una secuenciación adecuada) y por lo tanto fue útil para la validación de las deleciones halladas. En caso de hallar deleciones significativas dentro de las regiones del genoma con una secuenciación problemática, estas tendrán que ser descartadas por la posibilidad de ser un artificio metodológico. Después de normalizar los datos en bruto y

transformarlos a lecturas por kilobase encontramos una gran incidencia de deleciones cortas y una baja incidencia de las de mayor tamaño. En ambos grupos todas las muestras mostraron un máximo de longitud de deleción de 16 nucleótidos. Esta tendencia es consistente con los fenómenos biológicos celulares, a mayor longitud de deleción mayor riesgo de disfunción mitocondrial. Por esto mismo es que las deleciones largas no son favorecidas en el genoma mitocondrial. Después de haber transformado los números de lecturas divididas a una resolución por base pudimos observar una mayor presencia de deleciones en las muestras biológicas que en el control.

La mayor parte de las deleciones halladas tuvieron una longitud de un solo nucleótido y muy pocas más de dos nucleótidos. Esto indica que los sistemas de reparación del genoma mitocondrial tienen un altísimo grado de fidelidad.

En el control observamos varios picos en el número de lecturas divididas no esperados los cuales podrían deberse a una mayor amplificación de estas regiones. No obstante, no conocemos la causa exacta de la aparición de estos picos, pero sospechamos que podrían ser errores durante la amplificación de un mismo fragmento previamente clonado. En contraste con los controles, las muestras biológicas contienen deleciones en bajas frecuencias a lo largo del genoma. Esto sugiere que estas señales podrían ser ruido de fondo. Cada célula tiene miles de mitocondrias y, asumiendo la tasa de mutaciones espontáneas, podría haber mitocondrias con deleciones por probabilidad aleatoria. Estas deleciones ocurrirán en muy baja frecuencia porque al incrementar en números darán lugar a la muerte celular por falta de metabolismo energético. Así mismo, como estas deleciones son eventos aleatorios, no esperamos un sesgo posicional en el cromosoma. Por ello, es más probable que las deleciones que encontramos a lo largo del genoma mitocondrial en las muestras biológicas provengan de estas mitocondrias mutadas de manera espontánea. Por el otro lado, no observamos estas deleciones distribuidas uniformemente en las muestras control, sino picos en ciertas posiciones porque estas muestras fueron amplificadas a partir de los mismos fragmentos de genoma mitocondrial clonados en *E. coli*.

Como fue mencionado previamente, las regiones en las cuales encontramos muchas lecturas divididas son más probablemente debido a probabilidad aleatoria y podrían no tener ninguna consecuencia en las actividades celulares considerando su baja abundancia y corta longitud (menos de 5 nucleótidos). Corroboramos estos resultados utilizando el paquete “EdgeR” que procesa datos de conteos en bruto y utiliza modelos lineales para detectar diferencias (Robinson y col., 2009). Al utilizar una corrección múltiple no hallamos ninguna delección diferencial significativa. Las delecciones con los “p-value” más bajos podrían haberse generado por errores del mapeo, lo cual es más probable debido a que se encuentran en regiones con dímeros de nucleótidos (Figuras 14 y 15). Esto provee evidencia que apoya la idea de que el genoma mitocondrial es altamente resistente a delecciones y quizás a mutaciones en general. Sin embargo, en estudios a futuro se podrían analizar otros tipos de variantes, especialmente sustituciones de un solo nucleótido.

Una limitación importante en el desarrollo de algoritmos bioinformáticos para la detección de mutaciones mitocondriales es que la mayoría de las plataformas actuales han sido desarrolladas para su uso en genomas nucleares. Esto tiene un impacto importante en los resultados, sobre todo en análisis que requieren un alto grado de sensibilidad, como lo es la detección de heteroplasmia en el genoma mitocondrial. Existe una gran variabilidad de resultados entre métodos, e incluso en la de predicción de patogenicidad, al aplicar las herramientas de alineamiento actuales en el genoma mitocondrial. Por ello mismo, se deben dedicar esfuerzos en el desarrollo de herramientas específicas para su uso en el genoma mitocondrial (Bris y col., 2018). Aun tomando en consideración la posibilidad de limitaciones metodológicas, las delecciones halladas en esta investigación y su distribución son altamente consistentes con la biología de la mitocondria.

En la célula eucariota, las mitocondrias generan energía para su mantenimiento, supervivencia y también participan en sus vías de muerte celular programada. El genoma mitocondrial es altamente denso en genes y cualquier mutación podría tener una alta probabilidad de afectar directamente la supervivencia celular. Esta podría ser la razón de que no observamos ninguna delección significativa

en ninguna de las muestras biológicas provenientes de muestras heterogéneas de tejido cerebral en desarrollo. Todos los eventos de deleción reportados por nuestro análisis aparentan ser ruido de fondo por varias razones, de las cuales la más importante es que se encuentran muy por debajo del umbral estándar para ser clasificadas como eventos reales.

7. CONCLUSIONES.

En este estudio analizamos las deleciones en el genoma mitocondrial del ratón de manera sistemática. Mediante el uso de dos herramientas de mapeo completamente distintas no hallamos un solo evento de deleción significativo. Utilizando estadísticas descriptivas e inferenciales descartamos la posibilidad de que las deleciones recurrentes pueden ocurrir en el genoma mitocondrial. Las deleciones recurrentes detectadas en estudios previos podrían deberse a un bajo número de mitocondrias afectadas por mutaciones espontáneas. Basándonos en las longitudes de deleción encontradas, las células con una deleción mitocondrial de más de quince nucleótidos tienen una probabilidad mínima de supervivencia. Es altamente improbable que en el genoma mitocondrial ocurra una deleción de más de dieciseis nucleótidos. Basándonos en los datos analizados, las deleciones reportadas en este estudio podrían ser probablemente eventos aleatorios o errores de secuenciación. Las herramientas actuales para detección de variantes estructurales han mostrado resultados muy variados en algunos casos en el genoma mitocondrial. Además, se han reportado varios problemas en la interpretación a causa de cargas mutacionales muy bajas y el reportaje de variantes de baja significancia (Bris y col., 2018).

8. REFERENCIAS.

Alberts B, Johnson A, Leis J, Raff M, Roberts K y Walter P. Molecular biology of the cell. Quinta edición. Nueva York: Garland Science, **2008**:813-878.

Alexeyev M, Shokolenko I, Wilson G y Ledoux S. The Maintenance of Mitochondrial DNA Integrity--Critical Analysis and Update. **2013** Cold Spring Harbor Perspectives in Biology, 5(5).

Allen J y Paula WD. Mitochondrial genome function and maternal inheritance. **2013** Biochemical Society Transactions, 41(5), 1298-1304.

Ameur A, Stewart JB, Freyer C, Hagström E, Ingman M, Larsson N y Gyllensten U. **2011** Ultra-Deep Sequencing of Mouse Mitochondrial DNA: Mutational Patterns and Their Origins. **2011** PLoS Genetics, 7(3).

Anderson S, Bankier AT, Barrell BG, Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R y Young IG. Sequence and organization of the human mitochondrial genome. **1981** Nature, 290(5806), 457-465.

Bacman SR, Williams SL y Moraes CT. Intra- and intermolecular recombination of mitochondrial DNA after in vivo induction of multiple double-strand breaks. **2009** Nucleic Acids Res 37: 4218–4226.

Bayona-Bafaluy MP, Acín-Pérez R, Mullikin JC, Park JS, Moreno-Loshuertos R, Hu P, Pérez-Martos A, Fernández-Silva P, Bai Y y Enríquez JA. Revisiting the mouse mitochondrial DNA sequence. **2003** Nucleic Acids Research, 31(18), 5349-5355.

Bibb MJ, Etten RA, Wright CT, Walberg MW y Clayton, DA. Sequence and gene organization of mouse mitochondrial DNA. **1981** Cell, 26(2), 167-180.

Bosworth CM, Grandhi S, Gould MP y LaFramboise T. Detection and quantification of mitochondrial DNA deletions from next-generation sequence data. **2017** BMC bioinformatics, 18(Suppl 12), 407. doi:10.1186/s12859-017-1821-7.

Bris C, Goudenege D, Desquirit-Dumas V, Charif M, Colin E, Bonneau D, Amati-Bonneau P, Lenaers G, Reynier P y Procaccio V. Bioinformatics Tools and Databases to Assess the Pathogenicity of Mitochondrial DNA Variants in the Field of Next Generation Sequencing. **2018** *Frontiers in Genetics*, 9.

Cerritelli SM, Frolova EG, Feng C, Grinberg A, Love PE y Crouch RJ. Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. **2003** *Mol. Cell.*;11:807–815.

Chinnery PF y Zeviani M. 155th ENMC workshop: Polymerase gamma and disorders of mitochondrial DNA synthesis. **2008** *Neuromuscular Disorders*, 18(3), 259-267.

Chouchani ET, Kazak L y Spiegelman BM. Mitochondrial reactive oxygen species and adipose tissue thermogenesis: Bridging physiology and mechanisms. **2017** *Journal of Biological Chemistry*, 292(41), 16810-16816.

Ciesielski G, Oliveira M y Kaguni L. Animal Mitochondrial DNA Replication. **2016** *Enzymes*, 255-292.

Clayton DA, Doda JN y Friedberg EC. Absence of a pyrimidine dimer repair mechanism for mitochondrial DNA in mouse and human cells. **1975** *Basic Life Sci.*; 5B:589–591.

Coffey G, Lakshmipathy U y Campbell C. Mammalian mitochondrial extracts possess DNA end-binding activity. **1999** *Nucleic Acids Res* 27: 3348–3354.

Copeland WC y Longley MJ. Mitochondrial genome maintenance in health and disease. **2014** *DNA Repair*, 19, 190-198.

Davies AM, Hershman S, Stabley GJ, Hoek JB, Peterson J y Cahill A. A Ca²⁺-induced mitochondrial permeability transition causes complete release of rat liver endonuclease G activity from its exclusive location within the mitochondrial intermembrane space. Identification of a novel endo-exonuclease activity residing within the mitochondrial matrix. **2003** *Nucleic Acids Res* 31:1364–1373.

De Souza-Pinto NC, Mason PA, Hashiguchi K, Weissman L, Tian J, Guay D, Lebel M, Stevnsner TV, Rasmussen LJ y Bohr VA. Novel DNA mismatch-repair activity involving YB-1 in human mitochondria. **2009** DNA Repair (Amst).; 8:704–719.

De Stefani D, Patron M y Rizzuto R. Structure and function of the mitochondrial calcium uniporter complex. **2015** Biochim Biophys Acta 1853(9): 2006–2011.

Elliott HR, Samuels DC, Eden JA, Relton CL y Chinnery PF. Pathogenic Mitochondrial DNA Mutations Are Common in the General Population. **2008** The American Journal of Human Genetics 83, 254–260.

Eme L, Spang A, Lombard J, Stairs CW y Ettema TJG. Archaea and the origin of eukaryotes. **2017**; Nature Reviews Microbiology 15, 711–723.

Falkenberg M, Larsson NG y Gustafsson CM. DNA replication and transcription in mammalian mitochondria. **2007** Annu. Rev. Biochem. 76:679–99.

Friedman JR y Nunnari J. Mitochondrial form and function. **2014** Nature, 505(7483), 335-343.

Friedman JR, Lackner LL, West M, DiBenedetto JR, Nunnari J y Voeltz GK. ER tubules mark sites of mitochondrial division. **2011** Science 334:358–62.

Gabaldón T y Huynen MA. Shaping the mitochondrial proteome. **2004** Biochimica Et Biophysica Acta (BBA) - Bioenergetics, 1659(2-3), 212-220.

Gadaleta M, Rainaldi G, Lezza A, Milella F, Fracasso F y Cantatore P. Mitochondrial DNA copy number and mitochondrial DNA deletion in adult and senescent rats. **1992** Mutation Research/DNAging, 275(3-6), 181-193.

Grady JP, Campbell G, Ratnaïke T, Blakely EL, Falkous G, Nesbitt V, Schaefer AM, McNally RJ, Gorman GS, Taylor RW, Turnbull DM y McFarland R. Disease progression in patients with single, large-scale mitochondrial DNA deletions. **2013** Brain, 137(2), 323-334.

Grossman LI, Watson R y Vinograd J. The presence of ribonucleotides in mature closed-circular mitochondrial DNA. **1973** Proc. Nat. Acad. Sci. U.S.A.; 70:3339–3343.

Gustafsson CM, Falkenberg M y Larsson N. Maintenance and Expression of Mammalian Mitochondrial DNA. **2016** Annual Review of Biochemistry, 85(1), 133-160.

Havey MJ. Predominant paternal transmission of the mitochondrial genome in cucumber. **1997** J Hered.;88:232–5.

He L, Chinnery PF, Durham SE, Blakely EL, Wardell TM, Borthwick GM, Taylor RW y Turnbull D. Detection and quantification of mitochondrial DNA deletions in individual cells by real-time PCR. **2002** Nucleic acids research, 30(14), e68.

Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, Vogel J, Stadler PF y Hackermüller, J. Fast Mapping of Short Sequences with Mismatches, Insertions and Deletions Using Index Structures. **2009** PLoS Computational Biology, 5(9).

Holt IJ, Harding AE y Morgan-Hughes JA. Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. **1988** Nature, 331(6158), 717-719.

Holt IJ, Lorimer HE y Jacobs HT. Coupled Leading- and Lagging-Strand Synthesis of Mammalian Mitochondrial DNA. **2000** Cell, 100(5), 515-524.

Jeninga EH, Schoonjans K y Auwerx J. Reversible acetylation of PGC-1: connecting energy sensors and effectors to guarantee metabolic flexibility. **2010** Oncogene 29, 4617–4624.

Karp GC. Biología celular y molecular: conceptos y experimentos. Quinta edición, México DF: McGraw Hill Interamericana Editores, **2009**.

Knoll AH, Bergmann KD y Strauss JV. Life: the first two billion years. **2016** Phil Trans R Soc. B 371: 20150493.

Korhonen JA, Pham XH, Pellegrini M y Falkenberg M. Reconstitution of a minimal mtDNA replisome in vitro. **2004** EMBO J. 23:2423–29.

Ladoukakis ED y Zouros E. Evolution and inheritance of animal mitochondrial DNA: Rules and exceptions. **2017** Journal of Biological Research-Thessaloniki, 24(1).

Lane N. Serial endosymbiosis or singular event at the origin of eukaryotes? **2017**; J Theoret Biol 000 (2017) 1–10.

Larsson NG. Somatic mitochondrial DNA mutations in mammalian aging. **2010** Annu Rev Biochem 79: 683–706.

Lewin B. Genes VII. Séptima edición, Nueva York: Oxford University Press Inc. **2000**.

Liao Y, Smyth GK, Shi W. “The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote.” **2013** Nucleic Acids Research, 41, e108.

Ling-Ping L, Chien-Chen T, Ming-Jai S, Jiunn-Lee L, Yih-Sharng C, Yung-Zu T, Shoei K y Stephen H. Atrial fibrillation is associated with accumulation of aging-related common type mitochondrial DNA deletion mutation in human atrial tissue. **2003** Chest 123: 539–44.

Liu P y Demple B. DNA repair in mammalian mitochondria: Much more than we thought? **2010** Environmental and Molecular Mutagenesis.

Luo X, Budihardjo I, Zou H, Slaughter C, y Wang X. Bid, a Bcl2 Interacting Protein, Mediates Cytochrome c Release from Mitochondria in Response to Activation of Cell Surface Death Receptors. **1998** Cell, 94(4), 481-490.

Margulis L. On the origin of mitosing cells. **1967**; J Theoret Biol 14, 225-274.

Martin WF y Mentel M. The origin of mitochondria. **2010**; Nature Education 3(9): 58.

McCormick E, Place E y Falk MJ. Molecular genetic testing for mitochondrial disease: from one generation to the next. **2013** Neurotherapeutics; 10:251–261.

Morel F, Renoux M, Lachaume P y Alziari S. Bleomycin induced double-strand breaks in mitochondrial DNA of Drosophila cells are repaired. **2008** Mutat Res 637: 111 –117.

Nicholls DG, Bernson VSM, Heaton GM, Girardier L y Seydoux J. Effectors of Thermogenesis. Experientia Supplementum - The Identification of the Component in the Inner Membrane of Brown Adipose Tissue Mitochondria Responsible for

Regulating Energy Dissipation. Basilea: Birkhäuser, **1978**. Disponible en: https://link.springer.com/chapter/10.1007%2F978-3-0348-5559-4_9.

NIH: Human Genome Research Institute. 2000 Release: Mouse Sequencing Consortium – Why Mouse Matters [serie en internet] **2010** [consultado 2018 julio 17]. Disponible en: <https://www.genome.gov/10001345/importance-of-mouse-genome/>.

Parikh S, Goldstein A, Koenig MK, Scaglia F, Enns GM, Saneto R, Anselm I, Cohen BH, Falk MJm Greene C, Gropman AL, Haas R, Hirano M, Morgan P, Sims K, Tarnopolsky M, Van Hove JLK, Wolfe L y DiMauro S. Diagnosis and management of mitochondrial disease: A consensus statement from the Mitochondrial Medicine Society. **2014** Genetics in Medicine, 17(9), 689-701.

Peralta S, Wang X y Moraes CT. Mitochondrial transcription: Lessons from mouse models. **2012** Biochimica et Biophysica Acta 1819: 961–969.

Pernas L y Scorrano L. Mito-Morphosis: Mitochondrial Fusion, Fission, and Cristae Remodeling as Key Mediators of Cellular Function. **2016** Annual Review of Physiology, 78(1), 505-531.

Phillips NR, Simpkins JW y Roby RK. Mitochondrial DNA deletions in Alzheimers brains: A review. **2014** Alzheimers & Dementia, 10(3), 393-400.

Rahman S y Poulton J. Diagnosis of mitochondrial DNA depletion syndromes. **2009** Archives of Disease in Childhood, 94(1), 3-5.

Robinson MD, McCarthy DJ y Smyth GK. EdgeR: A Bioconductor package for differential expression analysis of digital gene expression data. **2009** Bioinformatics, 26(1), 139-140.

Schon EA, Dimauro S y Hirano M. Human mitochondrial DNA: Roles of inherited and somatic mutations. **2012** Nature Reviews Genetics, 13(12), 878-890.

Shutt TE y Gray MW. Bacteriophage origins of mitochondrial replication and transcription proteins. **2006** Trends Genet. 22:90–95.

Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C y Schatten G. Ubiquitin tag for sperm mitochondria. **1999** Nature 402:371–2.

Toyama EQ, Herzig s, Courchet J, Lewis Jr. TL, Losón OC, Hellberg K, Young NP, Chen H, Polleux F, Chan DC y Shaw RJ. AMP-activated protein kinase mediates mitochondrial fission in response to energy stress. **2016** Science, Vol. 351:6270.

Tuppen HA, Blakely EL, Turnbull DM y Taylor RW. Mitochondrial DNA mutations and human disease. **2010** Biochimica Et Biophysica Acta - Bioenergetics, 1797(2), 113-128.

Wali J, Masters S y Thomas H. Linking Metabolic Abnormalities to Apoptotic Pathways in Beta Cells in Type 2 Diabetes. **2013** Cells, 2(2), 266-283.

Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza AM, Elsas LJ y Nikoskelainen EK. Mitochondrial DNA mutation associated with Lebers hereditary optic neuropathy. **1988** Science, 242(4884), 1427-1430.

Wallace DC. Mitochondrial DNA sequence variation in human evolution and disease. **1994**, Proc. Nat. Acad. Sci. Vol. 91, pp. 8739-8746.

Wang C y Youle RJ. Form follows function for mitochondria. **2016** Nature Vol. 530, 288-289.

Wang C y Youle RJ. The Role of Mitochondria in Apoptosis. **2009** Annual Review of Genetics, 43(1), 95-118.

Woese CR y Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. **1977**; Proc Natl Acad Sci. 74, 5088–5090.

Woese CR, Kandler O y Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. **1990** Proc Natl Acad Sci. 87, 4576–4579.

Yang MY, Bowmaker M, Reyes A, Vergani L, Angeli P, Gringeri E, Jacobs HT y Holt IJ. Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strandasymmetric DNA replication. **2002** Cell.; 111:495–505.

Zhang H, Zhang YW, Yasukawa T, Dalla Rosa I, Khiati S y Pommier Y. Increased negative supercoiling of mtDNA in TOP1mt knockout mice and presence of topoisomerases II α and II β in vertebrate mitochondria. **2014** Nucleic Acids Res. 42:7259–67.

Zheng W, Khrapko K, Coller HA, Thilly WG y Copeland WC. Origins of human mitochondrial point mutations as DNA polymerase γ -mediated errors. **2006** Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 599(1-2), 11-20.